J. Vis. Commun. Image R. 43 (2017) 164–172

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

journal homepage: www.elsevier.com/locate/jvci

A privacy-preserving content-based image retrieval method in cloud environment ☆

Yanyan Xu [a,*], Jiaying Gong [a], Lizhi Xiong [b], Zhengquan Xu [a], Jinwei Wang [b], Yun-qing Shi [c]

[a] State Key Lab of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, 129 Luoyu Road, Wuhan 430079, China
[b] School of Computer and Software, Nanjing University of Science & Technology, Nanjing 210094, China
[c] Dept. of Electrical and Computer Engineering, New Jersey Institute of Technology, Newark 07102, USA

ARTICLE INFO

Article history:
Received 5 August 2016
Revised 14 December 2016
Accepted 2 January 2017
Available online 4 January 2017

Keywords:
Privacy-preserving
Image retrieval
Secure search
Orthogonal decomposition

ABSTRACT

In order to protect data privacy, image with sensitive or private information needs to be encrypted before being outsourced to a cloud service provider. However, this causes difficulties in image retrieval and data management. A privacy-preserving content-based image retrieval method based on orthogonal decomposition is proposed in the paper. The image is divided into two different components, for which encryption and feature extraction are executed separately. As a result, cloud server can extract features from an encrypted image directly and compare them with the features of the queried images, so that users can thus obtain the image. Different from other methods, the proposed method has no special requirements to encryption algorithms, which makes it more universal and can be applied in different scenarios. Experimental results prove that the proposed method can achieve better security and better retrieval performance.

1. Introduction

With the rapid development of Multimedia and Internet, massive images are generated and distributed, how to store and share such large amount of data efficiently becomes an important issues. It is a natural solution to outsource images in cloud service due to its tremendous advantages, such as on-demand self-service, ubiquitous network access, location independent resource pooling, rapid resource elasticity, usage-based pricing and transference of risk [1]. Reports show that Cloud services specifically designed for image storage and sharing, such as Instagram, are among the largest growing internet services today [2]. Additionally, how to efficiently search the image information from a massive image database is another challenging issues in large-scale images storing and sharing. Content-based image retrieval (CBIR), which involves extraction of visual features from image and search in the visual feature space for similar images, has grown rapidly in recent years. Progresses have been made in both the derivation of new features and the construction of signatures based on these features. The richness in the mathematical formulation of signatures grows alongside the invention of new methods for measuring similarity. In order to improve retrieval accuracy further, some methods are proposed to narrow down the "semantic gap" between the visual features and the richness of human semantics, such as machine learning and relevance feedback [3,4]. Yu et al. proposed to jointly considers visual features and click features in image retrieval to solve this semantic gap problem [5–8].

However, despite the tremendous advantages, privacy becomes the biggest concern about image storage and CBIR outsourcing in cloud. Under cloud environment, data owner, cloud service providers (CSP) and data user can be taken by different parties. Data owners no longer store their data locally so that they will lose physical control over it; CSP is considered "honest but curious" and the privacy should not be expected to be preserved by it, which means attackers may have more opportunities to gain unauthorized access to servers [9–11]. Under such circumstance, user's privacy information are more vulnerable to untrustworthy service providers and malicious intruders. In order to protect data privacy and to allow restrict access, sensitive images need to be encrypted before being uploaded to CSP. However, cipher-image will impede CBIR operations that are usually conducted on the plain-image. If users want to retrieve images from CSP, CSP needs to decrypt it first to make retrieval be operated on plaintext, which will make user's private information being exposed to attackers, break privacy and hence is not desired. Therefore it is necessary to develop technologies of privacy-preserving image retrieval over encrypted domain that can protect users' privacy without sacrificing the

☆ This paper has been recommended for acceptance by Zicheng Liu.

* Corresponding author.
E-mail address: xuyy@whu.edu.cn (Y. Xu).

http://dx.doi.org/10.1016/j.jvcir.2017.01.006
1047-3203/

usability and accessibility of the information, which is also called private/privacy-preserving CBIR (PCBIR) [12–14]. PCBIR is also a challenging issues that multimedia information security met in cloud environment, and it will become an indispensable part of future content-based search systems.

Information retrieval on encrypted domain originated from retrieval on text document. Song et al. [15] proposed a ciphertext scanning method based on streaming cipher to make sure whether the search term is existed in the ciphertext. Boneh et al. [16] proposed a keyword search method based on public-key encryption, where the server can identify whether messages encrypted by user's public key contain some specific keyword, but learn nothing else. Swaminathan et al. [17] explored techniques to securely rank-order the documents and extracted the most relevant documents from an encrypted collection based on the encrypted search queries. Wang el al. [18] utilized an order-preserving symmetric encryption (OPSE) to achieve both security and privacy-preserving, although the guarantee to security could be weakened by it. Cao et al. [19] proposed privacy-preserving multi-keyword ranked search over encrypted data. Xia et al. [20] proposed a secure multi-keyword ranked search scheme over encrypted cloud data, which simultaneously supports dynamic update operations like deletion and insertion of documents. Fu et al. [21] designed a searchable encryption scheme to support both multi-keyword ranked search and parallel search, which can take advantage of the powerful computing capacity and resources of the cloud server. Fu et al. [22] also present a secure multi-keyword ranked search scheme over encrypted cloud data, which simultaneously supports dynamic update operations like deletion and insertion of documents.

Although secure text search techniques 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 [23]. To the best of our knowledge, Lu et al. [24] proposed the first image retrieval scheme over encrypted domain, where secure indexing scheme for matching visual strings in the encrypted domain is constructed through order preserving encryption and randomized hash functions; Lu et al. [23] proposed three schemes to preserve the distances between encrypted image features approximately, including bit-plane randomization, random projection, and randomized unary encoding. Although it is efficient, the security is compromised. Karthik et al. [25] presented a transparent privacy preserving hashing scheme tailored to preserve the DCT-AC coefficient distributions, but its search results are inaccurate due to missing space-frequency information. In addition, its security is compromised because the constrained shuffling is used on part of AC coefficients. A similar method is proposed in [26], where AC coefficients are encrypted by using scrambling encryption, and its invariant histogram is used to measure the similarity between encrypted query image and database image. Ferreira et al. proposed a novel cryptographic scheme in [27], where color information is encrypted by deterministic encryption techniques to enable privacy-preserving image retrieval, and texture information is encrypted by probabilistic encryption algorithms for better security. Hsu et al. [28] proposed a homomorphic encryption-based secure SIFT for image feature extraction, but the size of ciphertext is expanded and the computing is laborious. Zhang et al. [29] proposed a secure image retrieval method based on Pallier encryption algorithm, however, the computational complexity and communication cost is very high due to characteristic of homomorphic encryption. Erkin et al. [30,31] proposed to use secure multiparty computation to retrieve private information, but it requires a large number of interactive rounds and communication complexity is too high, thus is not suitable for cloud environment. In [32], distance-preserving randomization encryption methods and homomorphic encryption are compared in terms of their search performance, security strength and computational efficiency. In [14], Xia et al. proposed to transform earth mover's distance (EMD) in a way that the cloud server can evaluate the similarity between images without learning sensitive information. Xia et al. [33] proposed a privacy-preserving and copy-deterrence CBIR scheme using encryption and watermarking techniques, where the image content is encrypted by a stream cipher so that the watermark can be directly embedded into the encrypted images by cloud server efficiently.

It should be noted that the above research results are all relying on specific encryption methods, such as shuffling and homomorphic encryption, which aims at preserving the distance of image features after images are encrypted and make the image retrieval in encrypted domain possible. However, these kinds of methods limit its universality. For example, multiparty computation and homomorphic encryption based methods are secure, but they are too computation and communication intensive to be used in low-profile devices and large-scale systems; distance-preserving randomization methods have the advantage of high efficient and less user-involvement, however, it is not suitable for some situations that have high requirements to security, etc. In order to solve these problems, a privacy-preserving content-based image retrieval method based on orthogonal decomposition is proposed in this paper. By using orthogonal decomposition, image can be divided into encryption field and feature extraction field, therefore encryption operation and feature extraction can be executed separately, which make it possible to let CSPs get image features of encrypted image directly without decrypt it and compare with features of queried image. Two kinds of operation results will be fused in the final results to guarantee image security by orthogonal composition. This method is quite different from other methods because it has no specific requirements of cipher algorithms. Experimental results show that the proposed scheme has good encryption security and can achieve better retrieval performance. As far as we know, this is the first time that orthogonal decomposition is used in privacy-preserving image retrieval.

The organization of this paper is as follows: Section 2 proposes our scheme. Section 3 provides experimental results and a performance analysis, and Section 4 presents conclusions.

2. Proposed scenarios

In this section, a privacy-preserving image retrieval method based on orthogonal decomposition is proposed, and its theory is given as follows.

2.1. Orthogonal decomposition

Orthogonal decomposition is a kind of vector representation method, any vector can be expressed as a sum of a set of component coefficients through orthogonal decomposition. Assume that $x \in R^m$ is a random vector and $\{\varnothing_i\}_{i=1}^m$ is a set of arbitrary orthogonal basis vectors, x can be expressed as:

$$x = \sum_{i=1}^{m} y_i \varnothing_i = y_1\varnothing_1 + y_2\varnothing_2 + \cdots + y_m\varnothing_m = \varnothing y \tag{1}$$

and $\{y_i\}_{i=1}^m$ is a set of component coefficient, y_i can be expressed as:

$$y_i = \sum_{i=1}^{m} x_i \varnothing_i^T \tag{2}$$

For any x, if there is a δ change of y_k, then x will be changed too:

$$x' = \sum_{i=1}^{m} y_i \varnothing_i = y_1\varnothing_1 + \cdot s + (y_k + \delta)\varnothing_k + \cdot s + y_m\varnothing_m \tag{3}$$

For y_i, it can be expressed as:

$$y_i = \begin{cases} y_k + \delta & \text{if } i = k \\ y_i & \text{if } i \neq k \end{cases} \quad i = 1, \ldots, m \tag{4}$$

Assuming q, p is n-dimensional vector, A is an orthogonal matrix of size n × n, then we can get Q = Aq, P = Ap, where Q, P is component vector after orthogonal decomposition. Supposing s = q − p, S = Q − P, then we can get S = As. if L is a kind of distance measuring method, then we can get:

$$\mathrm{L}(\mathrm{Q},\mathrm{P})^2 = S^T S = (As)^T(As) = s^T A^T As = s^T s = L(q,p)^2 \tag{5}$$

From (3)–(5), we can get three characteristics of orthogonal decomposition:

(1) Independence. Any change of a component coefficient will not affect the others, that is, they are mutually independent;
(2) Fusion. Composite vector is sensitive to any change of component coefficient that will be fused in the composite vector.
(3) Distance preserving. The distance between component vectors after orthogonal decomposition is equal to the distance between original vectors.

The first characteristics can be used to keep the independence of different operations, and the second one guarantees different operation results being fused in the final results. These two characteristics enable it suits for different scenarios. For example, we proposed a commutative encryption and watermarking (CEW) framework based on orthogonal decomposition (CEWod) in [34], where encryption and watermarking are combined to protect both the confidentiality and the ownership of digital media, while encryption and watermarking can be operated in any order to achieve commutativity. Similarly, orthogonal decomposition can be used to keep encryption operation and image retrieval operation independent, thus image features will not be changed by encryption. Meanwhile, with the third characteristic, orthogonal decomposition can solve the problem of privacy-preserving image retrieval. However, privacy-preserving image retrieval is quite different from CEW in several aspects, how to apply orthogonal decomposition in privacy-preserving image retrieval is still a challenging problem. In this paper, we will solve these problems and propose a privacy-preserving image retrieval method based on orthogonal decomposition.

2.2. Privacy-preserving content-based image retrieval based on orthogonal decomposition

Based on the theory mentioned above, a new privacy-preserving content-based image retrieval method is proposed. We assume that the original image data can be expressed as n-dimensional vector $\mathrm{X} = (x_1, x_2, \ldots, x_n)^T$. Denote $\mathrm{B} = (b_1, b_2, \ldots, b_n)$ an orthogonal matrix of size n × n, which satisfies

$$\begin{cases} \mathrm{b_i^T} * \mathrm{b_j} \neq 0 & \text{if } \mathrm{i} = \mathrm{j} \\ \mathrm{b_i^T} * \mathrm{b_j} = 0 & \text{otherwise} \end{cases} \quad \mathrm{i,j} \in [1, \mathrm{n}] \tag{6}$$

Using orthogonal decomposition based on B, X can be represented as:

$$X = B \cdot Y \tag{7}$$

where the component coefficient vector $\mathrm{Y} = (y_1, y_2, \ldots, y_n)^{\mathrm{T}}$ is calculated by:

$$Y = B^T \cdot X \tag{8}$$

Matrix B can be divided into two sub-matrixes, i.e., B = (R, S), where $\mathrm{R} = (b_1, b_2, \ldots, b_m)$ and $\mathrm{S} = (b_{m+1}, b_{m+2}, \ldots, b_n)$, then vector Y can be divided into two sub-vectors: $\mathrm{Y} = (Y_1, Y_2)^T$, and Y_1, Y_2 can be described as:

$$Y_1 = R^T \cdot X \tag{9}$$

$$Y_2 = S^T \cdot X \tag{10}$$

X can be expressed as:

$$X = R \cdot Y_1 + S \cdot Y_2 \tag{11}$$

If encryption is operated on Y_1 and feature is extracted from Y_2 respectively, encryption operation is defined as E(), K_e is a key for encryption, feature extraction is defined as F(), K_f is a key for feature extraction, then we can get X_{ef}:

$$X_{ef} = R \cdot E(Y_1, K_e) + S \cdot F(Y_2, K_f) = B \cdot Y_{ef} \tag{12}$$

Under the proposed framework, encryption and feature extraction are applied to orthogonal decomposition coefficients Y_1 and Y_2 instead of being directly applied to X, and these two different operations will be independent and will not interfere with each other. Because orthogonal based vector is mutual independent, the modification of orthogonal decomposition coefficients will not counteract each other on the original data domain after orthogonal composition, as a result, two different kinds of operation results are fused in original host vector X.

Assuming the encrypted Y_1 coefficient is Y_{1e}, the corresponding vector after orthogonal composition is X_{ef}, decryption operation is defined as $D()$, decryption key is K_d, feature extraction operation is defined as $F()$, feature extraction key is K_f, then we get:

$$D(X_{ef}, K_d) = R \cdot D(Y_{1e}, K_d) + S \cdot Y_2 \tag{13}$$

$$F(X_{ef}, K_f) = F(Y_2, K_f) \tag{14}$$

From (13) and (14), we can see that decryption operation is only related to Y_{1e} while feature extraction is only related to Y_2, even if Y_{1e} is encrypted, image features can be extracted from Y_2 directly.

However, there are two issues existed in the above process. First, the normal orthogonal matrix is in real field. Because of finite precision of computer, the orthogonal transform in floating point number may cause non-negligible errors that will have influence on correctness of operation. Secondly, R and S are both needed to participate in decryption and feature extraction process, which will reduce the independence of the two operations and make the retrieval process insecure. In order to solve these problems, we transform the above process to the following: compensation operations are introduced into the transform to obtain the correct results; R and S are defined as the key of different operand domain, which is only related to decryption and feature extraction separately. Then we transform the above equations into the following:

Assume the encryption domain is divided by $\mathrm{R_p}$, and the feature extraction domain by $\mathrm{S_p}$, which satisfy $(R_p, S_p)^T = P \cdot B^T$, where P is a compensation matrix, $P = (B^T B)^{-1} = diag(\lambda_1^{-1}, \lambda_2^{-1}, \ldots, \lambda_n^{-1})$, $P_R = (R^T R)^{-1} = diag(\lambda_1^{-1}, \lambda_2^{-1}, \ldots, \lambda_m^{-1})$, $P_S = (S^T S)^{-1} = diag(\lambda_{m+1}^{-1}, \lambda_{m+2}^{-1}, \ldots, \lambda_n^{-1})$, then we can get:

$$\begin{cases} Y_1 = R_p^T \cdot X = P_R \cdot R^T \cdot X \\ Y_2 = S_p^T \cdot X = P_s \cdot S^T \cdot X \end{cases} \tag{15}$$

$$\begin{cases} Y_{1e} = R_p^T \cdot X_{ef} = P_R \cdot R^T \cdot X_{ef} \\ Y_2 = S_p^T \cdot X_{ef} = P_s \cdot S^T \cdot X_{ef} \end{cases} \tag{16}$$

$$D(X_{ef}, K_d) = R \cdot D((R^T \cdot R)^{-1} \cdot R^T \cdot X_{ef}, K_d) + X_{ef} - R \cdot (R^T \cdot R)^{-1} \cdot R^T \cdot X_{ef} = D(X_{ef}, K_d, R) \quad (17)$$

$$F(X_{ef}, K_f) = F(Y_2, K_f) = F((S^T \cdot S)^{-1} \cdot S^T \cdot X_{ef}, K_f) = F(X_{ef}, K_f, S) \quad (18)$$

Eqs. (17) and (18) proves that decryption is only related to matrix R, and feature extraction is only determined by matrix S. If only S is given, then the feature can be extracted from cipher-image directly while it is still kept encrypted. That is, R and S, by which the operand domains of encryption and feature extraction, can also be defined as types of keys used to protect two different operands.

2.3. System model

According to Eqs. (7)–(18), system model for the proposed method is shown in Fig. 1. There are three entities involved in this model: content owner who owns the images, CSP who stores encrypted images and performs image retrieval, and users who want to get images.

The retrieval process is given as following:

(1) Content owner constructs operating data X from image, performs orthogonal decomposition on X with R, S and gets encryption operation field Y_1 and feature extraction operation field Y_2. Y_{1e} is available after Y_1 is encrypted, and feature Y_{2f} is extracted from Y_2. X_{ef} is calculated through orthogonal composition with B from Y_{1e} and Y_2, then the encrypted image is formed through X_{ef} and uploaded to CSP. Content owner only needs to save features Y_{2f} as search index.
(2) Users send request to content owner. After identity authentication, content owner sends Y_{2f} to users, users send it to cloud servers as search index.
(3) CSP gets matrix S and K_f securely from content owner, and operates orthogonal decomposition on encrypted images and encrypted query image. Features are extracted and compared with search index sent by users, and images with smallest distance will be returned to users.
(4) Users get R and decryption key K_d securely from content owner, decrypt the cipher-image and get plain image.

2.4. Implementation of the proposed method

The detailed implementation of the proposed method includes the following steps: operating data organization; orthogonal decomposition; encryption; image search; orthogonal composition; data postprocessing.

2.4.1. Operating data organization

Theoretically any data can be used as operating data. Considering most images need to be transmitted or stored in compressed format such as JPEG, in order to improve efficiency of data processing, we partially decode JPEG streaming and extract quantized coefficients to constitute operating data. Since combination of DC coefficients and AC coefficients in low frequency will yield best retrieval results [29], we extract quantized DC and low-middle frequency AC coefficients from each sub-block of image and constitute the host data matrix X.

2.4.2. Orthogonal decomposition

Selecting orthogonal matrix B is the key part of orthogonal decomposition. Any matrix satisfying Eq. (6) can be used as B, however, different orthogonal matrix will have different effects in applications. For example, DCT is usually used in data compression because it has the characteristics of de-correlation and energy concentration; DWT is often used in numerical analysis and functional analysis as it has a key advantage of capturing both frequency and location information. In the proposed method, we use a random Gaussian orthonormal matrix as B, which can not only be used to implement orthogonal decomposition on host data X but also to improve its security, because the elements of matrix are random variables. According to Eq. (15), X is decomposed into encryption field Y_1 and feature extraction field Y_2.

2.4.3. Encryption

In the proposed method, there is no specific requirement to cryptographic algorithms. We choose Advanced Encryption Standard (AES) to encrypt data. If encryption is only operated on X formed by selected DC and AC coefficients, then other coefficients are transparent which will decrease the security to a certain extent. To overcome these difficulties and achieve better security, we encrypt all of other information, i.e., all other AC coefficients. After encryption, we get encrypted component Y_{1e}.

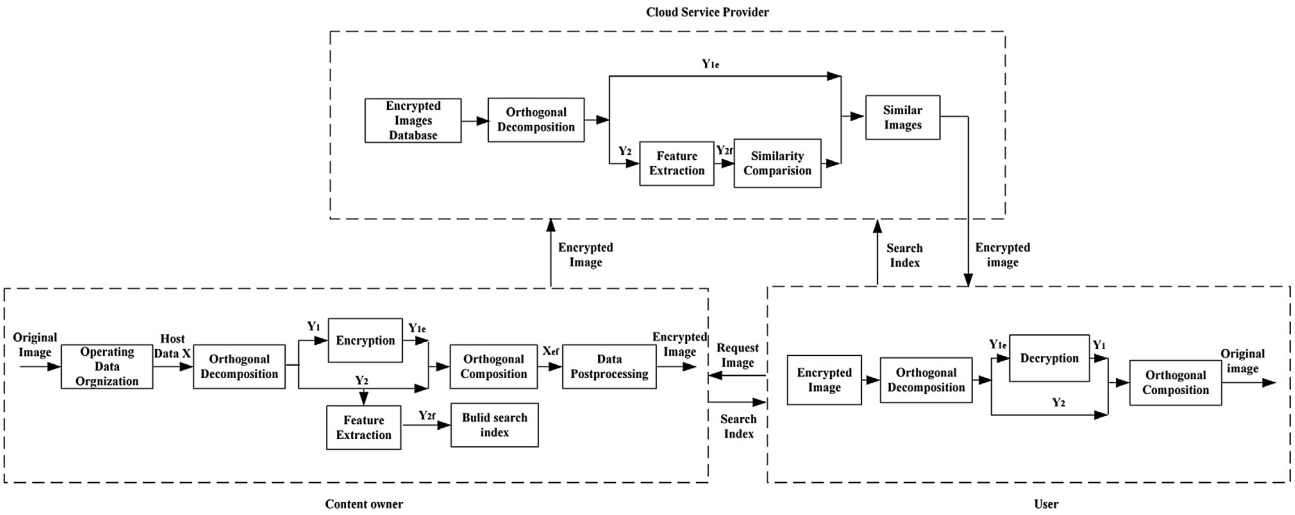

Fig. 1. System model of the proposed method.

2.4.4. Image search

Image search includes feature extraction and similarity comparison. Since orthogonal decomposition is implemented on DCT coefficients in the proposed method, features in compressed domain are used to represent image content. Many methods have been presented in image retrieval of compressed domain. Jiang etc. proposed a JPEG retrieval method in [35], where they extract an approximated image with smaller size for indexing and content browsing without incurring full decompression. Lu et al. extract color, texture and basic edge information from JPEG images in [36]. Lay et al. presented a method that using energy histograms of the low frequency DCT coefficients as features for the retrieval of DCT compressed images [37]. Eom et al. built an edge histogram detector extracted from AC coefficients that resembles the MPEG-7 non-homogeneous texture descriptor in [38]. Edmundson et al. compared performance of these JPEG compressed domain image retrieval techniques with pixel-domain CBIR techniques, such as color histograms and MPEG-7 visual descriptors, and concluded that several of the JPEG CBIR techniques allow much faster feature calculation and faster image retrieval, while providing retrieval performance similar to common pixel-domain algorithms [39]. Schaefer et al. utilize DC difference histogram and color histogram as image features in [40], and get better retrieval performance than methods mentioned above. Recently, two kinds of compact descriptor in compressed domain are proposed in [41,42], which can tackle contradictory aspects of efficiency and precision effectively.

We modified the method proposed in [40] to extract features. Most of energy concentrates in DC coefficients in JPEG image, the differences between them describe important information about the image gradient, image texture and about edges in the image as well as about uniform image areas, therefore a histogram of DC difference data in Y channel is utilized as an important image feature. Besides, since AC coefficients provides a simple description of image texture, so AC histogram of Y channel is used as another image feature. Combining color descriptor can also improve retrieval performance, thus DC and AC coefficients histogram of C_b and C_r channels are also used as feature components. L1 norm on normalized histograms is used for distance calculation. To integrate these three components, we utilize weighted average of three distances. Denote d_t as a distance calculated for DC difference histograms, d_h as a distance calculated for AC histogram, d_g as a distance calculated for color histogram, then the combined distance is established as:

$$d = \propto d_t + \beta d_h + \gamma d_g \tag{19}$$

The results are sorted and images with smallest distance will be returned as target images.

2.4.5. Orthogonal composition

The encrypted X_{ef} is calculated through the inverse orthogonal transform with B and the encrypted component Y_{1e} and Y_2. According to characteristics of orthogonal transform, two different kinds of operations are overlapped and distributed in original host vector X.

2.4.6. Data postprocessing

After being processed, X_{ef} are stored back to their original locations where they are extracted. Following the consequent encoding process such as entropy encoding, the encrypted image is constructed.

3. Experimental results

In this section, we present an experimental evaluation of the proposed method and compare results with other methods proposed in [23,32], that is, bit-plane randomization, random projection, and randomized unary encoding, homomorphic encryption. We perform experiment on three image datasets: Corel 1k dataset [43], containing 1000 images which are grouped into 10 categories; Corel 10k dataset [44], containing 10,000 images which are grouped into 100 categories; Inria Holidays dataset [45], 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, retrieval performance and computational complexity. All experiments are implemented by C/C++ and run on a Windows Server with Intel Core i3 Processor 3.07 GHz CPU and 6 GB RAM.

3.1. Security analysis

Different from text/binary encryption, multimedia information security requires not only cryptographic security, but also perceptual security. Both security against cryptographic attacks and perception unintelligibility should be satisfied. Therefore we will prove the security of the proposed method from these two aspects.

3.1.1. Perceptual security

We use Peak Signal to Noise Ratio (PSNR) to evaluate perceptual security of cipher-image. The encryption results of images are show in Fig. 2. From these figures, we can see that encrypted images are unintelligible and have low PSNR value, which means good perceptual security. Malicious attackers cannot infer image content from encrypted images.

In our scheme, feature extraction field are unencrypted in the cipher-image. There exists the possibility that attackers make use of the unencrypted data as a security leak to attack the protected information, the degree to which the protected information may be revealed by the unencrypted data must be assessed.

Assume the protected image is I, I can be expressed as:

$$I = f(X, Z) = f_1(X) + f_2(Z) \tag{20}$$

where $f(\cdot,\cdot)$ is a linear function that is determined by an image encoding algorithm, f1(.), f2(.) are different image encoding algorithm. X is a selected operating dataset, Z is other data besides X. X is encrypted by (12). The encrypted image, C_e can be reconstructed as:

$$C_e = f_2(Z_e) + f_1(X_e) = f_2(Z_e) + f_1(R, Y_{1e}) + f_1(S, Y_2) \tag{21}$$

In C_e, only $f_1(S,Y_2)$ is unencrypted, thus it is the only term that may lead to some information leakage. However, if X is properly defined and R, S is appropriately selected, the potential information leakage can be controlled within an acceptable range. c

To prove conclusion mentioned above, we encrypt each kind of image in Corel 1k dataset. Table 1 shows experimental results of encrypted image quality with different X and different size of R. From the table we can see if X is formed by less coefficients, perceptual security of cipher-image is better. This is because more coefficients involved in the orthogonal decomposition will lead to more information leakage. We also can see larger size of R, better perceptual security of image. But different X and different size of R only has a slight difference in PSNR of cipher-image, which means the information leakage is very limited and perceptual security of the proposed method is good.

3.1.2. Cryptographic security

In the proposed method, R and S are sub-matrixes of B, they are not fully stochastic and independent numbers like a single secret key, which reduces the valid value space of R and S. Therefore the system is not secure unless two conditions are validated:

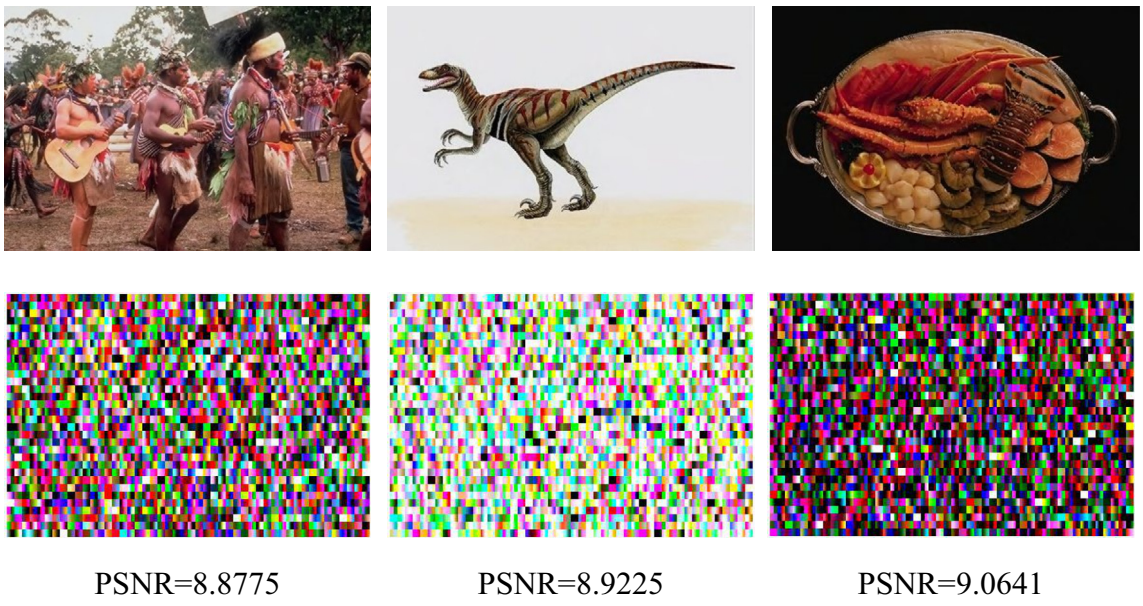

Fig. 2. Encrypted image quality.

Table 1
PSNR of cipher-image with different operated data X and different size of R.

Image name	PSNR(dB)								
	DC + 1AC			DC + 3AC			DC + 5AC		
	R = 16	R = 32	R = 48	R = 16	R = 32	R = 48	R = 16	R = 32	R = 48
African	8.9114	8.8400	8.4946	8.9840	8.8775	8.5079	8.9928	8.8691	8.5524
Beach	9.2442	9.1371	8.8138	9.2877	9.1766	8.8461	9.3716	9.2041	8.8571
Building	8.8518	8.7445	8.4917	8.8651	8.7721	8.4489	8.8677	8.8186	8.5036
Bus	8.5832	8.4952	8.2607	8.6205	8.5352	8.2798	8.6362	8.5772	8.3570
Dinosaur	9.0066	8.9025	8.5100	9.0662	8.9225	8.5359	9.0857	8.9790	8.5597
Elephant	8.8730	8.7436	8.5007	8.8847	8.7941	8.5221	8.9142	8.8871	8.5988
Flower	8.5521	8.4576	8.2371	8.5947	8.5539	8.2307	8.6055	8.5655	8.2390
Horse	9.3122	9.1794	8.8820	9.3354	9.1826	8.9156	9.3742	9.2454	8.9269
Mountain	8.5508	8.4809	8.2139	8.6456	8.5606	8.2824	8.6559	8.5683	8.2825
Food	9.1900	9.1170	8.7707	9.1911	9.0641	8.8352	9.2581	9.1257	8.8464

Condition1: as secret keys, R and S can be adjusted independently to satisfy any predetermined security threshold, i.e., its key space is large enough to resist the brute force attack.

Condition2: it is difficult for attackers to decrypt cipher-image unless R and K_d are known; it is difficult for the server to deduce R through S.

We proved in [28] that the complexity for attacking B is:

$$d_B \sim 2^{n^2k-k_m} + 2^{\frac{n(n+1)(k-1)}{2}+k_0} \tag{22}$$

where the size of B is n, elements of B is k-bit fixed word-length integer, the number of columns in S is m, k_m is the factor that includes the efficiency improvement achieved over an exhaustive search by using a different algorithm. Thus the complexity for attacking B tends to increase exponentially as $O(n^2k)$.

Eq. (22) shows that R and S, which are the B sub-matrixes, can be adjusted independently to satisfy any predetermined security threshold by set rational n and k. Therefore condition 1 is met.

We also proved the complexity for attacking R with a known S can be estimated as:

$$d_R \sim 2^{3k\log_2^{m(n-m)}+m^2k-k_m} + 2^{\frac{m(m+1)k}{2}+k_0} \tag{23}$$

The complexity for attacking S with a known R can be estimated as:

$$d_s \sim 2^{3k\log_2^{m(n-m)}+(n-m)^2k-k_m} + 2^{\frac{(n-m)(n-m+1)k}{2}+k_0} \tag{24}$$

According to (22)–(24), the complexity of attacking B and R, S are exponential, increasing with $O(n^2k)$, $O(m^2k)$ and $O((n-m)^2k)$, respectively. Thus provide the criteria for the proper choice of the parameters B, R, S to achieve computational security of the scheme. For example, if n = 8, m = 4, and k = 16, the complexity for B and R must be of order $2^{1024-k_m}+2^{476+k_0}$, $2^{448}+2^{160+k_0}$, respectively, which satisfies the security requirements of most applications. The above analysis proved that condition 2 could also be met.

Since condition 1 and 2 are all met, then we can get a conclusion that the system is secure.

3.2. Retrieval performance

Retrieval performance is evaluated by precision-recall curves that are widely used to measure image retrieval performance. Precision and recall are defined as:

$$\text{precision} = \frac{\text{number of relevant images among retrieved images}}{\text{number of retrieved images}}$$

$$\text{recall} = \frac{\text{number of relevant images among retrieved images}}{\text{number of relevant images in the data base}}$$

Besides, mean average precision (mAP) is used to evaluate the retrieval performance for a group of queries. It is defined as the means of average precision (AP) for a group of queries:

$$\text{mAP} = \frac{\sum_{q=1}^{Q} AveP(q)}{Q}$$

where Q is the number of queries, $AveP()$ is the average of all the precisions measured each time a new relevant image is retrieved.

A higher precision value at a given recall value indicates that the retrieval performance is better. We perform retrieval performance experiment on different dataset and compare experimental results with the methods proposed in [18,27]. Image in each category in the dataset is used as query over all others. Fig. 3(a) shows the experimental results conducted on Corel 1k dataset, from the figure we can see that our methods get a better retrieval performance than other methods. Fig. 3(a) also shows if a wrong S is given, the retrieval precision is reduced to around 10%. Therefore, without knowing the correct key, retrieval from an encrypted image database equals to picking images randomly from the database, which proves our retrieval process is secure. Fig. 3(b) shows the results performed on Corel 10k dataset. Although the precision is decreased due to large size of dataset, our method still achieves better retrieval results than other methods. However, because the proposed method has no special requirement to feature extraction, advanced techniques can be used to improve the retrieval performance, such as machine learning. We calculate the mAP of a group of 100 queries on different datasets and show the results in Table 2. In this experiment, our method achieves the best retrieval result, too.

In the proposed method, since features are extracted from the component vectors after orthogonal decomposition, and this will affect retrieval performance in some degree. To illustrate how far it is, we conducted image retrieval on Corel 10k dataset with/without orthogonal decomposition when S size is 48, recall is 0.1. Fig. 4 shows the results. We can see that there is only a slight difference between them, which also proves the distance preserving characteristics of the orthogonal decomposition we mentioned in Section 2.

In our method, features are extracted from DCT coefficients, operation field is divided into two parts, thus different host data X and size of R or S will have different influence on retrieval performance. Fig. 5 shows corresponding retrieval results on Corel 1k dataset when recall is 0.1. From figure we can see retrieval precision is higher when S size is larger, this is because more features can be extracted from a larger feature extraction field, which will improve retrieval performance. From figure we can also see that more AC coefficients will also lead to better retrieval performance. However, X formed by DC and first 5 AC coefficients in Zigzag

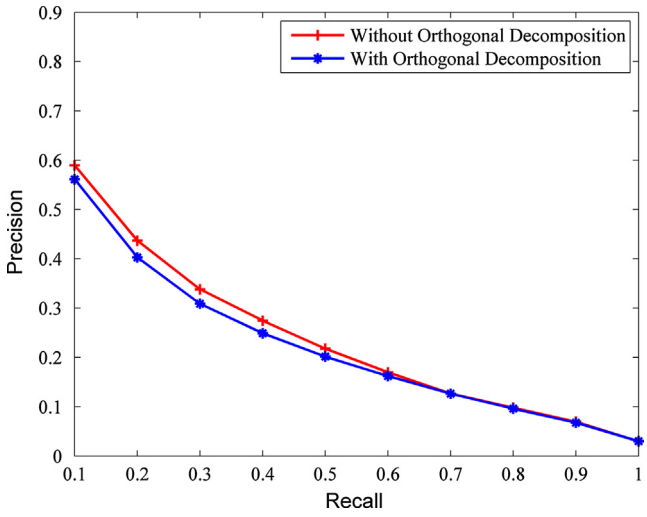

Fig. 4. Precision vs. recall graph with/without orthogonal decomposition on Corel 10k dataset.

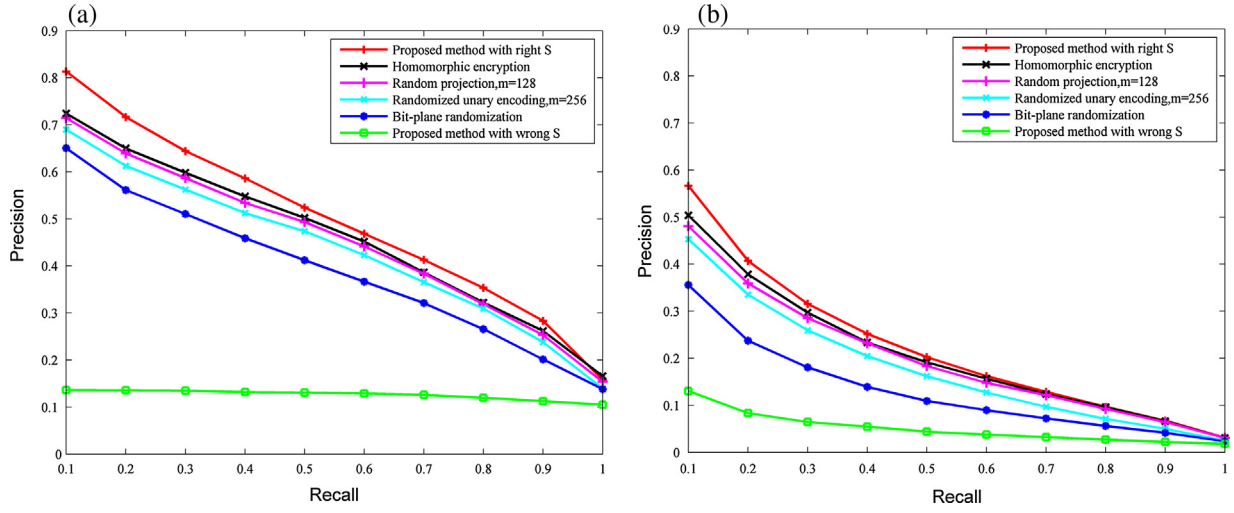

Fig. 3. Precision vs. recall graph.

Table 2
mAP for different datasets.

Dataset	The proposed method (%)	Homomorphic encryption (%)	Random projection (%)	Random unary encoding (%)	Bitplane randomization (%)
Corel 1 k	72.61	66.31	63.40	61.40	58.67
Corel 10k	46.62	41.11	40.19	39.12	30.79
Holiday	56.04	53.94	53.8	51.67	38.15

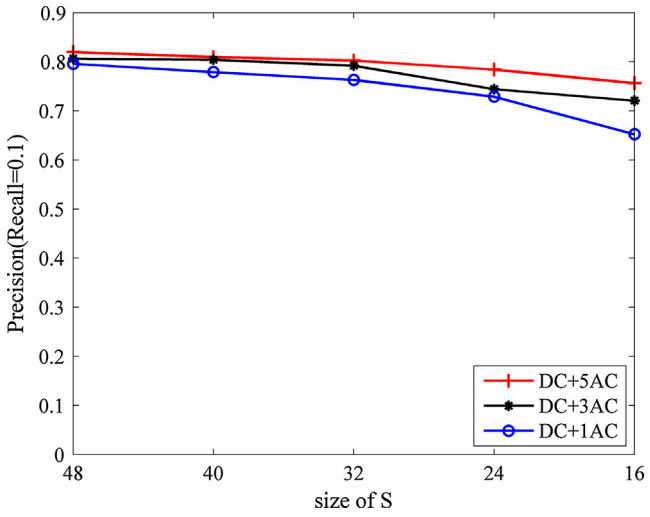

Fig. 5. Precision vs. recall graph with different host data X and different size of S on Corel 1k dataset.

Table 3
Comparison of computational complexity.

Methods	Time complexity	Precise computation time (s)
Paillier homomorphic	$O(n^4(n^n)^{-2})$	1478.75
Bit plane randomization	O(gn)	0.01
Random projection	O(mn)	0.12
Randomized unary encoding	O(mnM)	3.65
The proposed method	O(kn)	0.02

sequence only achieves minor advantages than by DC and 3 AC coefficients, this is because the first 3 AC coefficients carrying more texture and edge information, which affects retrieval performance more than other components. Considering more AC coefficients involved in feature extraction will result in more information leakage, DC and first 3 AC coefficients should be a good choice for forming X.

In the proposed method, suppose the size of X is n, the size of S is k ($k < n$), then the time complexity of getting feature extraction field is O(kn). The comparison results with other methods are given in Table 3, where g is the number of bit-planes to randomize, m is the dimension of the projected features, and M is the largest value of the feature vector. From Table 3 we can see that Paillier homomorphic method has highest time complexity; randomized unary encoding has higher time complexity; time complexity of our method is lower than these methods and is similar to bit plane randomization. Precise computation time is for all 1000 images in the Corel database, and the results are in accordance with the theory analysis.

4. Conclusion

A privacy-preserving content-based image retrieval method based on orthogonal decomposition in cloud environment is proposed in the paper. By orthogonal transform, an image is decomposed into components of two orthogonal fields, therefore encryption and feature extraction can be operated separately. Two independent components are integrated to form the final data after orthogonal composition. With this method, the CSP can retrieve image from encrypted image database directly without violating data privacy. Different from other methods reported in the literatures, the proposed method has no restrictions in using special encryption algorithms, which makes the proposed method more universal, thus can accommodate different kinds of applications. Applying more effective feature extraction algorithm to improve retrieval accuracy is still an open problem and will be further studied in our future work.

Acknowledgments

We thank the anonymous reviewers for their helpful suggestions. This work was supported by the National Natural Science Foundation of China under Grant 41571426, 41371402, 61232016, U1405254, PAPD, and LIESMARS Special Research Funding.

References

[1] P. Mell, T. Grance, Draft NIST Working Definition of Cloud Computing <http://csrc.nist.gov/groups/SNS/cloudcomputing/ index.html>, June 2009.
[2] Global Web Index, Instagram Tops the List of Social Network Growth <http://blog.globalwebindex.net/instagram-tops-list-of-growth>, 2013.
[3] R. Datta, D. Joshi, J. Li, J.Z. Wang, Image retrieval: ideas, influences, and trends of the new age, ACM Comput. Surv. 40 (2) (2008) 5–60.
[4] Liu, D. Zhang, G. Lu, W. Ma, A survey of content-based image retrieval with high-level semantics, Pattern Recogn. 40 (2007) 262–282.
[5] J. Yu, D. Tao, M. Wang, et al., Learning to rank using user clicks and visual features for image retrieval, IEEE Trans. Cybernet. 45 (4) (2015) 767–779.
[6] J. Yu, X. Yang, F. Gao, et al., Deep multimodal distance metric learning using click constraints for image ranking, IEEE Trans. Cybernet. (2016) 1–11.
[7] J. Yu, Y. Rui, B. Chen, Exploiting click constraints and multi-view features for image re-ranking, IEEE Trans. Multimedia 16 (1) (2014) 159–168.
[8] J. Yu, Y. Rui, D. Tao, Click prediction for web image reranking using multimodal sparse coding, IEEE Trans. Image Process. 23 (5) (2014) 2019–2032.
[9] S. Subashini, V. Kavitha, A survey on security issues in service delivery models of cloud computing, J. Netw. Comp. Appl. 34 (1) (2011) 1–11.
[10] L. Ferretti, F. Pierazzi, M. Colajanni, M. Marchetti. Security and confidentiality solutions for public cloud database services, in: The Seventh International Conference on Emerging Security Information, Systems and Technologies, 2013.
[11] R. Huang, X. Gui, S. Yu, et al., Study of privacy-preserving framework for cloud storage, Comp. Sci. Inf. Syst. 8 (3) (2011) 801–819.
[12] J. Shashank, P. Kowshik, K. Srinathan, et al., Private content based image retrieval, IEEE Conference on Computer Vision & Pattern Recognition (2008) 1–8.
[13] L. Weng, L. Amsaleg, A. Morton, et al., A privacy-preserving framework for large-scale content-based information retrieval, IEEE Trans. Inf. Forensics Secur. 10 (1) (2015) 152–167.
[14] Z. Xia, Y. Zhu, X. Sun, et al., Towards privacy-preserving content-based image retrieval in cloud computing, IEEE Trans. Cloud Comput. (2015) 1.
[15] D. Song, D. Wagner, A. Perrig, Practical techniques for searches in encrypted data, in: Proceedings of the IEEE Symposium on Research Security and Privacy, 2000, pp. 44–55.
[16] D. Boneh, G. Crescenzo, R. Ostrovsky, G. Persiano. Public-key encryption with keyword search, in: Advances in Cryptology - EUROCRYPT 2004. Springer, Berlin, Heidelberg, 2004, pp. 506–522.
[17] A. Swaminathan, Y. Mao, G.-M. Su, H. Gou, A. L. Varna, S. He, et al., Confidentiality preserving rank-ordered search, in: Proceedings of ACM Workshop Storage, Security, and Survivability, 2007, pp. 7–12.
[18] C. Wang, N. Cao, J. Li, K. Ren, W. Lou, Secure ranked keyword search over encrypted cloud data, in: Proceedings of IEEE 30th International Conference on Distributed Computing Systems (ICDCS) 2010, pp. 253–262.
[19] N. Cao, C. Wang, M. Li, K. Ren, W. Lou, Privacy-preserving multi-keyword ranked search over encrypted cloud data, IEEE Trans. Parallel Distrib. Syst. 25 (1) (2014) 222–233.
[20] Zhihua Xia, Xinhui Wang, Xingming Sun, Qian Wang, A secure and dynamic multi-keyword ranked search scheme over encrypted cloud data, IEEE Trans. Parallel Distrib. Syst. (2015), http://dx.doi.org/10.1109/TPDS.2015.2401003.
[21] Zhangjie Fu, Xingming Sun, Qi Liu, Lu Zhou, Jiangang Shu, Achieving efficient cloud search services: multi-keyword ranked search over encrypted cloud data supporting parallel computing, IEICE Trans. Commun. E98-B (1) (2015) 190–200.
[22] Zhangjie Fu, Xinle Wu, Chaowen Guan, Xingming Sun, Kui Ren, Towards efficient multi-keyword fuzzy search over encrypted outsourced data with accuracy improvement, IEEE Trans. Inf. Foren. Sec., doi:http://dx.doi.org/10.1109/TIFS.2016.2596138.
[23] W. Lu, A.L. Varna, A. Swaminathan, M. Wu, Secure image retrieval through feature protection, in: Proceedings of the IEEE Conference on Acoustics, Speech Signal Processing, 2009, pp. 1533–1536.
[24] W. Lu, A. Swaminathan, A.L. Varna, M. Wu. Enabling search over encrypted multimedia databases, in: IS&T/SPIE Electronic Imaging, International Society for Optics and Photonics, 2009, pp. 725418–725418-11.

[25] K. Karthik, S. Kashyap, Transparent hashing in the encrypted domain for privacy preserving image retrieval, SIViP 7 (4) (2013) 647–664.
[26] H. Cheng, X. Zhang, J. Yu. AC-coefficient histogram-based retrieval for encrypted JPEG images, Multim. Tools Appl., doi:http://dx.doi.org/10.1007/s11042-015-2741-z.
[27] B. Ferreira, J. Rodrigues, J. Leitao, et al., Privacy-preserving content-based image retrieval in the cloud, Reliab. Distrib. Syst. IEEE (2015) 11–20.
[28] C. Hsu, C. Lu, S. Pei, Image feature extraction in encrypted domain with privacy-preserving SIFT, IEEE Trans. Image Process. 21 (11) (2012) 4593–4607.
[29] Y. Zhang, Li. Zhuo, Y. Peng, J. Zhang, A secure image retrieval method based on homomorphic encryption for cloud computing, in: Proceedings of the 19th International Conference on Digital Signal Processing, IEEE, 2014, pp. 269–274.
[30] Z. Erkin, A. Piva, S. Katzenbeisser, et al., Protection and retrieval of encrypted multimedia content: when cryptography meets signal processing, EURASIP J. Inf. Sec. 2008(1) (2007).
[31] R.L. Lagendijk, Z. Erkin, M. Barni, Encrypted signal processing for privacy protection: conveying the utility of homomorphic encryption and multiparty computation, Sig. Process. Magaz. IEEE 30 (1) (2013) 82–105.
[32] W. Lu, A. Varna, M. Wu, Confidentiality-preserving image search: a comparative study between homomorphic encryption and distance-preserving randomization, IEEE Access 2 (2014) 125–141.
[33] Zhihua Xia, Xinhui Wang, Liangao Zhang, Zhan Qin, Xingming Sun, Kui Ren, A privacy-preserving and copy-deterrence content-based image retrieval scheme in cloud computing, IEEE Trans. Inf. Forensics Secur. (2016), http://dx.doi.org/10.1109/TIFS.2016.2590944.
[34] Z. Xu, L. Xiong, Y. Xu, On the provably secure CEW based on orthogonal decomposition, Sig. Process. Image Commun. 29 (5) (2014) 607–617.
[35] J. Jiang, A.J. Armstrong, G.-C. Feng, Direct content access and extraction from JPEG compressed images, Pattern Recogn. 35 (11) (2002) 2511–2519.
[36] Z.M. Lu, S.Z. Li, H. Burkhardt, A content-based image retrieval scheme in JPEG compressed domain, Int. J. Innov. Comput., Inf. Control 2 (4) (2006) 831–839.
[37] J.A. Lay, L. Guan, Image retrieval based on energy histograms of the low frequency DCT coefficients, IEEE International Conference on Acoustics, Speech, and Signal Processing 6 (1999) 3009–3012.
[38] M. Eom, Y. Choe, Fast extraction of edge histogram in DCT domain based on MPEG7, in: International Conference on Enformatika, Systems Sciences and Engineering, 2005.
[39] D. Edmundson, G. Schaefer, Performance comparison of JPEG compressed domain image retrieval techniques, in: IEEE International Conference on Signal Processing, Communication and Computing, 2012, pp. 587–592.
[40] G. Schaefer, D. Edmundson, DC stream based JPEG compressed domain image retrieval, in: International Conference on Active Media Technology, Springer-Verlag, 2012, pp. 318–327.
[41] Y. Wang, M. Shi, S. You, et al., DCT inspired feature transform for image retrieval and reconstruction, IEEE Trans. Image Process. 25 (9) (2016) 4406–4420.
[42] T. Song, H. Li, Local polar DCT features for image description, IEEE Signal Process. Lett. 20 (1) (2013) 59–62.
[43] J.Z. Wang, J. Li, G. Wiederhold, Simplicity: semantics sensitive integrated matching for picture libraries, IEEE Trans. Pattern Anal. Mach. Intell. 25 (9) (2001) 947–963.
[44] Jia Li, James Z. Wang, Automatic linguistic indexing of pictures by a statistical modeling approach, IEEE Trans. Pattern Anal. Mach. Intell. 25 (9) (2003) 1075–1088.
[45] H. Jegou, M. Douze, C. Schmid, Hamming embedding and weak geometric consistency for large scale image search, in: Computer Vision-ECCV, Springer, 2008, pp. 304–317.