BOEW A Content-Based Image Retrieval Scheme Using Bag-of-Encrypted-Words in Cloud Computing

ABSTRACT

Content-based Image Retrieval (CBIR) techniques have been extensively studied with the rapid growth of digital images. Generally, CBIR service is quite expensive in computational and storage resources. Thus, it is a good choice to outsource CBIR

service to the cloud server that is equipped with enormous resources. However, the privacy protection becomes a big problem, as the cloud server cannot be fully trusted. In this paper, we propose an outsourced CBIR scheme based on a novel bag-of-encrypted-words (BOEW) model. The image is encrypted by color value substitution, block permutation, and intra-block pixel permutation. Then, the

local histograms are calculated from the encrypted image blocks by the cloud server. All the local histograms are clustered together, and the cluster centers are used as the encrypted visual words. In this way, the bag-of-encrypted-words (BOEW) model is built to represent each image by a feature vector, i.e., a normalized histogram of the encrypted visual words. The similarity between images can be directly measured by the Manhattan distance between feature vectors on the cloud server side. Experimental results and security analysis on the proposed scheme demonstrate its search accuracy and security.

**EXISTING SYSTEM**

Searchable encryption (SE) enables the clients to store the encrypted data at the cloud, meanwhile supports data search over cipher-text domain [9]. However, many of the existing SE schemes are designed for text documents [10], [11], [12]. Lu *et al*. [13] proposed the first privacy-preserving CBIR scheme over the encrypted image database. The scheme utilized the set of visual words to represent images. The similarity between images was measured by Jaccard distance between the sets of visual words. The min-hash algorithm and order-preserving encryption were employed to protect the visual words.

In another work [14], Lu *et* *al*. investigated three image feature protection techniques including bitplane randomization, random projection, and randomized unary encoding. The bitplane randomization and random unary encoding support the calculation of Hamming distance in the encryption domain. The random projection supports the approximate calculation of L1 dis- tance in the encryption domain. In [15], Lu *et al*. compared the three mentioned methods with the homomorphic encryption and indicated that the homomorphic encryption consumed much more computation and communication resources.

Yuan *et al*. [16] protected the image features using local sensitive hashing and Cuckoo Hashing to support secure similarity search. This method was used to discovery the social connections between image owners. Xia *et al*. [17] proposed a privacy-preserving CBIR scheme based on Scale-Invariant Feature Transform (SIFT) features and Earth Mover’s Distance (EMD). The calculation of the EMD is in fact a linear program problem. The linear transformation was utilized to protect the privacy information during the solution process of EMD problem. Yuan *et al*. [18] designed an encrypted image search scheme based on the secure *k*NN (*k*-nearest neighbors) algorithm and constructed a tree index to improve the search efficiency. In [19], Chen *et al*. proposed a Markov process-based retrieval scheme over encrypted images. The image content was protected by encrypting the Huffman table in JPEG files. The Markov features were directly extracted from the DCT coefficients which were decoded with the encrypted Huffman table.

In [20], [21], Weng *et al*. proposed a framework for privacy-preserving multimedia retrieval. The media features were protected by the robust hashing and partial encryption by image owner. Then encrypted part of hash introduced search ambiguity to server. The similar images were retrieved using the unencrypted part of hash on the server side and refined with the whole plaintext hash on the query user side.

In [22], Xia *et al*. proposed a privacy-preserving CBIR scheme. Four MPEG descriptors were used to represent the images. Secure *k*NN algorithm was employed to protect the image features. Locality-sensitive hash was used to increase the search efficiency. In addition, the authors incorporated an encryption-domain watermarking method to the scheme so as to deter the image users’ illegal distribution. In [23], Zhang *et al*. proposed a secure outsourced CBIR scheme with fine-grained access control. A key-agent was intro- duced to identify which images can be accessed by a user.

**Disadvantages**

* An existing methodology doesn’t implement Privacy-preserving CBIR protocol method.
* The system not implemented Bag-of-Word model Technique.

Proposed System

In this paper, the system proposes an outsourced CBIR scheme where the image content is properly protected. The main contributions are summarized as follows:

1) A BOEW model is proposed for CBIR outsourcing. We propose to encrypt images by blocks and make sure that the secure and useful local features can be directly ex- tracted from the encrypted blocks. *k*-means clustering algorithm is deployed to generate the encrypted visual words. The final feature vectors, also the encrypted ones, are then constructed with the visual words. The similarity between the feature vectors can be directly measured by Euclidean or Manhattan distance. The

proposed BOEW could be a valuable model in encrypted image processing.

2) As a case study, we propose to encrypt image by color value substitution, block permutation, and intra-block pixel permutation. With the specially-designed encryp-

tion method, secure local histograms can be directly extracted from the encrypted images on cloud server side. The index construction can also be finished by cloud server. Compared with the scheme using secure global histogram [3], [4], our method achieves a much better retrieval accuracy.

**Advantages**

* Image owner encrypts the query image, and submits the encrypted image to cloud server as query trapdoor. After searching on the index, the cloud server returns the most similar images to the image owner.
* The information leaked here includes the encrypted query image and the similarity between the images in the database and Searchable encryption (SE) enables the clients to store the encrypted data at the cloud, meanwhile supports data search over cipher-text domain.

**SYSTEM REQUIREMENTS**

➢ **H/W System Configuration:-**

➢ Processor - Pentium –IV

➢ RAM - 4 GB (min)

➢ Hard Disk - 20 GB

➢ Key Board - Standard Windows Keyboard

➢ Mouse - Two or Three Button Mouse

➢ Monitor - SVGA

**Software Requirements:**

* Operating System - Windows XP
* Coding Language - Java/J2EE(JSP,Servlet)
* Front End - J2EE
* Back End - MySQL