IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received October 10, 2020, accepted October 25, 2020, date of publication November 3, 2020, date of current version November 13, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3035563

Privacy-Preserving Content-Based Image Retrieval Using Compressible Encrypted Images

KENTA IIDA, (Student Member, IEEE), AND HITOSHI KIYA, (Fellow, IEEE)
Department of Computer Science, Tokyo Metropolitan University, Tokyo 191-0065, Japan

Corresponding author: Hitoshi Kiya (kiya@tmu.ac.jp)

ABSTRACT In this article, we propose a novel content-based image-retrieval (CBIR) scheme using compressible encrypted images called "encryption-then-compression (EtC) images." The proposed scheme allows us not only to directly retrieve images from visually protected images but to also apply EtC images that can be compressed by using the JPEG standard for the first time. In addition, the sensitive management of secret keys is not required in our framework. The proposed retrieval scheme is carried out on the basis of weighted searching images with MPEG-7-powered localized descriptors (weighted SIMPLE descriptors) combining scalable color descriptor (SCD) or color and edge directivity descriptor (CEDD). Weighted SIMPLE descriptors are extended, and CEDD is also modified to avoid the influence of image encryption. In an experiment, the proposed scheme is demonstrated to have almost no degradation in retrieval performance compared with conventional content-based retrieval methods with plain images under the use of two datasets. In addition, the proposed scheme is shown to outperform conventional privacy-preserving CBIR schemes including state-of-the-art ones in terms of mean average precision (mAP) scores.

INDEX TERMS Content-based image retrieval, encryption-then-compression system, SIMPLE descriptors.

I. INTRODUCTION

With the rapid growth of cloud computing, outsourcing images to cloud storage services and sharing photos have greatly increased. Generally, images are uploaded and stored in a compressed form to reduce the amount of data. In addition, most images include sensitive information, such as personal data and copyright information [1], [2]. However, cloud providers are not trusted in general, so there is the possibility of data leakage and unauthorized use in cloud environments. Therefore, various privacy-preserving image identification, retrieval, and processing schemes have been studied for untrusted cloud environments [3]–[21].

For the above reasons, privacy-preserving image-retrieval methods should satisfy three requirements: 1) protecting visual information on plain images, 2) having a high retrieval performance in the encrypted domain, and 3) being applicable to compressible encrypted images. To satisfy requirement 1), full encryption with provable security, such as RSA and AES, is the most secure option for protecting multimedia data [4]–[7]. In retrieval schemes using this type of encryption, it is required that the descriptors used for the retrieval be extracted from plain images, and encrypted for privacy-preserving by other forms of searchable encryption in general. In contrast, perceptual encryption methods have been proposed that can be directly applied to a number of signal processing algorithms in the encrypted domain [14]–[17], [21]–[30]. Recently, a scheme with descriptors extracted from images encrypted by AES was also proposed [5], but the retrieval performance is not enough compared with that of using plain images. In addition, requirement 3) has never been considered for any content-based encrypted image retrieval schemes. As systems that satisfy both requirements 1) and 3), encryption-then-compression (EtC) systems have been developed [22]–[30] for untrusted cloud environments. In this article, we focus on a block scrambling-based image-encryption method that has been proposed for EtC systems [23]–[30], and encrypted images used in the systems are referred to as "EtC images."

So far, EtC images have never been applied to image retrieval, although image identification schemes have been proposed for detecting EtC images having the same plain

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

images [21] in the encrypted domain. Generally, image retrieval methods are classified into content-based image retrieval (CBIR) and text-based image retrieval (TBIR). CBIR methods extract descriptors from the content information of images, while images have to be manually annotated with keywords in TBIR methods. In this article, we focus on CBIR because it is required not only that images be manually annotated with keywords but also that extracted keywords be protected. For CBIR, various types of image descriptors have been proposed [31]–[45]. In regards to descriptors generated by using deep neural networks [7]–[9], [31], [32], there are schemes using networks fine-tuned for target image datasets and schemes using pre-trained networks such as AlexNet [46] and VGG [47]. However, a large number of images and huge computational costs are required for training a model, so their applications are limited. In contrast, using hand-crafted descriptors allows us to flexibly retrieve images even when we use a small dataset.

Hand-crafted descriptors are classified into two types in general: global image descriptors such as MPEG-7 and GIST image descriptors [33]–[40] and local image descriptors such as SIFT and SURF image descriptors [31], [41]–[45]. To reduce the effects of image encryption on retrieval performance, local image descriptors are chosen in this article. Regarding the type of local descriptor, searching images with MPEG-7-powered localized descriptors (SIMPLE descriptors) has been proposed [41] for improving the retrieval performance by combining global descriptors with a technique that uses local descriptors. In addition, weighted SIMPLE descriptors, which is an extension of SIMPLE descriptors, was also demonstrated to outperform state-of-the-art retrieval schemes with hand-crafted descriptors. Accordingly, the proposed scheme is carried out on the basis of weighted SIMPLE descriptors combining scalable color descriptor (SCD) or color and edge directivity descriptor (CEDD) to satisfy all three requirements.

In this article, we propose a novel content-based image-retrieval scheme that allows us not only to directly use EtC images but also to apply EtC images that can be compressed by using the JPEG standard for the first time. In addition, the sensitive management of secret keys is not required in our framework. In the proposed scheme, weighted SIMPLE descriptors are extended, and CEDD is also modified to avoid the influence of image encryption on retrieval performance. Simulation results show that the proposed scheme has almost the same retrieval performance in terms of mAP scores as conventional content-based retrieval methods with plain images under the use of two datasets.

The rest of the paper is organized as follows. In Section II, the procedure for generating EtC images and a review of related work are presented. Section III describes extended SIMPLE descriptors and modified CEDD, which are used in the proposed privacy-preserving image-retrieval scheme. Image retrieval performances under various conditions are evaluated in Section IV. Concluding remarks are given in Section V.

II. RELATED WORK

A. EtC IMAGE

Security mostly refers to protection from adversarial forces. In this article, image encryption aims to protect visual information that allows us to identify an individual, the time, and the location of a taken photograph. Untrusted cloud providers and unauthorized users are assumed to be adversaries.

We focus on EtC images, which have been proposed for encryption-then-compression (EtC) systems with JPEG compression [22]–[29]. EtC images not only have almost the same compression performance as that of plain images but also enough robustness against various ciphertext-only attacks including jigsaw puzzle solver attacks [24]–[29]. The procedure for generating EtC images is conducted as below (see Figs. 1 and 2) [23].

(a) Divide image I_i with $X \times Y$ pixels into non-over-lapping 16×16 blocks.

(b) Permute randomly $\lfloor \frac{X}{16} \rfloor \times \lfloor \frac{Y}{16} \rfloor$ divided blocks by using a random integer generated by secret key K_1.

(c) Rotate and invert randomly each divided block by using a random integer generated by secret key K_2 (see Fig. 3).

(d) Apply negative-positive transformation to each block by using a random binary integer generated by secret key K_3 to obtain encrypted image I_i^e. In this step, a transformed pixel value in the ith block B_i, p' is computed by

$$\begin{cases} p' = p, & r(i) = 0, \\ p' = 255 - p, & r(i) = 1, \end{cases} \tag{1}$$

where $r(i)$ is a random binary integer generated by K_3 under the probability $P(r(i)) = 0.5$, and p is a pixel value in a plain image with 8 bpp.

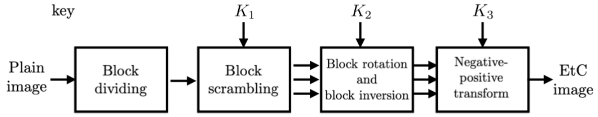


FIGURE 1. Generation of EtC images.

(a) Plain image

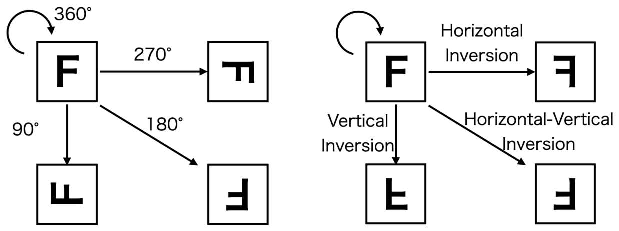
(b) EtC image

FIGURE 2. Example of plain image and encrypted one.

In this article, images encrypted by using these steps are referred to as "EtC images." K_1, K_2, and K_3 are stored as a key set, $\mathbf{K} = [K_1, K_2, K_3]$.

B. CONTENT-BASED IMAGE RETRIEVAL

CBIR schemes are categorized in accordance with the type of extracted descriptor, as shown in Fig. 4. Hand-crafted descriptors, which are generated without using neural networks, are classified into local descriptors and global descriptors in terms of the number of descriptors extracted from an



(a) Block rotation (b) Block inversion

FIGURE 3. Block rotation and inversion. (a) Block rotation, (b) block inversion.

image, where one descriptor is extracted from each image for the type of global descriptors. SIMPLE descriptors and weighted SIMPLE descriptors correspond to local descriptors, and they have variations depending on the method used to select patches and the type of descriptors extracted from patches. In this article, extended SIMPLE descriptors are proposed as an extension of weighted SIMPLE descriptors for privacy-preserving image retrieval. Descriptors extracted from patches are also referred to as patch descriptors.

C. PRIVACY-PRESERVING IMAGE RETRIEVAL

For privacy-preserving CBIR, there are two approaches in terms of how to obtain descriptors as below.

1) GENERATING DESCRIPTORS FROM PLAIN IMAGES

In this approach, descriptors such as SIFT-based [4], SURF-based [6], ORB [8], MPEG-7 [10], [11], and CNN-based [7], [9], [12] ones are calculated by using plain images. Descriptors based on DCT coefficients in JPEG images are also used in the case of extraction from JPEG images [13]. After the descriptors and plain images are encrypted by a data owner, the encrypted descriptors and images are sent to a cloud server. In this approach, data owners are required to extract descriptors and encrypt both the descriptors and plain images by themselves. Moreover, data owners and users have to share a common key in some schemes [7], [9], so their applications are limited due to the difficulty of safely managing keys.

2) GENERATING DESCRIPTORS FROM ENCRYPTED IMAGES

In the second approach, descriptors are directly extracted from encrypted images by a cloud provider as well as for CBIR methods for plain images, after data owners encrypt images and then send them and the owners' information to the cloud provider [5], [14]–[17], where each data owner only encrypts images. In addition, data owners and users do not need to share keys. Thus, we focus on this approach in this article. Bag-of-visual words (BOVW)-based retrieval methods [5], [14], [17] and a retrieval method with support vector machine [15] are used in this approach. However, no conventional methods have ever considered compressing encrypted images. In contrast, the proposed method allows both data owners and users to upload JPEG compressed images to a cloud.

D. WEIGHTED SIMPLE DESCRIPTORS

It is well-known that weighted SIMPLE descriptors outperform SIMPLE descriptors [41]. Thus, the proposed descriptor is an extension of weighted SIMPLE one.

The procedure for extracting weighted SIMPLE descriptors from plain images under the use of the BOVW model is summarized here.

a) Decide the positions and sizes of patches from every image by using a detector such as the SURF detector or random sampling.
b) Extract a patch descriptor from each patch.
c) Generate a codebook with a size of M from the extracted patch descriptors.
d) Calculate SIMPLE descriptors by using the codebook.
e) Obtain weighted SIMPLE descriptors by weighting the SIMPLE descriptors.

When the SURF detector is used in step a), the position and size of each patch are determined from detected feature points

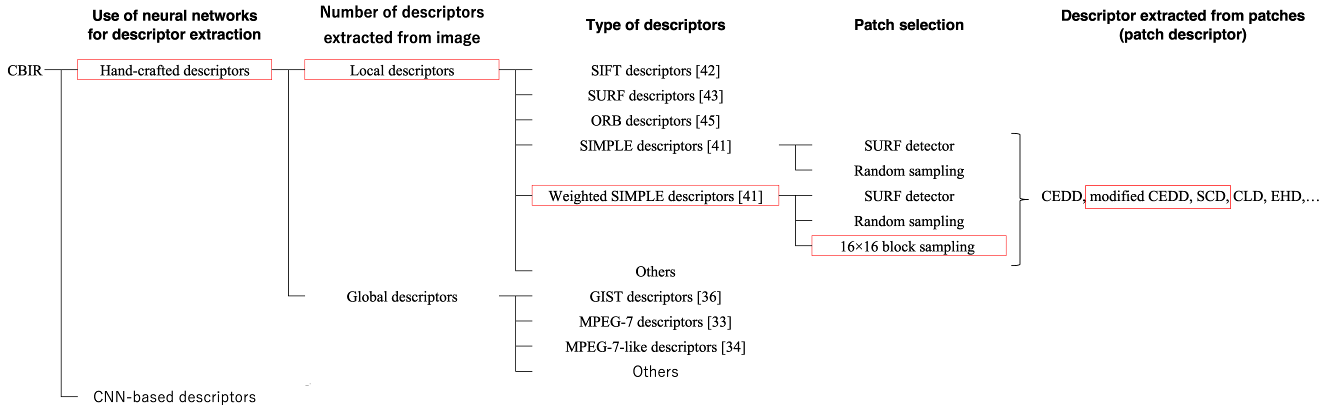


FIGURE 4. Categorization of image descriptors used in conventional content-based image retrieval schemes. Descriptors encapsulated in red boxes are used for proposed privacy-preserving retrieval.

and scales, respectively. In contrast, when using random sampling, they are randomly selected.

In step b), a patch descriptor is selected from MPEG-7 image descriptors or MPEG-7-like descriptors. It was reported in [41] that SIMPLE descriptors combining scalable color descriptor (SCD) [33] and color and edge directivity descriptor (CEDD) [34] have better retrieval performance than color layout descriptor (CLD) [33] and edge histogram descriptor (EHD) [33]. Thus, we use SCD and CEDD as patch descriptors.

A codebook is generated from patch descriptors by using k-means clustering in step c), where the size of the codebook corresponds to the number of classes in k-means clustering, and the center of each class is defined as a visual word in the codebook. When using the codebook, a SIMPLE descriptor is represented as a histogram of the frequencies of these visual words included in an image. Note that different visual words are selected for every generation process for codebooks because the center of each initial class is randomly selected, even if the same images are used to generate codebooks.

In step e), all descriptors are weighted in order to obtain weighted SIMPLE descriptors. When N SIMPLE descriptors are generated, the mth component of the nth weighted SIMPLE descriptor $v_n(m)$, $0 \leq m < M$, $0 \leq n < N$, is calculated as below in this article.

$$v_n(m) = (1 + log(tf_{(m,n)})) \times log \frac{N}{df_{(m)}}, \tag{2}$$

where $tf(m, n)$ represents the frequency of the mth visual word in the nth descriptor, and $df(m)$ denotes the number of extended SIMPLE descriptors containing the mth visual word in the N SIMPLE descriptors. After that, l_2 normalization is applied to every SIMPLE descriptor.

III. PROPOSED SCHEME

Our novel content-based image-retrieval scheme using EtC images is proposed here. For carrying out image retrieval in the encrypted domain, weighted SIMPLE descriptors are extended, and SCD or the modified CEDD is used as a global descriptor of each patch.

A. SYSTEM MODEL

The framework used in the proposed scheme is shown in Fig. 5. Each operation in Fig. 5 is explained as follows.

1) IMAGE DESCRIPTOR GENERATION PROCESS

a-1) A data owner encrypts plain image I_i with secret key set $\mathbf{K}_i$; then, the EtC image is compressed with JPEG compression and uploaded to a third party.

a-2) The third party generates a codebook from the uploaded EtC images after decompression, and then image descriptors are calculated from EtC images by using the codebook. After that, the codebook and the image descriptors are stored in a database.

2) RETRIEVAL PROCESS

b-1) A user sends query image Q_U^e encrypted by using key set $\mathbf{K}_U$ to the third party, where $\mathbf{K}_U$ can be prepared by the user.

b-2) The third party calculates an image descriptor from Q_U^e by using the stored codebook and the stored image descriptors.

b-3) The third party retrieves EtC images in the database similar to Q_U^e by using the image descriptor in the encrypted domain. The retrieved images and the owner's information are returned to the user.

b-4) The user requests the data owner to send key set $\mathbf{K}_i$ for decrypting the EtC images received from the third party.

In this framework, the third party not only has no visual information on images but also no secret keys. Moreover, each image can be encrypted by using different keys.

The proposed scheme is carried out on the basis of hand-crafted descriptors, so a codebook can be generated

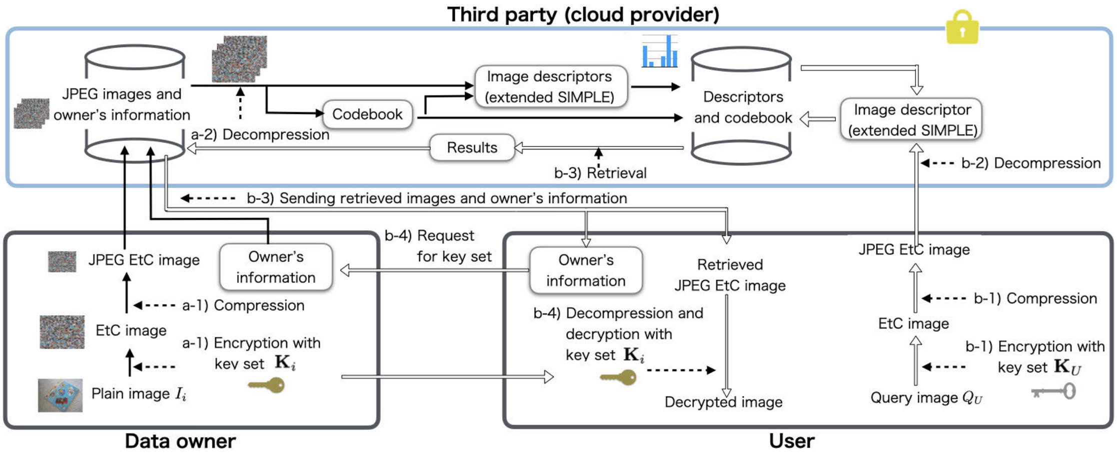


FIGURE 5. System model of proposed scheme, where image descriptors correspond to extended SIMPLE descriptors.

flexibly even when we use a small dataset. In contrast, a large number of images are required to train a model when DNNs are used.

JPEG compression is a lossy compression method, so retrieved images contain some distortions due to the influence of image compression. By slightly modifying the proposed framework, users can restore original plaintext images from the received encrypted images without any degradation of the image quality. It was confirmed that EtC images can be applied to lossless compression methods such as JPEG LS and JPEG 2000 as well as JPEG in [30]. Thus, users become able to restore original plaintext images when data owners upload EtC images in a lossless compression form.

B. THREAT MODEL

In the proposed system, there are three roles: image owner, third party and user, where the third party is not trusted. The third party has image owners' information and EtC images uploaded from image owners and uses, and moreover knows the encryption algorithm for generating EtC images. The proposed system is designed not only to achieve a high retrieval performance but also to protect the visual information of plain images against attacks by the third party.

The third party might try to restore visual information of plain images from the EtC ones. Therefore, the security of the system against such attacks is needed to be evaluated. Generally, cryptanalysis methods are classified into four attacks: ciphertext-only attack (COA), known-plaintext attack (KPA), chosen-plaintext attack (CPA) and chosen-ciphertext attack (CCA). We focus on robustness against COA because the proposed scheme is not an asymmetric cipher and all images can be encrypted by using independent keys. Although the security of EtC images was already analyzed as in [26]–[29], new attack methods for restoring visual information have been just recently proposed for privacy-preserving DNNs [48], [49]. In this article, the proposed scheme will be demonstrated to be robust enough even when the latest attacks are applied.

C. PROPOSED IMAGE-RETRIEVAL SCHEME

In the proposed scheme, there are three encryption operations for generating EtC images: block scrambling, block rotation and inversion, and negative-positive transformation as shown in Fig. 1. To maintain the retrieval performance that can be achieved when using plain images, an extension of weighted SIMPLE descriptors, referred to as extended SIMPLE descriptors, is first proposed to avoid the influence of block scrambling. Next, to avoid the influence of block rotation and inversion, the use of SCD or modified CEDD as patch descriptors is discussed. In addition, the influence of a negative-positive transform is shown to be able to be reduced by properly selecting the codebook size.

1) EXTENDED SIMPLE DESCRIPTORS

Extended SIMPLE descriptors are generated from EtC images as below.

a) Divide each EtC image into non-overlapping 16 × 16-blocks and use each 16 × 16-block as a patch, where 16 × 16 corresponds to the block size of EtC images.
b) Extract a patch descriptor from each patch.
c) Generate a codebook with a size of M from the extracted patch descriptors.
d) Calculate SIMPLE descriptors with a block size of 16 × 16 by using the codebook.
e) Obtain extended SIMPLE descriptors by weighting the SIMPLE descriptors.

Note that the difference between extend SIMPLE descriptors and weighted SIMPLE ones in Sec. II-D is step a). In step a), the method to select patches is modified to avoid the influence of block scrambling. In extended SIMPLE descriptors, considering block rotation and inversion, SCD or modified CEDD is also selected as a patch descriptor in step b). To avoid the effect of negative-positive transform, the selection of the proper codebook size is needed in step c). In the following sections, the selection of patches, patch descriptors and the codebook size are explained in more detail.

2) SELECTION OF PATCHES

When random sampling and SURF detectors are applied to EtC images, selected patches include boundaries caused by the block scrambling operation as shown in Fig. 6 (a). By using these detectors to select patches, the retrieval performance for EtC images heavily degrades in general, compared with the retrieval performances when using plain images. To avoid the effect of block scrambling, 16 × 16-blocks are selected as patches in step a) of the procedure mentioned in Sec. III-C1 (see Fig. 6 (b)). Each patch corresponds to a block of the block-based encryption, so no patch includes boundaries caused by the block scrambling operation. In addition, since modified SIMPLE descriptors are calculated by using a histogram of visual words contained in an image, block permutation in the block scrambling operation does not have any influence on the descriptors.

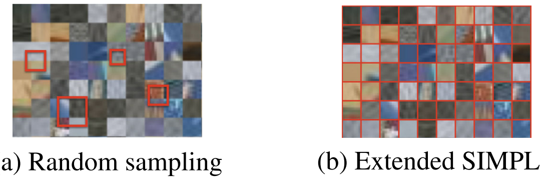

(a) Random sampling
(b) Extended SIMPLE

FIGURE 6. Examples of selected patches.

3) SELECTION OF PATCH DESCRIPTORS

Because of their good retrieval performance as in [41], we focus on SCD and CEDD in this article. SCD is represented as a vector of coefficients obtained by applying a Haar transform to a color histogram of a patch in the HSV color space, where the number of coefficients selected is between 16 and 256. Block rotation and block inversion operations for generating EtC images do not affect the color histogram in a patch, so SCD has no influence when using these operations.

In contrast, CEDD is calculated by using both the color information and edge information of images [34]. Therefore, CEDD should be modified to avoid the influence of image encryption if we want to use CEDD. To generate CEDD from a patch with a size of 16 × 16, the patch is divided into 2 × 2-blocks, and a temporary CEDD, $CEDD_{block} = [c_0, c_1, \cdots c_{143}]$, is calculated from each 2 × 2-block, where c_i represents a value of the ith component in $CEDD_{block}$ and has a value of zero initially. Five linear-phase digital filters with a size of 2 × 2 are applied to pixel values in each 2 × 2-block for classifying edges into six types: no edge and no directional, vertical, horizontal, 45-degree, and 135-degree edges, as shown in Table 1. A 24-bin histogram including color information, called "fuzzy linking histogram h_{color}," is generated by using the average value of every 2×2-block in the HSV color space, and the initial components of $CEDD_{block}$ are then replaced with the 24-bin histogram in accordance with the selected edge type (see Table 1). An average vector is calculated from all $CEDD_{block}$ vectors in each patch, and the average vector is then quantized to obtain the CEDD of the patch.

TABLE 1. Relation between components and edge types in CEDD.

Edge type	Corresponding components
No edge	$c_0, \cdots, c_{23}$
Non directional edge	$c_{24}, \cdots, c_{47}$
Vertical	$c_{48}, \cdots, c_{71}$
Horizontal	$c_{72}, \cdots, c_{95}$
45-degree	$c_{96}, \cdots, c_{119}$
135-degree	$c_{120}, \cdots, c_{143}$

CEDD is not robust against block rotation and inversion in principle because components corresponding to vertical and 45-degree edges are swapped for ones corresponding to horizontal and 135-degree edges, respectively, due to the block rotation operation. Thus, we propose modifying CEDD as shown in Table 2. In the modified CEDD, which consists of the information on edge types in Table 2, when a patch has a vertical edge, a horizontal edge, or both edges, components from the 48th to the 71st and from the 72nd to 95th are used for a histogram with color information. Similarly, components corresponding to 45-degree and 135-degree edges are not classified.

TABLE 2. Relation between components and edge types in modified CEDD.

Edge type	Corresponding components
No edge	$c_0, \cdots, c_{23}$
Non directional edge	$c_{24}, \cdots, c_{47}$
Vertical or horizontal	$c_{48}, \cdots, c_{71}$ and $c_{72}, \cdots, c_{95}$
45-degree or 135-degree	$c_{96}, \cdots, c_{119}$ and $c_{120}, \cdots, c_{143}$

In addition to the above modification, since edge types are determined on the basis of absolute values obtained by using the five filters, and, moreover, the filters are linear-phase ones with a symmetric weight, the modified CEDD is also robust against the block inversion operation used for image encryption, as shown in Fig. 3. Accordingly, when SCD or modified CEDD are used as patch descriptors, the influence of block rotation and block inversion can be avoided. In other words, the retrieval performance of using images encrypted by the block rotation and the block inversion operations is the same as that of using plain images under SCD or modified CEDD.

4) SELECTION OF CODEBOOK SIZE

The codebook size M has a trade-off relation between the processing time for creating codebooks and the retrieval performance because visual words in a codebook are determined by using a k-means classifier. In addition, the selection of a proper codebook size enables us to reduce the influence of the negative-positive transform.

Pixel values in each 16 × 16-block are randomly mapped in accordance with Eq. (1), where the probability is $P(r(i)) = 0.5$. Under this condition, both visual words for negative-transformed patches and visual words for positive-transformed patches are created in a codebook. Thus, when a codebook size $2M$ is chosen for EtC images, the retrieval performance is expected to be almost the same as that of using size M for images without negative transformation. In addition, it is well-known that the retrieval performance is saturated under a larger codebook size than a certain value. In that case, the performance for EtC images is almost the same as that for images without negative transformation, even when a common codebook size M is used. Therefore, the negative-positive transform has almost no influence on the retrieval performance when a proper codebook size is selected, as demonstrated later.

5) RETRIEVAL PROCESS FOR QUERY IMAGE

By using the extended SIMPLE descriptors and the codebook stored in the database, the retrieval for a query EtC image is performed as below (see Fig. 5).

a) Obtain patches from the EtC image, and extract a patch descriptor from each patch, where the type of the patch descriptor such as SCD is the same as that of the patch descriptors stored in the database.
b) Calculate an extended SIMPLE descriptor from the patch descriptors by using the codebook and the extended SIMPLE descriptors stored in the database.
c) Compute the l_2 distance between every extended SIMPLE descriptor stored in the database and the extended SIMPLE descriptor of the query, and then choose similar images from the database.
d) The retrieved images and the owner's information are returned to the user.
e) The user requests the data owner to send key set $\mathbf{K_i}$ for decrypting the EtC images received from the third party.

D. SECURITY ANALYSIS

The third party is assumed to be untrusted in this framework, so the third party may try to restore visual information of plain images from EtC images. We focus on robustness against ciphertext-only attack (COA). Following attacks are known as COAs for EtC images.

(a) Brute-force attack
(b) Jigsaw-solver attack
(c) Attack using edge information of images
(d) Attack using an inverse transformation model

In previous works [26]–[29], EtC images were demonstrated to be robust enough against attacks (a) and (b). In contrast, attacks (c) and (d) have been recently proposed for privacy-preserving DNNs [48], [49]. EtC images will be confirmed to be also robust against these attacks in experiments.

In the proposed scheme, codebooks are directly generated from EtC images stored in the cloud without decrypting encrypted images. Therefore, the codebooks are safely generated unless the cloud provider cannot restore the visual information of plain images from the EtC images. The codebook contains only representative patch descriptors directly extracted from blocks of EtC images. Thus, the codebook does not allow the third party to reveal the information of images. In addition, in the previous work regarding SIMPLE descriptors [41], it was confirmed that the retrieval performance cannot be greatly improved by reusing descriptors. Accordingly, the reuse of descriptors generated from the codebook is not effective in unauthorized image retrieval.

E. PROPERTIES OF PROPOSED SCHEME

The properties of the proposed scheme are summarized here.

- Tolerance against image encryption
  The proposed scheme allows us to use visually protected images, called "EtC images," for privacy-preserving retrieval.
- Availability for used with different encryption keys
  The proposed scheme enables us to avoid the influence of block scrambling, block rotation and block inversion, and the negative-positive transform, even when different key sets are used for image encryption. This property makes sensitive key management unnecessary.
- Availability for used with JPEG compressed images
  EtC images can be compressed by using the JPEG standard, so the proposed scheme allows us to use not only visually protected images but also JPEG compressed ones.

IV. EXPERIMENT

A. EXPERIMENT SETUP

In this experiment, the performance of the proposed image retrieval was evaluated by using LIRE [50], which is an open-source Java library for content-based image retrieval and supports various image descriptors. For comparison with the proposed scheme, the retrieval was conducted by using the descriptors shown in Tab. 3, where "W-SIMPLE rnd+SCD" represents weighted SIMPLE descriptors combining random sampling with SCD. In the case of retrieval with SURF, the BOVW model and weighting term frequencies were used.

The performances of these descriptors were evaluated in terms of mean average precision (mAP) scores [9], [41]. To obtain mAP scores, the average precision values were calculated for all query images. When the number of ground truth images is G, the average precision of the qth query image AP_q is calculated as,

$$AP_q = \frac{1}{G}\sum_{n=1}^{N}\frac{TP@n}{n} \times f(n), \tag{3}$$

where N is the number of images stored in a database, and $TP@n$ represents the number of true positive matches at the nth ranking and $f(n) = 1$ if the nth image is a ground truth one. Otherwise, $f(n) = 0$ if the nth image is not. After calculating the average precision values for all Q query images, mAP scores are calculated as

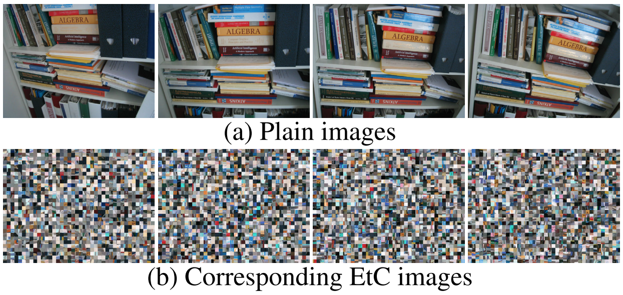
$$mAP = \frac{\sum_{q=0}^{Q-1} AP_q}{Q}. \tag{4}$$

B. EXPERIMENT RESULTS FOR UKbench

In this article, the retrieval performance of the proposed scheme was mainly evaluated by using the UKbench dataset [51]. This dataset consists of 10,200 images with a size of 640 × 480, and the images are classified into 2,250 groups. Each group has four images containing a single object captured from different viewpoints and lighting conditions. 1,000 images from No. 00000 to No. 00999 were chosen from the data set (see Fig. 7) in this experiment.

TABLE 3. Descriptors used in this experiment.

Abbreviation	Type of descriptors	Patch selection	Patch descriptor
E-SIMPLE+mCEDD	Weighted SIMPLE	16 × 16-block sampling	modified CEDD
E-SIMPLE+SCD	Weighted SIMPLE	16 × 16-block sampling	SCD
W-SIMPLE rnd+CEDD	Weighted SIMPLE	Random sampling	CEDD
W-SIMPLE rnd+SCD	Weighted SIMPLE	Random sampling	SCD
W-SIMPLE srf+CEDD	Weighted SIMPLE	SURF detector	CEDD
W-SIMPLE srf+SCD	Weighted SIMPLE	SURF detector	SCD
SURF	SURF	-	-
CEDD	MPEG-7-like	-	-
SCD	MPEG-7	-	-
CLD	MPEG-7	-	-
EHD	MPEG-7	-	-


(a) Plain images

(b) Corresponding EtC images

FIGURE 7. Image examples in group (UKbench dataset).

$N = 1,000$ images were uploaded to a third party by a data owner, where the number of groups was 250, and each group consisted of 4 images, i.e., $G = 4$.

1) EFFECT OF IMAGE ENCRYPTION WITHOUT NEGATIVE-POSITIVE TRANSFORM

To confirm the effect of image encryption without negative-positive transformation, an experiment was conducted by using E-SIMPLE+SCD. In Fig. 8, performances for plain images and EtC images generated without applying the negative-positive transform are shown, where "EtC images without NP" in the figure indicates that step (d) in Sec. II-A was not carried out. The retrieval performances for EtC images without the negative transform had the almost same scores as those for plain images, even when different secret keys were assigned to all images ($K_i \neq K_U$ in Fig. 8). This is because the proposed scheme is robust against block scrambling, block rotation, and block inversion.

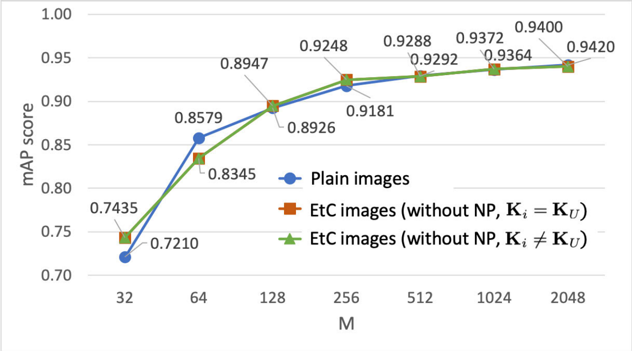


FIGURE 8. Retrieval performance of E-SIMPLE+SCD for EtC images generated without applying negative-positive transform (UKbench dataset).

2) EFFECT OF IMAGE ENCRYPTION WITH NEGATIVE-POSITIVE TRANSFORM

Next, the retrieval performances for EtC images generated with all steps mentioned in Sec. II-A were evaluated. Figure 9 shows that mAP scores for EtC images under a value of M were almost the same as those for plain images under a value of $\frac{1}{2}M$ due to the reason given in Sec. III-C4. Thus, by choosing a proper codebook size, the use of extended SIMPLE descriptors allows us to achieve almost the same retrieval accuracy as that for plain images, even when the negative transform is applied.

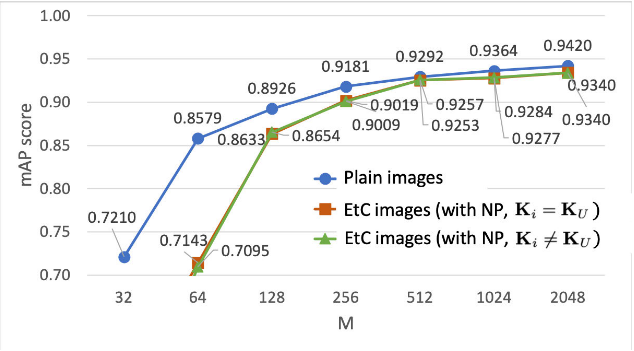


FIGURE 9. Retrieval performance of E-SIMPLE+SCD (UKbench dataset).

Similarly, the performances of E-SIMPLE+mCEDD are shown in Figs. 10 and 11. It was confirmed that the performance of E-SIMPLE+mCEDD had the same trend as that of E-SIMPLE+SCD. Therefore, the proposed CBIR scheme can have almost the same retrieval performance as that for plain images, even when images are encrypted by using different keys.

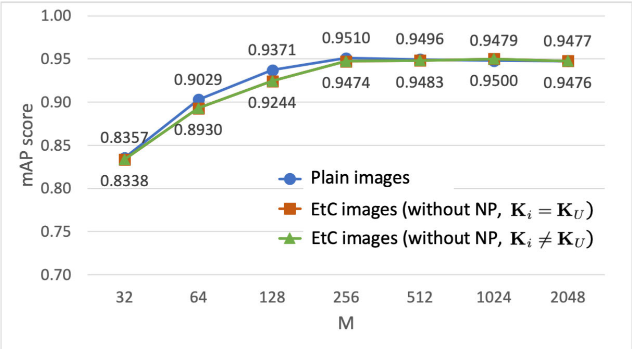


FIGURE 10. Retrieval performance of E-SIMPLE+mCEDD for EtC images without applying negative-positive transform (UKbench dataset).

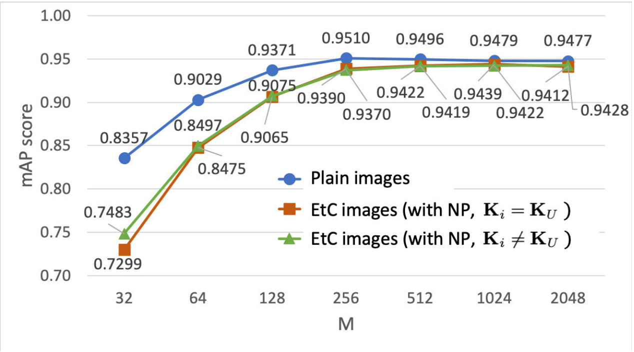


FIGURE 11. Retrieval performance of E-SIMPLE+mCEDD for EtC images (UKbench dataset).

3) COMPRESSIBILITY WITH JPEG COMPRESSION

Here, EtC images were evaluated in terms of compressibility with JPEG compression and the influence of compressed EtC images on the retrieval performance. For JPEG compression, a quality factor in the range of $1 \leq QF \leq 100$ was used to control image quality, in which a larger quality factor results in a higher-quality image. In this experiment, the values of $QF = 95$, 85, and 75 were used for both I_i and Q_U as shown in Fig. 5. After 1,000 EtC images were generated from plain images in UKbench, the EtC images were compressed with JPEG compression.

Figure 12 shows the mAP scores obtained using compressed EtC images under $\mathbf{K}_i \neq \mathbf{K}_U$, in which E-SIMPLE+SCD was carried out. The proposed scheme was confirmed to have almost no influence of image encryption on the retrieval accuracy, even when EtC images were compressed. In Fig. 13, the compression performance for EtC images was also compared with that for plain images, where the average PSNR (peak signal to noise ratio) of 1,000 images was calculated after decompressing and de-encrypting the compressed EtC images. From Fig. 13, EtC images were demonstrated to have almost the same compression performance as that of plain images.

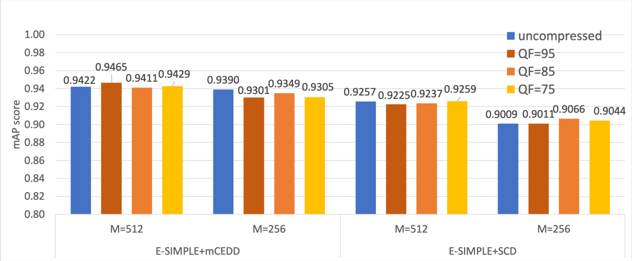


FIGURE 12. Retrieval performance of proposed scheme under compressed EtC images with $K_i \neq K_U$ (UKbench dataset).

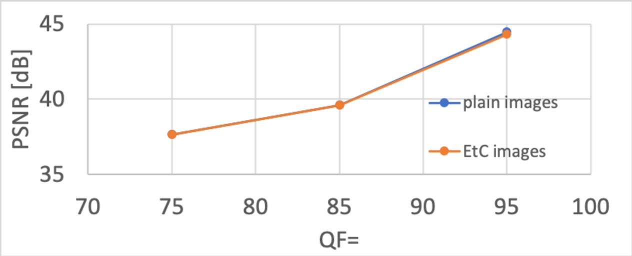


FIGURE 13. Average PSNR values of plain images and EtC images (UKbench dataset).

4) COMPARISON WITH CONVENTIONAL METHODS

To confirm whether the proposed scheme has sufficient retrieval performance, the scheme was compared with conventional CBIR methods using plain images. Retrieval results are shown in Tab. 4. It was confirmed that the mAP scores of the proposed scheme, i.e., E-SIMPLE, were almost the same as those of W-SIMPLE with plain images. In addition, the other CBIR methods had lower mAP values than W-SIMPLE.

Moreover, the proposed scheme was compared with W-SIMPLE using EtC images as shown in Figs. 14 and 15.

TABLE 4. Comparison with conventional CBIR methods using plain images (UKbench dataset).

Descriptor		$M =$	mAP score
EHD [33] (plain)		-	0.6040
CLD [33] (plain)		-	0.7815
SCD [33] (plain)		-	0.9179
CEDD [34] (plain)		-	0.8806
SURF [43] (plain)		256	0.8304
		512	0.8355
W-SIMPLE rnd+CEDD (plain)		256	0.9300
		512	0.9481
W-SIMPLE rnd+SCD (plain)		256	0.9110
		512	0.9262
W-SIMPLE srf+CEDD (plain)		256	0.9000
		512	0.9222
W-SIMPLE srf+SCD (plain)		256	0.8949
		512	0.9109
Proposed	E-SIMPLE+mCEDD (encrypted, $\mathbf{K}_i = \mathbf{K}_U$)	256	0.9370
		512	0.9419
	E-SIMPLE+SCD (encrypted, $\mathbf{K}_i = \mathbf{K}_U$)	256	0.9019
		512	0.9253

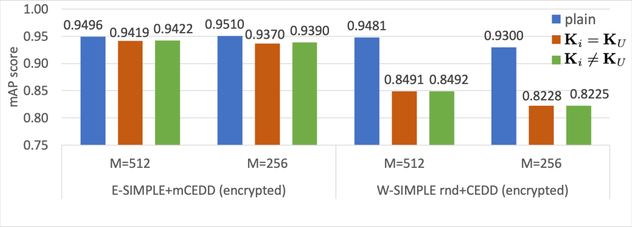


FIGURE 14. Comparison with performance of W-SIMPLE rnd+CEDD under EtC images (UKbench dataset).

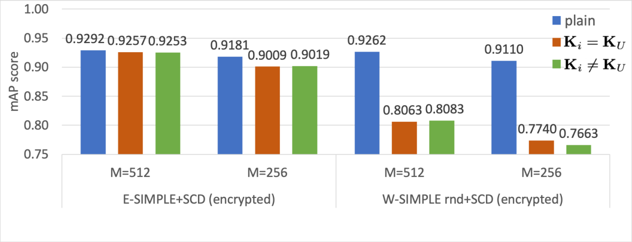


FIGURE 15. Comparison with performance of W-SIMPLE rnd+SCD under EtC images (UKbench dataset).

From the figures, the mAP scores of both W-SIMPLE rnd+CEDD and W-SIMPLE rnd+SCD were heavily degraded due to the effect of image encryption. In contrast, the proposed scheme was able to maintain high scores even under the use of encrypted images.

5) ROBUSTNESS AGAINST ATTACKS

As mentioned in Sec. III-D, we performed an experiment to confirm whether EtC images are robust against two state-of the-art attacks: attack using edge information of images [48], and attack using an inverse transformation model [49]. For the evaluation, PSNR and structural similarity (SSIM) values were measured between plain images and restored ones. In this experiment, 250 EtC images generated from the first image of each group in UKbench were used for evaluating the robustness. For the attack using an inverse transformation model [49], an inverse transformation model was trained with U-Net [52] as shown in Fig. 16, where 750 plain images in UKbench and the corresponding 750 EtC images generated with different keys were prepared as a pair, and other parameters for training a model were the same as those used in the previous work [49]. Note that the first images of the first 250 groups in UKbench dataset were used for evaluating the robustness, and other three images of each group were used for training a model respectively.

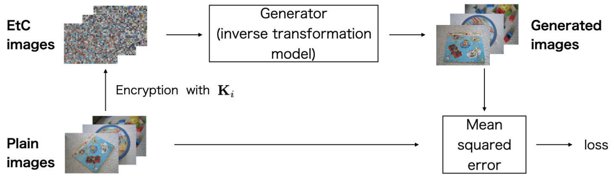

EtC images
Generator (inverse transformation model)
Generated images
Encryption with $\mathbf{K}_i$
Plain images
Mean squared error
loss

FIGURE 16. Training of inverse transformation model [49].

Table 5 shows average PSNR and SSIM values over 250 images and the maximum values in 250 images for each attack. Because of the low values, it was confirmed that visual information of plain images could not be restored from EtC images even when the state-of-the-art attacks were applied. Although some restored images had high values, they also had almost no visual information on plain images yet (see Fig. 17). We also confirmed that all restored images had no visual information on plain images. Accordingly, EtC images are still robust against these attacks.

TABLE 5. PSNR and SSIM values of restored images.

Attack	PSNR values[dB]		SSIM values	
	Average	Max	Average	Max
Edge information [48]	11.45	23.46	0.3249	0.6798
Inverse network [49]	11.18	15.10	0.3909	0.6574
EtC images	10.51	19.30	0.2947	0.6306

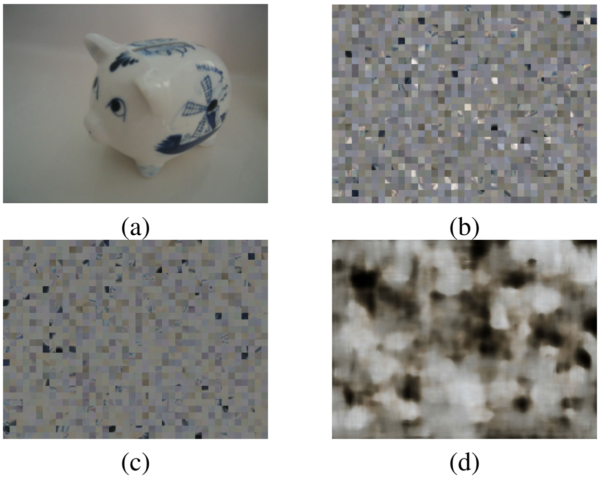


FIGURE 17. Example of restored images having high PSNR and SSIM values. (a) plain image, (b) EtC image (PSNR=17.62[dB], SSIM=0.5520), (c) image restored by attack using edge information [48] (PSNR=20.24[dB], SSIM=0.6675), (d) image restored by attack using inverse transformation model [49] (PSNR=14.11[dB], SSIM=0.6574).

6) EFFICIENCY OF PROPOSED SCHEME

The efficiency of the proposed scheme was evaluated in terms of processing time for image retrieval under the same condition as the experiments in Sec. IV-B2. The processing time was measured on an Ubuntu 18.04 LTS system with Intel(R) Xeon(R) W-2123 CPU 3.60 GHz and 64 GB memory.

Figure 21 (a) shows the processing time for generating a codebook and 1,000 E-SIMPLE descriptors. From the figure, the processing time for EtC images was almost the same as one for plain images under all codebook sizes. Similarly, the processing time for searching images similar to a query image had the same trend, as shown in Fig. 21 (b). Therefore, the proposed scheme enables us to avoid the influence of image encryption in terms of both the retrieval accuracy and the efficiency.

C. EXPERIMENT WITH INRIA HOLIDAYS DATASET

In addition to the UKbench dataset, the INRIA Holidays dataset [53] was also used for evaluating the performance of the proposed scheme. This dataset contains 1,491 images, which are classified into 500 groups, and each group has at least 2 images ($2 \leq G \leq 13$) (see Fig. 18). In this experiment, $N = 1,491$ images were stored in the third party, and $Q = 500$ images were used as query ones.

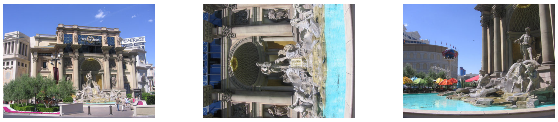


FIGURE 18. Image examples in group (INRIA Holidays dataset).

1) EFFECTS OF IMAGE ENCRYPTION

It was confirmed from Figs. 19 and 20 that extended SIMPLE descriptors achieved almost the same retrieval accuracy as that for plain images when a proper codebook size was chosen, as well as for the UKbench dataset. The sensitive management of secret keys was confirmed to not be required from the results with $\mathbf{K}_i = \mathbf{K}_U$. In addition, as shown in

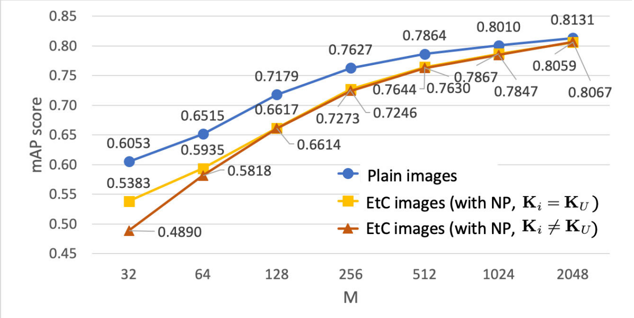


FIGURE 19. Retrieval performances of E-SIMPLE+SCD (INRIA Holidays dataset).

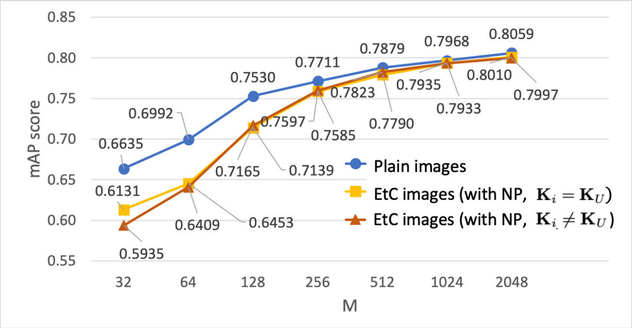
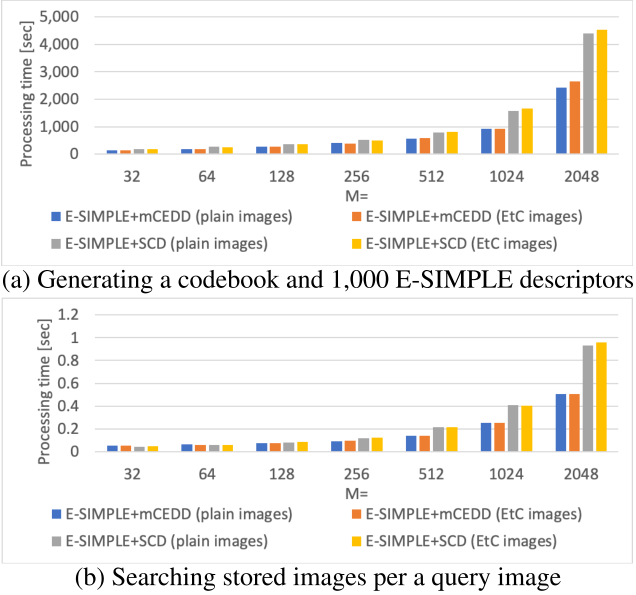


FIGURE 20. Retrieval performances of E-SIMPLE+mCEDD (INRIA Holidays dataset).



(a) Generating a codebook and 1,000 E-SIMPLE descriptors

(b) Searching stored images per a query image

FIGURE 21. Processing time (UKbench dataset).

Tab. 7, extended SIMPLE descriptors had higher mAP scores than those of conventional CBIR methods using plain images. These trends are similar to those for the UKbench dataset.

2) COMPARISON WITH CONVENTIONAL PRIVACY-PRESERVING CBIR SCHEMES

Finally, we compared the proposed CBIR scheme with conventional privacy-preserving CBIR schemes including state-of-the art ones under the use of the INRIA Holidays dataset. From Tab. 7, the proposed scheme was demonstrated not only to be applied to compressive encrypted images for the first time but also to achieve a higher retrieval accuracy than the conventional ones.

D. CHANGE IN FILE SIZE DUE TO IMAGE ENCRYPTION

In this experiment, the influence of image encryption was evaluated in terms of a change in the average size of files. Corel image database [54] was used for this evaluation, where the database consists of 1,000 JPEG images with a size of 384 × 256 or 256 × 384. EtC images were compared with conventional JPEG-based schemes [15], [16], [18], by using the file sizes of JPEG images and the encrypted JPEG images under the use of the same coding parameters.

TABLE 6. Comparison with conventional CBIR methods (INRIA Holidays dataset).

Descriptor		$M =$	mAP score
EHD [33] (plain)		-	0.5497
CLD [33] (plain)		-	0.6476
SCD [33] (plain)		-	0.7524
CEDD [34] (plain)		-	0.7247
SURF [43] (plain)		256	0.6858
		512	0.6986
W-SIMPLE rnd+CEDD (plain)		256	0.7890
		512	0.7987
W-SIMPLE rnd+SCD (plain)		256	0.7716
		512	0.7918
W-SIMPLE srf+CEDD (plain)		256	0.7316
		512	0.7449
W-SIMPLE srf+SCD (plain)		256	0.7156
		512	0.7338
Proposed	E-SIMPLE+mCEDD (encrypted, $\mathbf{K}_i = \mathbf{K}_U$)	256	0.7585
		512	0.7790
	E-SIMPLE+SCD (encrypted, $\mathbf{K}_i = \mathbf{K}_U$)	256	0.7273
		512	0.7644

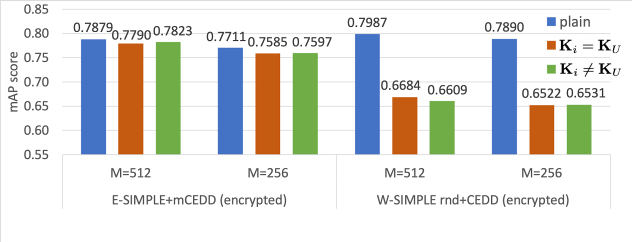


FIGURE 22. Comparison with performance of W-SIMPLE rnd+CEDD under EtC images (INRIA Holidays dataset).

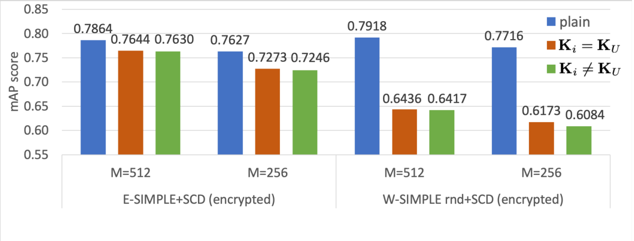


FIGURE 23. Comparison with performance of W-SIMPLE rnd+SCD under EtC images (INRIA Holidays dataset).

TABLE 7. Comparison with conventional privacy-preserving CBIR schemes using descriptors generated form encrypted images (INRIA Holidays dataset).

Scheme		mAP score
Proposed	E-SIMPLE+mCEDD	0.7790
	E-SIMPLE+SCD	0.7644
BOEW [14]		0.6424
IES [17](reported in [14])		0.5456
AES-based [5]		0.2872
Markov based [15](reported in [5])		0.54

TABLE 8. Change in average size of JPEG files due to image encryption, where difference = (average file size of encrypted JPEG images) - (average file size of unencrypted JPEG images).

Scheme	Difference [byte]
Proposed	671.96
[15]	8.61
[16]	-72
[18]	-0.88

Figure 8 shows changes in the average size of JPEG images. The file size of EtC images (proposed) was slightly increased (increased by about 7.91% compared to non-encrypted JPEG images). In contrast, changes in the file sizes for conventional schemes were less than that of EtC images. Although the proposed scheme increases the file size of encrypted images, it allows us to apply lossless compression methods to EtC images. Moreover, each EtC image is not needed to be compressed with the same coding parameters, such as a quantization parameter.

V. CONCLUSION

A novel CBIR scheme using EtC images was proposed for privacy-preserving image retrieval. By using the proposed scheme, we can directly retrieve visually protected images, and the visually protected images, called "EtC images," can be compressed with JPEG compression. We modified both weighted SIMPLE descriptors and CEDD so that the retrieval scheme is almost not influenced by image encryption. As a result, the proposed scheme enables us to retrieve compressible encrypted images with high accuracy for the first time, even when images encrypted by using different keys are used. In an experiment, the proposed scheme was demonstrated to have almost no degradation in retrieval performance under the use of two datasets. In addition, the proposed scheme was shown to outperform conventional privacy-preserving CBIR schemes including state-of-the-art ones in terms of mAP scores.

REFERENCES

[1] C. T. Huang, L. H. Z. Qin, H. Yuan, L. Zhou, V. Varadharajan, and C.-C. J. Kuo, "Survey on securing data storage in the cloud," *APSIPA Trans. Signal Inf. Process.*, vol. 3, p. e7, May 2014.

[2] R. L. Lagendijk 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.

[3] V. Itier, P. Puteaux, and W. Puech, "Recompression of JPEG crypto-compressed images without a key," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 3, pp. 646–660, Mar. 2020.

[4] Y. Xu, X. Zhao, and J. Gong, "A large-scale secure image retrieval method in cloud environment," *IEEE Access*, vol. 7, pp. 160082–160090, 2019.

[5] H. Wang, Z. Xia, J. Fei, and F. Xiao, "An AES-based secure image retrieval scheme using random mapping and BOW in cloud computing," *IEEE Access*, vol. 8, pp. 61138–61147, 2020.

[6] J. Qin, H. Li, X. Xiang, Y. Tan, 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.

[7] Z. Huang, M. Zhang, and Y. Zhang, "Toward efficient encrypted image retrieval in cloud environment," *IEEE Access*, vol. 7, pp. 174541–174550, 2019.

[8] Z. Zhang, F. Zhou, S. Qin, Q. Jia, and Z. Xu, "Privacy-preserving image retrieval and sharing in social multimedia applications," *IEEE Access*, vol. 8, pp. 66828–66838, 2020.

[9] X. Li, Q. Xue, and M. C. Chuah, "CASHEIRS: Cloud assisted scalable hierarchical encrypted based image retrieval system," in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, May 2017, pp. 1–9.

[10] Z. Xia, N. N. Xiong, A. V. Vasilakos, and X. Sun, "EPCBIR: An efficient and privacy-preserving content-based image retrieval scheme in cloud computing," *Inf. Sci.*, vol. 387, pp. 195–204, May 2017.

[11] Z. Xia, X. Wang, L. Zhang, Z. Qin, X. Sun, and K. Ren, "A privacy-preserving and copy-deterrence content-based image retrieval scheme in cloud computing," *IEEE Trans. Inf. Forensics Security*, vol. 11, no. 11, pp. 2594–2608, Nov. 2016.

[12] C. Zhang, L. Zhu, S. Zhang, and W. Yu, "TDHPPIR: An efficient deep hashing based privacy-preserving image retrieval method," *Neurocomputing*, vol. 406, pp. 386–398, Sep. 2020.

[13] 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.

[14] Z. Xia, L. Jiang, D. Liu, L. Lu, and B. Jeon, "BOEW: A content-based image retrieval scheme using bag-of-encrypted-words in cloud computing," *IEEE Trans. Services Comput.*, early access, Jul. 10, 2019, doi: 10.1109/TSC.2019.2927215.

[15] H. Cheng, X. Zhang, J. Yu, and F. Li, "Markov process-based retrieval for encrypted JPEG images," *EURASIP J. Inf. Secur.*, vol. 2016, no. 1, p. 1, Dec. 2016.

[16] H. Cheng, X. Zhang, and J. Yu, "AC-coefficient histogram-based retrieval for encrypted JPEG images," *Multimedia Tools Appl.*, vol. 75, no. 21, pp. 13791–13803, Nov. 2016.

[17] B. Ferreira, J. Rodrigues, J. Leitao, 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.

[18] H. Cheng, J. Wang, M. Wang, and S. Zhong, "Toward privacy-preserving jpeg image retrieval," *J. Electron. Imag.*, vol. 26, no. 4, pp. 043022-1–043022-5, 2017.

[19] W. Sirichotedumrong, T. Maekawa, Y. Kinoshita, and H. Kiya, "Privacy-preserving deep neural networks with pixel-based image encryption considering data augmentation in the encrypted domain," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2019, pp. 674–678.

[20] W. Sirichotedumrong, Y. Kinoshita, and H. Kiya, "Pixel-based image encryption without key management for privacy-preserving deep neural networks," *IEEE Access*, vol. 7, pp. 177844–177855, 2019.

[21] K. Iida and H. Kiya, "An image identification scheme of encrypted jpeg images for privacy-preserving photo sharing services," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2019, pp. 4564–4568.

[22] J. Zhou, X. Liu, O. C. Au, and Y. Y. Tang, "Designing an efficient image Encryption-Then-Compression system via prediction error clustering and random permutation," *IEEE Trans. Inf. Forensics Security*, vol. 9, no. 1, pp. 39–50, Jan. 2014.

[23] K. Kurihara, M. Kikuchi, S. Imaizumi, S. Shiota, and H. Kiya, "An encryption-then-compression system for JPEG/motion JPEG standard," *IEICE Trans. Fundamentals Electron., Commun. Comput. Sci.*, vol. E98.A, no. 11, pp. 2238–2245, 2015.

[24] T. Chuman and H. Kiya, "Security evaluation for block scrambling-based image encryption including JPEG distortion against jigsaw puzzle solver attacks," *IEICE Trans. Fundamentals Electron., Commun. Comput. Sci.*, vol. E101.A, no. 12, pp. 2405–2408, 2018.

[25] T. Chuman, W. Sirichotedumrong, and H. Kiya, "Encryption-then-compression systems using grayscale-based image encryption for JPEG images," *IEEE Trans. Inf. Forensics Security*, vol. 14, no. 6, pp. 1515–1525, Jun. 2019.

[26] T. Chuman, K. Kurihara, and H. Kiya, "On the security of block scrambling-based etc systems against jigsaw puzzle solver attacks," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Mar. 2017, pp. 2157–2161.

[27] T. Chuman, K. Kurihara, and H. Kiya, "On the security of block scrambling-based EtC systems against extended jigsaw puzzle solver attacks," *IEICE Trans. Inf. Syst.*, vol. E101.D, no. 1, pp. 37–44, 2018.

[28] T. Chuman, W. Sirichotedumrong, and H. Kiya, "Encryption-then-compression systems using grayscale-based image encryption for JPEG images," *IEEE Trans. Inf. Forensics Security*, vol. 14, no. 6, pp. 1515–1525, Jun. 2019.

[29] W. Sirichotedumrong and H. Kiya, "Grayscale-based block scrambling image encryption using YCbCr color space for encryption-then-compression systems," *APSIPA Trans. Signal Inf. Process.*, vol. 8, p. e7, Feb. 2019.

[30] K. Kurihara, S. Imaizumi, S. Shiota, and H. Kiya, "An encryption-then-compression system for lossless image compression standards," *IEICE Trans. Inf. Syst.*, vol. E100.D, no. 1, pp. 52–56, 2017.

[31] L. Zheng, Y. Yang, and Q. Tian, "SIFT meets CNN: A decade survey of instance retrieval," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 5, pp. 1224–1244, May 2018.

[32] J. Song, T. He, L. Gao, X. Xu, A. Hanjalic, and H. T. Shen, "Binary generative adversarial networks for image retrieval," in *Proc. AAAI Conf. Artif. Intell.*, 2018, pp. 394–401.

[33] T. Sikora, "The MPEG-7 visual standard for content description—An overview," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 6, pp. 696–702, Jun. 2001.

[34] S. A. Chatzichristofis and Y. S. Boutalis, "Cedd: Color and edge directivity descriptor: A compact descriptor for image indexing and retrieval," in *Proc. Springer Int. Conf. Comput. Vis. Syst.*, pp. 312–322, 2008.

[35] Y. Gong, S. Lazebnik, A. Gordo, and F. Perronnin, "Iterative quantization: A procrustean approach to learning binary codes for large-scale image retrieval," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 12, pp. 2916–2929, Dec. 2013.

[36] A. Oliva and A. Torralba, "Modeling the shape of the scene: A holistic representation of the spatial envelope," *Int. J. Comput. Vis.*, vol. 42, no. 3, pp. 145–175, 2001.

[37] Y. Li and P. Wang, "Robust image hashing based on low-rank and sparse decomposition," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2016, pp. 2154–2158.

[38] Y. N. Li, P. Wang, and Y. T. Su, "Robust image hashing based on selective quaternion invariance," *IEEE Signal Process. Lett.*, vol. 22, no. 12, pp. 2396–2400, Dec. 2015.

[39] Z. Tang, L. Chen, X. Zhang, and S. Zhang, "Robust image hashing with tensor decomposition," *IEEE Trans. Knowl. Data Eng.*, vol. 31, no. 3, pp. 549–560, Mar. 2019.

[40] Z. Tang, X. Zhang, X. Li, and S. Zhang, "Robust image hashing with ring partition and invariant vector distance," *IEEE Trans. Inf. Forensics Security*, vol. 11, no. 1, pp. 200–214, Jan. 2016.

[41] C. Iakovidou, N. Anagnostopoulos, A. Kapoutsis, Y. Boutalis, M. Lux, and S. A. Chatzichristofis, "Localizing global descriptors for content-based image retrieval," *EURASIP J. Adv. Signal Process.*, vol. 2015, no. 1, p. 80, Dec. 2015.

[42] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, Nov. 2004.

[43] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (SURF)," *Comput. Vis. Image Understand.*, vol. 110, no. 3, pp. 346–359, Jun. 2008.

[44] K. E. A. van de Sande, T. Gevers, and C. G. M. Snoek, "Evaluating color descriptors for object and scene recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 9, pp. 1582–1596, Sep. 2010.

[45] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: An efficient alternative to SIFT or SURF," in *Proc. Int. Conf. Comput. Vis.*, Nov. 2011, pp. 2564–2571.

[46] A. Krizhevsky, I. Sutskever, and G. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

[47] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*. [Online]. Available: http://arxiv.org/abs/1409.1556

[48] A. Habeen Chang and B. M. Case, "Attacks on image encryption schemes for privacy-preserving deep neural networks," 2020, *arXiv:2004.13263*. [Online]. Available: http://arxiv.org/abs/2004.13263

[49] H. Ito, Y. Kinoshita, and H. Kiya, "Image transformation network for privacy-preserving deep neural networks and its security evaluation," 2020, *arXiv:2008.03143*. [Online]. Available: http://arxiv.org/abs/2008.03143

[50] M. Lux and S. A. Chatzichristofis, "Lire: Lucene image retrieval: An extensible java CBIR library," in *Proc. 16th ACM Int. Conf. Multimedia (MM)*, 2008, pp. 1085–1088.

[51] D. Nister and H. Stewenius, "Scalable recognition with a vocabulary tree," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 2, Jun. 2006, pp. 2161–2168.

[52] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. Springer Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2015, pp. 234–241.

[53] H. Jegou, M. Douze, and C. Schmid, "Hamming embedding and weak geometric consistency for large scale image search," in *Proc. Springer Eur. Conf. Comput. Vis.*, 2008, pp. 304–317.

[54] J. Z. Wang, J. Li, and G. Wiederhold, "SIMPLIcity: Semantics-sensitive integrated matching for picture libraries," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 9, pp. 947–963, Sep. 2001.

KENTA IIDA (Student Member, IEEE) received the B.Eng. and M.Eng. degrees from Tokyo Metropolitan University, Japan, in 2016 and 2018, respectively, where he is currently pursuing the Ph.D. degree. His research interests include image/video processing and multimedia security. He is a member of IEICE.

HITOSHI KIYA (Fellow, IEEE) received the B.E. and M.E. degrees from the Nagaoka University of Technology, in 1980 and 1982, respectively, and the Dr.Eng. degree from Tokyo Metropolitan University, in 1987. In 1982, he joined Tokyo Metropolitan University, where he became a Full Professor, in 2000. From 1995 to 1996, he was a Visiting Fellow with The University of Sydney, Sydney, NSW, Australia. He is a Fellow of IEICE and ITE. He has received numerous awards, including ten best paper awards. He was an Editorial Board Member of eight journals, including the IEEE TRANSACTIONS ON SIGNAL PROCESSING, IEEE TRANSACTIONS ON IMAGE PROCESSING, and IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY. He was the Chair of two technical committees and a member of nine technical committees, including the APSIPA Image, Video, and Multimedia Technical Committee (TC) and the IEEE Information Forensics and Security TC. He has organized a lot of international conferences in roles, such as the TPC Chair of the IEEE ICASSP 2012 and the General Co-Chair of the IEEE ISCAS 2019. He served as the Inaugural Vice President (Technical Activities) for APSIPA, from 2009 to 2013, and the Regional Director-at-Large for Region 10 of the IEEE Signal Processing Society, from 2016 to 2017. He was also the President of the IEICE Engineering Sciences Society, from 2011 to 2012. He currently serves as the President for APSIPA. He has served as the Vice President and the Editor-in-Chief for *IEICE Communications Society Magazine* and Society Publications.