A PRIVACY-PRESERVING IMAGE RETRIEVAL SCHEME USING A CODEBOOK GENERATED FROM INDEPENDENT PLAIN-IMAGE DATASET

Kenta Iida* and Hitoshi Kiya*
* Tokyo Metropolitan University, Tokyo, Japan
E-mail: iida-kenta2@ed.tmu.ac.jp, kiya@tmu.ac.jp

Abstract—In this paper, we propose a privacy-preserving image-retrieval scheme using a codebook generated by using a plain-image dataset. Encryption-then-compression (EtC) images, which were proposed for EtC systems, have been used in conventional privacy-preserving image-retrieval schemes, in which a codebook is generated from EtC images uploaded by image owners, and extended SIMPLE descriptors are then calculated as image descriptors by using the codebook. In contrast, in the proposed scheme, a codebook is generated from a dataset independent of uploaded images. The use of an independent dataset enables us not only to use a codebook that does not require recalculation but also to constantly provide a high retrieval accuracy. In an experiment, the proposed scheme is demonstrated to maintain a high retrieval performance, even if codebooks are generated from a plain image dataset independent of image owners' encrypted images.

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 image identification, retrieval, and processing schemes have been studied for untrusted cloud environments [3]–[19].

For the above reasons, privacy-preserving image-retrieval methods should satisfy generally three requirements: 1) protecting visual information on plain images, 2) having a high retrieval performance for encrypted images, and 3) being applicable to compressible encrypted images. To satisfy both requirements 1) and 3), encryption-then-compression (EtC) systems have been developed for privacy-preserving image retrieval [20]–[24]. In this paper, we focus on a block scrambling-based image encryption method that was proposed for EtC systems, where images encrypted by the method are referred to as "EtC images."

To retrieve EtC images stored in a database, a codebook is generated from these EtC images, and extended SIMPLE descriptors (E-SIMPLEs) of the stored EtC images and a query are calculated by using the codebook. As a result, the retrieval performance of the methods depends on the EtC images

uploaded by image owners. Accordingly, the retrieval performance of the conventional methods degrades if the uploaded EtC images are not enough for calculating the codebook. In addition, the conventional methods require recalculation to handle each EtC image dataset in the encrypted domain.

Due to such a situation, in this paper, a novel privacy-preserving image retrieval scheme using a codebook generated by using a plain-image dataset independent of uploaded EtC images is investigated. Accordingly, codebooks can be prepared by using a suitable plain dataset, so the recalculation of codebooks is not required. In the proposed method, EtC images can be used as compressible encrypted images as well for the conventional schemes. In an experiment, the proposed scheme is demonstrated to maintain a high retrieval performance, even if codebooks are generated from a plain image dataset independent of image owners ' encrypted images.

II. RELATED WORK

A. Privacy-preserving content-based image retrieval

Privacy-preserving content-based image retrieval (CBIR) schemes have been proposed to carry out content-based image retrieval without visual information in cloud environments. The schemes are classified into two approaches in terms of how to obtain descriptors as below.

In the first approach, descriptors are extracted from plain images by an image owner [3], [4], [6], [8]–[11], and these descriptors are required to be encrypted by an image owner. In addition, some secret information is shared between the image owner and users [4], [11]. In contrast, in the second approach, descriptors are directly extracted from encrypted images without decryption by cloud providers [5], [12]–[16], and the image owner has no secret information that should be shared with users. The proposed scheme corresponds to this approach. Bag-of-visual words (BOVW) model-based [5], [12], [15]–[18] and JPEG-based [13], [14] schemes are also in this approach.

In conventional schemes for EtC images, to calculate image descriptors of EtC images stored in a database, a codebook is generated from the stored images. Thus, cloud providers can not control retrieval performances because the codebook depends on the stored images. However, the use of a plain-image dataset independent of the uploaded EtC images has
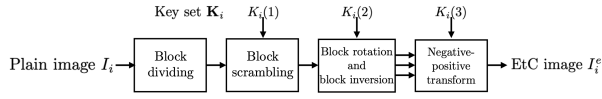
Fig. 1. Generation of EtC images

(a) Plain image　　　　　　　(b) EtC image

Fig. 2. Example of plain and encrypted images

never been considered. Therefore, the retrieval with such a dataset is considered in this paper.

B. EtC images

In this paper, we focus on EtC images, which are images encrypted by using a block-wise encryption method for EtC systems [21]–[24]. EtC images have almost the same compression performance as those of plain images but also enough robustness against various ciphertext-only attacks including jigsaw puzzle solver attacks [16], [25]–[28]. EtC images are generated by following the procedure as below (see Figs. 1 and 2).

(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 secret key $K_i(1)$.
(c) Rotate and invert randomly each divided block by using a random integer secret key $K_i(2)$.
(d) Apply negative-positive transformation to each block by using a random binary integer generated by secret key $K_i(3)$ to obtain encrypted image I_i^e. In this step, a transformed pixel value in the jth block B_j, p' is computed by

$$\begin{cases} p' = p, \ r(j) = 0, \\ p' = 255 - p, \ r(j) = 1, \end{cases} \quad (1)$$

where $r(j)$ is a random binary integer generated by $K_i(3)$ under the probability $P(r(j)) = 0.5$, and p is the pixel value of a plain image with 8 bpp.

In this paper, images encrypted by using these steps are referred to as "EtC images." $K_i(1)$, $K_i(2)$, and $K_i(3)$ are stored as a key set, $\mathbf{K_i} = [K_i(1), K_i(2), K_i(3)]$.

C. Extended SIMPLE descriptors

Extended SIMPLE descriptors (E-SIMPLEs) were proposed for image retrieval of EtC images, and high retrieval performances were reported in the conventional schemes [16]–[18]. In the proposed scheme, E-SIMPLEs are used as image descriptors, as well as conventional schemes.

E-SIMPLEs are designed for avoiding the influences of three encryption operations used for generating EtC images. E-SIMPLEs are calculated from images as below.

a) Divide each 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 modified color and edge directivity descriptor (mCEDD) as a patch descriptor from each patch.
c) Generate a codebook with a size of M by using the patch descriptors extracted from all images.
d) Calculate a histogram vector of each image by using the codebook and patch descriptors extracted from the image.
e) Obtain extended SIMPLE descriptors by weighting the histogram vectors and l_2 normalization.

To generate the codebook in step c), k-means clustering is applied to the extracted patch descriptors, where the size of the codebook corresponds to the number of classes in k-means clustering. The center of each class is defined as a visual word, and the set of visual words is stored as a codebook. By using this codebook, a histogram vector of the frequencies of these visual words included in the image is calculated in step d).

In step e), all histogram vectors are weighted in order to obtain extended SIMPLE descriptors. When N histogram vectors are generated from N images in step d), the mth component of the nth vectors $v_n(m)$, $0 \le m < M$, $0 \le n < N$, is calculated as below.

$$v_n(m) = (1 + log(v'_n(m))). \quad (2)$$

where $v'_n(m)$ represents the mth component of the nth histogram vector. After that, l_2 normalization is applied to every weighted histogram vector.

III. PROPOSED SCHEME

A. System model

In the proposed scheme, the system model shown in Fig. 3 is used. In this model, there are three roles: image owner, third party, and user, where the third party is not trusted. The third party may do data leakage and unauthorized use of stored images, so the privacy-preserving of uploaded images is required.

In this model, the third party generates a codebook by using a plain-image dataset in advance, which is independent of EtC images uploaded by image owners. An image owner uploads EtC images together with owner's information to the third party, and then the image descriptors are calculated with the pre-generated codebook. When a user sends a query image to the third party, the third party retrieves images by using image descriptors of the query image and the stored descriptors. The retrieved images are sent to the user with the owner's information, and the user decrypts the EtC images with the key set received from the image owner.

In this model, the third party has not only no visual information on plain images but also no key sets.

B. Proposed image retrieval scheme

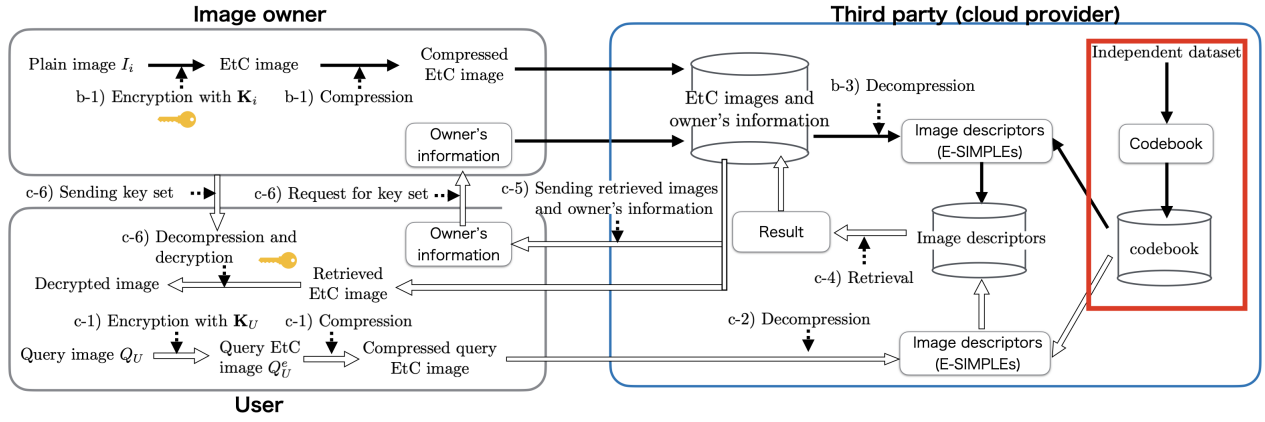Here, the details of each operation performed in the model are summarized as below.

Fig. 3. System model of proposed scheme

1) Generation process for codebook

A third party prepares a codebook with a size of M by using an independent training dataset as below.

(a-1) Divide each image of the training dataset into non-overlapping 16×16-blocks, where 16×16 corresponds to the block size of EtC images.

(a-2) Extract a mCEDD as a patch descriptor from each extracted patch.

(a-3) Apply k-means clustering to the extracted patch descriptors under M classes.

(a-4) Select each centroid vector as a visual word, and store the set of these visual words as a codebook.

2) Generation process for image descriptors of uploaded images

An image owner and a third party perform the following processes to store an image and the image descriptors in the third party.

(b-1) An image owner generates an EtC image from plain image I_i with secret key set $\mathbf{K}_i$, and then compresses the EtC image with JPEG compression/ a lossless-compression method.

(b-2) The image owner uploads the compressed EtC images to a third party, and the uploaded image is then stored in the third party.

(b-3) The third party obtains patches of the uploaded image by dividing the image into non-overlapping 16×16-blocks after decompression.

(b-4) mCEDDs are extracted as patch descriptors of the uploaded image from the obtained patches.

(b-5) A histogram vector of the uploaded image is calculated from extracted patch descriptors by using the codebook generated in step (a-4).

(b-6) The third party calculates an E-SIMPLE by weighting histogram vector in accordance with Eq. 2 and applying l_2 normalization to the histogram vectors.
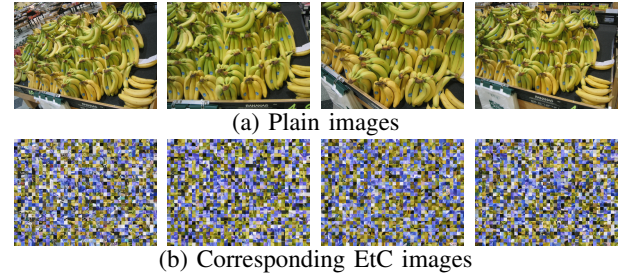

(a) Plain images

(b) Corresponding EtC images

Fig. 4. Image examples in group (UKbench dataset)

3) Retrieval process for query image

To retrieve stored EtC images similar to a query image uploaded by a user, the following processes are carried out.

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

(c-2) The third party obtains patches from the query image after decompression, and extracts a mCEDD as a patch descriptor from each patch.

(c-3) The third party calculates an E-SIMPLE from the patch descriptors by using the codebook stored in the database.

(c-4) The third party computes the l_2 distance between every E-SIMPLE stored in the database and the E-SIMPLE of the query, and then retrieves similar images from the database.

(c-5) The retrieved images and the owner's information are returned to the user.

(c-6) The user requests the data owner to send key set $\mathbf{K_i}$ for decrypting the EtC images received from the third party.

IV. EXPERIMENT

A. Experiment setup

In this experiment, we used three datasets: the UKBench dataset [29], the COREL dataset [30], and the UCID dataset [31]. The COREL and the UCID datasets were used for
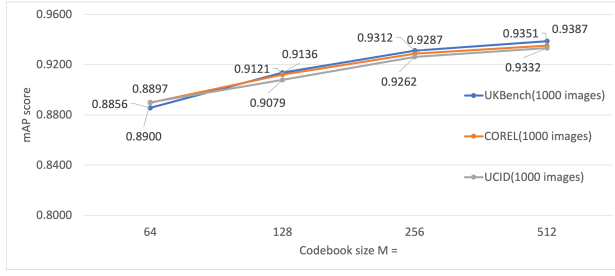
Fig. 5. Retrieval performance comparison under three codebooks generated from 1,000 images.
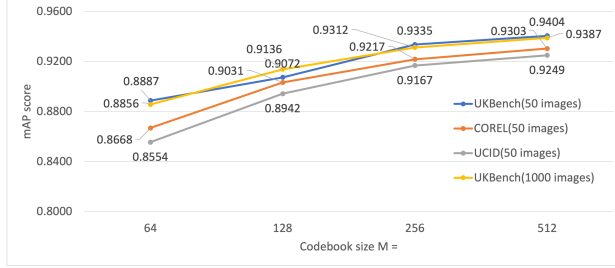

Fig. 6. Retrieval performance comparison under four codebooks generated from 50 or 1,000 images.

generating codebooks. 1,000 images from No. 00000 to No. 00999 in the UKBench dataset were selected as query and user's images. The 1,000 images in the UKBench dataset are classified into 250 groups and each group has four images (see Fig. 4). All the 1,000 images were used as images stored in the third party, and 250 query images were selected from the first image of each group.

The performance of the proposed scheme was evaluated in terms of mean average precision (mAP) scores in this experiment. mAP scores are obtained from average precision values of Q query images. The average precision value of the qth query image is calculated as below, when G ground truth images of this query are contained in N images stored in a database.

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

where $TP@n$ represents the number of the true positive matches at the rank n and $f(n)$ is defined by

$$f(n) = \begin{cases} 1, & \text{if the } n\text{th image is a ground truth image,} \\ 0, & \text{otherwise.} \end{cases}$$

$$(4)$$

By using the average precision values for all Q query images, mAP score is calculated as

$$mAP = \frac{\sum_{q=0}^{Q-1} AP_q}{Q}. \qquad (5)$$

TABLE I
COMPARISON WITH CONVENTIONAL CBIR METHODS USING PLAIN IMAGES.

Descriptor	$M =$	mAP score
SCD [32] (plain)	-	0.9179
CEDD [33] (plain)	-	0.8806
SURF [34]	256	0.8304
(plain)	512	0.8355
Weighted SIMPLE with CEDD	256	0.9300
and random sampling (plain) [35]	512	0.9481
Weighted SIMPLE with CEDD	256	0.9000
and SURF detector (plain) [35]	512	0.9222
Proposed scheme	256	0.9287
with COREL (EtC)	512	0.9351
Proposed scheme	256	0.9262
with UCID (EtC)	512	0.9332

B. Performance evaluation of proposed scheme

In Fig. 5, retrieval performances under three codebooks generated from 1,000 images in each dataset are shown. It was confirmed that mAP scores under the use of codebooks generated from images in the COREL and the UCID datasets were almost the same as those with codebooks generated from the uploaded images (UKBench), where the COREL and UCID datasets are independent of the UKBench dataset used by the image owner. From this figure, the proposed scheme was demonstrated to maintain a high retrieval performance even when the codebook was calculated by using a plain-image dataset independent of the uploaded dataset.

Next, retrieval performances under the use of codebooks generated from a smaller number of images were evaluated as shown in Fig. 6. Compared with Fig. 5, mAP scores for the COREL and UCID datasets degraded, but those for the UKBench dataset were maintained. From these results, when the codebook calculated from a small number of images was applied to images in other datasets, the reuse of the codebook provided a degraded performance. In contrast, the use of an independent dataset containing the enough number of images maintained a high performance. Accordingly, the proposed scheme that uses an independent dataset containing the enough number of images can maintain a high retrieval performance without the recalculation of codebooks.

C. Comparison with conventional image retrieval scheme

To compare the proposed scheme with conventional CBIR schemes that do not consider privacy protection, five image descriptors: scalable color descriptor (SCD) [32], color and edge directivity descriptor (CEDD) [33], SURF [34], weighted SIMPLE descriptor with random sampling [35], and weighted SIMPLE descriptor with SURF detector [35]. From Table I, the proposed scheme was verified to have almost the same performance as those of the conventional ones, even when codebooks were generated from an independent dataset.

V. CONCLUSION

A novel CBIR scheme using codebooks generated from a plain-image dataset independent of uploaded images was proposed for privacy-preserving image retrieval. In the proposed

scheme, EtC images can be used to reduce the amount of data, and extended SIMPLE descriptors are applied to avoid the influence of image encryption steps. In addition, codebooks are generated from a suitable plain-image dataset. Experimental results showed that the proposed scheme achieved a reasonable high retrieval performance, even when codebooks were generated from a plain image dataset independent of uploaded images.

REFERENCES

[1] C. T. Huang, L. Huangand Z. Qin, H. Yuan, L. Zhou, V. Varadharajan, and C-C Jay Kuo, "Survey on securing data storage in the cloud," *APSIPA Trans. Signal and Information Processing*, vol. 3, 2014.

[2] R. Lagendijk, Z. Erkin, and M. Barni, "Encrypted signal processing for privacy protection: Conveying the utility of homomorphic encryption and multiparty computation," *IEEE Signal Processing Magazine*, vol. 30, no. 1, pp. 82–105, 2013.

[3] 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.

[4] Z. Huang, M. Zhang, and Y. Zhang, "Toward efficient encrypted image retrieval in cloud environment," *IEEE Access*, vol. 7, pp. 174541–174550, 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*, 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. Xia, N. N. Xiong, A.V. Vasilakos, and X. Sun, "Epcbir: An efficient and privacy-preserving content-based image retrieval scheme in cloud computing," *Elsevier Information Sciences*, vol. 387, pp. 195–204, 2017.

[8] 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. Information Forensics and Security*, vol. 11, no. 11, pp. 2594–2608, 2016.

[9] Chengyuan Zhang, Lei Zhu, Shichao Zhang, and Weiren Yu, "Tdhppir: An efficient deep hashing based privacy-preserving image retrieval method," *Elsevier Neurocomputing*, 2020.

[10] Zongye Zhang, Fucai Zhou, Shiyue Qin, Qiang Jia, and Zifeng Xu, "Privacy-preserving image retrieval and sharing in social multimedia applications," *IEEE Access*, vol. 8, pp. 66828–66838, 2020.

[11] X. Li, Q. Xue, and M. C. Chuah, "Casheirs: Cloud assisted scalable hierarchical encrypted based image retrieval system," *in Proc. IEEE Conf. Computer Communications*, pp. 1–9, 2017.

[12] 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 Computing*, pp. 1–1, 2019.

[13] H. Cheng, X. Zhang, J. Yu, and F. Li, "Markov process-based retrieval for encrypted jpeg images," *EURASIP Journal on Information Security*, vol. 2016, no. 1, pp. 1, 2016.

[14] H. Cheng, X. Zhang, and J. Yu, "Ac-coefficient histogram-based retrieval for encrypted jpeg images," *Springer Multimedia Tools and Applications*, vol. 75, no. 21, pp. 13791–13803, 2016.

[15] B. Ferreira, J. Rodrigues, J. Leitao, and H. Domingos, "Practical privacy-preserving content-based retrieval in cloud image repositories," *IEEE Trans. Cloud Computing*, 2017.

[16] K. Iida and H. Kiya, "Privacy-preserving content-based image retrieval using compressible encrypted images," *IEEE Access*, vol. 8, pp. 200038–200050, 2020.

[17] K. Iida and H. Kiya, "A privacy-preserving content-based image retrieval scheme allowing mixed use of encrypted and plain images," *in Proc. APSIPA ASC*, pp. 1436–1441, 2020.

[18] K. Iida and H. Kiya, "Privacy-preserving image retrieval scheme allowing mixed use of lossless and jpeg compressed images," *in Proc. IEEE Global Conf. on Life Sciences and Technologies*, 2021.

[19] F. Arnia, I. Iizuka, M. Fujiyoshi, and H. Kiya, "Fast and robust identification methods for jpeg images with various compression ratios," *in Proc. IEEE Int'l Conf. Acoustics Speech and Signal Processing Proceedings*, vol. 2, pp. II–II, 2006.

[20] 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. Information Forensics and Security*, vol. 9, no. 1, pp. 39–50, 2014.

[21] K. Kurihara, M. Kikuchi, S. Imaizumi, S. Shiota, and H. Kiya, "An encryption-then-compression system for jpegmotion jpeg standard," *IEICE Trans. Fundamentals.*, vol. 98, no. 11, pp. 2238–2245, 2015.

[22] K. Kurihara, S. Imaizumi, S. Shiota, and H. Kiya, "An encryption-then-compression system for lossless image compression standards," *IEICE Trans. Information and Systems*, vol. 100, no. 1, pp. 52–56, 2017.

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

[24] W. Sirichotedumrong and H. Kiya, "Grayscale-based block scrambling image encryption using ycbcr color space for encryption-then-compression systems," *APSIPA Trans. Signal and Information Processing*, vol. 8, 2019.

[25] T. Chuman, K. Kurihara, and H. Kiya, "On the security of block scrambling-based etc systems against jigsaw puzzle solver attacks," pp. 2157–2161, 2017.

[26] T. Chuman, K. Kurihara, and H. Kiya, "On the security of block scrambling-based etc systems against extended jigsaw puzzle solver attacks," *IEICE Trans. Information and Systems*, vol. 101, no. 1, pp. 37–44, 2018.

[27] A. H. Chang and B. M. Case, "Attacks on image encryption schemes for privacy-preserving deep neural networks," *arXiv eprints, arXiv:2004.13263*, 2020.

[28] H. Ito, Y. Kinoshita, M. Aprilpyone, and H. Kiya, "Image to perturbation: An image transformation network for generating visually protected images for privacy-preserving deep neural networks," *IEEE Access*, vol. 9, pp. 64629–64638, 2021.

[29] D. Nister and H. Stewenius, "Scalable recognition with a vocabulary tree," *in Proc. IEEE Computer Society Conf. Computer Vision and Pattern Recognition*, vol. 2, pp. 2161–2168, 2006.

[30] J. Z. Wang, J. Li, and G. Wiederhold, "Simplicity: Semantics-sensitive integrated matching for picture libraries," *IEEE Trans. pattern analysis and machine intelligence*, vol. 23, no. 9, pp. 947–963, 2001.

[31] G. Schaefer and M. Stich, "UCID: an uncompressed color image database," vol. 5307, pp. 472 – 480, 2003.

[32] T. Sikora, "The mpeg-7 visual standard for content description an overview," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 11, no. 6, pp. 696–702, 2001.

[33] Savvas A Chatzichristofis and Yiannis S Boutalis, "Cedd: color and edge directivity descriptor: a compact descriptor for image indexing and retrieval," *in Proc. Springer International Conf. Computer Vision Systems*, pp. 312–322, 2008.

[34] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (surf)," *Elsevier Computer Vision and Image Understanding*, vol. 110, no. 3, pp. 346–359, 2008.

[35] C. Iakovidou, N. Anagnostopoulos, A. Kapoutsis, Y. Boutalis, M. Lux, and Savvas A. Chatzichristofis, "Localizing global descriptors for content-based image retrieval," *EURASIP Journal on Advances in Signal Processing*, vol. 2015, no. 1, pp. 80, 2015.