A Content-based Image Retrieval Scheme Using Compressible Encrypted Images

Kenta Iida and Hitoshi Kiya
Tokyo Metropolitan University, Tokyo, Japan

***Abstract*—In this paper, we propose a novel content based-image retrieval 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 also to make the sensitive management of secret keys unnecessary. In addition, encrypted images can be compressed by using JPEG compression. Weighted SIMPLE image descriptors, which are generated from global descriptors of localized regions, are extended, and then the extended descriptors are applied to EtC images. In an experiment, the proposed scheme is demonstrated to have almost the same accuracy as conventional retrieval methods with plain images.**

***Index Terms*—Content-based image retrieval, encryption-then-compression system, SIMPLE image descriptors**

I. Introduction

With the rapid growth of cloud computing, outsourcing images to cloud storage services and photo sharing have greatly increased. Generally, images are uploaded and stored in a compressed form to reduce the amount of data. In addition, most of images include sensitive information, such as personal data and copyrights [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 in untrusted cloud environments [3]–[14].

Due to the above reason, privacy-preserving image retrieval methods are required to meet the following three requirements: 1) protecting visual information on plain images, 2) having high retrieval performance in the encrypted domain, 3) compressing encrypted images.

To satisfy requirement 1), full encryption with provable security, such as RSA and AES, is the most secure option for the protection of multimedia data [3]–[6]. In contrast, some perceptual encryption methods have been proposed, which can be directly applied to some signal processing algorithms in the encrypted domain [7]–[10], [15]–[18]. However, requirement 3) has never been considered in conventional content-based encrypted image retrieval schemes. As systems that satisfy both requirements 1) and 3), Encryption-then-Compression (EtC) systems have been developed [15]–[18]. In this paper, we focus on a block scrambling-based image encryption method that has been proposed for the EtC systems [16]–[18], and images encrypted by this method are referred to as "EtC images".

Image retrieval schemes for EtC images have been never considered, but image identification schemes have been proposed for detecting EtC images having the same plain image [9], [10]. In general, 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 some keywords in TBIR methods. Therefore, we focus on CBIR in this paper.

For CBIR, various types of image descriptors have been proposed [19]–[28]. The schemes with descriptors generated by using deep neural networks, which include the pre-trained networks, such as AlexNet [29] and VGG [30], and generative adversarial networks [27], have high retrieval performances. However, not only a large number of training images but huge computational costs are required to train a model in general.

In contrast, using handcrafted descriptors allows us to efficiently retrieve images. They are classified into two types in general: global image descriptors such as MPEG-7 and GIST image descriptor [19]–[24], [28], [31] and local image descriptors such as SIFT and SURF image descriptors [25], [26]. In addition, by using both global descriptors and the technique to generate local image descriptors, searching images with MPEG-7-powered localized descriptors (SIMPLE image descriptors) have been proposed [19] to improve retrieval performance. The retrieval scheme using weighted SIMPLE image descriptors was demonstrated to outperform conventional retrieval schemes with handcrafted descriptors. Therefore, we apply weighted SIMPLE descriptors to EtC images for satisfying all requirements.

Due to such a situation, we propose a novel content-based image retrieval scheme for EtC images. In the proposed scheme, weighted SIMPLE image descriptors are extended to avoid the effect of the image encryption. For the retrieval, the extended SIMPLE image descriptors are extracted from EtC images. Simulation results show that the proposed scheme has almost the same accuracy as those of using plain images, even when EtC images are generated by using different keys.

II. Related Work

A. EtC image

We focus on EtC images which have been proposed for Encryption-then-Compression (EtC) systems with JPEG compression [15]–[17]. EtC images have not only almost the same compression performance as that of plain images, but also enough robustness against various ciphertext-only attacks including jigsaw puzzle solver attacks [17], [18]. The procedure

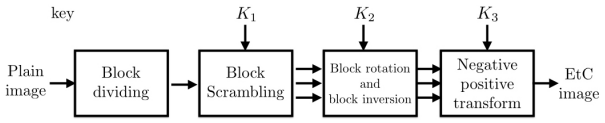

Fig. 1. Generation of EtC images

(a) Plain image (b) EtC image

Fig. 2. Example of plain image and encrypted one

of generating EtC images is conducted as below (see Figs. 1 and 2) [16].

(a) Divide an image 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_1.
(c) Rotate and invert randomly each divided block by using a random integer secret key K_2.
(d) Apply negative-positive transformation to each block by using a random binary integer generated by a key K_3. 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 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".

B. Image retrieval in encrypted domain

Image retrieval methods in the encrypted domain are classified into two classes in accordance with how to obtain descriptors for image retrieval.

1) Generating descriptors from plain images

In methods in the first class, descriptors such as SURF-based [3], SIFT-based [4], CNN-based [5], [12] and MPEG-7 descriptors [6], [13] are calculated by using plain images, and then the descriptors and the plain images are encrypted by a data owner. After that, the encrypted descriptors and images are sent to a cloud server. In this class, data owners are required to extract descriptors and encrypt both the descriptors and the plain images by their selves. Moreover, data owners and users have to share a common key, so their applications are limited due to the difficulty of safe key management.

2) Generating descriptors from encrypted images

In methods in the second class, descriptors are directly extracted from encrypted images by a cloud provider as well as content-based image retrieval methods for plain images, after data owners encrypt images and then send the encrypted ones and owners' information to the cloud provider [7], [8], [11]. In this class, the data owner performs only the encryption of the images. In addition, sharing secret keys with data owners and users is not required. Thus, we focus on this class in this paper.

A bag-of-visual words (BOVW) based retrieval method [7] and a support vector machine-based retrieval method [8] are in this class. However, all conventional methods in this class have never considered compressing encrypted images.

III. Proposed scheme

A novel content-based image retrieval scheme using EtC images is described here. Weighted SIMPLE image descriptors will be extended for EtC images.

A. System model

The framework used in the proposed scheme is shown in Fig. 3. Each operation in Fig. 3 is explained as follows.

1) A data owner encrypts a plain image I_i with a secret key K_i, then the EtC image is compressed with JPEG compression, and the compressed one is uploaded to a third party.
2) The third party generates a codebook from the EtC images after decompressing the JPEG compressed ones, and then image descriptors are calculated by using the codebook. After that, the codebook and the image descriptors are stored in a database.
3) A user sends a query image Q_U encrypted by using a key K_U, to the third party, where K_U can be prepared by the user.
4) The third party retrieves EtC images of being similar to the query image in the encrypted domain. The retrieved images and the owner's information are returned to the user.
5) The user requests the secret key to the data owner for decrypting the EtC images received from the third party.

In this framework, the third party does not have not only any visual information on images, but also secret keys. Moreover, each image can be encrypted by using a different key.

B. Proposed image retrieval scheme

B.1) Weighted SIMPLE image descriptors

It is well-known that weighted SIMPLE image descriptors outperform non-weighted ones [19]. Thus, the weighted SIMPLE image descriptors are considered to be applied to EtC images in this paper. The procedure for extracting weighted SIMPLE image descriptors from plain images by using the BOVW model is summarized here.

a) Decide the positions and the sizes of patches from every image by using a detector such as SURF detector or random sampling.

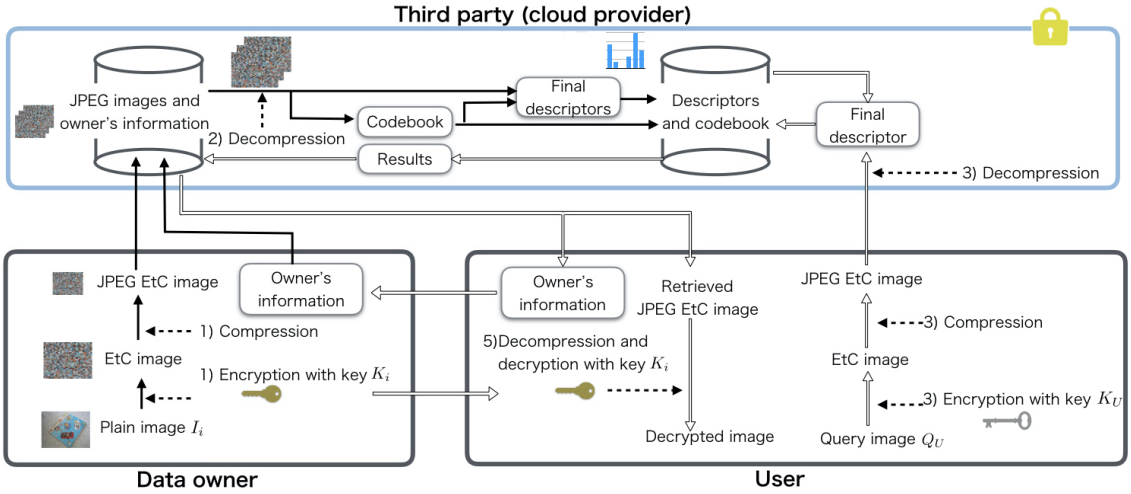

Fig. 3. System model

(a) Random sampling (b) Extended SIMPLE

Fig. 4. Examples of selected patches

b) Extract a global image descriptor from each patch.
c) Generate a codebook with a size of M from the extracted global image descriptors.
d) Calculate SIMPLE image descriptors by using the the codebook.
e) Obtain a weighted SIMPLE image descriptor by weighting the SIMPLE image descriptors.

When SURF detector is used in step a), the positions and the sizes are determined from detected feature points and scales respectively. In contrast, when using random sampling, they are randomly selected (see Fig.4 (a)).

In step c), a codebook is generated from global descriptors by k-means clustering, where the center of each class is defined as a visual word in the codebook. By using the codebook, a SIMPLE image descriptor is represented as a histogram of the frequencies of these visual words included in an image.

B.2) Extended SIMPLE image descriptors

Weighted SIMPLE image descriptors mentioned above are extended in this paper. To generate extended SIMPLE image descriptors, a) in the above step is replaced with a') as below.

a') Divide each image into non-overlapping 16×16-blocks to be selected as patches, where 16×16 corresponds to the block size of the image encryption (see Fig.4 (b)).

After that, steps from b) to e) are carried out to calculate descriptors, called extended SIMPLE image descriptors. By using the extended SIMPLE image descriptors, the influence of the image encryption can be avoided due to the following reasons.

- Block scrambling
  Each patch corresponds to a block of the block-based encryption, so every patch does not include the boundaries generated by the block scrambling operation. In addition, the positions of selected patches are not used in the calculation of SIMPLE image descriptors, because the SIMPLE image descriptor of an image is based on a histogram of visual words contained in the image.
- Block rotation and inversion
  In step b), the influence of the block rotation and inversion operation can be avoided by extracting a rotation-invariant descriptor such as color and edge directivity descriptor (CEDD) [28] .
- Negative-positive transform
  Pixel values in each block are randomly mapped in accordance with Eq. (1), where the probability is $P(r(i)) = 0.5$. Therefore, the negative-positive transform gives almost no influence to the accuracy of retrieval under the use of extended SIMPLE descriptors as demonstrated later.

B.3) Generating descriptors from encrypted query images and retrieval

As shown in Fig. 3, a query image Q_U is encrypted by a user, and the encrypted query one is sent to the third party. Next, the third party calculates an extended SIMPLE descriptor from the encrypted query as well as in B.2), as below.

a) Divide the encrypted query image into non-overlapping 16×16-blocks to be selected as patches.
b) Extract a global image descriptor from each patch, where the type of the global image descriptor is the same as that of the stored descriptors.
c) Calculate an extended SIMPLE image descriptor by using the stored codebook and the stored descriptors.

d) Compute the l_2 distance between every descriptor stored in the database and the query descriptor, and then decide similar images.

In this paper, CEDD [28] is used as the global descriptor, due to the good retrieval performance as reported in [19]. When N SIMPLE descriptors are generated, the mth component of the nth SIMPLE image descriptor $v_n(m)$ is calculated as below $(0 \leq m < M, 0 \leq n < N)$.

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

where $tf_{(m,n)}$ and $df_{(m)}$ represent the frequency of the mth visual word in the nth descriptor and the number of the SIMPLE descriptors containing the mth visual word in the N SIMPLE descriptors respectively. After that, l_2 normalization is applied to every SIMPLE image descriptor to obtain the weighted SIMPLE image descriptor.

IV. Experiment

A. Experiment setup

In this experiment, the performance of the proposed image retrieval was evaluated by using LIRE [32], which is an open source Java library for content-based image retrieval and supports various image descriptors. We used images with a size of 480×640 in UKbench dataset [33], which consists of 10,200 images (four images per a group). 1,000 images from No.00000 to No.00999 were chosen from the data set (see Fig.5). $N = 1,000$ images were uploaded to the third party by a data owner, and $Q = 250$ images which are the first images of 250 groups were used as query ones. In addition to weighted SIMPLE srf and weighted SIMPLE rnd, which combine CEDD with SURF detector and CEDD with random sampling respectively, two global image descriptors: CEDD and Opponent histogram [34] were used for the comparison. The retrieval with the extended SIMPLE image descriptors were carried out under the use of two codebook sizes ($M = 128, 512$).

The performance was evaluated in terms of mean average precision (mAP). To obtain mAP scores, the average of precision values was 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 the images stored in the database, and $TP@n$ represents the number of the true positive matches at the rank n and $f(n) = 1$ if the nth image is a ground truth one. Otherwise, $f(n) = 0$, if the nth image is not. After the calculation of average precision values for all Q query images, mAP score were calculated as

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

In this simulation, the mAP scores were calculated with $N = 1000$, $G = 4$ and $Q = 250$.

(a) No.00144 (b) No.00145 (c) No.00146 (d) No.00147

Fig. 5. Image examples of the same group (UKbench)

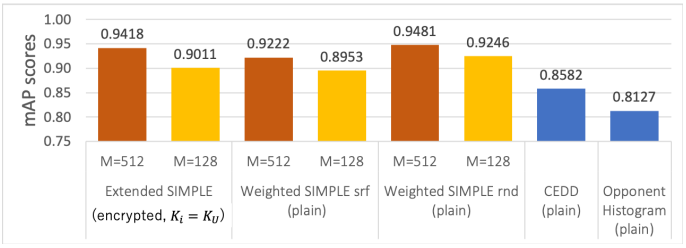

Fig. 6. Comparison with conventional methods using plain images ($K_i = K_U$)

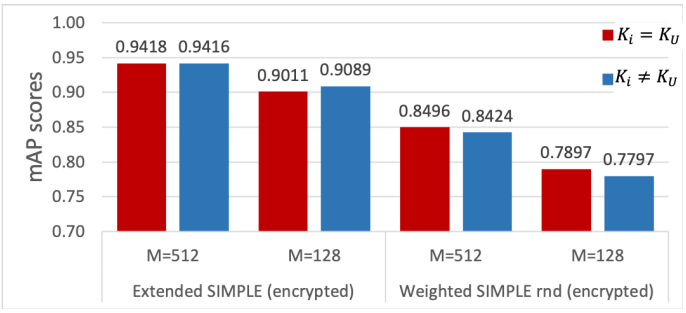

Fig. 7. Comparison of two key conditions

B. Experiment results

1) Comparison with conventional methods ($K_i = K_U$)

At first, the proposed scheme was compared with conventional methods using plain images under the condition of $K_i = K_U$, where all images including query ones were encrypted by using a common random integer. Retrieval results are illustrated in Fig.6. The proposed scheme is confirmed to have higher mAP scores than those of the other descriptors except for SIMPLE rnd, although the conventional methods were carried out by using plain images. The image retrieval of using plain images with the same descriptors, i.e. extended SIMPLE descriptors, as those of the proposed scheme was performed to confirm the influence of using EtC images. The mAP scores under $M = 128$ and $M = 512$ were 0.9321 and 0.9498 respectively. Therefore, there was almost no influence of the image encryption.

2) Comparison of two conditions: $K_i = K_U$ and $K_i \neq K_U$

Next, two key conditions were compared under the use of the proposed scheme. Figure 7 shows retrieval results for EtC images. The proposed scheme (Extended SIMPLE (encrypted) in Fig. 7) is demonstrated to provide almost the same retrieval performance, even when all image were encrypted by using different keys.

In Fig.7, the retrieval performances of SIMPLE rnd using EtC images (Weighted SIMPLE rnd (encrypted) in Fig. 7), which provided the highest scores in Fig. 6 for plain images, were also shown. The mAP scores of SIMPLE rnd degraded greatly, compared to the performances for plain images. This is because SIMPLE rnd does not consider avoiding the influence of the image encryption in the selection of patches. In contrast, extended SIMPLE descriptors can avoid the influence of the encryption.

V. Conclusion

A novel content-based image retrieval scheme using EtC images was proposed for privacy-preserving image retrieval. The image encryption is based on a block-scrambling method, which is known as a compressible encryption method. The proposed scheme satisfies all three requirements, although conventional ones do not.

For the image retrieval of EtC images, the weighted SIMPLE image descriptors are extended. In this extension, the non-overlapping blocks having the same size as that of the image encryption are selected as patches to avoid the effect of the block scrambling operation. In addition, the global image descriptors robust against the block rotation operation are extracted from these patches to generate extended SIMPLE image descriptors. Experiment results show that the proposed scheme enables us to avoid the influence of the image encryption, even if images were encrypted with different secret keys.

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] 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.

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

[6] 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.

[7] 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. on Services Computing*, pp. 1–1, 2019.

[8] 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.

[9] K. Iida and H. Kiya, "An image identification scheme of encrypted jpeg images for privacy-preserving photo sharing services," *Proc. IEEE International Conf. on Image Processing*, pp. 4564–4568, 2019.

[10] K. Iida and H. Kiya, "Image identification of encrypted jpeg images for privacy-preserving photo sharing services," *IEICE Trans. on Information and Systems*, vol. 103, no. 1, pp. 25–32, 2020.

[11] 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.

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

[13] 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 transactions on information forensics and security*, vol. 11, no. 11, pp. 2594–2608, 2016.

[14] S. Ferdowsi, B. Razeghi, T. Holotyak, F. P Calmon, and S. Voloshynovskiy, "Privacy-preserving image sharing via sparsifying layers on convolutional groups," *arXiv preprint arXiv:2002.01469*, 2020.

[15] 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.

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

[17] T. Chuman and H. Kiya, "Security evaluation for block scrambling-based image encryption including jpeg distortion against jigsaw puzzle solver attacks," *IEICE Trans. Fundamentals*, vol. E101-A, no. 12, 2018.

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

[19] 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.

[20] Y. Li and P. Wang, "Robust image hashing based on low-rank and sparse decomposition," *in Proc. IEEE Int'l Conf. on Acoustics, Speech and Signal Processing*, pp. 2154–2158, 2016.

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

[22] Z. Tang, L. Chen, X. Zhang, and S. Zhang, "Robust image hashing with tensor decomposition," *IEEE Trans. on Knowledge and Data Engineering*, vol. 31, no. 3, pp. 549–560, 2018.

[23] Z. Tang, X. Zhang, X. Li, and S. Zhang, "Robust image hashing with ring partition and invariant vector distance," *IEEE transactions on information forensics and security*, vol. 11, no. 1, pp. 200–214, 2015.

[24] 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. on Pattern Analysis and Machine Intelligence*, vol. 35, no. 12, pp. 2916–2929, 2013.

[25] 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.

[26] L. Zheng, Y. Yang, and Q. Tian, "Sift meets cnn: A decade survey of instance retrieval," *IEEE Trans, on pattern analysis and machine intelligence*, vol. 40, no. 5, pp. 1224–1244, 2017.

[27] J. Song, T. He, L. Gao, X. Xu, A. Hanjalic, and H. T. Shen, "Binary generative adversarial networks for image retrieval," *in Proc. AAAI Conference on Artificial Intelligence*, 2018.

[28] 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. on Computer Vision Systems*, pp. 312–322, 2008.

[29] A. Krizhevsky, I. Sutskever, and G.E Hinton, "Imagenet classification with deep convolutional neural networks," *in Proc. Advances in neural information processing systems*, pp. 1097–1105, 2012.

[30] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[31] 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.

[32] M. Lux and Savvas A Chatzichristofis, "Lire: lucene image retrieval: an extensible java cbir library," *in Proc. ACM international Conf. on Multimedia*, pp. 1085–1088, 2008.

[33] "Ukbench dataset," https://archive.org/details/ukbench.

[34] K. Van De Sande, T. Gevers, and C. Snoek, "Evaluating color descriptors for object and scene recognition," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 32, no. 9, pp. 1582–1596, 2009.