A Fast and Secure Content-based Image Retrieval Scheme using Clustering and Searchable Encryption

Md Shahriar Uzzal[a], Ijaz Ahmad[b], and Seokjoo Shin[a]
[a]Dept. of Computer Engineering, Chosun University, Gwangju, Korea
[b]Dept. of Electrical and Computer Engineering, Korea University, Seoul, Korea
shahriar@chosun.ac.kr, ijaz@korea.ac.kr, sjshin@chosun.ac.kr (Corresponding author)

***Abstract*— The rapid expansion of cloud-based image storage has intensified the need for secure and privacy-preserving content-based image retrieval systems. However, achieving an effective balance among retrieval accuracy, security robustness, and computational efficiency remains a significant challenge. Existing approaches often compromise one of these aspects either exhibiting reduced retrieval performance, limited resistance to cryptographic attacks, or increased computational overhead. To address these limitations, this paper presents a fast and secure content-based image retrieval (FSCBIR) framework that integrates sub-block–based perceptual encryption with a lightweight histogram-based feature descriptor and k-means clustering. The block-based encryption allows feature computation directly in the encrypted domain, thereby preserving data privacy while facilitating efficient retrieval. Experimental evaluations on the Corel-1K dataset demonstrate that proposed scheme achieve comparable accuracy, while significantly enhances the retrieval speed. Moreover, a detailed analysis under varying sub-block configurations optimizes the overall performance, confirming the robustness and practicality of the proposed framework.**

Keywords—content-based image retrieval, searchable encryption, CEDD, clustering, Corel-1K, k-means

I. INTRODUCTION

Content-Based Image Retrieval (CBIR) enables efficient image searching from cloud servers based on visual contents [1], [2]. However, outsourcing image data raises serious privacy concerns, as sensitive information may be exposed or misused [3]. A straightforward countermeasure is to encrypt the images prior to outsourcing them to a cloud server. Therefore, encryption-based secure CBIR techniques are proposed, which can be mainly divided into feature encryption–based and image encryption–based approaches. The former extracts features from plaintext images, followed by encryption of both the images and their corresponding features before uploading them to the cloud [3], [4]. However such methods often rely on shared encryption keys, and this reliance increases privacy risks. In contrast, image encryption–based schemes (such as the one proposed in [5] and [6]), perform feature extraction directly in the encrypted domain; therefore, eliminating the need for encrypting the features. However, in this case, the image encryption should preserve searchability for retrieval purposes while ensuring adequate security and computational efficiency.

In this context, searchable perceptual encryption has emerged as a promising approach, as it conceals visual content while retaining the structural information necessary for retrieval. Such encryption is typically implemented through pixel-based or block-based mechanism. For example, a pixel-based searchable encryption has been explored for medical image retrieval in [7]. However, for large-scale cloud environments, maintaining compression efficiency is often desirable to reduce both storage and computational overhead [8], making block-based encryption (for example, [9]) a more practical choice. In block-based schemes, the image is divided into non-overlapping blocks, which then undergo several geometric and color transformations to obfuscate their details. Although large block sizes improve computational efficiency, they also reduce the effective key space, increasing vulnerability to cryptographic and structural attacks. To mitigate this, sub-block processing was introduced in [10], which significantly expands the key space and enhances robustness against brute-force and Jigsaw Puzzle Solver (JPS) attacks. The applications of sub-block-based encryption schemes were extended to secure image retrieval domain in [6]. Specifically, they proposed several configurations of sub-block level processing to find a better tradeoff between retrieval accuracy and security robustness. Despite these advancements, performing feature extraction and similarity matching in an efficient way remains a key challenge for secure content-based image retrieval. Conventional matching techniques (for example, [6]) compute similarities between a query and all stored features individually, which becomes computationally expensive for large datasets.

To address the joint challenges of retrieval efficiency, privacy preservation, and computational overhead, this study proposes a fast and secure content-based image retrieval (FSCBIR) framework. The proposed approach integrates sub-block-based searchable encryption, and a lightweight color and edge directivity descriptor (CEDD) feature extraction mechanism. To alleviate the computational overhead of feature matching, we employee a feature grouping strategy based on k-means clustering, which matches the query only against cluster centroids instead of all individual features, thus reducing the overall retrieval complexity. Experiments conducted on the Corel-1K dataset demonstrate that the proposed method delivers fast retrieval performance with favorable accuracy and strong security resilience.

II. PROPOSED FSCBIR METHOD

Fig. 1. illustrates the proposed FSCBIR framework, which is composed of three primary modules: *image preprocessing*, *feature extraction and clustering*, and *similarity matching and retrieval*. In a cloud-assisted image retrieval application scenario, the first step is implemented on the client-side (image owner and image user) while the remaining two steps are carried out on the cloud server. A detailed description of each module is provided below.

A. Image Preprocessing

In the preprocessing stage, all the stored and query images undergo encryption using a searchable image encryption

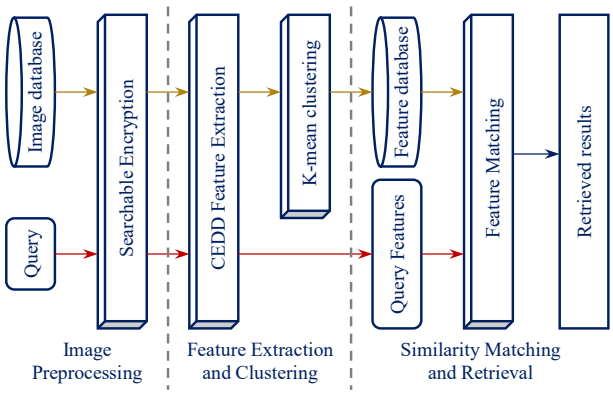


Fig. 1. An illustration of proposed FSCBIR framework.

scheme. For this purpose, we extend the applications of a perceptual encryption algorithm proposed in [10] to secure image retrieval domain. Following [6], the searchable encryption procedure is described as below:

Step 1). The input image $I^{(W\times H\times C)}$ is divided into non-overlapping blocks of size $(E_B \times E_B)$. These blocks are shuffled according to a randomly generated key K_1.

Step 2). Pixel substitution is applied within each block using a uniformly distributed binary keyK_2. The encrypted pixel value $p'(j,k)$ is computed as,

$$p'(j,k) = \begin{cases} p(j,k) & \text{if } K_2(i) = 0, \\ 255 - p(j,k) & \text{if } K_2(i) = 1. \end{cases} \quad (1)$$

Step 3). Each block is further divided into sub-blocks of size $\acute{E}_B \times \acute{E}_B$, on which inversion-rotation operations are performed according to a random key $\acute{K}_3$.

Step 4). Finally, all the encrypted blocks are reassembled to generate the cipher image $\acute{I}$.

Overall, this multi-level encryption scheme provides strong confidentiality by integrating hierarchical key dependencies with both global and local structural obfuscation. Also, this adaptability of block and sub-block processing enables users to fine-tune the trade-off between cryptographic robustness and computational overhead.

B. Feature Extraction and Clustering

In this module, the CEDD-based feature extraction process is adopted from [11], whereas the clustering stage is newly incorporated to enhance the retrieval speed of the proposed FSCBIR scheme compared to existing secure retrieval schemes (such as [6]), as described below.

Step 1). First each encrypted image $\acute{I}$ is transformed from the RGB colorspace into HSV and YIQ colorspaces. Then the resulting images are divided into non-overlapping blocks of size $(F_B \times F_B)$ for feature extraction.

Step 2). For every block, features are computed as color (ϕ^C) and texture (ϕ^T) histograms in the HSV and YIQ representations, respectively. Also, the size of the color histogram is 24 bins and texture histogram is 6 bins.

Step 3). The CEDD descriptor for the i^{th} block is defined as the concatenation of both ϕ^C and ϕ^T feature vectors that is, $\varphi_i = \phi_b^C \times \phi_b^T$, where b denotes a block index. The block-level descriptors are then quantized and aggregated across all blocks to form the global CEDD feature vector of dimension 144 for an image with B number of blocks as,

$$\varphi = \bigcup_{b=1}^{B} \varphi_b \,. \quad (2)$$

Both stored and query images are represented in a feature space using (2) and their corresponding feature vectors are denoted as φ^s and φ^q.

Step 4). The stored feature vectors (φ^s) are subsequently grouped into C_{N_c} clusters using the k-means clustering algorithm, where N_c denotes the number of clusters. Each cluster is represented by a cluster head or centroid, which serves as a reference point to enable faster similarity matching and retrieval as shown in Fig. 2. and described below.

C. Similarity Matching and Retrieval

In the final stage of the proposed FSCBIR framework, similarity matching is performed between the query feature vector and the stored feature vectors to retrieve visually similar images. To accelerate the retrieval process, the search is confined to a single cluster that exhibits the minimum centroid distance to the query feature, rather than searching the entire dataset as shown in Fig. 2. This process is as follows,

Step 1). For a query feature vector φ^q, the Euclidean distance between the query and each cluster centroid C_j, where $j = 1, 2, \ldots\ldots, N_c$, is computed. The nearest cluster is determined by the minimum centroid distance d_j^c, defined as

$$d_j^c = \min_{j\in[1,\,N_c\,]} \sqrt{\sum_{j=1}^{N_c} \left(\varphi^q - C_j\right)^2}. \quad (3)$$

Here, N_c denotes the total number of clusters.

Step 2). Once the nearest cluster is identified, the retrieval distances d_k^r between the query feature and all stored feature vectors φ_k^s within that cluster, where $k = 1, 2, \ldots\ldots, n_c$, are computed as

$$d_k^r = \sqrt{\sum_{k=1}^{n_c} (\varphi^q - \varphi_k^s)^2} \,. \quad (4)$$

Here, n_c is the total number of images containing in the selected cluster.

Step 3). The images within the chosen cluster are then ranked in ascending order based on their distance values d_k^r. For a query q, the top-n_q images with the smallest distances are returned as the final retrieval results.

It is noteworthy that our FSCBIR scheme significantly differs from [6]'s scheme in the similarity matching and retrieval module as we incorporated a clustering strategy for faster retrieval. Restricting similarity computation to the most relevant cluster, the proposed method significantly reduces computational complexity and retrieval time compared to [6], as demonstrated in Section III. A.

III. Results and Discussion

To evaluate the effectiveness and robustness of the proposed FSCBIR scheme, experiments were conducted on the publicly available Corel-1K dataset [12]. The dataset comprises 1,000 natural images uniformly distributed across

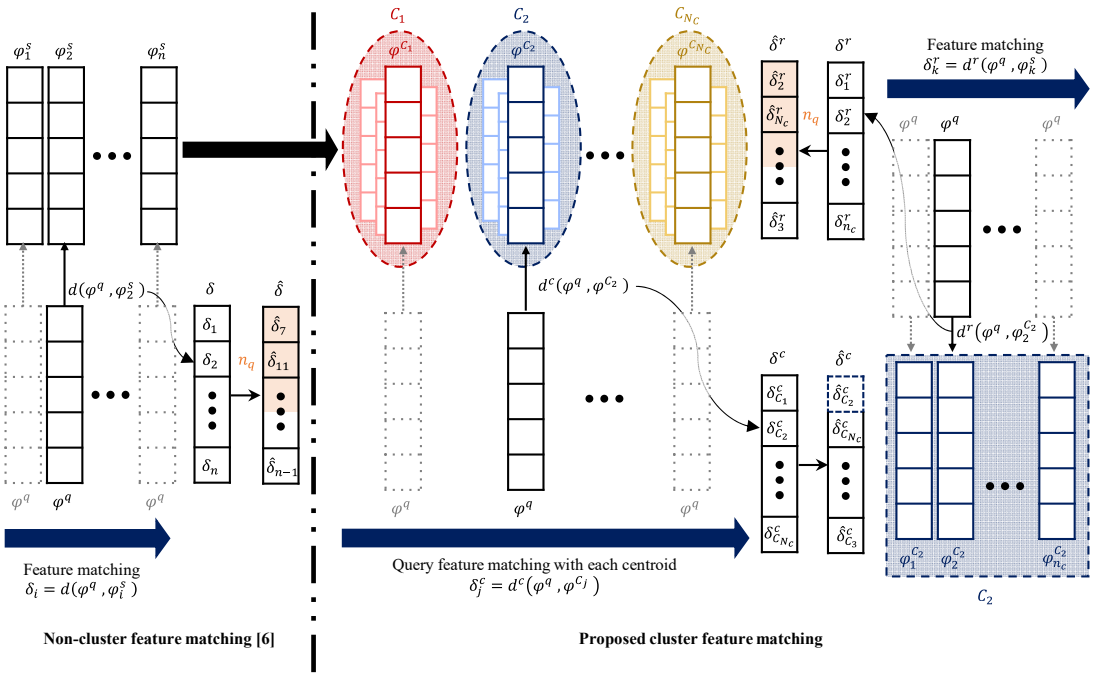


Fig. 2. A comparison of proposed cluster and conventional non-cluster [6] feature matching techniques.

10 semantic categories, namely Africans, Beaches, Buildings, Buses, Dinosaurs, Elephants, Flowers, Food, Horses, and Mountains, with 100 images per class. For performance evaluation, 80% of the images from each class were used for training, while the remaining 20% were used for testing. The training images serve as stored images, while the testing images act as query images. Furthermore, a 5-fold cross-validation strategy was employed to ensure comprehensive evaluation, where the dataset was partitioned into five distinct sets of training and testing images such that all images were used for testing exactly once across different folds.

The performance of our proposed FSCBIR scheme was tested using average retrieval precision (*ARP*), average retrieval recall (*ARR*) and mean average precision *(mAP)* metrics. In general, precision measures the proportion of retrieved images that are relevant to a given query, whereas recall quantifies the system's ability to retrieve all relevant images. The *mAP* score, on the other hand, evaluates the overall retrieval effectiveness by incorporating both the relevance and ranking of the retrieved results across all queries. The three-evaluation metrics are defined in (5), (6) and (7), respectively.

$$ARP = \frac{1}{Q}\sum_{q=1}^{Q}\frac{\left|\mathcal{D}_q^l \cap \mathcal{D}_q^r\right|}{\left|\mathcal{D}_q^r\right|}. \tag{5}$$

$$ARR = \frac{1}{Q}\sum_{q=1}^{Q}\frac{\left|\mathcal{D}_q^l \cap \mathcal{D}_q^r\right|}{\left|\mathcal{D}_q^l\right|}. \tag{6}$$

$$mAP = \frac{1}{Q}\sum_{q=1}^{Q}\left(\frac{1}{n_q}\sum\nolimits_{k=1}^{n_q}\frac{P_q@k}{k}\right). \tag{7}$$

Where, Q is the total number of query images; $\mathcal{D}_q^r$ and $\mathcal{D}_q^l$ respectively represent the sets of retrieved and relevant images for the q^{th} query; n_q is the number of retrieved images for a query q; and $P@k$ is the number of true matches at k^{th} position in a selected cluster which contains n_c images. A higher value of each metric indicates superior retrieval performance of a scheme.

A. Retrieval Performance Analysis

In this subsection, we compared our FSCBIR scheme with existing methods to show its efficacy. First, we considered a baseline secure image retrieval algorithm proposed in [6]. Then, to analyze the impact of sub-block processing, we carried out a comparison with a block-level encryption method proposed in [9] by extending its applications to secure image retrieval task. Furthermore, to assess the trade-off between security and retrieval accuracy, each secure approach was compared with a non-secure image retrieval scheme proposed in [11]. Finally, the retrieval performance was evaluated under both cluster-based and non-cluster-based retrieval settings. In all secure methods employing a block-based encryption structure, we used a fixed block size of (16×16) pixels. For sub-block processing, we considered block sizes $\acute{E}_B \in \{(8\times 8), (4\times 4)\}$, as these are the optimal sub-block sizes for secure image retrieval application demonstrated in [6]. Furthermore, the results were computed using feature extraction block size $F_B = (8\times 8)$ in the CEDD algorithm.

Table 1. presents a comprehensive comparison of retrieval performance across cluster-based and non-cluster-based approaches under both secure and non-secure settings. The table reports the *mAP* scores for ten retrieved images (*mAP@10*) per query across five different folds, along with their averaged results. Additionally, the average retrieval time (in milliseconds) is included to assess computational efficiency. The results indicate a consistent performance gap of approximately 2~5% between secure and non-secure methods in both cluster and non-cluster configurations,

Table 1. Comparison of plain and various encryption schemes using the Corel 1K with cluster and non-cluster approaches.

	Schemes	$\acute{E}_B$	mAP@10						Time (ms)
			K=1	*K=2*	*K=3*	*K=4*	*K=5*	*Mean*	
Non-cluster	Plain [11]	_	0.822	0.775	0.786	0.787	0.793	0.793	0.058
	[9]	_	0.763	0.732	0.737	0.742	0.747	0.744	0.057
	[6]	8×8	0.763	0.731	0.743	0.742	0.750	0.746	0.057
		4×4	0.785	0.754	0.770	0.767	0.776	0.770	0.058
Cluster	Plain [11]	_	0.765	0.725	0.736	0.745	0.749	0.7440	0.021
	[9]	_	0.742	0.667	0.674	0.691	0.690	0.6929	0.021
	FSCBIR	8×8	0.739	0.672	0.679	0.689	0.695	0.6947	0.020
		4×4	0.753	0.712	0.715	0.712	0.706	0.7195	0.020

reflecting the expected trade-off between security and retrieval accuracy. Furthermore, when comparing cluster-based to non-cluster-based approaches, a marginal reduction in *mAP* (around 5%) is observed across all methods. However, this slight decrease in accuracy is offset by a threefold improvement in retrieval speed, confirming the effectiveness of the clustering strategy in significantly reducing search time without substantial accuracy loss. Importantly, when compared with the baseline block-based encryption technique [9], the proposed method demonstrated superior retrieval accuracy. For example, the best recorded *mAP* of 0.7195 (Ours 4×4) outperforms the baseline by approximately 3%.

In addition, the average precision-recall curves are presented in Fig. 3., for the cluster-based configuration across three methods: the non-secure [11], baseline encryption [9], and our FSCBIR scheme. The number of retrieved images varied from 1 to 12 in Fig. 3. It can be observed that as the number of retrieved images increased, the performance differences among different methods became more pronounced. Specifically, both configurations of the proposed method using (4×4) and (8×8) sub-block sizes achieved higher retrieval accuracy compared to [9] as the number of retrieved images increased.

Overall, these findings highlight that the encryption adjustment introduces only a marginal impact on retrieval performance, confirming that the integration of sub-block processing strengthens security without significantly degrading accuracy. Although the proposed cluster-based approach results in a slight reduction in retrieval accuracy compared to [6], it provides a substantial improvement in search efficiency, making it a highly practical solution for secure and scalable image retrieval applications.

B. Security Analysis

The security of the proposed scheme is evaluated in terms of key space size and resilience against JPS attacks as below.

Key space Analysis:

Let an image I with dimension $H \times W$ is divided into N non-overlapping blocks, each of size $E_B \times E_B$. Each block is further partitioned into sub-block of size $S_B \times S_B$ resulting N_s sub-blocks per block. Now the total number of blocks and sub-blocks, along with the corresponding key spaces for the baseline method [9] and the proposed scheme, are defined in (8), (9) and (10), respectively.

$$N = \frac{H}{E_B} \times \frac{W}{E_B}, N_s = \frac{E_B}{S_B} \times \frac{E_B}{S_B}. \tag{8}$$

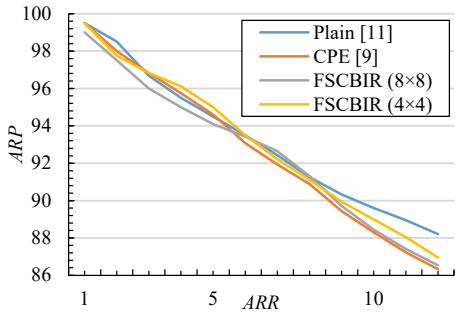


Fig. 3. Precision-Recall curves for different methods of cluster approach (retrieved images varied between 1 to 12).

$$K_{[9]} = K_1 \cdot K_2 \cdot K_3, \\ = N! \cdot 8^N \cdot 2^N. \tag{9}$$

$$K_{[ours]} = K_1 \cdot K_2 \cdot \acute{K}_3, \\ = N! \cdot 2^N \cdot (8^N \cdot 8^{N_s}). \tag{10}$$

It is evident that the proposed scheme key space is 8^{N_s} times larger than that of [9], significantly increasing resistance against brute force attacks.

JPS Attack Analysis:

JPS attacks [10] aim to reconstruct encrypted images by reordering shuffled blocks based on boundary and texture similarities [6] [10]. To evaluate the resistance of the proposed scheme, three standard metrics are employed: Direct Comparison (Dc), Neighbor Comparison (Nc), and Largest Component Comparison (Lc). Let I_r denote the reconstructed image. Then these metrics can be defined as below.

Direct Comparison (Dc): Measures the proportion of blocks correctly positioned as,

$$D_c(I_r) = \frac{1}{n}\sum_{i=1}^{n} d_c(i),$$

$$d_c(i) = \begin{cases} 1, & \text{if } I_r(i) \text{ is in correct position,} \\ 0, & \text{otherwise.} \end{cases} \tag{11}$$

Neighbor Comparison (Nc): Evaluates the ratio of correctly matched neighboring pairs as ,

$$N_c(I_r) = \frac{1}{B}\sum_{k=1}^{B} n_c(k),$$

$$n_c(k) = \begin{cases} 1, & \text{if } B(k) \text{ is joined correctly,} \\ 0, & \text{otherwise.} \end{cases} \tag{12}$$

Largest Component Comparison (Lc): Represents the proportion of blocks in the largest correctly assembled component as,

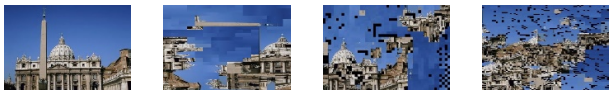
$$L_c(I_r) = \frac{1}{n}\max_j\{I_c(I_r, j)\}, j = 1,2, \dots, m \tag{13}$$

The three measures satisfy $D_c, N_c, L_c \in [0,1]$; lower values indicate stronger encryption resistance. Following [6], the evaluation considered only permutation and rotation operations to simplify computation, as robustness against these transformations implies resilience against more complex JPS attack that take other geometric and color transformations into account.

Table 2. summarizes the average values of D_c, N_c, L_c and recovery time for ten test images evaluated under five different encryption keys. The results indicate that the proposed scheme consistently yields lower D_c, N_c, L_c values than the conventional block-based method [9], reflecting enhanced resistance against JPS-based reconstruction. Furthermore, the proposed method exhibited a higher average recovery time per image, which corresponds to its increased structural complexity and stronger security properties. To support the quantitative findings, Fig. 4. presents visual comparisons of the reconstructed outputs, illustrating that our encrypted images remained visually disordered and resistant to reassembly, unlike those produced by the baseline method.

Table 2. Analysis of searchable encryption resilience against JPS attack (Dc: Direct comparison, Nc: Neighbor comparison, Lc: Large component comparison).

Schemes	$\acute{E}_B$	D_c	N_c	L_c	Time (Sec.)
[9]	_	0.799	0.897	0.888	9.43
FSCBIR	8×8	0.191	0.359	0.374	11.14
	4×4	0.090	0.231	0.218	35.35



(a) (b) (c) (d)

Fig. 4. Visual comparison of JPS recovery: (a) is plain image; (b) is result with the existing encryption scheme [9]; (c) and (d) are the recovered images from our FSCBIR method at sub-block sizes of (8×8) and (4×4), respectively.

IV. Conclusion

This study proposed a computationally efficient and secure image retrieval scheme that integrated a lightweight histogram-based feature extraction method with a sub-block processing–based encryption technique. Also, it incorporated a clustering strategy to reduce the feature matching and retrieval time. Experimental results demonstrated that the proposed approach achieved retrieval performance comparable to non-secure methods while significantly reducing search time and maintaining strong data confidentiality. The efficient clustering mechanism and optimized sub-block encryption collectively made the system a promising solution for secure and scalable image retrieval in cloud environments.

Although the proposed clustering strategy efficiently dealt with the computational complexity of secure image retrieval schemes, it introduced a drop in retrieval accuracy. Therefore, future work will focus on further optimizing the trade-off between retrieval accuracy, computational efficiency, and security robustness to enhance applicability in large-scale multimedia systems.

Acknowledgment

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government. (MSIT) (RS-2023-00278294).

References

[1] D. Srivastava, S. S. Singh, B. Rajitha, M. Verma, M. Kaur, and H.-N. Lee, "Content-Based Image Retrieval: A Survey on Local and Global Features Selection, Extraction, Representation, and Evaluation Parameters," IEEE Access, vol. 11, pp. 95410–95431, 2023, doi: 10.1109/ACCESS.2023.3308911.

[2] E. Akbacak, A. Toktas, U. Erkan, and S. Gao, "MLMQ-IR: Multi-label multi-query image retrieval based on the variance of Hamming distance," Knowl.-Based Syst., vol. 283, p. 111193, Jan. 2024, doi: 10.1016/j.knosys.2023.111193.

[3] W. Tang, X. Zhang, D. Gu, C. Huang, J. Xue, and X. Liang, "Enabling Authorized Fine-Grained Data Retrieval Over Aggregated Encrypted Medical Data in Cloud-Assisted E-Health Systems," IEEE Trans. Cloud Comput., vol. 12, no. 4, pp. 1131–1144, Oct. 2024, doi: 10.1109/TCC.2024.3445430.

[4] M. Li, Y. Zhu, R. Du, and C. Jia, "Verifiable Encrypted Image Retrieval With Reversible Data Hiding in Cloud Environment," IEEE Trans. Cloud Comput., vol. 13, no. 1, pp. 397–410, Jan. 2025, doi: 10.1109/TCC.2025.3535937.

[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, doi: 10.1109/ACCESS.2020.2983194.

[6] M. S. Uzzal, I. Ahmad, and S. Shin, "SCBIR-PE: Secure Content-Based Image Retrieval With Perceptual Encryption," IEEE Trans. Dependable Secure Comput., pp. 1–15, 2025, doi: 10.1109/TDSC.2025.3622246.

[7] I. Ahmad, M. S. Uzzal, and S. Shin, "Secure Retrieval of Brain Tumor Images Using Perceptual Encryption in Cloud-Assisted Scenario," Electronics, vol. 14, no. 9, p. 1759, Apr. 2025, doi: 10.3390/electronics14091759.

[8] P. Yu, J. Tang, Z. Xia, Z. Li, and J. Weng, "A Privacy-Preserving JPEG Image Retrieval Scheme Using the Local Markov Feature and Bag-of-Words Model in Cloud Computing," IEEE Trans. Cloud Comput., vol. 11, no. 3, pp. 2885–2896, July 2023, doi: 10.1109/TCC.2022.3233421.

[9] K. Iida and H. Kiya, "Privacy-Preserving Content-Based Image Retrieval Using Compressible Encrypted Images," IEEE Access, vol. 8, pp. 200038–200050, 2020, doi: 10.1109/ACCESS.2020.3035563.

[10] I. Ahmad and S. Shin, "IIB–CPE: Inter and Intra Block Processing-Based Compressible Perceptual Encryption Method for Privacy-Preserving Deep Learning," Sensors, vol. 22, no. 20, p. 8074, Oct. 2022, doi: 10.3390/s22208074.

[11] C. Iakovidou, L. Bampis, S. A. Chatzichristofis, Y. S. Boutalis, and A. Amanatiadis, "Color and Edge Directivity Descriptor on GPGPU," in 2015 23rd Euromicro International Conference on Parallel, Distributed, and Network-Based Processing, Turku, Finland: IEEE, Mar. 2015, pp. 301–308. doi: 10.1109/PDP.2015.105.

[12] J. Z. Wang, Jia 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, Sept. 2001, doi: 10.1109/34.955109.