International Journal of Informatics and Communication Technology (IJ-ICT)
Vol. 15, No. 3, September 2026, pp. 1376~1384
ISSN: 2252-8776, DOI: 10.11591/ijict.v15i3.pp1376-1384

A multi-expert approach to content-based image retrieval using feature fusion and late re-ranking

Ali Abdulazeez Mohammed Baqer Qazzaz[1], Yousif Samer Mudhafar[1,2]
[1]Department of Computer Science, Faculty of Education, University of Kufa, Najaf, Iraq
[2]Department of Computer Techniques Engineering, Faculty of Technical Engineering, The Islamic University, Najaf, Iraq

Article Info

Article history:

Received Sep 25, 2025
Revised Apr 18, 2026
Accepted Jul 5, 2026

Keywords:

Content-based image retrieval
Discrete cosine transform
Principal component analysis
Scattering wavelet transform
Singular value decomposition

ABSTRACT

As digital data rapidly grows, content-based image retrieval (CBIR) has become important for optimizing collections of visual data. This work proposes a retrieval framework which operates in two stages and improves accuracy by using systematic fusion of features. In the first stage, first-stage wide-scope descriptors called bag-of-visual-words (BoVW), scattering wavelet transform (SWT), discrete cosine transform (DCT), and principal component analysis (PCA) retrieve initial candidate images. The second stage undertakes detailed re-ordering of candidate images by implementing the local binary pattern (LBP), histogram of oriented gradients (HOG), and singular value decomposition (SVD) descriptors to re-evaluate similarity scores. Each individual descriptor returned results for mean average precision for the top 10 retrieved images (mAP, top-10) of between 0.63 and 0.79 and the fused framework achieved 0.88, which is evidence of the viability of complementary feature integration. These findings support the hypothesis that while multiple descriptors performed well and delivered high retrieval accuracy, hierarchical fusion of multiple handcrafted descriptors does not involve the computational costs associated with deep learning methods.

Corresponding Author:

Yousif Samer Mudhafar
Department of Computer Science, Faculty of Education, University of Kufa
Al-Kufa Street, Najaf, Kufa, Iraq
Email: yousifs.mudhafar@uokufa.edu.iq

1. INTRODUCTION

As we enter a visual data-centered world, efficiently accessing images from an enormous, unstructured database in response to a user query has become a pressing problem. Content-based image retrieval (CBIR) systems confront this problem by allowing the user to search images based on visual content-based features such as color, texture, and shape instead of relying on human annotated meta-data. CBIR is important in applications like digital libraries, medical imaging, e-commerce, and law enforcement, where the need for large-scale and accurate retrieval of images continues to grow [1]. CBIR, and especially CBIR that uses deep learning methods based on convolutional neural networks (CNNs), have become a versatile and dominant approach that can learn hierarchical feature representations from very large datasets and can be considered state-of-the-art [2]. However, traditional CNN models function as black boxes and lack explanation of their output: this is especially concerning in high-stakes applications, e.g. medical diagnostics, where an explanation is warranted. Furthermore, CNNs require massive amounts of labeled image data and computational power to train, which makes them impractical in environments with limited access to resources [3].

This research returns to classical handcrafted feature-based image retrieval systems citing desirable properties of efficiency and explainability over black-box deep learning approaches for retrieval. The authors suggest that aggregating several complementary feature extractors, even in a single retrieval application, will provide a richer, more comprehensive, and more interpretable understanding of imagery than any single extracted-descriptor could [4]. Each expert feature extractors describes unique feature characteristics, including texture, structure, and frequency representation, within a simple multi-stage, training-free user improvement framework [5].

The proposed multi-expert system employs several key descriptors: the bag-of-visual-words (BoVW) model captures local keypoints through clustered visual vocabularies; the scattering wavelet transform (SWT) extracts multi-scale structure; the discrete cosine transform (DCT) encodes frequency components; and principal component analysis (PCA) compresses color histograms. Local binary pattern (LBP) and singular value decomposition (SVD) procedures conduct dimensionality reduction and fine ranking. The late-fusion phase employs a weighted vote to aggregate the experts' outputs, including a fine-grained re-ranking to achieve a final accurate retrieval. The key contributions are: i) a hybrid CBIR framework fusing diverse handcrafted descriptors; ii) a weighted voting and re-ranking mechanism for robust fusion; iii) an evaluation demonstrating competitive performance on a large public dataset.

2. RELATED WORK

Many studies have shown progress towards improving CBIR using various methods, particularly through the use of deep learning and hybrid methods. In reference to study [5], a dual-CNN framework used red, green, blue (RGB) and texture-based representations from the various LBP variants, namely orthogonal combination local binary pattern (OC-LBP), center-symmetric local binary pattern (CSLBP), and local directional pattern (LDP), achieved average precision scores of 94.5%, 89.7% and 88.7% respectively on the Corel-1K, Caltech-256 and Oxford 102 Flowers image datasets. Using a greater number of encoder types significantly increases computational complexity.

Study by Magliani *et al.* [6] focused on creating an encryption based solution for providing privacy as well as receiving image returns through a privacy preserving CBIR system which allows users to retrieve image results against encrypted images using aggregation of local image feature data and use of an asymmetric scalar product preserving encryption. While these addresses provide privacy considerations, it generates additional computational overhead. Study by Gayathri [7] proposed a sketch based process CBIR using edge histogram descriptors and scale-invariant feature transform (SIFT) refinements to obtain related image results; however, no quantitative evaluation on a large dataset was provided to show accuracy or viability of this method, thus limiting its overall capabilities to 110 images.

Deep learning-based methods have also shown strong performance. Visual geometry group 16-layer network (VGG16) and residual network-50 (ResNet-50) were both paired with a similarity matching methodology for an overall result of 81.68% precision and 90.18% precision, respectively [8]. Though scalability remains limited. A new loss function called class anchor margin (CAM) introduced by Ghita and Ionescu in [9] is able to enhance feature embedding among non-homogeneous datasets (e.g., CIFAR-100 and ImageNet-200) and offers improved results when compared to contrastive and cross-entropy losses, but has an increased sensitivity to anchor initialization.

Hybrid approaches attempt to balance accuracy and efficiency. The EfficientNet-histogram of oriented gradients (HOG) model in [10] combines deep learning features with hand-crafted features; achieving mAP values of 0.89, 0.85, and 0.83 on the Corel-1K, Oxford5K, and Paris6K datasets, respectively. There are still limitations in scaling the hybrid solution to high-resolution images. Implementing a vision transformer (ViT)-based CBIR system in [11], along with natural language processing (NLP)-driven search and user feedback, achieved an F1-score of 0.7851 on Corel-1K, but this resulted in greater computational complexity and dependence on user interaction.

Overall, deep learning-based CBIR methods, including CNNs and ViTs, achieve high retrieval accuracy but require large labeled datasets and significant computational resources [12]. These methods are associated with high retrieval accuracy, but at a cost of requiring large, labelled datasets and significant computing power [12]. Hand-crafted feature CBIR methods on the other hand typically demonstrate exploit efficiency and interpretability however they often lack semantic richness when considered as standalone CBIR methods. These limitations motivate the proposed multi expert CBIR framework which integrates complementary handcrafted descriptors within a hierarchical fusion and re-ranking structure to balance accuracy efficiency and interpretability. Table 1 summarizes the key differences among existing approaches in terms of datasets, methodologies, performance metrics, advantages, and limitations.

Table 1. Summary of the related works

#	Dataset(s) used	Methodology summary	Best results and metrics	Advantages	Limitations
[5]	Corel-1K, Caltech-256, Oxford 102	Feature fusion enhanced (LDP) images using dual CNN encoders.	Avg. Precision: 94.5% (on Corel-1K)	Robustness to class imbalance; superior precision via feature fusion.	Multiple encoders increase computational cost and training complexity.
[6]	Oxford5K, Paris6K	Deep CNN features with DDR and locVLAD aggregation.	mAP: 0.7188 (on Paris6K)	Enhanced feature aggregation and accuracy	Higher computational cost
[7]	local data (110) images	Sketch-based using Edge Histogram with SIFT.	No metrics provided.	Interface: robust to noise.	Lack of rigorous evaluation; very small dataset.
[8]	Custom data (2,100) images	VGG16/ResNet-50 with similarity matching.	Precision: 90.18% (ResNet-50)	Superior performance; efficient retrieval.	Image downsizing; limited dataset scope.
[9]	CIFAR-100, Food-101, SVHN, ImageNet-200	Introduces CAM loss to optimize embedding.	Superior mAP compared to contrastive and cross-entropy losses.	Eliminates costly pair mining; enforces compact and separable features; efficient.	Sensitive to anchor initialization; marginal gains in low-data regimes.
[10]	Corel-1K, Oxford5K, Paris6K	A hybrid model combining HOG and EfficientNet co-attention mechanism.	mAP: 0.89 (on Corel-1K)	Adaptive feature fusion; computational efficiency.	Reliance on HOG for local features.
[11]	Corel-1K	ViT with NLP and user feedback mechanisms.	F1-score: 0.7851	Semantic query understanding; adaptive learning from user feedback.	High computational overhead; dependency on feedback quality.

3. RESEARCH METHOD

The proposed CBIR framework is established as a complete multi-expert system that does not require data-heavy archival deep learning training. In this process, the proposed framework employs two levels: feature extraction, where the five complementary visual descriptors examine each image under different aspects, and fusion and re-ranking, where the outputs for the five image descriptors are combined to arrive at a final retrieval list that is highly precise, accurate, and robust. This hierarchical architecture leverages the strengths of classical methods while hedging against the weaknesses by use of a simple decision-making process. The complete model architecture is depicted in Figure 1.

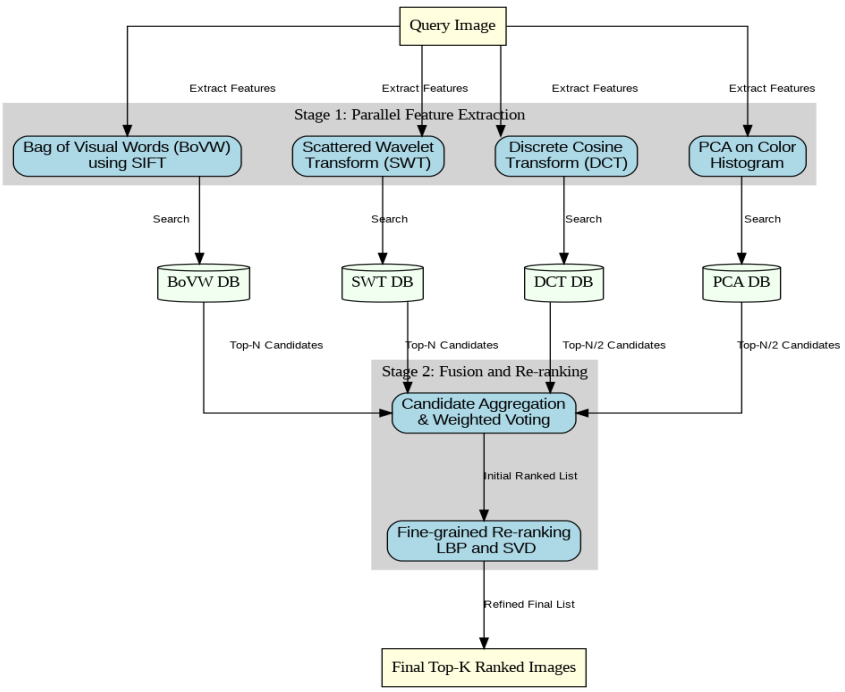


Figure 1. The block diagram of the proposed system

3.1. Feature descriptors overview

The proposed approach combines multiple classical feature descriptors to encode the various visual behaviors in images. For example, SIFT uses gradient-based signatures to identify salient keypoints across multiple scales and unfavorable lighting conditions while ensuring that the process remains invariant to geometric changes and illumination [13]. BoVW approaches images as compact representations of visual-word distributions and generates histogram representations from local feature vectors that can be clustered and are then better suited to representing the content of images [14], [15]. DCT uses cosine functions to decompose images into frequency components, highlighting the low-frequency coefficients that capture the global structure of an image while filtering out noise [16]. LBP encodes relationships between pixels (neighborhoods) into binary patterns based on intensity comparisons alongside their neighbors, producing histograms that are robust to illumination changes [17]. SVD decomposes the image matrix into singular values that correspond to the dominant structural components of the images, producing compact, discriminative descriptors [18], [19]. HOG uses gradient-orientations from cells, while capturing shapes of objects through spatially-aggregated histogram-based gradient orientations to encode the features, and normalizing the spatial gradient orientations histogram to assure robustness to illumination changes [20].

After computing these complementary descriptors, to further improve retrieval robustness, the process of feature fusion integrates multiple descriptors by either concatenating the resultant feature vector or combining the decision of directing models, creating the ability for one descriptor to balance or compensate for the limitations/effects of another descriptor [21], [22]. For the system evaluation, we adopt the mean average precision (mAP) evaluated on the topped ten, averaged across queries and images for reliability, computation of mAP metric (mAP, top-10). The model is evaluated on the Intel Image Classification dataset, which consists of 14 034 training images and 3000 testing images across segregated into seven categories-buildings, forest, glacier, mountain, sea, and street-capturing varied textures, lighting, and viewpoints [23]. Figure 2 shows representative sample images from each of these categories.

Figure 2. Images from the intel image classification dataset

3.2. Feature extraction

To ensure a comprehensive representation, seven feature extraction modules are applied to each image, each capturing distinct visual properties:

- The BoVW model uses local features that were obtained by extracting SIFT features to detect extrema within the DoG space, generating 128-dimensional descriptors. The BoVW model uses Mini-Batch K-means clustering (k=200) to create a visual vocabulary, with each image subsequently represented by a normalised histogram of visual word occurrences.
- The SWT extracts translation-invariant features via the use of multi-scale filtering (approximately J=4) across multiple orientations to produce features. Coefficients are produced and averaged to produce a feature vector that represents the structural patterns contained within the original images.
- The DCT transforms images to a frequency-based representation. When working with RGB channels, the 16x16 low-frequency coefficients in the top left corner were selected from each of the RGB channels, averaged and concatenated to produce a compact descriptor. The compact descriptor was created by ensuring that it is robust to noise.
- PCA was performed on a colour histogram (RGB histogram, 512-dimensional space created via 8x8x8 bins); using PCA the histogram was reduced from a 512-dimensional space to a 64-dimensional space by selecting the principal components that preserve the majority of the variance found within that space to create a compact colour descriptor.
- LBP provides a way to encode the local texture via thresholding of the neighbourhood pixels in order to produce binary patterns, and subsequently aggregate those into histograms that are robust to illumination changes, thus producing an LBP feature.
- SVD decomposes the input image matrix into singular values; SVD allows for the identification of the dominant structural characteristics of the original matrix by retaining the largest singular values. As such, SVD produces a compact feature descriptor.

- HOG is used to determine the shape of an object using the direction (orientation) of the gradient by calculating an orientation-histogram for areas near a localized region and normalizing the values so that changes in ambient light do not interfere with the captured values.

These descriptors collectively encode complementary aspects of visual content, including local structure, global frequency, color distribution, and texture patterns.

3.3. Fusion and re-ranking framework

The central mechanism to this system is a late fusion technique whereby data from multiple descriptors contribute toward improving accuracy associated with retrieved images.

- Weighted voting is the basis for the implementation of feature-based descriptors (BoVW, SWT, DCT, PCA), where four features are originally ranked to produce initial lists of ten ranking images (N=10). These lists are merged into a candidate pool. Each image is assigned a score based on weighted voting, where weights are derived from the empirical performance (mAP, top-10) of each descriptor (e.g., SWT = 0.43, PCA = 0.17), and the final score is computed as the sum of contributing weights.
- To further improve precision, a second-stage re-ranking is performed using a combined descriptor consisting of LBP, HOG, and SVD. Each descriptor facilitates the identification of local texture (LBP), shape (HOG), and structure (SVD) characteristics. Cosine similarity is computed between the concatenated LBP-HOG-SVD descriptors of the query image and candidate images. Final ranking is determined using a two-level criterion: (i) descending weighted-vote score and (ii) ascending similarity distance. This ensures that images consistently selected by multiple descriptors and exhibiting high visual similarity are prioritized.

4. RESULTS AND DISCUSSION

This section explains the visual outcomes of the proposed system with the performance results of the system as illustrated in Table 2, comparison with the related systems, an ablation study, analysis of the results, advantages of the proposed system, limitations of the system, and finally, the suggested related work.

Table 2. Results of the proposed system

Class	Input image	Retrieval images	Efficiency metrics
Street			Precision of 6: 1.00
Sea			Final precision of 5: 0.80
Mountain			Final precision of 8: 0.75
Glacier			Final precision of 7: 1.00
Forest			Final precision of 9: 1.00
Street			Precision of 6: 1.00
Buildings			Final precision of 5: 1.00

The proposed hierarchical fusion system demonstrated highly competitive performance, achieving a mean average precision for the top 10 retrieved images of 0.88 on the intel image classification dataset.

The performance of each of the five core feature descriptors will be evaluated. This isolates the inherent discriminative power of each feature. The results, presented in Table 3, establish a performance baseline for each method and indicate a robust ability to retrieve semantically relevant images consistently within the top results.

Table 3. Performance of individual feature descriptors

Feature descriptor	Role in system	mAP, top-10 (Individual performance)
BoVW	Broad filtering	0.72
SWT	Broad filtering	0.68
DCT	Broad filtering	0.65
PCA on color histograms	Broad filtering	0.79
SVD	Fine-grained ranking	0.76
LBP	Fine-grained ranking	0.63

4.1. Comparison

In order to appropriately locate the proposed model within the current state of activity taking place utilizing CBIR systems, a comparison was performed against other currently available techniques active within this area employing various model architecture types as seen in Table 4.

Table 4. Comparative analysis of retrieval performance with state-of-the-art methods

#	Methodology summary	Dataset	Best results
[5]	Dual-CNN encoders on enhanced LDP images.	Corel-1K	Average precision: 94.5%
[6]	Deep CNN features with DDR and locVLAD aggregation.	Paris6K	mAP: 0.7188
[8]	Fine-tuned ResNet-50 with similarity matching.	Custom (2,100)	Precision: 90.18%
[10]	Hybrid model: EfficientNet with HOG and co-attention.	Corel-1K	mAP: 0.89
Proposed system	Hierarchical fusion: (BoVW+SWT+DCT+LBP) and (PCA+SVD)	Intel image	mAP, top-10: 0.88

4.2. Analysis of comparison

The results demonstrate that the proposed multi-expert framework achieves a strong balance between accuracy, efficiency, and interpretability. The proposed multi-expert framework allows for the generation of candidate images from broad descriptors (e.g., BoVW, SWT, DCT), and the refinement of image ranking from fine-grained descriptors (e.g., LBP, HOG, SVD). A multi-expert CBIR system based on hierarchical fusion and sub-ranking allows for an effective means by which to capture complementary visual features. Although hybrid and deep models may achieve slightly higher accuracy in some cases [5], [10], their performance depends on high computational resources and training data.

4.3. Ablation study

To assess the contribution of various individual components, as well as to assess efficacy as full, we conducted a full ablation study utilizing the the image classification dataset from the intel dataset. All evaluations were performed using the mean average precision for top 10 retrieved images metrics. The results are briefly summarized in Table 4, with fine-grained ranking features (PCA and SVD) having the best accuracy performance on their own, which confirms their use in a final refinement step. Table 5 shows individual contributions and incremental improvements at each hierarchy level, which shows the cumulative changes from architectural decisions of size and scale (objectively broad-filtering models to a final consensus step) that achieve optimal retrieval performance. The results yield two important observations: First, as shown in Table 3, no single descriptor performs well enough as a stand-alone topic descriptor – even the best individual feature, PCA, still performs significantly worse than the entire system; Second, the results highlight a clear hierarchy of strength with combining three methods for broad-filtering to build an initial pool of candidates, which had the largest performance increase (+0.09 mAP) justifying that these features are capturing visual cues that are complementary to each other. After the constructed pool of candidates was filtered and the candidates were fine-grained re-ranked based on pairing PCA, this performed better by another +0.04 mAP, to the peak performance of 0.88 from a rank consensus between PCA and SVD rankings.

4.4. Limitations

Despite attaining notably high retrieval precision in our proposed system, existing limitations exist. First, fusion weights among descriptors are based on empirical rather than adaptive optimization, likely constraining generalization to unseen datasets. Second, combinations of hand-crafted features may not capture semantic variance or high-level contextual cues as well as django-style or other deep-learning based tasks. Third, the proposed framework relies upon fixed feature dimensionality for every descriptor,

which may limit scalability to much larger image datasets. Finally, while the architecture is training free, it is incapable of online adaptability - and slight shifts in dataset domain, or visual conditions, may require manual features normalization and descriptor weight updates.

Table 5. Incremental performance of the hierarchical system

System configuration	Description	mAP, top-10	Performance gain
Stage 1 only	Results from the best single broad-filter (BoVW).	0.72	(Baseline)
Stage 1 + pool	Merged candidate pool from all three broad-filters (BoVW + SWT + DCT).	0.81	+0.09
Stage 1 + Re-rank (PCA only)	Candidate pool re-ranked using only the PCA feature.	0.85	+0.04
Proposed system	Final hierarchical model with consensus (PCA ∩ SVD).	0.88	+0.03

4.5. Future works

Future work will place an emphasis on extending the proposed framework to facilitate greater flexibility and scalability. A potential way forward is to create an adaptive weighting approach to adjust the contributions of every descriptor based on the individual query images. As well, another direction of future work will be to evaluate whether or not lightweight deep features and/or self-supervised representations could be brought into the hybrid fusion process in an interpretable way. In addition, a third future direction is to pursue real-time integration on edge devices and demonstrations over a variety of CBIR benchmarks to show utility ability in real-world scenarios. Future frameworks also may include mechanisms that allow for users to control anything from guided to interactive retrieval of images and refinement. From an implementation perspective, utilizing FPGA architectures to classify hardware acceleration holds promise for enabling real-time features and processing as highlighted in some domains of recent neural systems identification work [24]. Also, cost-effective paradigms that utilize learning at the edge would assist in extending CBIR applications in education, or embedded settings [25].

5. CONCLUSION

This paper proposed a multi-expert framework for CBIR based on hierarchical fusion and re-ranking of seven hand-crafted descriptors (BoVW, SWT, DCT, PCA, LBP, HOG, and SVD). Evaluations indicated that some individual descriptors yielded mAP (top-10) scores ranging from 0.63 to 0.79, with PCA as the best single descriptor. The proposed fused system returned an mAP score of 0.88. The proposed study offers a competitive level of precision compared with state-of-the-art deep and hybrid CBIR models while remaining transparent, cost-efficient, and independent of large-scale training data. The preceding results substantiate the ability of training-free, interpretable retrieval systems to be utilized in constrained environments, such as embedded or mobile platforms.

FUNDING INFORMATION

The authors state no funding was involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Ali Abdulazeez Mohammed Baqer Qazzaz	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓
Yousif Samer Mudhafar		✓				✓		✓	✓	✓	✓			✓

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

Not applicable, as no new data were generated or analyzed in this study.

REFERENCES

[1] M. S. Sayed, A. A. A. Gad-Elrab, K. A. Fathy, and K. R. Raslan, "Unsupervised content based image retrieval using pre-trained CNN and PCNN features extractors," *International Journal of Intelligent Engineering and Systems*, vol. 16, no. 1, pp. 584–596, Feb. 2023, doi: 10.22266/ijies2023.0228.50.

[2] S. R. Dubey, "A decade survey of content based image retrieval using deep learning," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 5, pp. 2687–2704, May 2022, doi: 10.1109/TCSVT.2021.3080920.

[3] V. K. Indu, M. Charmila, A. Swaroop, and C. L. Sucharitha, "Content-based image retrieval system using machine learning," *International Journal of Creative Research Thoughts*, vol. 13, no. 3, pp. 556–563, 2025.

[4] S. M. Chavda and M. M. Goyani, "Robust content-based image retrieval using ICCV, GLCM, and DWT-MSLBP descriptors," *Computer Science*, vol. 23, no. 1, Mar. 2022, doi: 10.7494/csci.2022.23.1.3821.

[5] C. Palai, P. K. Jena, S. R. Pattanaik, T. Panigrahi, and T. K. Mishra, "Content-based image retrieval using encoder based RGB and texture feature fusion," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 3, 2023, doi: 10.14569/IJACSA.2023.0140328.

[6] F. Magliani, T. Fontanini, and A. Prati, "A dense-depth representation for VLAD descriptors in content-based image retrieval," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 11241 LNCS, 2018, pp. 662–671. doi: 10.1007/978-3-030-03801-4_58.

[7] V. Gayathri, "An analysis of content-based image retrieval system," *International Journal of Research in Advanced Engineering and Technolog*, vol. 9, no. 4, pp. 19–23, 2023.

[8] G. Gautam and A. Khanna, "Content based image retrieval system using CNN based deep learning models," *Procedia Computer Science*, vol. 235, pp. 3131–3141, 2024, doi: 10.1016/j.procs.2024.04.296.

[9] A. Ghita and R. Ionescu, "Class anchor margin loss for content-based image retrieval," in *Proceedings of the 16th International Conference on Agents and Artificial Intelligence*, SCITEPRESS - Science and Technology Publications, 2024, pp. 848–853, doi: 10.5220/0012400500003636.

[10] R. Kumar and N. M. M S, "Relevance-aware content-based image retrieval using deep hybrid feature extraction," *Engineering, Technology & Applied Science Research*, vol. 15, no. 3, pp. 22976–22982, Jun. 2025, doi: 10.48084/etasr.10767.

[11] A. Mahajan, N. Rane, and P. Patil, "Content-based image retrieval system utilizing vision transformer," *International Journal for Research Trends and Innovation*, vol. 10, no. 2, pp. 786–790, 2025.

[12] C. Zhang and J. Liu, "Content based deep learning image retrieval: A survey," in *Proceedings of the 2023 9th International Conference on Communication and Information Processing*, New York, NY, USA: ACM, Dec. 2023, pp. 158–163, doi: 10.1145/3638884.3638908.

[13] 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.

[14] A. Dong, Y. Zhang, W. Li, and M. Chen, "A weighted bag of visual words model for predicting fetal growth restriction at an early stage," *Frontiers in Medicine*, vol. 12, Jun. 2025, doi: 10.3389/fmed.2025.1529666.

[15] Pooja and M. Arora, "Text and image classification using shape context and bag of visual words," *International Journal of Scientific Research & Engineering Trends*, vol. 10, no. 3, pp. 785–789, 2024.

[16] R. A. Asmara, R. Agustina, and Hidayatulloh, "Comparison of discrete cosine transforms (DCT), discrete Fourier transforms (DFT), and discrete wavelet transforms (DWT) in digital image watermarking," *International Journal of Advanced Computer Science and Applications*, vol. 8, no. 2, 2017, doi: 10.14569/IJACSA.2017.080232.

[17] S. A. B. Pacheco, M. Goyani, Z. G. Rehman, S. F. Rehman, T. Champaneria, and S. Goyani, "Enhanced content-based image retrieval using multivisual features fusion," *International Journal of Computers and Applications*, vol. 47, no. 10, pp. 835–856, 2025, doi: 10.1080/1206212X.2025.2536577.

[18] S. Chavda and M. Goyani, "Multi-label remote sensing scene classification using two-level double channel spatial attention residual blocks," *Journal of Applied Remote Sensing*, vol. 18, no. 3, p. 036511, Sep. 2024, doi: 10.1117/1.JRS.18.036511.

[19] A. A. M. B. Qazzaz and N. E. Kadhim, "Watermark based on singular value decomposition," *Baghdad Science Journal*, vol. 20, no. 5, pp. 1797–1807, Feb. 2023, doi: 10.21123/bsj.2023.7168.

[20] J. D. Soler *et al.*, "Histogram of oriented gradients: a technique for the study of molecular cloud formation," *Astronomy & Astrophysics*, vol. 622, p. A166, Feb. 2019, doi: 10.1051/0004-6361/201834300.

[21] C.-T. Yen, J.-X. Liao, and Y.-K. Huang, "Evaluating feature fusion techniques with deep learning models for coronavirus disease 2019 chest X-ray sensor image identification," *Sensors and Materials*, vol. 36, no. 2, pp. 683–699, Feb. 2024, doi: 10.18494/SAM4685.

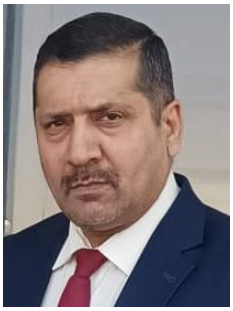
[22] T. A. Al-Asadi and A. A. A. M. Baqer, "Fusion for multiple light sources in texture mapping object," *Journal of Telecommunication, Electronic and Computer Engineering*, vol. 9, no. 2–11, pp. 7–12, 2017.

[23] P. Bansal, "Intel image classification," Kaggle. Accessed: Apr. 18, 2026. [Online]. Available: https://www.kaggle.com/datasets/puneet6060/intel-image-classification

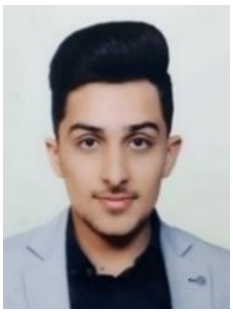
[24] S. H. Abdulnabi, Y. S. Mudhafar, A. A. Kadhim, M. B. Mahdi, and H. H. Sojar, "Neural network-based system identification: a comprehensive FPGA design and implementation," in *2024 IEEE International Conference on Artificial Intelligence and Mechatronics Systems (AIMS)*, IEEE, Feb. 2024, pp. 1–7. doi: 10.1109/AIMS61812.2024.10512531.

[25] A. M. A. Al-muqarm, Y. Mudhafar, A. M. Shakir, M. Kazem, R. Abdel-Yahiya, and B. S. A. Zahra, "Low-cost smart learning with Moodle-based Raspberry Pi 4 for university students," in *2023 6th International Conference on Engineering Technology and its Applications (IICETA)*, IEEE, Jul. 2023, pp. 603–608, doi: 10.1109/IICETA57613.2023.10351266.

BIOGRAPHIES OF AUTHORS



Ali Abdulazeez Mohammed Baqer Qazzaz received the M.Sc. and Ph.D. degrees in Computer Science from Babylon University, Iraq, in 2012 and 2018, respectively. He worked as a lecturer at the University of Kufa, College of Education, Department of Computer Science. His research interests include image processing, computer vision, information security, deep learning, AI, and data mining. He can be contacted at email: alia.qazzaz@uokufa.edu.iq.



Yousif Samer Mudhafar earned his B.Sc. in Computer Techniques Engineering from the Islamic University in Najaf in 2018. He completed his M.Sc. in Computer Science Engineering at the University of Debrecen in 2022, graduating with honors and receiving the Outstanding Student certificate. He works at the University of Kufa, Faculty of Education, Department of Computer Science. His research interests include computer networks, IoT, and AI. He can be contacted at email: yousifs.mudhafar@uokufa.edu.iq.