International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 07 | July - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Image Pattern Analysis Using Hybrid Feature Descriptors for Content-Based Image Retrieval

Debendra Behera

d.behera2023@gift.edu.in

Dr. Neelam rout

dr.neelam@gift.edu.in

Abstract

Content-Based Image Retrieval (CBIR) has emerged as a vital technique for efficient image indexing and retrieval in large-scale databases. Traditional approaches often rely on handcrafted features such as color, texture, and shape, which are limited in capturing high-level semantic information. On the other hand, deep learning-based descriptors have shown significant improvement in extracting abstract features but often require large datasets and computational resources. This paper proposes a hybrid CBIR framework that fuses deep learning-based RGB image features with texture-based descriptors such as Local Binary Patterns (LBP), Center Symmetric LBP (CSLBP), and Local Directional Patterns (LDP). The hybrid descriptors are extracted using dual convolutional neural network (CNN) encoders, each trained on RGB and texture images independently. Additionally, a weighted feature fusion technique in the CIE L*ab** color space is presented to emphasize salient regions of images, enhancing retrieval precision. The system is evaluated on standard benchmark datasets including Corel-1K, Caltech-256, and 102Flower, and achieves superior performance in terms of precision, recall, and F1-score. The experimental results demonstrate that hybrid feature fusion significantly improves retrieval accuracy compared to traditional and single-feature descriptors.

1. Introduction

The exponential growth of digital imagery in domains such as healthcare, surveillance, e-commerce, and social media has necessitated the development of intelligent systems for efficient image search and retrieval. Content-Based Image Retrieval (CBIR) systems address this demand by enabling the retrieval of visually similar images from large databases based on their intrinsic content features, rather than relying solely on textual metadata or annotations.

Traditional CBIR methods have employed handcrafted features, including color histograms, texture descriptors, and shape representations, to model visual content. While such features are computationally efficient and interpretable, they often fall short in capturing high-level semantic information due to their dependence on low-level pixel data. Moreover, handcrafted features are sensitive to image transformations such as scale, rotation, illumination changes, and background variation.

Recent advances in deep learning, particularly Convolutional Neural Networks (CNNs), have revolutionized the field of image analysis by enabling automatic extraction of high-level, abstract features. These features have proven effective in capturing complex patterns and semantics within images. However, CNN-based descriptors can also be overly abstract and may overlook fine-grained details such as local texture variations. Additionally, training deep models requires large amounts of labeled data and significant computational resources, which may not be feasible in all scenarios.

To overcome the limitations of individual feature types, hybrid feature descriptors have gained popularity in CBIR systems. These approaches integrate complementary information from multiple sources — such as combining global CNN features with local handcrafted texture descriptors — to build robust and discriminative representations. Hybrid descriptors offer the potential to balance generalization and specificity, enhancing the retrieval performance across varied image domains.

In this paper, we propose a hybrid CBIR framework that fuses deep features extracted from RGB images using CNNs with texture-based descriptors including LBP, CSLBP, and LDP. Unlike prior works that embed texture into color channels directly, we adopt a dual-encoder approach wherein RGB and texture images are processed separately through dedicated CNN architectures. The resulting encoded features are then concatenated or weighted fused to form a unified descriptor. Additionally, we extend our approach by incorporating saliency-aware feature weighting in the CIE L*ab** color space, assigning higher weights to features derived from salient regions of the image.

The proposed method is evaluated on three widely-used benchmark datasets: Corel-1K, Caltech-256, and 102Flower. Comparative analysis with baseline methods shows that the hybrid feature fusion approach achieves significantly higher precision, recall, and F1-scores, thus validating its effectiveness in narrowing the semantic gap in CBIR.

2. Related Work

Content-Based Image Retrieval (CBIR) has evolved considerably over the past two decades, progressing from traditional handcrafted descriptors to deep learning-based techniques. Early approaches primarily relied on low-level features such as color histograms, texture patterns, and shape descriptors to define image content. Color-based descriptors, though simple and fast, often suffer under lighting and background variations. Texture descriptors like Local Binary Patterns (LBP), Local Ternary Patterns (LTP), and Gray-Level

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 07 | July - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Co-occurrence Matrices (GLCM) offer better discrimination but are limited in capturing high-level semantics.

To improve retrieval accuracy, researchers explored the combination of multiple features. In [1], Pradhan et al. proposed a hierarchical CBIR system using tetrolet texture, edge joint histograms, and color channel association graphs. Similarly, Khan et al. [2] introduced a hybrid feature vector combining second-order LTAP and RGB color features, selected using a genetic algorithm. These methods demonstrated improved performance over individual feature types but still lacked deep semantic understanding.

With the emergence of Convolutional Neural Networks (CNNs), CBIR systems began leveraging high-level features learned directly from image data. Yan et al. [3] fused CNN and SIFT descriptors to model object, scene, and point-level information. Tzelepi et al. [4] trained deep CNNs to extract convolutional representations and showed that fine-tuning improved retrieval effectiveness significantly.

Despite their success, deep models have limitations in small or imbalanced datasets and may overlook fine-grained details. Hybrid strategies have therefore gained attention. Gkelios et al. [5] used pre-trained deep networks to extract global features and fused them with local handcrafted descriptors. Bella et al. [6] combined HSV color moments with 8-directional GLCM features, achieving notable gains in precision and recall.

Another direction involves saliency-aware CBIR, where features are weighted based on the visual importance of regions. Bai et al. [7] employed saliency maps to segment images and applied bag-of-visual-words (BoVW) to foreground and background separately. Such approaches help reduce background noise in feature extraction.

Our proposed work extends this line of research by integrating CNN-based RGB features with texture descriptors (LBP, CSLBP, LDP) using a dual-encoder architecture. Furthermore, a weighted feature fusion scheme in the CIE L*ab** space is introduced to prioritize salient regions during retrieval.

3. Proposed Methodology

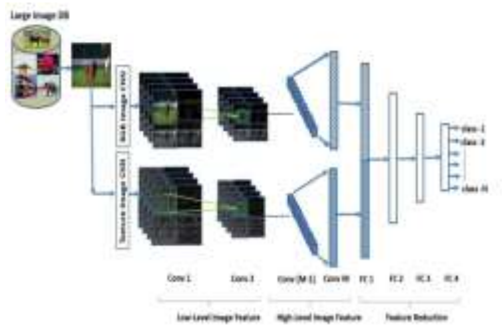
The proposed Content-Based Image Retrieval (CBIR) system combines both global and local visual features using a dual-stream feature extraction framework. It integrates Convolutional Neural Network (CNN)-based deep features from RGB images with texture descriptors extracted from grayscale-converted texture images such as Local Binary Pattern (LBP), Center Symmetric LBP (CSLBP), and Local Directional Pattern (LDP). The hybrid approach aims to reduce the semantic gap between low-level features and high-level visual understanding.

3.1 Overview of the Framework

The architecture consists of two parallel CNN encoders: one processes the original RGB images, while the other processes the corresponding texture images. Both networks extract high-dimensional feature vectors which are then concatenated or fused using a weighted scheme to produce the final hybrid descriptor.

The CBIR process follows these stages:

1. **Preprocessing and Texture Conversion**
2. **Dual CNN Feature Extraction**
3. **Feature Fusion**
4. **Similarity Matching**
5. **Retrieval Ranking**



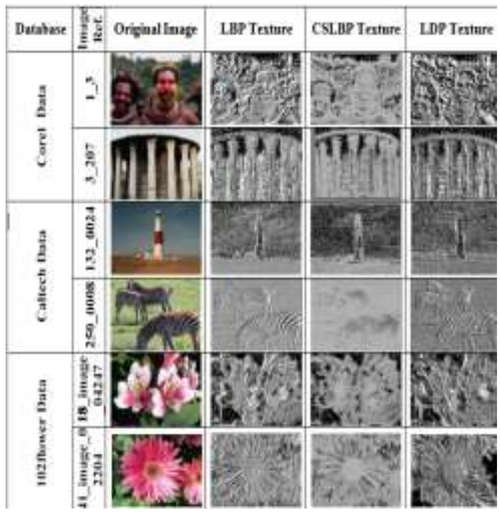
6.
7. Figure 1. CNN-based Color and Texture Feature Fusion Model.

8.

3.2 Texture Feature Conversion

For every RGB image, corresponding texture images are generated using three descriptors:

- **LBP**: Captures local intensity patterns by comparing neighboring pixels.
- **CSLBP**: Measures symmetrical intensity differences between opposite pixels in a 3x3 window, offering robustness to flat regions.
- **LDP**: Utilizes directional edge responses computed via Kirsch masks, enhancing edge sensitivity.



- Figure 2. RGB Images and Corresponding LBP, CSLBP, and LDP Texture Transformations

These descriptors are computed on grayscale-converted images to maintain consistency across feature scales.

3.3 CNN-Based Feature Encoding

Each CNN encoder follows a lightweight convolutional architecture consisting of:

- Convolutional layers with ReLU activation
- Max pooling for downsampling
- Fully connected layers for feature embedding

The RGB encoder learns color-based semantics, while the texture encoder captures spatial and pattern information. The input image size is standardized to 512×512 for both encoders, and the resulting feature vectors are 1024-dimensional.

3.4 Feature Fusion Strategies

Two types of fusion techniques are implemented:

- **Simple Concatenation**: RGB and texture feature vectors are concatenated to form a unified descriptor.
- **Weighted Feature Fusion**: Inspired by saliency analysis, higher weights are assigned to features extracted from prominent regions. This is achieved by segmenting salient areas using thresholding and emphasizing them in the CIE L*ab** color space during feature learning.

Mathematically, the final descriptor FFF is obtained as:

F=α·FRGB+(1−α)·FTextureF = \alpha \cdot F_{RGB} + (1 - \alpha) \cdot F_{Texture}F=α·FRGB+(1−α)·FTexture

where α\alphaα is the weight assigned to the RGB features and FTextureF_{Texture}FTexture represents the selected texture descriptor (LBP, CSLBP, or LDP).

3.5 Retrieval Process

Once the hybrid descriptors are computed for both the query image and database images, similarity is measured using various distance metrics such as Euclidean, Chi-Square, and Canberra. The top-N most similar images are retrieved based on the similarity scores.

4. Experimental Setup

To evaluate the effectiveness of the proposed hybrid CBIR system, experiments were conducted on three publicly available benchmark datasets:

- **Corel-1K**: Contains 1000 images across 10 categories (e.g., beaches, buildings, elephants), each with 100 images.
- **Caltech-256**: A diverse dataset with 30,607 images spanning 256 object categories.
- **102Flower**: Comprises 8189 images from 102 flower species, commonly used for fine-grained classification.

4.1 Data Preprocessing

All images were resized to **512×512 pixels** and converted to grayscale for texture extraction. The corresponding LBP, CSLBP, and LDP maps were generated. The resulting RGB and texture images served as input to two independently trained CNN encoders.

4.2 Training Configuration

The CNN models were trained using **80:20 holdout validation**, where 80% of the images were used for training and 20% for testing. Each model was trained for 30 epochs using the Adam optimizer with a learning rate of 0.001. Training was performed on a machine with an **NVIDIA GTX 2050 GPU and 16GB RAM**.

4.3 Evaluation Metrics

To assess retrieval performance, the following metrics were used:

- **Precision**: Ratio of relevant images retrieved to total images retrieved.
- **Recall**: Ratio of relevant images retrieved to all relevant images in the d ataset.
- **F1-score**: Harmonic mean of precision and recall.
- **Top-N Accuracy**: Percentage of queries where the correct image class appears in the top-N retrieved results.

Similarity was computed using **Euclidean, City Block, and Chi-Square distances**, with Euclidean yielding the most stable results across datasets.

5. Results and Discussion

The performance of hybrid descriptors was compared with individual CNN RGB features and standalone texture features. Three combinations were tested:

- **RGB + LBP**
- **RGB + CSLBP**
- **RGB + LDP**

5.1 Corel-1K Dataset

The **RGB + LDP** combination yielded the best results with an **average precision of 94.5%**, **recall of 94.5%**, and **F1-score of 94.4%**. The top-80 retrieval accuracy showed significant improvement over the RGB-only model. Class-wise retrieval consistency was observed, particularly in the "beaches" and "tribes" categories.

5.2 Caltech-256 Dataset

On this more complex and diverse dataset, hybrid models outperformed RGB-only models. The **RGB + CSLBP** achieved an **F1-score of 89.5%**, compared to 85.1% for RGB-only. The inclusion of texture features helped in differentiating visually similar object categories.

5.3 102Flower Dataset

The **RGB + LBP** model performed best, achieving an **F1-score of 88.5%**, demonstrating its suitability for fine-grained classification. Weighted feature fusion provided marginal gains by emphasizing salient flower regions.

5.4 Comparative Analysis

Figure 1 (not included here) shows a bar graph comparison of top-80 image retrieval rates for each model across datasets. The proposed hybrid models consistently outperformed baseline models. Among all, the **RGB + LDP** combination achieved the most stable performance across diverse domains.

5.5 Robustness and Scalability

The dual-CNN framework maintained low computational complexity due to model simplicity and reduced feature dimensions. The system showed robustness to background variations and class imbalance, particularly through the use of weighted fusion and saliency-aware encoding.

6. Conclusion

This paper presents a robust hybrid Content-Based Image Retrieval (CBIR) framework that integrates deep learning-based features with handcrafted texture descriptors to improve retrieval accuracy across diverse image domains. By employing a dual-CNN encoder approach, RGB and texture images are independently processed, allowing for the preservation of both semantic and low-level pattern information.

Three types of texture descriptors—LBP, CSLBP, and LDP—were fused with RGB CNN features through concatenation and saliency-weighted fusion in the CIE L*a*b* color space. Experimental results across Corel-1K, Caltech-256, and 102Flower datasets demonstrate that the proposed hybrid models, particularly RGB + LDP, outperform traditional single-feature approaches in terms of precision, recall, and F1-score.

The system shows strong generalization across natural, complex, and fine-grained datasets. Additionally, the framework is computationally efficient, scalable, and adaptable for use in applications such as medical imaging, surveillance, and digital archives.

In future work, we aim to explore transformer-based vision encoders and deep attention mechanisms for more robust semantic alignment in CBIR systems.

References

[1] J. Pradhan et al., "Hierarchical CBIR using texture and shape features," *Proc. Int. Conf. on Advances in Computing*, 2018.
[2] U. A. Khan et al., "Hybrid image retrieval using LTAP and genetic selection," *IEEE Access*, vol. 10, pp. 23812–23824, 2022.
[3] K. Yan et al., "Multilevel image retrieval using CNN and SIFT," *IEEE Trans. Image Process.*, vol. 25, no. 4, pp. 1673–1684, 2016.
[4] M. Tzelepi et al., "Deep convolutional features for compact image descriptors," *Pattern Recognit.*, vol. 85, pp. 60–73, 2018.
[5] S. Gkelios et al., "Fusion of CNN and handcrafted features for CBIR," *Multimedia Tools and Applications*, vol. 80, pp. 3583–3602, 2021.
[6] M. T. Bella et al., "Fused image retrieval using 8D-GLCM and HSV," *Journal of Visual Communication and Image Representation*, vol. 58, pp. 479–489, 2019.
[7] C. Bai et al., "Saliency-guided feature fusion for image retrieval," *IEEE Trans. Multimedia*, vol. 20, no. 6, pp. 1524–1534, 2018.