Content-Based Image Retrieval Using Color, Texture, Shape, and Deep Features: A Review

Rani Lodhi[1], Ashish Kumar Tiwari[2]
[1]Research Scholar, [2] Assistant Professor
Department of CSE, SAM Global University, Raisen, India

Abstract: The rapid growth of digital images in fields such as medicine, education, surveillance, remote sensing, social media, digital libraries, and e-commerce has created a strong need for efficient image storage, indexing, and retrieval systems. Traditional image retrieval systems mainly depend on text-based annotations, keywords, or metadata. However, manual annotation is time-consuming, subjective, and often unable to fully describe the visual content of images. Content-Based Image Retrieval (CBIR) addresses this limitation by retrieving images based on their visual features rather than textual descriptions. The major low-level features used in CBIR include color, texture, and shape. These features help represent the visual content of an image in numerical form and allow the system to compare images using similarity measures. In recent years, deep learning techniques, especially Convolutional Neural Networks (CNNs), have significantly improved CBIR performance by automatically learning high-level semantic features from images. This paper presents a review of CBIR techniques based on color, texture, shape, and deep features. It discusses major feature extraction methods, similarity measurement techniques, applications, challenges, and future research directions. The review concludes that hybrid CBIR approaches combining handcrafted features with deep learning features can improve retrieval accuracy, reduce the semantic gap, and support large-scale image retrieval applications.

Keywords: Content-Based Image Retrieval, CBIR, color features, texture features, shape features, deep learning, CNN, feature extraction, similarity measurement.

How to cite this article: Rani Lodhi, Ashish Kumar Tiwari. (2026). Content-Based Image Retrieval Using Color, Texture, Shape, and Deep Features: A Review, International Journal of Scientific Modern Research and Technology (IJSMRT), ISSN: 2582-8150, Volume-24, Issue-01, Number-02, July-2026, pp. 10-16, URL: https://www.ijsmrt.com/wp-content/uploads/2026/07/IJSMRT-26070102.pdf

IJSMRT-26070102

I. INTRODUCTION

In the modern digital era, images are generated and stored in massive quantities through smartphones, satellites, medical imaging devices, surveillance systems, social media platforms, and web-based applications. Efficient retrieval of relevant images from large image databases has therefore become an important research problem in computer vision and information retrieval. Traditional image retrieval systems generally depend on text-based indexing, where images are searched using keywords, captions, file names, or manually assigned tags. Although text-based retrieval is simple, it suffers from several limitations. Manual annotation is time-consuming, expensive, and subjective. Different users may describe the same image using different words, which can reduce retrieval accuracy [1].

Content-Based Image Retrieval (CBIR) provides an alternative approach by retrieving images based on their actual visual content. Instead of depending only on text, CBIR systems analyze image features such as color, texture, shape, edges, regions, and spatial patterns. These visual features are extracted automatically and converted into feature vectors. When a user submits a query image, the system compares its feature vector with the feature vectors of images stored in the database and retrieves the most similar images [2].

Early CBIR systems mainly focused on low-level visual features such as color histograms, texture descriptors, and shape descriptors. These handcrafted features are simple and computationally efficient, but they often fail to capture the semantic meaning of an image. This limitation is commonly known as the semantic gap, which refers to the difference between low-level visual features and high-level human interpretation. For example, two images may have similar colors and textures but may represent completely different objects or scenes [3].

With the advancement of artificial intelligence, machine learning, and deep learning, CBIR has moved from handcrafted feature-based retrieval to deep feature-based retrieval. Deep learning models, particularly CNNs, can automatically learn hierarchical features from image data. Lower layers capture edges and basic textures, while deeper layers capture object parts and semantic patterns. As a result, deep features have improved the accuracy and robustness of modern CBIR systems [4].

The objective of this review paper is to examine the role of color, texture, shape, and deep features in CBIR. The paper also highlights major challenges and future directions in the development of efficient and intelligent image retrieval systems.

II. CONCEPT OF CONTENT-BASED IMAGE RETRIEVAL

Content-Based Image Retrieval is a technique used to search and retrieve images from a database based on the visual content of images. A typical CBIR system consists of the following steps:

- Image database creation
- Feature extraction from images
- Feature vector storage
- Query image submission
- Query feature extraction
- Similarity matching
- Ranking and retrieval of relevant images

The general workflow of CBIR begins with the extraction of image features from all database images. These features are stored in the form of numerical vectors. When a user provides a query image, the same feature extraction process is applied to the query image. The query feature vector is then compared with the stored feature vectors using similarity or distance measures such as Euclidean distance, Manhattan distance, cosine similarity, or histogram intersection. Finally, images with the highest similarity or lowest distance are retrieved and displayed to the user [5].

CBIR differs from text-based image retrieval because it does not depend mainly on image captions or keywords. Instead, it uses measurable visual characteristics. This makes CBIR useful in applications where textual annotation is unavailable, incomplete, or unreliable.

III. COLOR-BASED IMAGE RETRIEVAL

Color is one of the most widely used features in CBIR because it is simple, intuitive, and relatively stable. Color features describe the distribution of colors in an image and are useful for retrieving images with similar visual appearance.

Color Histogram

A color histogram represents the number of pixels belonging to different color bins. It is one of the earliest and most common methods used in CBIR. In this method, the color space is divided into bins, and the frequency of each color bin is calculated. Images with similar color distributions are considered visually similar [6].

Advantages: Easy to compute, Independent of image size and orientation, Useful for general image retrieval

Limitations: Ignores spatial arrangement of colors, Different images may have similar histograms, Sensitive to lighting variations

Color Moments

Color moments describe the statistical properties of color distribution. The most commonly used moments are mean, standard deviation, and skewness. These features provide compact representation compared to full histograms.

Advantages: Low-dimensional representation, computationally efficient, Suitable for large databases

Limitations: May lose detailed color information, Not suitable for complex image scenes.

Color Correlogram

A color correlogram captures both color distribution and spatial correlation between colors. It measures how often a pixel of one color occurs at a certain distance from another pixel of the same or different color.

Advantages: Includes spatial color information, more effective than simple histograms.

Limitations: Higher computational complexity, larger feature vector size.

Common Color Spaces

CBIR systems commonly use RGB, HSV, YCbCr, and Lab color spaces. RGB is simple and commonly used, but it is sensitive to illumination. HSV is closer to human color perception and is often preferred for image retrieval tasks.

IV. TEXTURE-BASED IMAGE RETRIEVAL

Texture refers to the visual pattern or surface structure of an image region. It describes properties such as smoothness, roughness, regularity, coarseness, and directionality. Texture is important in retrieving images such as fabrics, natural scenes, medical images, satellite images, and biological images [7].

Gray-Level Co-occurrence Matrix

The Gray-Level Co-occurrence Matrix (GLCM) is a statistical method used to analyze the spatial relationship between pixels. It calculates how often pairs of pixel intensities occur at a given distance and direction. From the GLCM, features such as contrast, correlation, energy, and homogeneity are extracted.

Advantages: Captures spatial relationship between pixels, Effective for texture classification

Limitations: Sensitive to image rotation and scale, computationally expensive for large gray-level ranges

Local Binary Pattern

Local Binary Pattern (LBP) is a simple and effective texture descriptor. It compares each pixel with its neighboring pixels and creates a binary pattern. LBP is widely used because of its simplicity and robustness to illumination changes.

Advantages: Simple and fast, Robust to monotonic illumination changes, Effective for local texture description

Limitations: Sensitive to noise, Limited ability to capture large-scale texture patterns

Gabor Filter

Gabor filters are used to extract texture information at different scales and orientations. They are useful because they resemble the response of the human visual system.

Advantages: Captures frequency and orientation information, Effective for texture-rich images

Limitations: High computational cost, Requires careful parameter selection

Wavelet Transform

Wavelet transform decomposes an image into different frequency components. It is useful for multi-resolution texture analysis and can capture both local and global texture information.

Advantages: Multi-resolution analysis, Useful for medical and remote sensing image retrieval

Limitations: Feature vector may become large, Performance depends on wavelet type and decomposition level

V. SHAPE-BASED IMAGE RETRIEVAL

Shape is an important feature in CBIR because it represents the structure and boundaries of objects within an image. Shape-based retrieval is especially useful when the object outline or geometry is more important than color or texture [8].

Edge-Based Shape Descriptors

Edge detection methods such as Sobel, Canny, and Prewitt are used to identify object boundaries. The extracted edges are used to describe the shape of objects.

Advantages: Useful for object-based retrieval, Simple and interpretable

Limitations: Sensitive to noise, Requires proper segmentation

Moment Invariants

Moment invariants describe the shape of an object in a way that is invariant to translation, rotation, and scaling. Hu moments are widely used for shape representation.

Advantages: Invariant to basic geometric transformations, Compact representation

Limitation: May not handle complex shapes effectively, Sensitive to segmentation errors

Fourier Descriptors

Fourier descriptors represent the boundary of an object using Fourier transform coefficients. They are useful for capturing shape information in a compact form.

Advantages: Suitable for closed boundary shapes, Can be made invariant to rotation and scale

Limitation: Requires accurate boundary extraction, Less effective for irregular or overlapping objects

Region-Based Shape Descriptors

Region-based descriptors use the entire object region rather than only its boundary. They are useful for describing complex shapes where boundary information alone is insufficient.

VI. DEEP FEATURE-BASED IMAGE RETRIEVAL

Deep learning has transformed CBIR by enabling automatic feature learning from images. Instead of manually designing color, texture, and shape descriptors, deep learning models learn features directly from data [9].

Convolutional Neural Networks

Convolutional Neural Networks are widely used in image retrieval because they can extract hierarchical features from images. The lower layers of a CNN detect edges, corners, and textures, while deeper layers capture object parts and semantic information. Features extracted from pre-trained CNN models such as AlexNet, VGG, ResNet, Inception, and EfficientNet are commonly used for image retrieval.

Advantages: Automatic feature learning, Strong semantic representation, High retrieval accuracy, Robust to complex image variations

Limitations: Requires large datasets, Computationally expensive, May require GPU support, Interpretability remains limited.

Transfer Learning

Transfer learning allows a model trained on a large dataset to be used for feature extraction in another domain. In CBIR, pre-trained CNNs are often used to extract deep features from images. This reduces the need for training a model from scratch.

Advantages: Reduces training time, Useful for small datasets, Improves retrieval performance

Limitations: Features may not be domain-specific, Fine-tuning may be needed for specialized applications

Autoencoders

Autoencoders are neural networks used to learn compact representations of images. They encode images into lower-dimensional feature vectors and reconstruct them. The encoded vectors can be used for image retrieval.

Advantages: Learns compact image representation, Useful for unsupervised learning

Limitations: Reconstruction quality may not always match retrieval quality, Requires careful architecture design

Deep Metric Learning

Deep metric learning aims to learn an embedding space where similar images are placed close together and dissimilar images are placed far apart. Methods such as Siamese networks, triplet loss, and contrastive loss are widely used in modern retrieval systems.

Advantages: Directly optimized for similarity learning, Suitable for fine-grained retrieval

Limitations: Requires carefully selected image pairs or triplets, Training can be computationally expensive

VII. SIMILARITY MEASURES IN CBIR

Similarity measurement is a critical step in CBIR. After feature extraction, the similarity between query image features and database image features is calculated. Common similarity measures include [10]:

Euclidean Distance: Euclidean distance measures the straight-line distance between two feature vectors. Smaller distance indicates higher similarity.

Manhattan Distance: Manhattan distance calculates the sum of absolute differences between feature vector components. It is simple and useful for high-dimensional data.

Cosine Similarity: Cosine similarity measures the angle between two vectors. It is commonly used for deep feature comparison because it focuses on direction rather than magnitude.

Histogram Intersection: Histogram intersection is used to compare color histograms. It calculates the common area between two histograms.

Chi-Square Distance: Chi-square distance is often used for histogram-based features and is useful in comparing color and texture distributions.

VIII. COMPARATIVE OVERVIEW OF CBIR FEATURES

Table 1: Study of Features Types

Feature Type	Common Methods	Strengths	Limitations
Color	Histogram, color moments, correlogram	Simple, fast, useful for general retrieval	Ignores semantics and may be affected by illumination
Texture	GLCM, LBP, Gabor, wavelet	Captures surface patterns and local structure	Sensitive to noise, scale, and rotation
Shape	Edge, moments, Fourier descriptors	Useful for object-based retrieval	Requires accurate segmentation
Deep Features	CNN, autoencoder, Siamese network	High accuracy and semantic representation	Requires large data and computational resources
Hybrid Features	Combination of handcrafted and deep features	Better robustness and accuracy	Feature fusion can increase complexity

IX. APPLICATIONS OF CBIR

CBIR has several important applications across different fields.

Medical Image Retrieval: CBIR is used to retrieve similar medical images such as X-rays, CT scans, MRI scans, ultrasound images, and histopathology images. It helps doctors compare current cases with previous similar cases and supports diagnosis.

Remote Sensing: Satellite image retrieval is useful for land use classification, urban planning, environmental monitoring, and disaster management.

Digital Libraries: CBIR helps organize and retrieve images from large digital archives, museums, and historical collections.

E-Commerce: Online shopping platforms use image-based retrieval to find visually similar products such as clothes, shoes, furniture, and accessories.

Crime Investigation and Surveillance: CBIR can support face retrieval, object matching, vehicle identification, and forensic image analysis.

Education and Research: Researchers and students use CBIR systems to search scientific images, diagrams, artworks, and visual teaching materials.

X. CHALLENGES IN CBIR

Despite major progress, CBIR still faces several challenges.

Semantic Gap: The semantic gap is the most important challenge in CBIR. Low-level features such as color and texture may not fully represent the meaning of an image as understood by humans.

High-Dimensional Feature Vectors: Deep features and combined descriptors may produce high-dimensional vectors, increasing storage and computation requirements.

Illumination and Viewpoint Variations: Images captured under different lighting conditions, scales, rotations, and viewpoints may reduce retrieval accuracy.

Dataset Diversity: A retrieval model trained on one dataset may not perform well on another dataset due to differences in image quality, object categories, and domain characteristics.

Computational Complexity: Large-scale image retrieval requires efficient indexing and fast similarity search techniques. Deep learning models may require high computational power.

Lack of Explainability: Deep learning-based CBIR systems often work as black boxes, making it difficult to understand why certain images are retrieved.

Relevance Feedback: User feedback can improve retrieval results, but designing effective and user-friendly feedback mechanisms remains challenging.

XI. FUTURE DIRECTIONS

Future CBIR research can focus on the following directions:

Hybrid Feature Fusion: Combining color, texture, shape, and deep features can improve retrieval accuracy and robustness.

Explainable CBIR: Explainable AI methods can help users understand why certain images are retrieved.

Large-Scale Retrieval: Efficient indexing methods such as hashing, approximate nearest neighbor search, and vector databases can support large image collections.

Domain-Specific CBIR: Specialized CBIR systems for medical imaging, agriculture, remote sensing, and e-commerce can improve practical usefulness.

Multimodal Retrieval: Combining image features with text, audio, metadata, and user feedback can reduce the semantic gap.

Self-Supervised Learning: Self-supervised models can learn useful visual representations from unlabeled image data.

Real-Time Retrieval Systems: Faster models and optimized search algorithms can support real-time image retrieval applications.

Privacy-Preserving CBIR: Secure image retrieval methods are needed for sensitive domains such as healthcare and surveillance.

XII. DISCUSSION

The development of CBIR has moved through several stages. Early systems relied heavily on handcrafted features such as color histograms, texture descriptors, and shape descriptors. These features are computationally efficient and easy to interpret, but they are limited in representing complex image semantics. Later, machine learning methods improved retrieval by selecting important features and learning better similarity functions.

Deep learning has greatly improved CBIR by learning hierarchical and semantic features directly from images. CNN-based features are now widely used because they provide stronger representation than traditional handcrafted descriptors. However, deep learning also introduces challenges such as high computational cost, large data requirements, and limited interpretability.

A practical CBIR system should balance accuracy, speed, storage efficiency, and user interpretability. In many applications, hybrid approaches may provide better results than using a single feature type. For example, color and texture features may be useful for natural scene retrieval, while deep features may be more suitable for object and semantic retrieval. Similarly, shape features are valuable when object boundaries are important.

Therefore, the future of CBIR lies in intelligent feature fusion, deep semantic learning, efficient indexing, and user-centered retrieval design.

XIII. CONCLUSION

Content-Based Image Retrieval is an important research area in computer vision and multimedia information retrieval. It enables users to search images based on visual content rather than textual descriptions. Traditional CBIR methods use handcrafted features such as color, texture, and shape. Color features are simple and effective for representing visual appearance, texture features capture surface patterns, and shape features describe object structure. However, these low-level features often fail to capture high-level semantic meaning.

Deep learning has significantly improved CBIR by automatically learning meaningful image representations. CNN-based deep features have shown strong performance in many image retrieval tasks because they capture both local and semantic patterns. Nevertheless, challenges such as semantic gap, computational complexity, dataset variation, and explainability remain important.

The review shows that no single feature type is sufficient for all retrieval tasks. Color, texture, shape, and deep features each have their own strengths and limitations. Hybrid CBIR systems that combine handcrafted and deep features are likely to provide better retrieval accuracy and robustness. Future research should focus on explainable, scalable, domain-specific, and multimodal CBIR systems.

REFERENCES

[1] Ahmad, J., Sajjad, M., Mehmood, I., Rho, S., & Baik, S. W. (2015). Describing colors, textures and shapes for content-based image retrieval: A survey.

[2] Datta, R., Joshi, D., Li, J., & Wang, J. Z. (2008). Image retrieval: Ideas, influences, and trends of the new age. *ACM Computing Surveys, 40*(2), 1–60.

[3] Dubey, S. R. (2020). A decade survey of content-based image retrieval using deep learning.

[4] El-Henawy, I. M., & Ahmed, K. (2016). Content-based image retrieval using multiresolution analysis of shape-based classified images.

[5] Flickner, M., Sawhney, H., Niblack, W., Ashley, J., Huang, Q., Dom, B., Gorkani, M., Hafner, J., Lee, D., Petkovic, D., Steele, D., & Yanker, P. (1995). Query by image and video content: The QBIC system. *Computer, 28*(9), 23–32.

[6] Kato, T. (1992). Database architecture for content-based image retrieval. In *Image Storage and Retrieval Systems*.

[7] Lew, M. S., Sebe, N., Djeraba, C., & Jain, R. (2006). Content-based multimedia information retrieval: State of the art and challenges. *ACM Transactions on Multimedia Computing, Communications, and Applications, 2*(1), 1–19.

[8] Rashno, A., & Rashno, E. (2019). Content-based image retrieval system with most relevant features among wavelet and color features.

[9] Rui, Y., Huang, T. S., & Chang, S.-F. (1999). Image retrieval: Current techniques, promising directions, and open issues. *Journal of Visual Communication and Image Representation, 10*(1), 39–62.

[10] Smeulders, A. W. M., Worring, M., Santini, S., Gupta, A., & Jain, R. (2000). Content-based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 22*(12), 1349–1380.

[11] Tamura, H., Mori, S., & Yamawaki, T. (1978). Textural features corresponding to visual perception. *IEEE Transactions on Systems, Man, and Cybernetics, 8*(6), 460–473.