SHODH SAGAR®
International Journal for Research Publication and Seminar
ISSN: 2278-6848 | Vol. 15 | Issue 3 | Jul - Sep 2024 | Peer Reviewed & Refereed

Content-Based Image Retrieval and Feature Extraction: Analysing the Literature

[1]Shilpa Jaitly
Research Scholar in University college of Computer Application,
Guru Kashi University Talwandi sabo.

[2]Dr. Vijay Laxmi
Professor in University College of Computer Applications,
Guru Kashi University Talwandi Sabo

[3]Dr. Gagan Jindal
Professor in Chandigarh Engineering College
CGC Landran Mohali Punjab

Check for updates

DOI: https://doi.org/10.36676/jrps.v15.i3.1520

Accepted 16/09/2024 Published 18/09/2024 * Corresponding author

Abstract
A significant amount of multimedia data consists of digital images, and multimedia content analysis is used in many real-world computer vision applications. Multimedia information, especially photos, has become much more complicated in the last several years. Every day, millions of photos are posted to various websites, such as Instagram, Facebook, and Twitter. Finding a suitable image in an archive is a difficult research subject for the field of computer vision. Most search engines use standard text-based techniques that depend on metadata and captions in order to fetch photos. Over the past 20 years, a great deal of research has been conducted on content-based image retrieval (CBIR), picture categorization, and analysis. In image classification models and CBIR, high-level picture representations are represented as feature vectors made up of numerical values. Empirical evidence indicates a considerable disparity between picture feature representation and human visual understanding. Reducing the semantic gap between human visual understanding and picture feature representation is the aim of this study. This study aims to do a thorough analysis of the latest advancements in the domains of Content-Based picture Retrieval and picture representation. We performed a comprehensive analysis of many models for image retrieval and picture representation, encompassing the most recent advancements in semantic deep-learning methods and feature extraction. This paper provides an in-depth analysis of the key ideas and important studies related to image representation and content-based picture retrieval. In an effort to stimulate more research in this field, it also offers a preview of potential future study topics.
Keywords: Content-based image retrieval (CBIR), image retrieval, images.

1. Introduction

The proliferation of digital cameras, smartphones, and the internet has give rise to an escalating issue in the field of image retrieval. Archival retrieval of pertinent photographs is difficult because of the disparity between human visual perception and manual tagging. Content-based image retrieval (CBIR) is a methodology that tackles these challenges by employing visual analysis of the request picture. Comparative Bayesian Image Retrieval (CBIR) assesses the visual characteristics of the query image by comparing it with the images stored in the archive and determining the level of visual similarity based on image attributes. A vector functions as a fundamental basis for correctly recognizing photos that have similar content. The query-by-image content (QBIC) and Simplicity models are image retrieval models that rely on extracting low-level visual semantics. Cross-Border Image Retrieval (CBIR) and feature extraction methods find use in many domains such as medical image analysis, remote sensing, criminal detection, video analysis, military surveillance, and the textile industry. The selection of visual characteristics is contingent upon the specific needs of the end user and the depiction of discriminative features. The effectiveness of CBIR may be improved by utilizing the image feature vector as a data for machine learning methods via the implementation of training and test models. Advancements in image retrieval now focus on deep neural networks (DNNs), which have the ability to generate exceptional outcomes but at a significant computational expense. Figure 1 provides an overview of the basic concepts and mechanism of image retrieval.

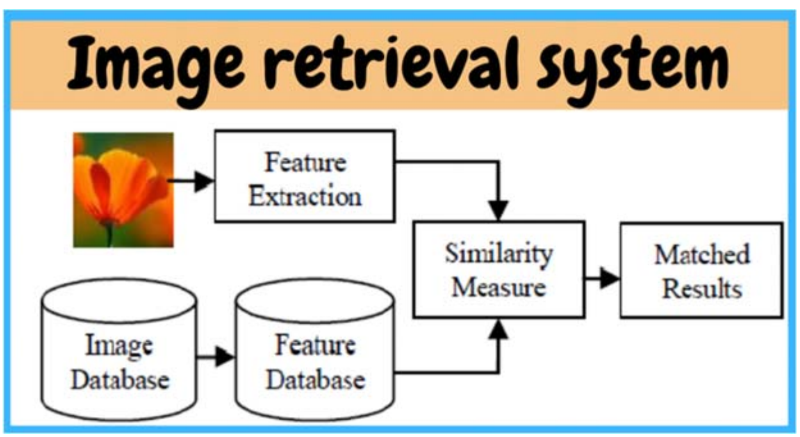

Figure 1: Overview of the basic concepts and mechanism of image retrieval

2. Characteristics of Color

Colour is a fundamental low-level visual attribute that differentiates across pictures by variations in colour. The human visual system has the ability to differentiate between different visual stimuli by analyzing their color. Thus, the complete color information of an image may be substituted by a limited number of color representations using a dominant color descriptor (DCD). DCD, one of the color descriptors in MPEG-7, utilizes a concise and user-friendly structure to define the representative color distribution and its characteristics. [1] proposed a new approach for CBIR based on the MPEG-7 descriptor, employing the histogram intersection technique to assess the characteristics of each picture. [2] Contended that traditional methods are unable to provide consumer satisfaction since they depend on labeling and annotations. Researchers investigated an alternate approach to picture retrieval: recovering images by analyzing their content. The suggested approach utilizes a two-stage clustering protocol to adapt a compact image descriptor based on the contextual information of the picture. In their study, [3] introduced a technique for extracting pictures by analyzing color and texture characteristics, therefore offering a reliable and efficient assessment of the early human capacity to comprehend visual information. Numerous research groups have examined the completeness property of invariant descriptors, and [4] has proposed a new approach to derive a comprehensive collection of pseudo-Zernike moment invariants.

A new approach was introduced in [5] for indexing pictures by utilizing characteristics derived from the error diffusion block truncation coding (EDBTC). This method demonstrated superior performance compared to existing image retrieval systems and previous BTC-based image indexing efforts. An innovative method for region-based picture learning was presented by [6], which utilizes a decision tree called DT-ST. The proposed method addresses the problem of feature discretization by using semantic templates based on low-level characteristics to label specific areas of an image. The authors of reference [7] proposed an advanced color-based vector quantization technique that can automatically categorize picture components, while optimally preserving the variable feature vector, including the dominating color descriptors.

The introduction of a multiscale distance coherence vector (MDCV) for CBIR by [8] can not guarantee the total elimination of noise, even when distinct shapes possess the same descriptor. A comprehensive overview of the aforementioned color characteristics [24–31] is included in Table 1.

TABLE 1: A summary of the performance of color features.

Author	Application	Method	Dataset	Accuracy
Duanmu [25]	Image retrieval	Color moment invariant	COIL-100	0.985
Wang et al. [26]	Content-based image retrieval	Integrated color and texture features	Corel	0.613
Zhang et al. [27]	Object recognition	Complete set of pseudo-Zernike invariants	COIL-100	—
Guo et al. [28]	Content-based image retrieval	Error diffusion block truncation coding features	Corel	0.797
Shao et al. [24]	Image retrieval	MPEG-7 dominant color descriptor	Corel	0.8964
Liu et al. [29]	Region-based image retrieval	High-level semantics using decision tree learning	Corel	0.768
Islam et al. [30]	Automatic categorization of image regions	Dominant color-based vector quantization	Corel	0.9767
Jiexian et al. [31]	Content-based image retrieval	Multiscale distance coherence vector algorithm	MPEG-7 image database	0.97

3. **Texture Features**

In order to showcase the selective powers of wavelet moments, [9] performed tests on four datasets. The suggested model (WMs) was assessed using two specifically designed wavelet configurations: WMS-1 and WMs-2. Comparative analysis was conducted on the performance of the model against Zernike, pseudo-Zernike, Fourier-Mellin, Legendre, and two more models. The machine learning approach, which incorporates wavelet moments and moment invariants, demonstrated exceptional classification accuracy for moment descriptors. [10] Experimental evaluation of the suggested model (MSD) for image retrieval using Corel datasets. They employed Corel-5000 and Corel-10000 software to analyze a dataset of 15,000 photos. The retrieval performance was evaluated using the HSV, RGB, and Lab colour spaces. The model exhibited superior performance in the HSV and Lab color spaces, but showed subpar performance in the RGB color space. This can be attributed to the inconsistent use of different color quantization levels and texture orientation quantization levels. The results indicated that MSD consistently surpasses other models in terms of retrieval accuracy, storage capacity, and computational speed.

The efficiency of the suggested approach (SVM-based architecture) with the Corel picture gallery was demonstrated by trials described in reference [11]. The suggested model was shown to provide improved performance and more consistency in picture retrieval. A series of experiments were carried out using the Brodatz and Vistex datasets to investigate content-based picture retrieval. The findings demonstrated that the suggested model (LDRP) had exceptional performance and average accuracy rates. Additionally, it demonstrated greater efficiency in extracting features and a decreased latency in matching features. Due to their representation of a set of pixels, texture characteristics have greater semantic significance than color features. Nevertheless, the main drawback of these systems is their vulnerability to visual noise, and their semantic representation relies on the shapes of the objects in the scenes. Table 2 provides a comprehensive overview of the specific textural characteristics described above.

TABLE 2: A summary of the performance of texture features.

Authors	Datasets	Purpose	Model	Performance/accuracy
Papakostas et al. [32]	COIL, ORL, JAFFE, TRIESCH I	Wavelet moments and their corresponding invariants in machine vision system	Wavelet moments and moment invariants	Classification performances on (100%) percent of entire data are 0.3083, 0.2425, 0.1784, and 0.1500, respectively, for datasets
Wang et al. [34]	Corel-1000 and Corel-10000	Image retrieval	SED	Similarity between query image and image database is 3.9198, 9.92209, and 8.86239 for dragons, busses, and landscapes, and there will be high precision rate when the query image has noteworthy regions or texture
Liu et al. [33]	Corel datasets (Corel-5000 and Corel-10000)	Image retrieval	MSD	Average retrieval precision and recall ratios on Corel-5000 and Corel-10000 are 55.92%, 6.71% and 41.44%, 5.48%
Lasmar and Berthoumieu [40]	Vistex, Brodatz, ALOT	Texture image retrieval	GC-MGG and GC-MWbl	Improvement in average retrieval rate on Brodatz (EB2) by our model is 6.86% and 5.23%, respectively, with Daubechies filter db4 and dual-tree complex wavelet transform
Fadaei et al. [38]	Brodatz and Vistex	Content-based image retrieval	LDRP	80.81% and 91.91% are average precision rates of the first-order LDRP($P = 6$, $K = 4$) for the respective datasets

4. **Characteristics of Shape**

Shape is also regarded as a critical low-level feature, as it aids in the identification of real-world shapes and objects. [13] conducted an exhaustive examination of the utilization of shape features in the fields of image retrieval and image representation. The primary classifications of shape features are contour-based and region-based [14]. One of the specific domains in which shape features are employed for image representation is trademark-based image retrieval [15].

5. **Spatial Characteristics**

Spatial characteristics of an image are essential for comprehending the positions of objects in a two-dimensional picture-space. The Bag of Visual Words (BoVW) is a commonly employed paradigm that encodes pictures as a histogram, disregarding the spatial distribution. Spatial Pyramid Matching (SPM) is a method that accurately reconstructs the spatial characteristics of a picture without being affected by scale or rotation. We presented a technique for storing spatial information required to describe the BoVW model. This method begins by computing the universal geometric correlation between sets of comparable visual words. Evaluation of the performance of the proposed strategy was conducted using five datasets. The Hybrid Geometric Spatial Image Representation (HGSIR) method was introduced by [17] using an image classification-based framework. By integrating spatial information into the inverted index of the BoVW model, the approach calculates the universal spatial inclination of visual words in a manner that is invariant to gyration. Four datasets were used to evaluate the efficacy of the suggested approach.

The suggested Pairs of Identical Visual Word (PIW) methodology aims to enhance the integration of global information, powerful geometric transformation, effective extraction of spatial information, complexity reduction, and classification rate enhancement by combining differentiating information. A model was proposed in reference [19] that utilizes scale-invariant feature transform-based BoVW for symbol identification. To integrate spatial information into BoVW, circular tilings were used, and the angle histograms of an existing approach were adjusted to guarantee rotation invariance. A novel soft pairwise spatial angle-distance histogram was introduced by [20], including both distance and angle data of visual words throughout an image. Achieving an effective representation that can incorporate relative spatial information is the goal.

6. Low-Level Function Fusion

The CBIR model proposed by [21] utilizes discrete wavelet transform (DWT) and color to leverage low-level characteristics of color, texture, and shape for the purpose of retrieving comparable pictures. These traits exert a substantial impact on the retrieval process. This study investigates several classifications of features and methods for extracting features, and clarifies the situations in which these strategies are successful. The extraction of color characteristics is performed using the RGB and YCbCr color spaces. The YCbCr transformation is selected because of the human visual system's capacity to detect different colors and degrees of brightness. In order to extract edge characteristics, the Canny edge detector is utilized, while the viewfinder guarantees the best possible form in every dimension.

A unique Color-Based Image Retrieval (CBIR) method utilizes both color and texture characteristics to extract local vectors, which are then employed as feature vectors. In order to extract color characteristics, color moments are utilized, while texture features are extracted using discrete wavelet transform and Gabor wavelet approaches. Application of a directory descriptor enhances the color and edge of the feature vector. The study on CBIR, done by [22], included hybrid features and distance metrics. These metrics encompassed spatial features, frequency, binaryized statistical image features (BSIF), and color and edge directivity descriptors (CEDD). The CEDD technique incorporates the HSV color two-stage fuzzy linkage scheme for feature extraction.

The work done by [23] focused on Image-Based Image Retrieval (CBIR) using the combination of spatial color characteristics with shape features and item detection. Enhancements in image retrieval are achieved by including the spatial color feature into the feature vector. Additionally, the descriptor becomes more effective when the corners and boundaries of the form are identified. The Bag of Word (BoW) network is supplied with compressed data characteristics to enable efficient indexing or retrieval of the picture.

In their study, [24] introduced a novel approach to picture classification and search by integrating the local basis pattern (LBP) with the color information feature (CIF). The textural feature is derived from the Local Binary Pattern (LBP) to generate the picture descriptor. However, the LBP does not provide sufficient performance in terms of the descriptor for the color feature. In order to effectively extract color photos from an extensive database, both color characteristics and texture characteristics are utilized. The proposed approach utilizes a unique color feature called CIF, which is derived from the LBP feature. CIF combines the color and textural information of an image. This study explored the feasibility of combining the indexing of SIFT features with deep convolutional neural networks (d-CNN) for the purpose of picture retrieval. The authors introduced collaborative index embedding methods that dynamically adjust the index proportion of CNN and SIFT features to validate the common image-level neighborhood structure and seamlessly combine the CNN and SIFT features. In comparison to the original CNN and SIFT index, this approach has a retrieval accuracy that is 10% more efficient.

Reference [26] proposed a technique for extracting pictures in both color and texture context by utilizing the Gaussian copula model of Gabor wavelets. The proposed model effectively captures the dependence patterns present in variables that exhibit dependencies. The Gaussian copula approach is employed to examine and capture these existence dependencies. Three distinct techniques have been developed for Gaussian copulas, leading to the emergence of four Kullback-Leibler distances (KLD) for processing color retrieval pictures.

In their work, [27] investigated the application of Multi-Resolution Multi-Directional (MRMD) filters to integrate color and texture characteristics obtained from the picture while analyzing CBIR. The HSV color space was chosen because of its close similarity to the human visual system, allowing for the extraction of local and global characteristics from both low- and high-frequency domains. [28] presented a research on Color-Based Image Retrieval (CBIR) by integrating color and texture characteristics, suggesting many approaches to tackle the difficulty of recovering photos from extensive datasets.

7. Methods Based on Local Features

The work undertaken by [29] focused on picture similarity utilizing sparse feature representation. The objective of the study was to independently analyze the similarities between different photos. The problem of information fidelity is tackled by a feature-based method that assesses the amount of information obtained from the test picture and extracts information from the reference image. This technique is applied in three widely used applications: picture copy-move detection, retrieval, and character recognition. [30] introduced a cooperative sparse representation in two contrasting orientations for semisupervised image annotation. This approach has the capability to enhance the quantity of annotated pictures used in training image classifiers. Their primary focus was on the collaborative implementation of super-resolution (SR) in semisupervised image annotation.

This approach has the capacity to augment the quantity of labeled pictures in the training image classifiers for future use. In their work, [31] investigated the application of supervised local sparse coding of subimage features for image retrieval. This approach incorporates several global/local features. In comparison to spatial pyramid features derived using local descriptors, the suggested features exhibit superior performance. [32] introduced an innovative ranking technique using Quantum Bayesian Machine Learning (QBME) to enhance picture retrieval speed, using a unique learning framework. A semantic correlation hypergraph was constructed to model the association between photos in the dataset. A multiple probing technique was employed to rank the images based on several query cases. The suggested method shown effectiveness in terms of retrieval performance and processing efficiency.

The study [33] investigated the use of weak label regularized local coordinate coding for retrieval-based face annotation. One approach to tackle the difficulties related to collecting large quantities of online face photos is the implementation of a system that utilizes content-based image retrieval and labels for automated annotating. The proposed approach, known as WLRLCC, incorporates the concepts of graph-based weak label organization and local coordinate coding. Experimental investigations were carried out on many online databases of facial images, and the primitive approximation technique (AWLRLCC) was proposed to further improve efficiency and scalability. A research was undertaken by [34] to investigate the retrieval of content-based medical pictures using dictionary learning. The study proposed a clustering technique that leverages dictionaries and an orthogonal matching pursuit algorithm to organize extensive medical datasets. The suggested approach is compatible with a wide range of medical databases and does not need any training. The study conducted by [35] examined a content-based picture retrieval system that utilises sparse representation, a crucial attribute in multimedia information processing systems and applications.

Image retrieval is an essential application in multimedia information processing systems, and scholars have investigated two separate approaches: text-based and content-based image retrieval. The suggested architecture utilizes sparse representation for image retrieval, integrating the IDWT feature with sparse representation. The findings demonstrate that the suggested approach exhibits greater retrieval accuracy compared to other traditional approaches and achieves optimal performance across five datasets, leading to a decrease in the size of feature vectors and storage space.

An innovative sketch-based image retrieval technique was introduced by [36-39], which utilizes product quantization with sparse coding to create the codebook. The efficiency of this approach in computation and its better effectiveness in comparison to various widely utilized SBIRs are evident.

Image retrieval is a technique that facilitates the exploration, browsing, and retrieval of pictures from a wide range of databases, therefore providing enhanced ease to human existence. The use of machine learning is highly efficient in the tasks of picture annotation,

SHODH SAGAR®
International Journal for Research Publication and Seminar
ISSN: 2278-6848 | Vol. 15 | Issue 3 | Jul - Sep 2024 | Peer Reviewed & Refereed

categorization, and recognition. Diverse methodologies are used to extract pictures by exploiting color and texture characteristics.
The sparse coding-based few learning examples approach proposed for image retrieval aims to tackle the challenge of extracting high-level semantic information in picture retrieval. By combining a sparse coding-based instance distance, cross-validation sparse coding representation, and an enhanced KNN model, this method effectively decreases the number of learning instances without compromising retrieval accuracy. Facial recognition has been a prominent field of focus in computer vision for the last twenty years. Facial recognition involves two main processes: collecting discriminative characteristics from the face to differentiate between photos of various persons and creating efficient classifiers to identify distinct individuals. Recently, there has been a proliferation of face recognition techniques, mostly categorized as either holistic or local feature representation mechanisms. The superiority of local features over holistic features is mostly attributed to their resilience and consistency in the presence of localized variations in feature description within an image. Using a contextual-aware local binary feature learning (CA-LBFL) approach, the authors propose a technique for face rearrangement that directly extracts context-aware binary data from raw pixels and compares it to the current model. By restricting the amount of bitwise changes in each descriptor and using the contextual knowledge of neighboring bits, CA-LBFL obtains more resilient local binary features. A pre-computed codebook, or visual language, is required for histogram-based image description to extract and encapsulate local information. The computational expense of this procedure is substantial and its integrity may be jeopardized if the generated codebook is biassed as a result of a restricted number of training examples. This study introduces a new approach to transfer implicit codebooks for visual representation. This method allows for the use of pre-learned codebooks to create new visual applications by means of implicit learning.
We present a novel fine-grained image classification model that integrates codebook generation with low-rank sparse coding (LRSC), resulting in enhanced discriminative fine-grained image classification. Class specific and generic codebooks are generated by optimizing the accumulative reconstruction error, sparsity restrictions, and codebook incoherence. In order to address the inconsistent handling of visual features and the construction of an explicit semantic space, the authors propose a structured, weak semantic space for the image classification problem, which is influenced significantly by the visual characteristics of the image. The implementation of a pre-learned classifier involves the computation of correlations for each candidate region with high confidence ratings. Any parts of the pictures that do not include an item are classified as representing the background.
Supervised learning is mostly used for the classification and categorization of digital photographs. However, handling the labelling process may be somewhat difficult when dealing with a large quantity of photos. Using exemplar classifiers to predict weak semantic correlations, the authors provide a unique weak semantic consistency constrained (WSCC)

technique for image classification, where each picture is treated as a distinct class. The diagram in Figure 2 provides a summary of fundamental machine learning methods for CBIR.

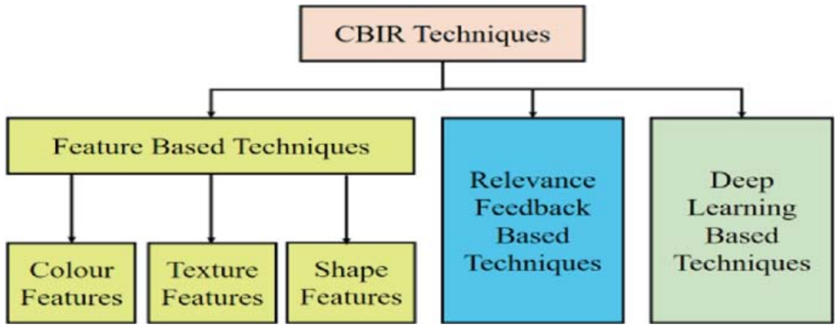

Figure 2: An overview of basic machine learning techniques for CBIR

8. **CBIR Research Using Deep-Learning Techniques**

Content-based image retrieval (CBIR) is a widely used technique for retrieving digital pictures from extensive storage or electronic databases. The application of deep convolutional neural networks (CNN) for this objective has been extensively explored, with the introduction of a unique term frequency-inverse document frequency (TF-IDF) approach for CDIR. This method utilizes the acquired filters from convolutional layers of convolution neuron models to identify visual words. It calculates the intensity of the visual pattern by activating each filter as the Task Function (TF) component. This study proposes three methods for calculating the IDF component by integrating TF-IDF with CNN analysis for visual material.

A hashing technique which collects characteristics from photos and learns their binary representations was proposed by [40-42]. The efficacy of these algorithms was demonstrated by conducting experiments on hundreds of histopathological pictures, resulting in a classification accuracy of 97.34%. Proposed by [43-44], semantics-assisted visual hashing (SAVH) is an unsupervised visual hashing method that utilizes offline and online learning to extract features without requiring previous knowledge of the semantics of picture content. CNN has demonstrated exceptional performance in computer vision applications, particularly in Continuous Bayesian Inference (CBIR) models. Nevertheless, the majority of CNN models restrict the use of intermediate convolutional layers during the detection of local picture patterns. This work presents the binary CNN-based architecture, a novel approach that use two concurrent CNN models to extract features without requiring previous knowledge of the semantics of visual information. This approach optimizes picture representation by reducing it to a compact length, thereby enhancing retrieval performance, search time, and storage cost. The hashing function plays a vital role in facilitating effective image search in Computer-Based Image Recognition (CBIR). It generates a binary code that closely resembles the picture content, hence transforming high-dimensional visual data into a

SHODH SAGAR®
International Journal for Research Publication and Seminar
ISSN: 2278-6848 | Vol. 15 | Issue 3 | Jul - Sep 2024 | Peer Reviewed & Refereed

relatively low-dimensional binary space. This technique depends on Convolutional Neural Networks (CNN) and presupposes that semantic labels be encoded as binary code in latent layer characteristics. The supervised deep hashing method generates a hash function by manipulating a latent layer in a deep neural network. The binary code is then acquired by learning objective functions that account for classification error and other desired characteristics of the binary code. An open research challenge is the effective analysis and categorization of images utilizing discriminative information. The current approaches face difficulties in calculating discriminative information at the boundaries of images and determining similarity consistency restrictions. A novel multiview label sharing technique (MVLS) was introduced to preserve and guarantee similarity by integrating transformation matrix learning with classifier training. Empirical findings demonstrated the efficacy of the suggested MVLS method. CNN models have difficulties in achieving accurate item categorization, particularly when there is a scarcity of training data and labels. A novel multiview technique, MVFL-VC, was introduced. This approach leverages both labeled and unlabeled pictures to ensure image view consistency while incorporating multiview information. The MVFL-VC algorithm, as suggested, is compatible with several image categorization and representation methods.

The multiview semantic representation (MVSR) technique is a suggested approach that partitions pictures according to their semantic and visual similarities. This technique exhibits more discriminativeness compared to previous semantics approaches, as it calculates semantic information for future use from each individual frame and from distinct sets of pictures and diverse perspectives.

9. Feature extraction techniques for face recognition

Face recognition is a crucial application of computer vision, used to identify individuals based on facial features. However, it is considered a challenging problem due to the complex nature of the facial manifold. In various studies, researchers have proposed various algorithms for face recognition, such as pose and expression-invariant algorithms, two-pass face alignment methods, Kernel Fisher analysis (KFA), deeply learned pose-invariant image analysis algorithms, hybrid models, demographic traits, facial asymmetry-based anthropometric dimensions, and data augmentation. In some studies, the pose of the probe face image is corrected using an intrinsic coordinate system (ICS)-based approach, region-based principal component analysis (PCA), Mahalanobis Cosine (MahCos) distance metric, and weighted Borda count method through the re-ranking stage. The methodology is validated by using two face recognition datasets, GavabDB and FRGC v2.0. In another study, the authors introduced a novel approach for aligning facial faces and transformed the pose of face acquisition into an aligned frontal view based on the three-dimensional variance of the facial data. Facial features are extracted using Kernel Fisher analysis (KFA) in a subject-specific perspective based on isodepth curves, and classification is performed using four classification algorithms.

In another study, a hybrid model for age-invariant face recognition was presented, which uses generative and discriminative models, deep networks to extract discriminative features, and deeply learned matching scores to get final recognition accuracies. Demographic traits, such as age group, gender, and race, were used to enhance recognition accuracy across challenging aging variations. Facial asymmetry-based anthropometric dimensions were used to estimate the gender and ethnicity of a given face image, and a matching-scores space-based face recognition scheme was presented. Asymmetric left and right face images were also studied for accurate age estimation. 3D face recognition is an active area of research, but it is still a challenge due to the complex nature of the facial manifold. Existing methods based on holistic, local, and hybrid features show competitive performance but are still short of what is needed. Novel and accurate alignment algorithms may further enhance face recognition accuracies, while deep-learning algorithms successfully employed in various image processing applications are needed to improve 3D face recognition performance.

10. Distance Measures

Different distance measures are applied to feature vectors to compute similarity between query images and archived images. The distance measure is chosen based on the feature vector structure and indicates similarity. Effective image retrieval depends on the type of applied similarity, matching object regions, background, and objects. Finding an adequate and robust distance measure is challenging, but popular distance measures are commonly used in CBIR.

11. Performance Evaluation Criteria

There are various performance evaluation criteria for CBIR and they are handled in a predefined standard. It is important to mention here that there is no single standard rule/criterion to evaluate the CBIR performance. There are set of some common measures that are reported in the literature. The selection of any measure among the criteria mentioned below depends on the application domain, user requirement, and the nature of the algorithm itself. The following performance evaluation criteria are commonly used.

11.1. Precision and Recall. Precision (P) and recall (R) are commonly used for performance evaluation of CBIR research. Precision is the ratio of the number of relevant images within the first k results to the total number of images that are retrieved and is expressed as follows: precision (P) is equivalent to the ratio of relevant images retrieved to the total number of images retrieved (N_{TR}):

$$P = \frac{\mathrm{tp}}{N_{\mathrm{TR}}} = \frac{\mathrm{tp}}{\mathrm{tp} + \mathrm{fp}}, \qquad (1)$$

where tp refers to the relevant images retrieved and fp refers to the false positive, i.e., the images misclassified as relevant images.

11.2. Recall. Recall (R) is stated as the ratio of relevant images retrieved to the number of relevant images in the database:

(2)

$$R = \frac{\text{tp}}{N_{\text{RI}}} = \frac{\text{tp}}{\text{tp} + \text{fn}},$$

where tp refers to the relevant images retrieved, N_{RI} refers to the number of relevant images in the database. NRI is obtained as tp + fn, where fn refers to the false negative, i.e., the images that actually belonged to the relevant class, but misclassified as belonging to some other class.

11.3. F-Measure. It is the harmonic mean of P and R; the higher F-measure values indicate better predictive power:

$$F = 2\frac{P \cdot R}{P + R}, \qquad (3)$$

where P and R refer to precision and recall, respectively.

11.4. Average Precision.)e average precision (AP) for a single query k is obtained by taking the mean over the precision values at each relevant image:

$$\text{AP} = \frac{\sum_{k=1}^{\text{NRI}} (P(k) \times R(k))}{\text{NRI}}. \qquad (4)$$

11.5. Mean Average Precision. For a set of queries S, the mean average precision (MAP) is the mean of AP values for each query and is given by

$$\text{MAP} = \frac{\sum_{q=1}^{S} \text{AP}(q)}{S}, \qquad (5)$$

where S is the number of queries.

11.6. Precision-Recall Curve. Rank-based retrieval systems display appropriate sets of top-k retrieved images.)e P and R values for each set are demonstrated graphically by the PRcurve. The PRcurve shows the trade-off between P and R under different thresholds. Many other evaluation measures have also been proposed in the literature as averaged normalized modified retrieval rank (ANMRR) [12]. It has been applied for MPEG-7 color experiments. ANMRR produces results in the range [0-1], where smaller values indicate better performance. Mean normalized retrieval order (MNRO) proposed by [11] used a metric to represent the scaled-up behavior of the system without bias for top-k retrievals. For more details on performance evaluation metrics, the readers are referred to the article [31].

12. **Conclusion and Future Directions**

This literature review examines several methods for classification and image representation (CBIR) including those developed in the last 10-12 years. This emphasizes the significance of low-level visual characteristics such as color, texture, spatial arrangement, and form in the representation of picture information. However, the absence of uniformity in picture collections and the heterogeneity of image characteristics make single feature representation ineffective. To enhance efficiency in CBIR and picture representation, it is possible to merge low-level characteristics, hence minimizing semantic gaps. Subsequent investigations should prioritize the integration of local and global characteristics. Although traditional machine learning methods have demonstrated successful outcomes in many fields, optimizing the representation of features in terms of their dimensions can establish a robust foundation for building classification-based models. Deep neural networks have emerged as the focus of recent research, demonstrating strong performance across several datasets and surpassing manually designed features. Effectively handling extensive picture datasets for supervised training is a difficult and time-consuming task.

13. References

[1] D. Zhang, M. M. Islam, and G. Lu, "A review on automatic image annotation techniques," Pattern Recognition, vol. 45, no. 1, pp. 346–362, 2012.
[2] Y. Liu, D. Zhang, G. Lu, and W.-Y. Ma, "A survey of contentbased image retrieval with high-level semantics," Pattern Recognition, vol. 40, no. 1, pp. 262–282, 2007.
[3] T. Khalil, M. U. Akram, H. Raja, A. Jameel, and I. Basit, "Detection of glaucoma using cup to disc ratio from spectral domain optical coherence tomography images," IEEE Access, vol. 6, pp. 4560–4576, 2018.
[4] S. Yang, L. Li, S. Wang, W. Zhang, Q. Huang, and Q. Tian, "SkeletonNet: a hybrid network with a skeleton-embedding process for multi-view image representation learning," IEEE Transactions on Multimedia, vol. 1, no. 1, 2019.
[5] W. Zhao, L. Yan, and Y. Zhang, "Geometric-constrained multi-view image matching method based on semi-global optimization," Geo-Spatial Information Science, vol. 21, no. 2, pp. 115–126, 2018.
[6] W. Zhou, H. Li, and Q. Tian, "Recent advance in contentbased image retrieval: a literature survey," 2017, https://arxiv. org/abs/1706.06064.
[7] A. Amelio, "A new axiomatic methodology for the image similarity," Applied Soft Computing, vol. 81, p. 105474, 2019.
[8] C. Celik and H. S. Bilge, "Content based image retrieval with sparse representations and local feature descriptors: a comparative study," Pattern Recognition, vol. 68, pp. 1–13, 2017.
[9] T. Khalil, M. Usman Akram, S. Khalid, and A. Jameel,"Improved automated detection of glaucoma from fundus image using hybrid structural and textural features," IET Image Processing, vol. 11, no. 9, pp. 693–700, 2017.

[10] L. Amelio, R. Jankovi´c, and A. Amelio, "A new dissimilarity measure for clustering with application to dermoscopic images," in Proceedings of the 2018 9th International Conference on Information, Intelligence, Systems and Applications (IISA), pp. 1–8, IEEE, Zakynthos, Greece, July 2018.
[11] S. Susan, P. Agrawal, M. Mittal, and S. Bansal, "New shape descriptor in the context of edge continuity," CAAI Transactions on Intelligence Technology, vol. 4, no. 2, pp. 101–109, 2019.
[12] L. Piras and G. Giacinto, "Information fusion in content based image retrieval: a comprehensive overview," Information Fusion, vol. 37, pp. 50–60, 2017.
[13] L. Amelio and A. Amelio, "Classification methods in image analysis with a special focus on medical analytics machine Learning Paradigms, pp. 31–69, Springer, Basel, Switzerland, 2019.
[14] D. Ping Tian, "A review on image feature extraction and representation techniques," International Journal of Multimedia and Ubiquitous Engineering, vol. 8, no. 4, pp. 385–396, 2013.
[15] D. Zhang and G. Lu, "Review of shape representation and description techniques," Pattern Recognition, vol. 37, no. 1, pp. 1–19, 2004.
[16] R. Datta, J. Li, and J. Z. Wang, "Content-based image retrieval: approaches and trends of the new age," in Proceedings of the 7th ACM SIGMM International Workshop on Multimedia Information Retrieval, pp. 253–262, ACM, Singapore, November 2005.
[17] Z. Yu and W. Wang, "Learning DALTS for cross-modal retrieval," CAAI Transactions on Intelligence Technology, vol. 4, no. 1, pp. 9–16, 2019.
[18] N. Ali, D. A. Mazhar, Z. Iqbal, R. Ashraf, J. Ahmed, and F. Zeeshan, "Content-based image retrieval based on late fusion of binary and local descriptors," International Journal of Computer Science and Information Security (IJCSIS), vol. 14, no. 11, 2016.
[19] N. Ali, Image Retrieval Using Visual Image Features and Automatic Image Annotation, University of Engineering and Technology, Taxila, Pakistan, 2016.
[20] B. Zafar, R. Ashraf, N. Ali et al., "Intelligent image classification-based on spatial weighted histograms of concentric circles," Computer Science and Information Systems, vol. 15, no. 3, pp. 615–633, 2018.
[21] Brar, R., & Billing, S. S. (2016). Certain sufficient conditions for parabolic starlike and uniformly close-to-convex functions. Stud. Univ. Babes-Bolyai Math, 61, 53-62.
[22] U. Markowska-Kaczmar and H. Kwa´snicka, "Deep learning—a new era in bridging the semantic gap," in Bridging the SemanticGap in Image and Video Analysis, pp. 123–159, Springer, Basel, Switzerland, 2018.
[23] F. Riaz, S. Jabbar, M. Sajid, M. Ahmad, K. Naseer, and N. Ali, "A collision avoidance scheme for autonomous vehicles inspired by human social norms," Computers & Electrical Engineering, vol. 69, pp. 690–704, 2018.

[24] H. Shao, Y. Wu, W. Cui, and J. Zhang, "Image retrieval based on MPEG-7 dominant color descriptor," in Proceedings of the 9th International Conference for Young Computer ScientistsICYCS 2008, pp. 753–757, IEEE, Hunan, China,November 2008.
[25] X. Duanmu, "Image retrieval using color moment invariant," in Proceedings of the 2010 Seventh International Conference on Information Technology: New Generations (ITNG), pp. 200–203, IEEE, Las Vegas, NV, USA, April 2010.
[26] X.-Y. Wang, B.-B. Zhang, and H.-Y. Yang, "Content-based image retrieval by integrating color and texture features," Multimedia Tools and Applications, vol. 68, no. 3, pp. 545–569, 2014.
[27] H. Zhang, Z. Dong, and H. Shu, "Object recognition by a complete set of pseudo-Zernike moment invariants," in Proceedings of the 2010 IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP), pp. 930–933, IEEE, Dallas, TX, USA, March 2010.
[28] Brar, T. P. S. (2016). Green banking adoption: A comparative study of Indian public and private sector banks. Envision Journal of Commerce Department of ACFA, 10.
[29] Y. Liu, D. Zhang, and G. Lu, "Region-based image retrieval with high-level semantics using decision tree learning,"Pattern Recognition, vol. 41, no. 8, pp. 2554–2570, 2008.
[30] M. M. Islam, D. Zhang, and G. Lu, "Automatic categorization of image regions using dominant color based vector quantization," in Proceedings of the Digital Image Computing: Techniques and Applications, pp. 191–198, IEEE, Canberra, Australia, December 2008.
[31] Z. Jiexian, L. Xiupeng, and F. Yu, "Multiscale distance coherence vector algorithm for content-based image retrieval," @e Scientific World Journal, vol. 2014, Article ID 615973, 13 pages, 2014.
[32] G. Papakostas, D. Koulouriotis, and V. Tourassis, "Feature extraction based on wavelet moments and moment invariants in machine vision systems," in Human-Centric Machine Vision, InTech, London, UK, 2012.
[33] G.-H. Liu, Z.-Y. Li, L. Zhang, and Y. Xu, "Image retrieval based on micro-structure descriptor," Pattern Recognition, vol. 44, no. 9, pp. 2123–2133, 2011.
[34] X.-Y. Wang, Z.-F. Chen, and J.-J. Yun, "An effective method for color image retrieval based on texture," Computer Standards & Interfaces, vol. 34, no. 1, pp. 31–35, 2012.
[35] R. Ashraf, K. Bashir, A. Irtaza, and M. Mahmood, "Content based image retrieval using embedded neural networks with bandletized regions," Entropy, vol. 17, no. 6, pp. 3552–3580,2015.
[36] A. Irtaza and M. A. Jaffar, "Categorical image retrieval through genetically optimized support vector machines (GOSVM) and hybrid texture features," Signal, Image and Video Processing, vol. 9, no. 7, pp. 1503–1519, 2015.
[37] C. C. Chang and C. J. Lin, "LIBSVM: a library for support vector machines," ACM Transactions on Intelligent Systems and Technology (TIST), vol. 2, no. 3, pp. 1–27, 2011.

[38] S. Fadaei, R. Amirfattahi, and M. R. Ahmadzadeh, “Local derivative radial patterns: a new texture descriptor for content-based image retrieval,” Signal Processing, vol. 137,pp. 274–286, 2017.
[39] X. Wang and Z. Wang, “A novel method for image retrieva l based on structure elements’ descriptor,” Journal of Visual Communication and Image Representation, vol. 24, no. 1, pp. 63–74, 2013.
[40] Brar, R., & Billing, S. S. (2018). On Certain Results Involving a Multiplier Transformation in a Parabolic Region. Khayyam Journal of Mathematics, 4(2), 123-143.
[41] J. M. Guo, H. Prasetyo, and J. H. Chen, “Content-based image retrieval using error diffusion block truncation coding features,” IEEE Transactions on Circuits and Systems for Video Technology, vol. 25, no. 3, pp. 466–481, 2015.
[42] Brar, T. P. S. (2021). Digital marketing performance: Understanding the challenges and measuring the outcomes. In Big Data Analytics for Improved Accuracy, Efficiency, and Decision Making in Digital Marketing (pp. 51-63). IGI Global.
[43] G. Qi, H. Wang, M. Haner, C. Weng, S. Chen, and Z. Zhu,“Convolutional neural network based detection and judgement of environmental obstacle in vehicle operation,” CAAITransactions on Intelligence Technology, vol. 4, no. 2, pp. 80–91, 2019.
[44] Kumar, Y., Brar, T. P. S., Kaur, C., & Singh, C. (2024). A Comprehensive Study of Deep Learning Methods for Kidney Tumor, Cyst, and Stone Diagnostics and Detection Using CT Images. Archives of Computational Methods in Engineering, 1-26.