Sagar Chavda
Mahesh Goyani

ROBUST CONTENT-BASED IMAGE RETRIEVAL USING ICCV, GLCM, AND DWT-MSLBP DESCRIPTORS

Abstract *Content-based image retrieval (CBIR) retrieves visually similar images from a dataset based on a specified query. A CBIR system measures the similarities between a query and the image contents in a dataset and ranks the dataset images. This work presents a novel framework for retrieving similar images based on color and texture features. We have computed color features with an improved color coherence vector (ICCV) and texture features with a gray-level co-occurrence matrix (GLCM) along with DWT-MSLBP (which is derived from applying a modified multi-scale local binary pattern [MS-LBP] over a discrete wavelet transform [DWT], resulting in powerful textural features). The optimal features are computed with the help of principal component analysis (PCA) and linear discriminant analysis (LDA). The proposed work uses a variance-based approach for choosing the number of principal components/eigenvectors in PCA. PCA with a 99.99% variance preserves healthy features, and LDA selects robust ones from the set of features. The proposed method was tested on four benchmark datasets with Euclidean and city-block distances. The proposed method outshines all of the identified state-of-the-art literature methods.*

Keywords content-based image retrieval, improved color coherence vector, gray-level co-occurrence matrix, discrete wavelet transform, multi-scale local binary pattern, principal component analysis, linear discriminant analysis

Citation Computer Science 23(1) 2022: 5–36

1. Introduction

CBIR is a method for solving the image retrieval problem by utilizing the contents of images. CBIR retrieves visually similar images from a dataset as per a given query's image contents [79]. The requirement for retrieving images with identical visual information from large datasets is the need of the hour; nevertheless, there has to be some explanation and a rapid solution for resolving this challenge. We can retrieve identical images based on query image metadata or visual contents. A traditional image-retrieval system is tiresome; as time proceeds, retrieving, processing, searching, browsing, and managing images become hard. We can solve this problem of image retrieval by using low-level, mid-level, and high-level feature-extraction methods. Researchers combine these approaches for feature extraction to improve their performance [66, 79]. Such a retrieval becomes even harder if a specified query is text-based. This type of image-retrieval system is known as text-based image retrieval (TBIR). TBIR uses text-based metadata for a query. The automatic or manual annotation of images is required in order to perform this type of query. TBIR extracts images from a dataset according to the given query information (metadata). Annotating images is not an easy task, as it may require laboring costs. A query specification will need deeper query information for searching for a relevant images from a dataset. Human perception can also create more challenges to the existing annotation problems [66]. Multiple CBIR research practices have been proposed to overcome these issues of the conventional TBIR approach. Rather than using metadata or annotation, CBIR uses various feature-extraction techniques from the vision domain. CBIR indexes a dataset's images based on their query contents; these contents include color, texture, shape, and spatial information. Progress in the CBIR domain has given many sophisticated algorithms for utilizing these contents; such algorithms are not able to adequately model image semantics and have several limitations when dealing with broad-content image datasets [44]. Some well-known CBIR systems are Netra, QBIC, SIMPLIcity, MetaSEEK, VisualSeek, Blobworld, PicHunter, Google Image Search, Camfind, Pixolution, and DRAWSEARCH; other creative examples are Jet.com (elastic search-based e-commerce), PictPicks (Google's material design interface), and Veracity (reverse-image search on iOS) [7].

CBIR suffers from two issues: an intention gap, and a semantic gap. Sometimes users are unable to precisely express the expected query contents; this is known as an intention gap. A semantic gap is the challenge of describing a high-level semantic concept with low-level visual information [44, 79, 95]. The image-retrieval realm has the scope of research, as the research progress has not provided a satisfactory performance on large datasets as of yet. An open challenge in the CBIR domain is choosing the building blocks of the system. The CBIR system has various techniques at its core, namely, feature extraction, feature reduction, feature selection, feature fusion, classification, ranking, etc. Determining the proper feature-extraction strategy to capture features from an image is the most important. The extracted features should be lower in numbers, and the system should preserve the most dominant ones

without losing any prominent details. A single descriptor-based strategy will fail, as the image content can have variable scales, color information, illumination changes, and background cluttering as well as high intra-class variation and low inter-class variation. All in one, a single descriptor may not provide all of the color, texture, shape, and spatial information. Thus, we need to provide a set of techniques that captures all of the features with the above-mentioned information, or there must be some method or framework that can control these issues. Deciding on a set of descriptors is hard, but what about the minimal features and robustness of those features? This can be solved by using various methods of statistics, machine learning, and graph visualization. These methods generally fall in the area of feature reduction and feature selection. Feature reduction can be done in a supervised or unsupervised manner. Various techniques (viz., PCA [32], non-negative matrix factorization (NMF) [37], and independent component analysis (ICA) [15]) can be applied for this task. Furthermore, we can use data-visualization techniques like t-distributed stochastic neighbor embedding (t-SNE). In addition to this, selecting the appropriate features for a given vision task is also challenging. We can use LDA [88], a genetic algorithm, and other feature-selection techniques from evolutionary computing (EC), swarm optimization (SO), and artificial intelligence (AI). We can further assess suitable-feature-fusion techniques as well. Classification is as significant as any other phase of a CBIR system. We can do this with a template that matches the classifiers or machine-learning classifiers. This task helps us rank images based on similarity. As a user is likely interested in a few of the top retrieval results, we can also explore ranking strategies. Last but not least, achieving higher retrieval rates in less than a second is most critical. Several applications of CBIR are art collections [26, 29], crime prevention [42,43,68], geographical information and remote sensing systems [24], intellectual property [39,78,83,97], medical imaging [3,13,33,67,69,96], military and defense [49], photograph archives and retail catalogs [14,20,23], nudity-detection filters [11,45,46], and face-finding systems [1,36,92].

As discussed, single descriptor-based approaches will not provide maximal performance; so, we have designed the framework of the three descriptors (ICCV, GLCM, and DWT-MSLBP) to enhance the performance. The ICCV descriptor covers more color spatial information than a conventional CCV operator. GLCM is computationally efficient, and DWT-MSLBP is a firm-texture descriptor. ICCV is applied for color descriptions, and GLCM and DWT-MSLBP are applied for texture-feature descriptions. DWT-MSLBP is a composition of DWT and MS-LBP. These features are reduced with PCA and provided to the LDA for robust feature extraction. The main highlights of this paper are the optimal feature processing, high precision-recall, and lower retrieval time of the proposed work.

In this section, we have outlined the essential information of the topic, the scope of the work, the challenges, and the potential applications of the CBIR system. Furthermore, it also reviews our contribution to the existing knowledge. Section 2 summarizes the related work that has been done in the domain. Section 3 covers the

various feature-extraction methods that were used to form the proposed framework. Section 4 explains the proposed work with figures and algorithms. The section also reports on the feature complexity and benefits of the proposed work. Experimental datasets are introduced and discussed in Section 5 (with sample images). Section 6 reports the experiments on these datasets and a state-of-the-art comparison. We conclude the article in Section 7.

2. Related work

A feature or piece of visual information is derived from image contents to differentiate two images from each other. Feature extraction is the process of converting input images into a set of feature values or feature vectors [66]. There are four main kinds of visual information: (1) color, (2) texture, (3) shape, and (4) spatial. Low-level features are those that can be retrieved automatically from an image without any spatial information. The color, texture, and shape descriptors contribute low-level features. We can extract these features by applying the operator locally or globally. Global methods consider complete images, while local ones divide images into patches and employ descriptors locally.

We can retrieve the color contents of an image with color histograms [57], color moments [16], color coherence vectors (CCV) [58], color correlograms [27, 28], dominant color descriptors (DCD) [47, 65], fuzzy color histograms (FCH) [35], dominant color structure descriptors (DCSD) [85], color difference histograms (CDH) [40], and weighted DCDs (WDCD) [76].

We can represent shapes with stochastic modeling [51], shape signatures with Fourier [91], chain code distribution [75], Hu moments [87], Legendre moments [71], convex hulls [48], shape decomposition [80], and shape matrices [51].

For textural representation, we can use gray-level co-occurrence matrices (GLCM) [25], local binary pattern (LBP) [54], multi-scale LBP (MS-LBP) [55], local ternary patterns (LTP) [77], local directional patterns (LDP) [30], local tetra patterns (LTrP) [52], multi-scale local spatial binary patterns (ML-SBP) [86], bag-of-features models [89], and bag-of-filters-LBP models (BOF-LBP) [17].

The researchers prefer the use of two or more feature descriptors as mentioned here. The approaches with multiple descriptors include CH + LDP [94], color directional local quinary patterns (CDLQP) [82], MS-LBP + GLCM [73], CCV + LBP [56], CH + LDP + SIFTBOF [93], multi-color channel local extrema patterns (MCLEP) [62], wavelet + LBP + Legendre moments (LM) [71], correlation histogram + adaptive tetrolet transform + joint histogram [61], modified micro structure descriptor (MMSD) + LBP + weighted adjacent structure (WAS) [53], square texton histograms (STH) [63], composite micro structure descriptors (CMSD) [81], and multi-level colored directional motif histograms (MLCDMH) [60].

Bhunia et al. [6] used the V component of the HSV color space to compute the diagonally symmetric co-occurrence pattern (DSCoP) and H-S components for CH.

After this, they applied GLCM over DSCoP and merged these features with the CH features. The authors experimented on the Corel-1k, Corel-5k, Corel-10k, MIT VisTex, and Salzburg texture datasets. Jeena Jacob et al. [31] proposed a deep color-texture-based approach titled as deep inter-channel colored-texture pattern (ICCTP) for CBIR and face recognition. The authors used a convolutional neural network (CNN) to acquire the features and estimated retrieval performance on the CIFAR-10, Corel-1k, Corel-10k, and Facescrub datasets. Ashraf et al. [5] introduced an integrated approach of the color moments, discrete wavelet transform (DWT), Gabor wavelet transform (GWT), and color edge directivity Descriptor (CEDD) for CBIR. The authors tested these on the Corel-1k, Corel-1500, Corel-5k, and Ghim-10k datasets. Hong Chen et al. [10] presented a color histogram, overall structure feature (OSF), and Canny edge descriptors. OSF represents the edge and luminance structure descriptions. The authors provided the feature processing with vector quantization and block truncation coding and verified them with Corel-1k and Caltech-101 using four different distance measures (Euclidean, Manhattan, Cosine, and modified Canberra). Niu et al. [53] modified MSD and used it with uniform LBP for feature extraction. The authors created a microstructure map by examining the relationships of shape features with textural and color features. The manuscript covers experiments with five different benchmark datasets (viz., Corel-k, Corel-5k, Corel-10k, Ghim-10k, and CIFAR-10) with various numbers of retrieval images. The authors used a weighted adjacent structure (WAS) for similarity computation. Xie et al. [87] introduced a consistent zone for a DCD operator and used these features with Hu moments; the Hu moments gave the shape information, and the DCD provided the color information. The authors carried out their work on the Corel-1k, Corel-5k, and Corel-10k datasets. Joseph et al. [34] used color moments, HSV color histogram, color correlogram, GLCM, wavelet transform, dominant color, and region-based descriptors. The author introduced a hybrid k-means moth-flame optimization algorithm for the feature-selection process and tested this approach on the Corel-1k dataset. Ghozzi et al. [22] constructed a Type-2 beta fuzzy membership for the descriptor vectors and measured the similarity with the interval Type-2 beta fuzzy near sets (IT2FNMS) approach. Furthermore, the authors introduced three fuzzy similarity measures and tested them on the Corel-1k, SIMPLIcity, Caltech-101, and ImageNet datasets.

3. Methodology overview

3.1. Improved Color Coherence Vector (ICCV)

Pass et al. [58] proposed a color coherence vector (CCV) to improve the color histogram (CH) by utilizing the cohesiveness of the pixels. The idea was to store more spatial information. The descriptor identified the same colored large pixel formations that were recognized as the regions. The number of colors and the percentage of the minimum number of pixels for region formation was specified initially. Usually, the most researchers prefer 1% of an image's size as the minimum number of

pixels for formulating cohesive components (regions). This unique number is called τ. The regions give strong color spatial knowledge of an image. Suppose α is the number of coherent pixels, β is the number of incoherent pixels, and the image is discretized in n colors. We can define CCV as a set of vectors, and each vector is for distinct discretized color. Then, it can be mathematically written as follows:

$$\langle(\alpha_1, \beta_1), (\alpha_2, \beta_2), ..., (\alpha_n, \beta_n)\rangle \tag{1}$$

CCV uses a two-dimension vector to describe the color-feature distribution; it focuses on coherent (α) and incoherent pixels (β). If we analyze CCV, we can see that it covers more spatial information than CH but fails to take care of position information. After calculating α and β, we can compute the mean of the coordinate of the pixels of the maximum connected region in the coherent pixels. Let us say that this information is γ; then, we can define this improved version of CCV (ICCV) as follows:

$$\langle(\alpha_1, \beta_1, \gamma_1), (\alpha_2, \beta_2, \gamma_2), ..., (\alpha_n, \beta_n, \gamma_n)\rangle \tag{2}$$

Chen et al. [9] proposed an ICCV method in their conference work for CBIR. The authors described CCV and ICCV with the appropriate figures and explanations in their article. γ gives the midpoint of the max-connected coherent region; it is easy to compute this from α. CCV has two dimensions, and γ is also a two-dimensional coordinate value. Thus, there will be a total of 4n features for each image. The proposed work utilized $\tau = 1\%$ and n = 27 colors to compute the 108 features from the ICCV descriptor.

3.2. Gray-Level Co-occurrance Matrix (GLCM)

The gray-level co-occurrence matrix (GLCM) was proposed by Haralick et al. [25]. GLCM (or the gray-tone spatial-dependence matrix) is a statistical method for the texture-feature analysis of an image. This matrix computes frequently occurring pixel pairs with distinct values in a specified spatial direction based on the relationships between them. The occurrence of the pixels has a particular direction and distance. GLCM expresses properties (viz., uniformity, energy, correlation, contrast, and homogeneity) about the spatial relationship of the gray-level intensity values and their distribution in an image. The structural arrangement of the pixels helps when analyzing texture features and determining the pixel-pair value distribution. This helps us understand significant textural properties such as the coarseness, smoothness, and roughness of any surface.
We can define GLCM as follows:

$$P(i, j, d, a) = nu[(k, l)(m, n) \in (M \times N) \times (M \times N)] \tag{3}$$

where $I(k, l) = i, I(m, n) = j$; nu = the number of elements in the set; k, l, m, n show the horizontal and vertical positions of the pixel, and the $P(i, j) = (i, j)th$ entry in the GLCM matrix, where:

$$a = \begin{cases} 0^\circ & \text{if } k-m=0, |l-n|=d \\ 45^\circ & \text{if } k-m=d, l-n=-d \\ & \text{or } k-m=-d, l-n=d \\ 90^\circ & \text{if } |k-m|=d, l-n=0 \\ 135^\circ & \text{if } k-m=d, l-n=d \\ & \text{or } k-m=-d, l-n=-d \end{cases} \tag{4}$$

Figure 1 shows an example of an image along with its GLCM matrix. The GLCM measures are summarized below, along with their mathematical equations.

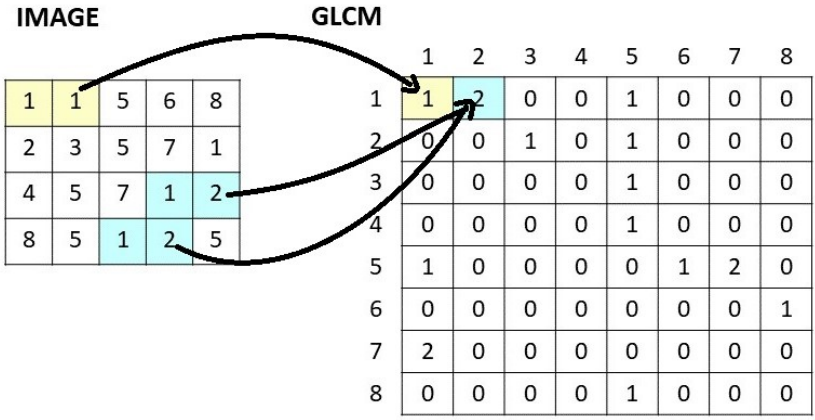

Figure 1. GLCM matrix [25]

Uniformity or Energy: Estimate the number of repeated pairs – range = [0 1].

$$Ene = \sum_{i,j} P_{ij}^2 \tag{5}$$

The energy will be high if the occurrence of repeated pixel pairs is high.

Entropy: Measures how randomly the gray-levels are distributed.

$$Ent = \sum_{i,j} P_{ij} \log P_{ij} \tag{6}$$

The values are high if the gray levels are randomly distributed.

Contrast: Measures the local contrast of an image. If the pixel pair seems to be similar, then the contrast is expected to be low.

$$Con = \sum_{i,j} |i-j|^k P_{ij}^l \tag{7}$$

Homogeneity: Measures the smoothness of the image. If the pixel pair seems to be similar, then the homogeneity is expected to be high.

$$Hom = \sum_{\substack{i,j \\ i \neq j}} \frac{P_{ij}^{l}}{|i-j|^{k}} \tag{8}$$

Correlation: Measure of gray-level line dependencies.

$$Cor = \sum_{i,j} \frac{(i-\mu)(j-\mu)P_{ij}}{\sigma^{2}} \tag{9}$$

GLCM is one of the important texture descriptors for image retrieval. In the proposed work, the offset is taken concerning three pixels in each of the four directions.

3.3. Multi-scale Local Binary Patterns (MS-LBP) of Discrete Wavelet Transform (DWT)

The discrete wavelet transform (DWT) is a multi-resolution analysis of an image at the spatial and frequency domains. This is the technique of decomposing an image into the sum of its wavelet functions at various scales and variations. The 2D wavelet transform decomposes the image into four sub-bands, viz., approximation (A or LL), horizontal (H or LH), vertical (V or HL), and diagonal (D or HH) [4]. The high-frequency components sharpen the image's details, while the low-frequency components smooth the image. The dwt2 function is used from the wavelet toolbox of MATLAB 2018a to compute the 2D DWT of the image. Arai et al. [4] propounded a wavelet-based image-retrieval system using a wavelet, Gabor, and CM for feature extraction. Joseph et al. [34] followed a multi-feature-based approach by using a wavelet transform for CBIR that used a hybrid k-means moth-flame optimization algorithm.

3.3.1. DWT-MSLBP formation

First, the 2D DWT is applied over the image to capture the features; then, these features are supplied to the modified MS-LBP version of the MS-LBP. We have proposed a 7×7 scale MS-LBP as given in [8,73] for the feature description due to its ability to cover more spatial details at various scales. A larger patch size makes it less noise- and illumination-invariant than with a conventional LBP operator. This composite feature-extraction strategy of two texture descriptors (DWT and MS-LBP) delivers a powerful texture-feature representation that we call DWT-MSLBP. A sample image from the Corel-1k dataset was used for the DWT-MSLBP feature visuals. Figure 2 shows the visualization of the DWT-MSLBP. The histogram features of DWT-MSLBP are computed for the proposed framework.

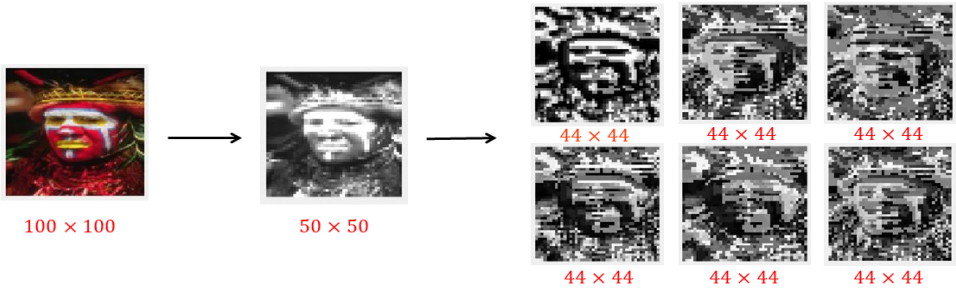

Figure 2. DWT-MSLBP feature visualization

4. Proposed methodology for image retrieval

4.1. Proposed framework

The ICCV operator extracts the color details, and GLCM and DWT-MSLBP extract the texture features. The proposed image-retrieval framework uses the color and texture descriptors for feature extraction. Figure 3 shows the building blocks of the proposed framework.

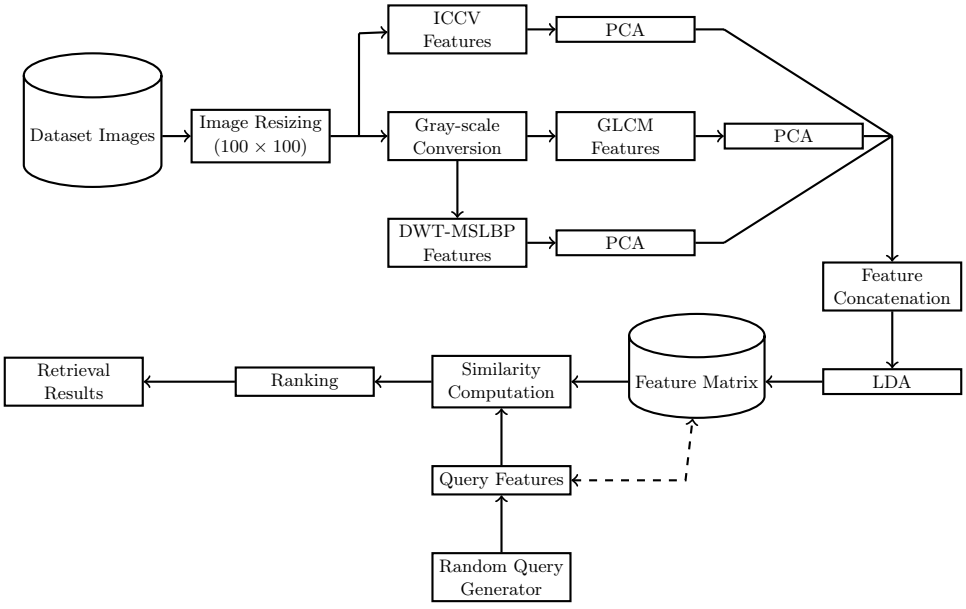

Figure 3. Proposed framework

The proposed framework is summarized here: first, the dataset images are resized into a 100×100 scale. The resizing function reduces the image size and applies

the same feature framework for all of the datasets. The GLCM and DWT-MSLBP descriptors use the gray-scaled images as input to compute the feature vectors. All three feature descriptors (ICCV, GLCM, and DWT-MSLBP) are computed separately; then, PCA is utilized to reduce the feature dimensionality. We used features with a total variance of 99.99% in order to reduce the feature vector size. After this, all three PCA feature vectors that were computed from the previous stage are concatenated and passed to LDA for optimal feature selection. In the last phase, the similarity is computed between the query and dataset images using two different metrics to achieve the most relevant results.

4.2. Proposed algorithm

- `Part 1: Feature Vector Formation`
  1. *Pre-process dataset images.*
  2. *Compute ICCV, GLCM, and DWT-MSLBP features.*
  3. *Apply PCA with 99.99% variance.*
  4. *Concatenate features.*
  5. *Apply LDA to generate final feature vector.*
     - If there are C classes in the image dataset then, the LDA will produce $C - 1$ optimal features.

- `Part 2: Similarity Computation and Retrieval`
  1. *Generate random query indices.*
  2. *Measure similarities.*
  3. *Rank images.*
     - Rank images based on minimum sorted distance.
  4. *Measure performance with precision and recall.*

The query generator creates the random image query indices for each class of the dataset. We have supplied all the dataset query samples to the proposed framework simultaneously. The query features have been loaded from the feature matrix based on the image indices, and computed the distance between the query and the dataset features. The proposed work utilizes two distance measures to assess the retrieval outcomes.

4.3. Feature Vector Dimensions

We have shown the number of features at each stage of the proposed framework for various datasets (Corel-1k, Corel-5k, Corel-10k, and Ghim-10k). The researchers prefer the graphical representation or variance computation to decide the number of eigenvectors. The number of principal components/eigenvectors is selected with the

variance-based approach. Those components that have a total of 99.99% variance are selected for feature reduction. Table 1 reports the feature-vector size for the different datasets at various stages of the proposed framework. It is clear from the table that PCA with a 99.99% variance preserves the healthy features, and LDA selects the robust ones from the set of concatenated features. We have highlighted the number of LDA features that were selected for similarity matching and retrieval by the proposed work.

Table 1
Proposed feature-vector sizes at various stages of framework

Dataset	Descriptor	Size of feature vector after PCA	Size of feature vector after Concatenation	Size of feature vector after LDA
Corel-1k	ICCV	75	1015	**9**
	GLCM	19		
	DWT-MSLBP	921		
Corel-5k	ICCV	85	1549	**49**
	GLCM	20		
	DWT-MSLBP	1444		
Corel-10k	ICCV	86	1571	**99**
	GLCM	20		
	DWT-MSLBP	1465		
Ghim-10k	ICCV	83	1534	**19**
	GLCM	21		
	DWT-MSLBP	1430		

*ICCV —> 108 features; GLCM —> 48 features; DWT-MSLBP —> 1536 features

4.4. Similarity measures

The proposed work incorporates two distinct template-matcher classifiers: one is Euclidean distance, and the other is city block distance. The performance of the classifier depends on the types of features that are retrieved. Initially, we checked the perfor-

mance of the proposed method with other similarity measures; however, these two are the most convincing for our features.

`Euclidean:` A straight-line distance between two corresponding elements of a feature vector. It is a bin-by-bin distance (also known as Pythagoras distance). Each pixel in a converted image has a value that corresponds to the distance to the nearest pixel in the image.

$$L_2(M,N) = \left(\sum_{i=1}^{k} |m_i - n_i|^2\right)^{\frac{1}{2}} \tag{10}$$

`City block:` Apart from Euclidean distance, another popular distance metric of the L_P family is Manhattan distance (also known as taxi-cab distance or city block distance). This is basically the sum of the absolute differences of the corresponding elements of two feature vectors.

$$L_1(M,N) = \sum_{i=1}^{k} |m_i - n_i| \tag{11}$$

4.5. Performance measures

Precision and recall metrics are used to evaluate CBIR's performance. Both of these measures depend on the number of relevant images that are retrieved by the image-retrieval system.

$$\texttt{Precision}, P = \frac{R_{ID}}{R_{ID} + N_{ID}} = \frac{Total\ Number\ of\ Relevant\ Images\ Retrieved}{Total\ Number\ of\ Retrieved\ Images} \tag{12}$$

$$\texttt{Recall}, R = \frac{R_{ID}}{C_{ID}} = \frac{Total\ Number\ of\ Relevant\ Images\ Retrieved}{Total\ Number\ of\ Relevant\ Images\ in\ Dataset} \tag{13}$$

where R_{ID} is the total number of relevant images, and N_{ID} is the total number of irrelevant images that are obtained in the retrieval results. C_{ID} is the total number of relevant images in the dataset for a particular object class (images from the same dataset class are relevant to each other); i.e., if there are 100 images in a particular object category of a dataset, then C_{ID} will be 100 for that category (C_{ID} = samples per class in the dataset). The user is solely interested in the top-most images (generally, the top 10 or 100). The total number of top-most images in the retrieval results can vary from 1 to N. We can decide this number N (where $1 \leq N \leq C_{ID}$). In the proposed study, we utilized the average retrieval precision (ARP) and average retrieval rate (ARR) as the performance metrics (which are the average values of the accuracy and recall, respectively). The proposed algorithm is tested ten times, and the values of the ARP and ARR measures are averaged to ensure steady performance. ARP and ARR are assessed for ten distinct N values.

4.6. Advantages of proposed framework

Each descriptor that was used in the proposed work has its pros and cons. ICCV is a stronger operator than CCV, as it calculates more color spatial information with the help of the mid-point coordinate of the max-connected coherent region. While CCV has only 2n features, ICCV captures 4n (n is the number of colors). GLCM is computationally efficient when compared to the ICCV operator and captures texture details in various directions; alone, it will not serve the purpose of maximal performance. Furthermore, DWT-MSLBP suffers from the problems of high dimensionality and being computationally inefficient; however, it produces a rich set of textural information.

When only using color or texture, the description will not yield a potent image retrieval system. When utilizing both, we can absorb the benefits of the texture and color representations. Furthermore, we must reduce the size of the feature vector to make the system efficient without sacrificing the effectiveness of the features. We have taken care of these things to produce an efficient and robust system. As the number of optimal features falls, this can make an efficient and effective image-retrieval system. We have combined the advantages of all three descriptors and taken care of the high dimensionality of the features; thus, the proposed features for image retrieval give a powerful feature description by summing up the benefits from all operators. Furthermore, it is a computationally efficient approach without suffering from low ARP or ARR.

5. Benchmark datasets

We focused on four different datasets (namely, Corel-1k [38, 84], Corel-5k [41], Corel-10k [41], and Ghim-10k [41]) to carry out the proposed work. The images in the Corel-1k dataset have ten diverse categories (viz., Africans, beaches, buildings, buses, dinosaurs, elephants, flowers, horses, mountains, and food). Each image has a size of either 256×384 or 384×256. There are 100 images in each image category. The images are in the JPEG format. The 5000 images from the Corel-10k dataset are included in the Corel-5k dataset. The first 50 item categories from the source Corel-10k dataset were used to create it.

The Corel-10k dataset consists of 100 classes and 10,000 images of different subjects, including sunsets, beaches, flowers, buildings, cars, horses, mountains, fish, cuisine, doors, etc. Each category has 100 JPEG images with sizes of 192×128 or 128×192. The Corel-10k dataset consists of 10,000 images. The Ghim-10k dataset has 20 categories. There are 10,000 / images from such diverse content as sunsets, ships, flowers, buildings, cars, mountains, insects, etc. Each category involves 500 images with sizes of 400×300 or 300×400 in the JPEG format. Some samples from these datasets are as shown in Figure 4.

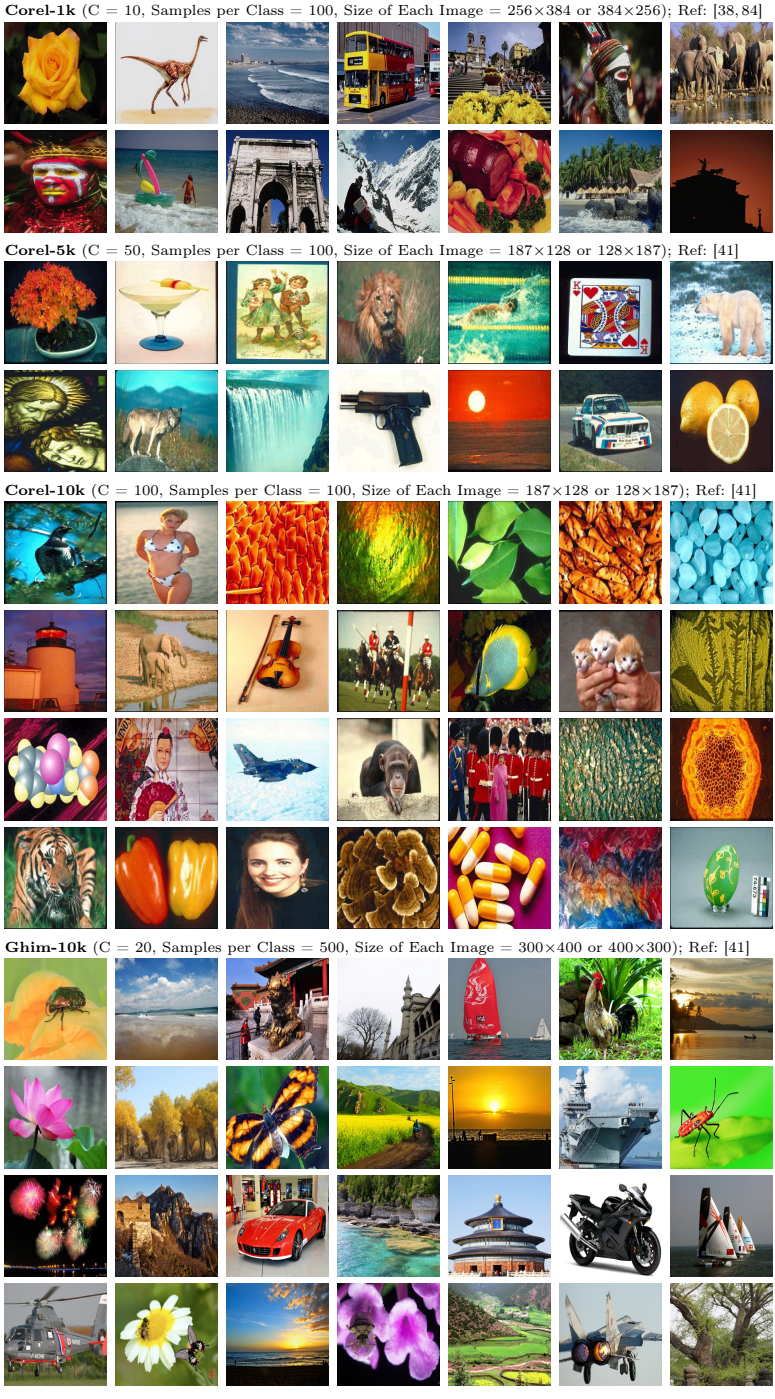

Figure 4. Sample images from experimental datasets

6. Experiments

We analyzed the effectiveness of the proposed algorithm by employing ARP and ARR for various numbers of top images (N). This section demonstrates the outcomes of four benchmark datasets on the proposed method. Furthermore, this section also compares and contrasts the performance of the proposed approach with several existing state--of-the-art methods. In addition, we discuss the proposed method's time complexity and relative performance. We conducted all of the tests on a laptop with an i5-2430M CPU that had an L3 cache of 3MB, 256GB SSD, 500GB HDD, and 6GB RAM using MATLAB 2018a.

6.1. Performance on Corel-1k

We evaluated the proposed methodology using the various numbers of top-N images. The proposed method reports 100% ARP for all of the measured values of N (ranging from 10–100) with both distance measures (Euclidean and city-block). Furthermore, we show the ARR of the proposed work over Corel-1k in Figure 5 for different values of N. This indicates that the presentation of Corel-1k gives an equivalent ARR with the Euclidean and city-block distances. The same is true for the reported ARP values of the proposed method over a particular dataset.

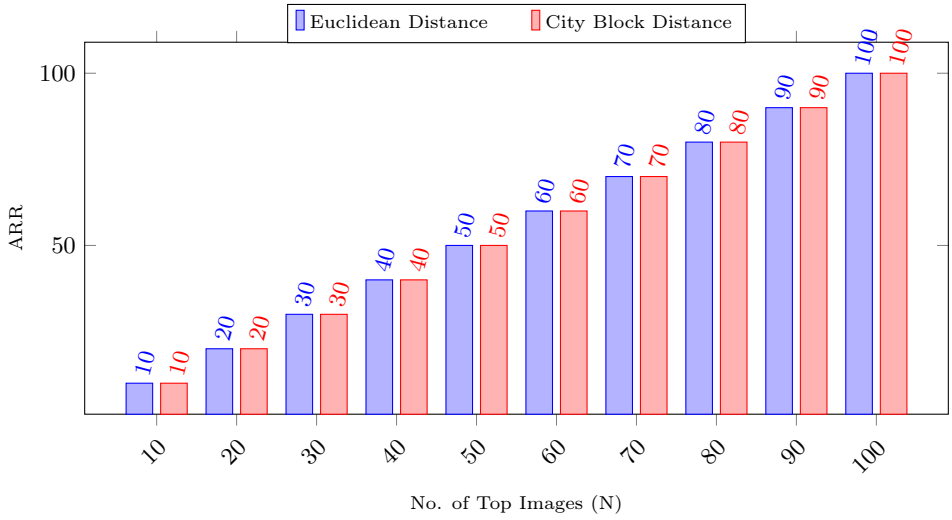

Figure 5. Average retrieval rate (ARR) on Corel-1k

A state-of-the-art comparison of the proposed method over the Corel-1k dataset is shown in Table 2. We assessed the performance for the top 10, 20, and 100 images in Table 2. The ARP of the proposed method performed marginally better than methodologies such as MLCDMH [60], LCP [74], CMSD [81], MLRDEP [59], and GSH [90] respectively.

Table 2
Comparison with state-of-the-art methods on Corel-1k dataset*

Reference	Method	$N = 10$		$N = 20$		$N = 100$	
		ARP[1]	ARR[2]	ARP[1]	ARR[2]	ARP[1]	ARR[2]
Pradhan et al. [60]	MLCDMH	64.00	6.40	59.60	11.90	–	–
Chen et al. [10]	CH + OSF	–	–	72.80	–	–	–
Subash Kumar and Nagarajan [74]	LCP	78.30	8.00	73.90	15.00	–	–
Bhunia et al. [6]	DSCoP + Modified CH	79.00	7.90	76.00	15.00	55.00	55.00
Xie et al. [87]	DCD + Hu Moments	80.40	8.04	74.05	14.81	–	–
Fadaei et al. [21]	DCD + Wavelet + Curvelet	–	–	76.50	13.09	–	–
Umamaheswaran et al. [81]	CMSD	81.44	9.75	–	–	–	–
Pavithra and Sharmila [59]	MLRDEP	81.80	–	–	–	–	–
Mohiuddin et al. [50]	CCV + LBP	82.52	8.00	77.00	15.00	55.50	55.00
Yuan and Liu [90]	GSH	82.80	8.28	–	–	–	–
Niu et al. [53]	MMSD + LBP + WAS	83.30	8.30	78.78	15.76	–	–
Joseph et al. [34]	KMFO	–	–	81.30	16.20	–	–
Srivastava and Khare [72]	DWT + SURF + GLCM	84.00	–	–	–	–	61.73
Pardede et al. [56]	Low-level + RF	85.59	–	–	–	–	27.70
Srivastava and Khare [73]	MS-LBP + GLCM	85.72	–	–	–	–	55.17
Ahmed et al. [2]	Key points + CH	85.90	–	–	–	–	31.60
Zhou et al. [93]	CH + LDP + SIFTBOF	88.00	8.00	85.00	17.50	71.00	71.00
Ashraf et al. [5]	CM + DWT + GWD + CEDD	91.00	–	87.50	–	62.00	87.50
Ghozzi et al. [22]	IT2FBNS	–	–	88.65	19.14	–	–
Chavda and Goyani [8]	MS-LBP + CH	92.90	9.29	91.05	18.21	–	–
Singh and Batra [70]	Bi-CBIR	–	–	92.00	18.40	–	–
Srivastava and Khare [71]	DWT + LBP + LM	99.95	–	–	–	–	92.04
Proposed method		**100**	**10**	**100**	**20**	**100**	**100**

*Blank cells indicate that data was not clear or authors did not test for those values of N where N was number of top images. [1]ARP: average retrieval precision [%]. [2]ARR: average retrieval rate [%].

The proposed method's ARP and ARR were 18.7 and 3.8% higher, respectively, than hybrid k-means moth-flame optimization algorithm-based CBIR (KMFO) [34] for $N = 20$. Furthermore, the proposed method's ARP and ARR led by 11.35 and 0.86%, respectively, over the Type-2 beta fuzzy near set approach (IT2FBNS) [22].

Thus, the proposed method performed excellently by giving the maximal outcomes of ARP and ARR at varying N values. Furthermore, it was also more effective (with a total of nine features) than the identified state-of-the-art literature methods.

6.2. Performance on Corel-5k

The performance of the proposed method on the Corel-5k dataset is shown in Figures 6 and 7 for different values of N with both similarity measures.

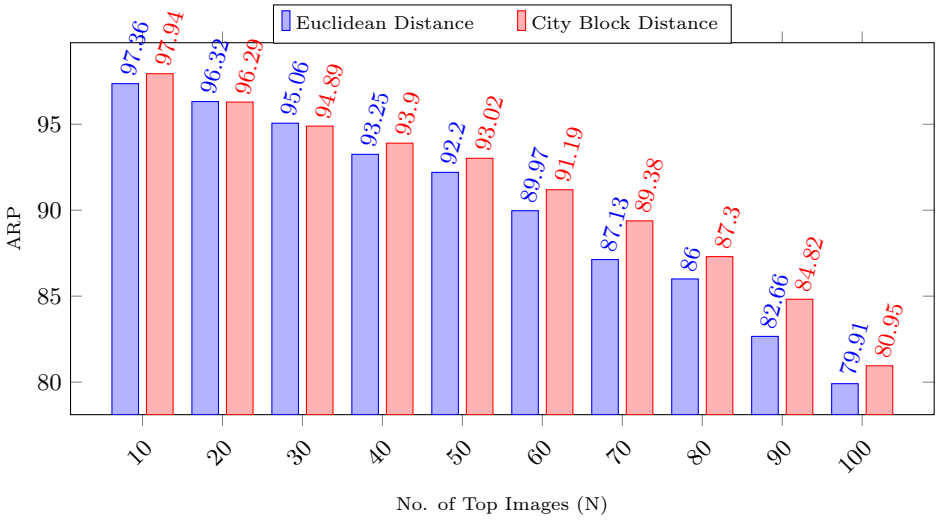

Figure 6. Average retrieval precision (ARP) on Corel-5k

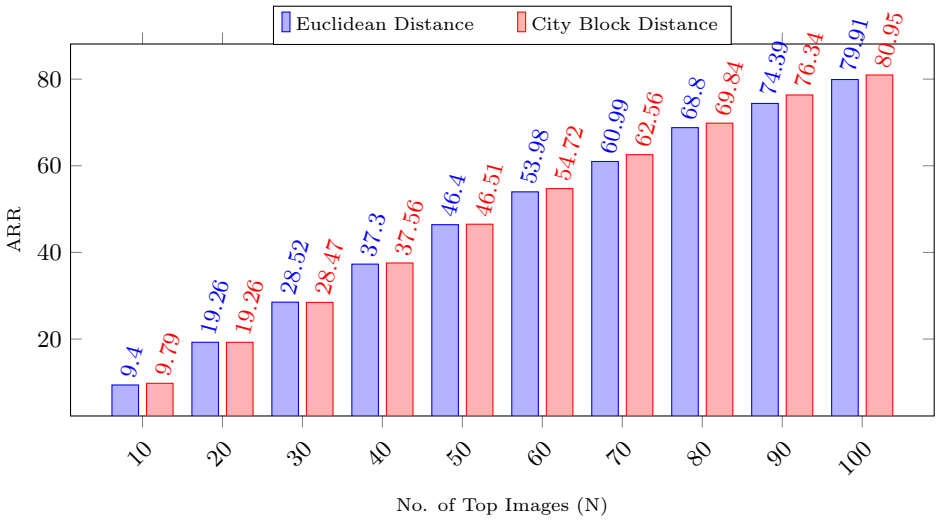

Figure 7. Average retrieval rate (ARR) on Corel-5k

City-block gave better results as compared to Euclidean for $N = 10$. As we move from left to right, ARP decreases while ARR increases for both measures. City-block has the advantage of a low time complexity as it computes an absolute difference, while Euclidean goes for the squared norm difference. We can prefer to choose city-block distance, as it was faster and provided similar (or better) performance than Euclidean for the Corel-5k dataset samples.

The state-of-the-art comparison of the proposed method over the Corel-5k dataset is shown in Table 3. We assessed the performance for the top 10, 12, 20, and 100 images in Table 3. The ARP of the proposed method marginally surpassed the examined state-of-the-art methods (viz., DWT + LBP + LM [71], MLCDMH [60], MS-LBP + GLCM [73], STH [63], DCD + Hu moments [87], MIFH [12], CH + LDP + SIFTBOF [93], DSCoP + modified CH [6], MMSD + LBP + WAS [53], MS-LBP + CH [8], and CM + DWT + GWD + CEDD [5]) As described, the proposed method performed better (with only 49 features) than the investigated state-of-the-art methods on the Corel-5k dataset.

Table 3

Comparison with state-of-the-art methods on Corel-5k dataset*

Reference	Method	$N = 10$		$N = 12$		$N = 20$		$N = 100$	
		ARP[1]	ARR[2]	ARP[1]	ARR[2]	ARP[1]	ARR[2]	ARP[1]	ARR[2]
Srivastava and Khare [71]	DWT + LBP + LM	56.76	–	–	–	–	–	–	42.53
Pradhan et al. [60]	MLCDMH	59.10	5.91	–	–	51.30	10.26	–	–
Srivastava and Khare [73]	MS-LBP + GLCM	59.50	–	–	–	–	–	–	27.28
Raza et al. [63]	STH	60.28	7.23	–	–	–	–	–	–
Xie et al. [87]	DCD + Hu Moments	63.47	6.35	60.05	7.26	53.92	10.78	–	–
Chu and Liu [12]	MIFH	–	–	60.16	7.21	–	–	–	–
Zhou et al. [93]	CH + LDP + SIFTBOF	65.90	7.00	–	–	58.00	12.00	36.90	36.90
Bhunia et al. [6]	DSCoP + Modified CH	66.00	6.60	–	–	58.00	13.00	35.00	35.00
Niu et al. [53]	MMSD + LBP + WAS	70.20	7.00	67.93	8.15	–	–	–	–
Chavda and Goyani [8]	MS-LBP + CH	70.92	7.09	–	–	64.53	12.91	–	–
Ashraf et al. [5]	CM + DWT + GWD + CEDD	–	–	79.83	9.58	–	–	–	–
Proposed Method		**97.94**	**9.79**	**97.68**	**11.72**	**96.32**	**19.26**	**80.95**	**80.95**

*Blank cells indicate that data was not clear or authors did not test for those values of N where N was number of top images. [1]ARP: average retrieval precision [%]. [2]ARR: average retrieval rate [%].

6.3. Performance on Corel-10k

The performance of the proposed method on the Corel-10k dataset is shown in Figures 8 and 9 for different values of N with both similarity measures.

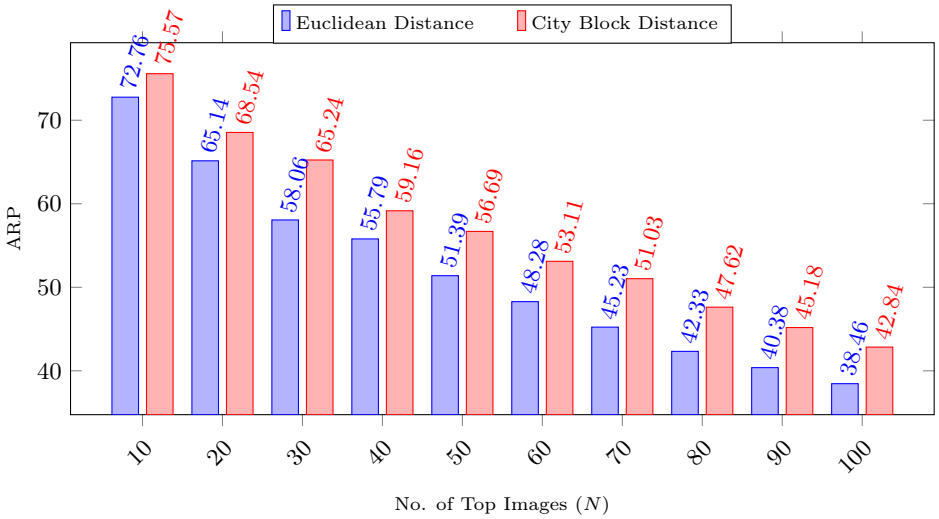

Figure 8. Average retrieval precision (ARP) on Corel-10k

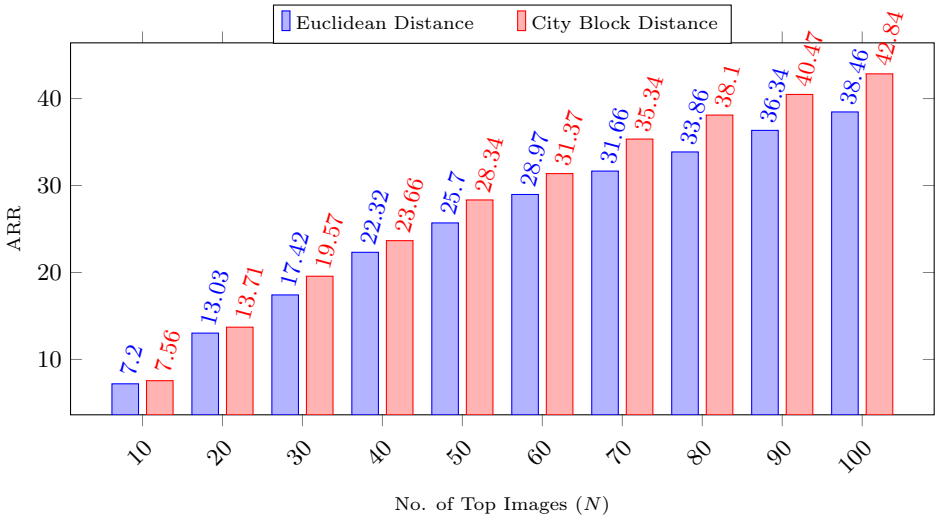

Figure 9. Average retrieval rate (ARR) on Corel-10k

The proposed method achieved a maximum ARP of 75.57% at $N = 10$ and a maximum ARR of 42.84% at $N = 100$ for the Corel-10k dataset. City-block distance

performed better in terms of ARP and ARR than Euclidean on the Corel-10k dataset. Thus, the feature representation of the Corel-10k dataset is more suitable to city-block distance than Euclidean. The state-of-the-art comparison of the proposed method over the Corel-10k dataset is shown in Table 4. We assessed the performance for the top 10, 12, 20, and 100 images in Table 4. Thus, the ARP and ARR of the proposed method on Corel-10k were better than all of the examined state-of-the-art methods.

Table 4
Comparison with state-of-the-art methods on Corel-10k dataset

Reference	Method	$N = 10$		$N = 12$		$N = 20$		$N = 100$	
		ARP[1]	ARR[2]	ARP[1]	ARR[2]	ARP[1]	ARR[2]	ARP[1]	ARR[2]
Rohini and Bindu [64]	QLTP	30.10	–	–	–	–	–	–	18.20
Srivastava and Khare [71]	DWT + LBP + LM	35.37	–	–	–	–	–	–	24.86
Pardede et al. [56]	Low-level + RF	37.16	–	–	–	–	–	–	23.59
Rao et al. [62]	MCLEP	46.00	5.00	–	–	33.50	8.00	18.00	25.60
Ghozzi et al. [22]	IT2FBNS	46.00	5.52	–	–	–	–	–	–
Pavithra and Sharmila [59]	MLRDEP	47.00	–	–	–	–	–	–	–
Raza et al. [63]	STH	48.03	5.76	–	–	–	–	–	–
Srivastava and Khare [73]	MS-LBP + GLCM	48.69	–	–	–	–	–	–	20.35
Chavda and Goyani [8]	MS-LBP + CH	54.46	5.44	–	–	45.22	9.04	–	–
Chu and Liu [12]	MIFH	–	–	52.96	6.36	–	–	–	–
Xie et al. [87]	DCD + Hu Moments	55.90	5.59	53.17	6.38	46.17	9.23	–	–
Yuan and Liu [90]	GSH	–	–	54.84	6.58	–	–	–	–
Ahmeda et al. [2]	Key points + CH	56.70	–	–	–	–	–	–	21.00
Pradhan et al. [60]	MLCDMH	57.00	5.70	–	–	51.10	10.24	–	–
Zhou et al. [93]	CH + LDP +SIFTBOF	57.70	5.50	–	–	49.00	8.00	28.80	28.80
Bhunia et al. [6]	DSCoP + Modified CH	57.00	5.70	–	–	50.00	11.00	28.00	28.00
Niu et al. [53]	MMSD + LBP + WAS	61.10	6.10	58.52	7.02	–	–	–	–
Proposed Method		**75.57**	**7.56**	**74.50**	**8.94**	**68.54**	**13.71**	**42.84**	**42.84**

*Blank cells indicate that data was not clear or authors did not test for those values of N where N was no. of top images. [1]ARP: average retrieval precision [%]. [2]ARR: average retrieval rate [%].

6.4. Performance on Ghim-10k

The performance of the proposed method on the Ghim-10k dataset is shown in Figures 10 and 11 for different values of N with both similarity measures.

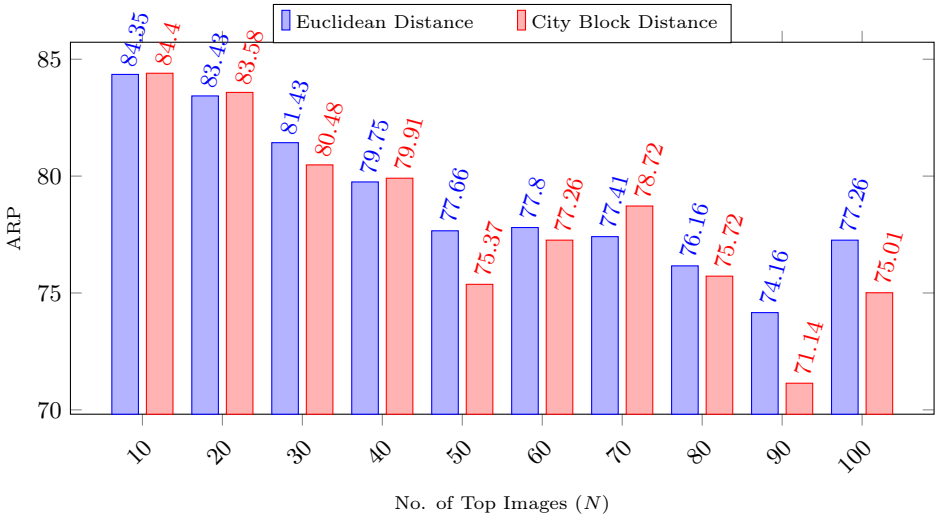

Figure 10. Average retrieval precision (ARP) on Ghim-10k

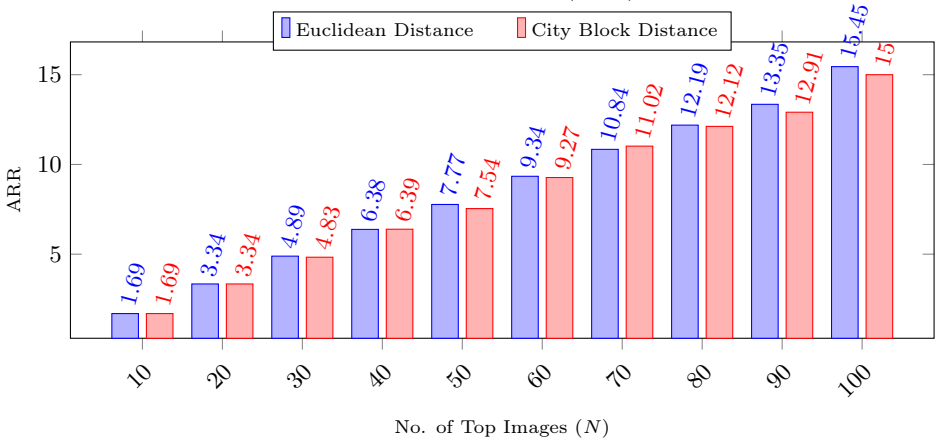

Figure 11. Average retrieval rate (ARR) on Ghim-10k

We achieved a maximum ARP of 84.4% at $N = 10$ and a maximum ARR of 15.45% at $N = 100$ on this complex dataset. The retrieval performance on the Ghim-10k dataset based on both distance measures is nearly the same. Still, city-block should be preferred, as it has faster execution. Thus, the ARP and ARR values of the proposed method on Ghim-10k are higher than the identified state-of-the-art methods in the literature.

The state-of-the-art comparison of the proposed method over the Ghim-10k dataset is shown in Table 5. The ARP of the proposed method outclassed several state-of-the-art methods such as low-level descriptors with relevance feedback [56], MIFH [12], CH + LDP [94], SSH [41], DWT + SURF + GLCM [72], GSH [90], MS-LBP + CH [8], CH + LDP + SIFTBOF [93], MMSD + LBP + WAS [53], MS-LBP + GLCM [73], MLRDEP [59], and CM + DWT + GWD + CEDD [5].

Table 5
Comparison with state-of-the-art methods on Ghim-10k dataset*

Reference	Method	$N = 10$		$N = 12$		$N = 20$		$N = 100$	
		ARP[1]	ARR[2]	ARP[1]	ARR[2]	ARP[1]	ARR[2]	ARP[1]	ARR[2]
Pardede et al. [56]	Low-level + RF	50.78	–	–	–	–	–	–	7.95
Chu and Liu [12]	MIFH	–	–	56.48	1.36	–	–	–	–
Zhou et al. [94]	CH + LDP	–	–	60.00	1.44	–	–	–	–
Liu and Yang [41]	SSH	–	–	61.20	1.47	–	–	–	–
Srivastava and Khare [72]	DWT + SURF + GLCM	63.66	–	–	–	–	–	–	–
Yuan and Liu [90]	GSH	–	–	63.38	1.52	–	–	–	–
Chavda and Goyani [8]	MS-LBP + CH	67.95	1.36	–	–	62.03	2.48	–	–
Zhou et al. [93]	CH + LDP + SIFTBOF	70.00	–	67.10	1.61	63.00	–	49.00	–
Niu et al. [53]	MMSD + LBP + WAS	–	–	68.50	1.65	–	–	–	–
Srivastava and Khare [73]	MS-LBP + GLCM	76.99	–	–	–	–	–	–	–
Pavithra and Sharmila [59]	MLRDEP	80.60	–	–	–	–	–	–	–
Ashraf et al. [5]	CM + DWT + GWD + CEDD	–	–	–	–	76.50	–	–	–
Proposed Method		**84.40**	**1.69**	**84.29**	**2.02**	**83.58**	**3.34**	**77.26**	**15.45**

*Blank cells indicate that data was not clear or authors did not test for those values of N where N was number of top images. [1]ARP: average retrieval precision [%]. [2]ARR: average retrieval rate [%].

6.5. Comparison on various datasets

Figures 12 and 13 report the relativistic performance of the proposed method over several tested benchmark datasets. Figure 12 provides a comparison based on the ARP data, and Figure 13 shows a comparison with ARR. The figures indicate that ARR increases as N increases; however, ARP degrades for three out of the four datasets. The proposed method produces better retrieval rates on the Corel-1k dataset than the other three benchmarks. The reason for this is the lower number of classes

and images as compared to the other three datasets (thus, it makes LDA compute robust features). So, we can say that, when the number of sample classes is lower, LDA performs very well. Furthermore, the proposed approach provides stable ARP on the Ghim-10k dataset as N goes up. The Corel-10k and Ghim-10k datasets are much more complex as compared to the other datasets. Still, the ARP and ARR of the proposed method were marginally better than the identified state-of-the-art methods in the research literature. The Ghim-10k dataset has lower ARR values because of the 500 images in each category of the given dataset. Altogether, the proposed approach provides notable and steady performance based on the ARP and ARR values.

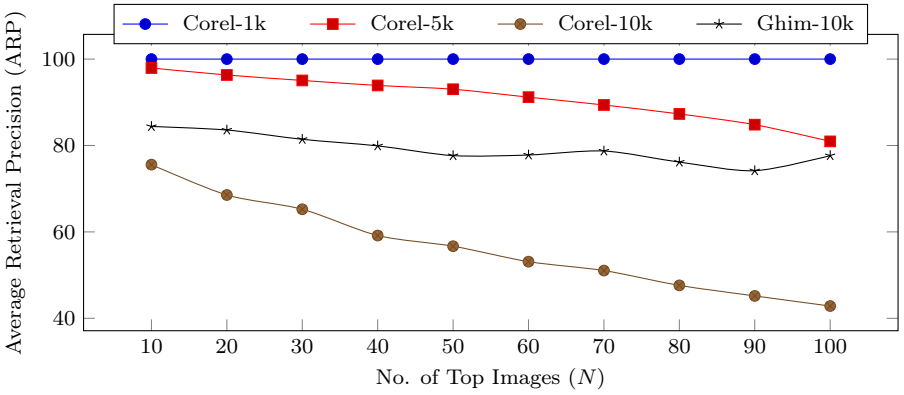

Figure 12. Relative ARP of proposed method on various benchmark datasets

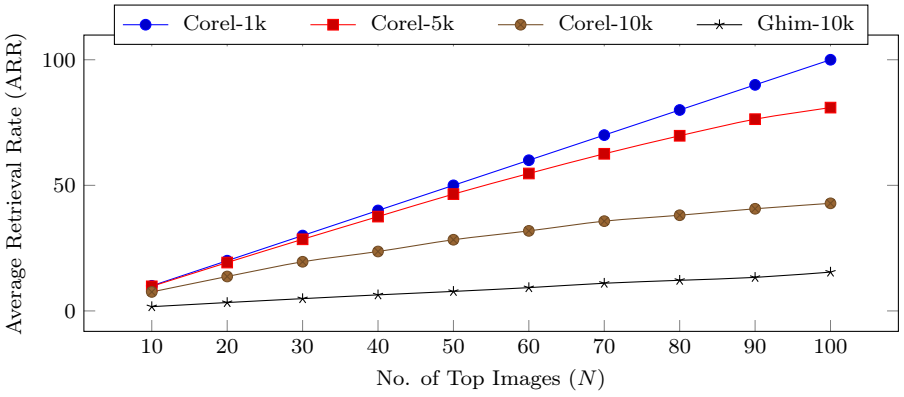

Figure 13. Relative ARR of proposed method on various benchmark datasets

6.6. Time complexity

We must measure the retrieval time (RT) to check the efficiency of the image retrieval system. The RT of the proposed algorithm reported 0.54, 2.1, 4, and 3 seconds for the Corel-1k, Corel-5k, Corel-10k, and Ghim-10k datasets, respectively. We noticed that

the RT of the Corel-1k dataset was lower than in the other three datasets. This is because it used only nine features for retrieval. We compared the RT and ARP of the proposed method with the state-of-the-art methods shown in Table 6. Furthermore, it is clear from the table that the proposed method outperformed several identified method in the literature in terms of ARP and RT.

Table 6
Retrieval time (RT) comparison to state-of-the-art methods on Corel-1k dataset

Reference	Method	ARP[1]	RT[2]
ElAlami [19]	3D Color Histogram + Gabor	73.90	2.65
Fadaei et al. [21]	DCD + Wavelet + Curvelet	76.50	1.74
SR Dubey et al. [17]	BOF of LBP	≈ 68	0.69
SR Dubey et al. [18]	RSHD	≈ 78	0.64
Proposed Method		**100**	**0.54**

[1]ARP: average retrieval precision [%]. [2]RT: retrieval time (seconds)

7. Conclusions

In this paper, we have proposed ICCV, GLCM, and DWT-MSLBP for the CBIR system. The proposed work improves the extracted features by using PCA and LDA; PCA reduces the dimensions, and LDA proffers the most robust features. The 99.99% variance-based principal components that are used in the PCA give the most beneficial features for robust feature selection. We supplied the DWT features to the modified MS-LBP and constructed DWT-MSLBP. This combination serves as a powerful texture-feature description and enhances the proposed method's outcome. Furthermore, the paper practices the ICCV descriptor instead of the conventional CCV operator to improve the feature matrix. We measured the similarities of the images with two different distance measures (Euclidean and city-block). We evaluated the proposed method at diverse numbers of top images on various datasets. The proposed system yielded 100% of the ARP and ARR on the Corel-1k dataset. The comparison to the state-of-the-art research methods on the various benchmark datasets shows the dominance of the proposed work.

References

[1] Abate A.F., Nappi M., Ricciardi S., Tortora G.: FACES: 3D FAcial reConstruction from anciEnt Skulls using content based image retrieval, *Journal of Visual Languages and Computing*, vol. 15(5), pp. 373–389, 2004. doi: 10.1016/j.jvlc.2003.11.004.

[2] Ahmed K.T., Ummesafi S., Iqbal A.: Content based image retrieval using image features information fusion, *Information Fusion*, vol. 51, pp. 76–99, 2019. doi: 10.1016/j.inffus.2018.11.004.

[3] Antani S., Long L.R., Thoma G.R.: Bridging the Gap: Enabling CBIR in Medical Applications. In: *Proceedings of the Twenty-First IEEE International Symposium on Computer-Based Medical Systems*, pp. 4–6, IEEE, 2008. doi: 10.1109/CBMS.2008.133.

[4] Arai K., Rahmad C.: Wavelet Based Image Retrieval Method. In: *Reports of the 260th Technical Conference of the Institute of Image Electronics Engineers of Japan*, pp. 243–247, 2012. doi: 10.11371/wiieej.11.04.0_243.

[5] Ashraf R., Ahmed M., Ahmad U., Habib M.A., Jabbar S., Naseer K.: MDCBIR-MF: multimedia data for content-based image retrieval by using multiple features, *Multimedia Tools and Applications*, vol. 79, pp. 8553–8579, 2020. doi: 10.1007/s11042-018-5961-1.

[6] Bhunia A.K., Bhattacharyya A., Banerjee P., Roy P.P., Murala S.: A novel feature descriptor for image retrieval by combining modified color histogram and diagonally symmetric co-occurrence texture pattern, *Pattern Analysis and Applications*, vol. 23(2), pp. 703–723, 2020. doi: 10.1007/s10044-019-00827-x.

[7] Chavda S., Goyani M.: Content-Based Image Retrieval: The State of the Art, *International Journal of Next-Generation Computing*, vol. 10(3), pp. 193–212, 2019. https://ijngc.perpetualinnovation.net/index.php/ijngc/article/view/166.

[8] Chavda S., Goyani M.: Hybrid Approach to Content-Based Image Retrieval Using Modified Multi-Scale LBP and Color Features, *SN Computer Science*, vol. 1, pp. 1–15, 2020. doi: 10.1007/s42979-020-00321-w.

[9] Chen X., Gu X., Xu H.: An Improved Color Coherence Vector Method for CBIR. In: *Communication and Information Technology Conference*, pp. 33–37, Beijing University Graduate Academic Exchange, 2008.

[10] Chen Y.H., Chang C.C., Hsu C.Y.: Content-based image retrieval using block truncation coding based on edge quantization, *Connection Science*, vol. 32, pp. 431–448, 2020. doi: 10.1080/09540091.2020.1753174.

[11] Choraś R.S.: CBIR system for detecting and blocking adult images. In: *The 9th World Scientific and Engineering Academy and Society International Conference on Signal processing*, pp. 52–57, 2010. https://dl.acm.org/doi/abs/10.5555/1844625.1844636.

[12] Chu K., Liu G.H.: Image Retrieval Based on a Multi-Integration Features Model, *Mathematical Problems in Engineering*, vol. 2020, pp. 1–10, 2020. doi: 10.1155/2020/1461459.

[13] Chuctaya H., Portugal C., Beltran C., Gutierrez J., Lopez C., Tupac Y.: M-CBIR: A medical content-based image retrieval system using metric data-structures. In: *30th International Conference of the Chilean Computer Science Society*, pp. 135–141, IEEE, 2011. doi: 10.1109/SCCC.2011.18.

[14] Colombo C., Alberto D.B.: Visible Image Retrieval. In: V. Castelli, L.D. Bergman (eds.) *Image Databases: Search and Retrieval of Digital Imagery*, pp. 11–33, John Wiley & Sons, Ltd, 2001. doi: 10.1002/0471224634.ch2.

[15] Comon P.: Independent component analysis, a new concept?, *Signal Processing*, vol. 36(3), pp. 287–314, 1994.

[16] Duanmu X.: Image Retrieval Using Color Moment Invariant. In: *Seventh International Conference on Information Technology*, pp. 200–209, IEEE, 2019. doi: 10.1109/ITNG.2010.231.

[17] Dubey S.R., Singh S.K., Singh R.K.: Boosting local binary pattern with bag--of-filters for content based image retrieval, *IEEE UP Section Conference on Electrical Computer and Electronics*, pp. 1–6, 2015. doi: 10.1109/UPCON.2015.7456703.

[18] Dubey S.R., Singh S.K., Singh R.K.: Rotation and scale invariant hybrid image descriptor and retrieval, *Computers & Electrical Engineering*, vol. 46, pp. 288–302, 2015. doi: 10.1016/j.compeleceng.2015.04.011.

[19] ElAlami M.E.: A novel image retrieval model based on the most relevant features, *Knowledge-Based Systems*, vol. 24(1), pp. 23–32, 2011. doi: 10.1016/j.knosys.2010.06.001.

[20] Enser P.G., Sandom C.J., Lewis P.H.: Surveying the reality of semantic image retrieval. In: *International Conference on Advances in Visual Information Systems*, pp. 177–188, Springer, 2005. doi: 10.1007/11590064_16.

[21] Fadaei S., Amirfattahi R., Ahmadzadeh M.R.: New content-based image retrieval system based on optimised integration of DCD, wavelet and curvelet features, *IET Image Processing*, vol. 11, pp. 89–98, 2017. doi: 10.1049/iet-ipr.2016.0542.

[22] Ghozzi Y., Baklouti N., Hagras H., Ayed M.B., Alimi A.M.: Interval Type-2 Beta Fuzzy Near Sets Approach to Content Based Image Retrieval, *IEEE Transactions on Fuzzy Systems*, vol. 30(3), pp. 805–817, 2021. doi: 10.1109/TFUZZ.2021.3049900.

[23] Graham M.E.: The cataloguing and indexing of images: time for a new paradigm?, *Art Libraries Journal*, vol. 26(1), pp. 22–27, 2001. doi: 10.1017/S0307472200012001.

[24] Hafiane A., Chaudhuri S., Seetharaman G., Zavidovique B.: Region-based CBIR in GIS with local space filling curves to spatial representation, *Pattern Recognition Letters*, vol. 27(4), pp. 259–267, 2006. doi: 10.1016/j.patrec.2005.08.007.

[25] Haralick R., Shanmugam K.: Textural Features for Image Classification, *IEEE Transactions on Systems, Man, and Cybernetics*, vol. SMC-3, pp. 610–621, 1973. doi: 10.1109/TSMC.1973.4309314.

[26] Holt B., Hartwick L.: Retrieving art images by image content: the UC Davis QBIC project, *Aslib Proceedings*, vol. 46(10), pp. 243–248, 1994. doi: 10.1108/eb051371.

[27] Huang J., Kumar S.R., Mitra M.: Combining supervised learning with color correlograms for content-based image retrieval. In: *The Fifth ACM International Conference on Multimedia*, pp. 325–334, 1997. doi: 10.1145/266180.266383.

[28] Huang J., Kumar S.R., Mitra M., Zhu W.J., Zabih R.: Image indexing using color correlograms. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 762–768, 1997. doi: 10.1109/CVPR.1997.609412.

[29] Ivanova K., Stanchev P.: Color harmonies and contrasts search in art image collections. In: *Advances in Multimedia*, pp. 180–187, IEEE, 2009. doi: 10.1109/MMEDIA.2009.41.

[30] Jabid T., Kabir M.H., Chae O.: Local Directional Pattern (LDP) – A Robust Image Descriptor for Object Recognition. In: *International Conference on Advanced Video and Signal Based Surveillance*, pp. 482–487, IEEE, 2010. doi: 10.1109/AVSS.2010.17.

[31] Jacob I.J., Srinivasagan K.G., Darney P.E., Jayapriya K.: Deep learned Inter-Channel Colored Texture Pattern: a new chromatic-texture descriptor, *Pattern Analysis and Applications*, vol. 23, pp. 239–251, 2020. doi: 10.1007/s10044-019-00780-9.

[32] Jolliffe I.: Principal Component Analysis. In: *International Encyclopedia of Statistical Science*, pp. 1094–1096, Springer Berlin Heidelberg, 2011. doi: 10.1007/978-3-642-04898-2_455.

[33] Jones B., Schaefer G., Zhu S.: Content-based image retrieval for medical infrared images. In: *26th International Conference on Engineering in Medicine and Biology Society*, pp. 1186–1187, IEEE, 2004. doi: 10.1109/IEMBS.2004.1403379.

[34] Joseph A., Rex E.S., Christopher S., Jose J.: Content-based image retrieval using hybrid k-means moth flame optimization algorithm, *Arabian Journal of Geosciences*, vol. 14(8), pp. 1–14, 2021. doi: 10.1007/s12517-021-06990-y.

[35] Ju H., Ma K.K.: Fuzzy color histogram and its use in color image retrieval, *IEEE Transactions on Image Processing*, vol. 11(8), pp. 944–952, 2002. doi: 10.1109/TIP.2002.801585.

[36] Kumar N., Berg A., Belhumeur P.N., Nayar S.: Describable visual attributes for face verification and image search, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33(10), pp. 1962–1977, 2011. doi: 10.1109/TPAMI.2011.48.

[37] Lee D.D., Seung H.S.: Learning the parts of objects by non-negative matrix factorization, *Nature Science Journal*, vol. 401, pp. 788–791, 1999.

[38] Li J., Wang J.Z.: Automatic Linguistic Indexing of Pictures by a statistical modeling approach, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25(9), pp. 1075–1088, 2003. doi: 10.1109/TPAMI.2003.1227984.

[39] List J.: How drawings could enhance retrieval in mechanical and device patent searching, *World Patent Information*, vol. 29(3), pp. 210–218, 2007. doi: 10.1016/j.wpi.2007.01.001.

[40] Liu G.H., Yang J.Y.: Content-based image retrieval using color difference histogram, *Pattern Recognition*, vol. 46(1), pp. 188–198, 2013. doi: 10.1016/j.patcog.2012.06.001.

[41] Liu G.H., Yang J.Y.: Content-based image retrieval using computational visual attention model, *Pattern Recognition*, vol. 48, pp. 2554–2566, 2015. doi: 10.1016/j.patcog.2015.02.005.

[42] Liu Y., Huang Y., Gao Z.: Feature extraction and similarity measure for crime scene investigation image retrieval, *Journal of Xian University of Posts and Telecommunications*, vol. 19, pp. 11–16, 2014.

[43] Liu Y., Huang Y., Zhang S., Zhang D., Ling N.: Integrating object ontology and region semantic template for crime scene investigation image retrieval. In: *Industrial Electronics and Applications*, pp. 149–153, IEEE, 2017. doi: 10.1109/ICIEA.2017.8282831.

[44] Liu Y., Zhang D., Lu G., Ma W.Y.: A survey of content-based image retrieval with high-level semantics, *Pattern Recognition*, vol. 40(1), pp. 262–282, 2007. doi: 10.1016/j.patcog.2006.04.045.

[45] Lopes A.P.B., de Avila S.E.F., Peixoto A.N.A., Oliveira R.S., de Araújo A.A.: A bag-of-features approach based on Hue-SIFT descriptor for nude detection. In: *17th European Signal Processing Conference*, pp. 1552–1556, IEEE, 2009.

[46] Lopes A.P.B., de Avila S.E.F., Peixoto A.N.A., Oliveira R.S., de Coelho M.M., de Araújo A.A.: Nude Detection in Video Using Bag-of-Visual-Features. In: *XXII Brazilian Symposium on Computer Graphics and Image Processing*, pp. 224–231, IEEE, 2009. doi: 10.1109/SIBGRAPI.2009.32.

[47] Manjunath B.S., Ohm J.R., Vasudevan V.V., Yamada A.: Color and texture descriptors, *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 11(6), pp. 703–715, 2001. doi: 10.1109/76.927424.

[48] Mathew S.P., Balas V.E., Zachariah K.: A content-based image retrieval system based on convex hull geometry, *Acta Polytechnica Hungarica*, vol. 12(1), pp. 103–116, 2015.

[49] Merwe J.S. van der, Ferreira H.C., Clarke W.A.: Towards Detecting Man-made Objects in Natural environments for a Man-made Object MPEG-7 CBIR Descriptor – SANDF Application. In: *16th Annual Symposium of the Pattern Recognition Association of South Africa*, vol. 1, pp. 19–24, 2005.

[50] Mohiuddin F., Hossain I., Kabir M.W.U.: A noble color-texture hybrid method for content-based image retrieval. In: *20th International Conference of Computer and Information Technology*, pp. 1–6, IEEE, 2017. doi: 10.1109/ICCITECHN.2017.8281841.

[51] Muller S., Rigoll G.: Improved stochastic modeling of shapes for content-based image retrieval. In: *Content-Based Access of Image and Video Libraries*, pp. 23–27, IEEE, 1999.

[52] Murala S., Maheshwari R.P., Balasubramanian R.: Local tetra patterns: A new feature descriptor for content-based image retrieval, *IEEE Transactions on Image Processing*, vol. 21, pp. 2874–2886, 2012. doi: 10.1109/TIP.2012.2188809.

[53] Niu D., Zhao X., Lin X., Zhang C.: A novel image retrieval method based on multi-features fusion, *Signal Processing: Image Communication*, vol. 87, p. 115911, 2020. doi: 10.1016/j.image.2020.115911.

[54] Ojala T., Pietikainen M., Harwood D.: Performance evaluation of texture measures with classification based on Kullback discrimination of distributions. In: *The 12th International Conference on Pattern Recognition, Computer Vision and Image Processing*, vol. 1, pp. 582–585, IEEE, 1994. doi: 10.1109/ICPR.1994.576366.

[55] Ojala T., Pietikainen M., Maenpaa T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24(7), pp. 971–987, 2002. doi: 10.1109/TPAMI.2002.1017623.

[56] Pardede J., Sitohang B., Akbar S., Khodra M.L.: Re-weighting Relevance Feedback in HSV Quantization for CBIR. In: *The 19th International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing*, pp. 58–63, IEEE/ACIS, 2018. doi: 10.1007/978-3-540-75690-3_13.

[57] Pass G., Zabith R.: Histogram refinement for content-based image retrieval. In: *Proceedings Third IEEE Workshop on Applications of Computer Vision. WACV'96*, pp. 96–102, IEEE, 1996. doi: 10.1109/ACV.1996.572008.

[58] Pass G., Zabith R., Miller J.: Comparing images using color coherence vectors. In: *The Fourth ACM International Conference on Multimedia*, pp. 65–73, ACM, 1997. doi: https://dl.acm.org/doi/10.1145/244130.244148.

[59] Pavithra L.K., Sharmila T.S.: A new multi-level radial difference encoded pattern for image classification and retrieval, *Multidimensional Systems and Signal Processing*, vol. 31(4), pp. 1411–1433, 2020. doi: 10.1007/s11045-020-00713-4.

[60] Pradhan J., Ajad A., Pal A.K., Banka H.: Multi-level colored directional motif histograms for content-based image retrieval, *The Visual Computer*, vol. 36, pp. 1847–1868, 2019. doi: 10.1007/s00371-019-01773-9.

[61] Pradhan J., Kumar S., Pal A.K., Banka H.: A hierarchical CBIR framework using adaptive tetrolet transform and novel histograms from color and shape features, *Digital Signal Processing*, vol. 82, pp. 258–281, 2018. doi: 10.1016/j.dsp.2018.07.016.

[62] Rao L.K., Rohini P., Reddy L.P.: Multiple Color Channel Local Extrema Patterns for Image Retrieval. In: *Innovations in Electronics and Communication Engineering. Lecture Notes in Networks and Systems*, vol. 65, pp. 115–123, 2019. doi: 10.1007/978-981-13-3765-9_13.

[63] Raza A., Nawaz T., Dawood H., Dawood H.: Square texton histogram features for image retrieval, *Multimedia Tools and Applications*, vol. 78, pp. 2719–2746, 2019. doi: 10.1007/s11042-018-5795-x.

[64] Rohini P., Bindu C.S.: Quantized Local Trio Patterns for Multimedia Image Retrieval System. In: *Lecture Notes in Networks and Systems*, pp. 107–113, Springer, 2019. doi: 10.1007/978-981-13-3765-9_12.

[65] Rui M., Cheng H.D.A.: Effective image retrieval using dominant color descriptor and fuzzy support vector machine, *Pattern Recognition*, vol. 42, pp. 147–157, 2009. doi: 10.1016/j.patcog.2008.07.001.

[66] Rui Y., Huang T.S., Chang S.F.: Image retrieval: Current techniques, promising directions, and open issues, *Journal of Visual Communication and Image Representation*, vol. 10, pp. 39–62, 1999. doi: 10.1006/jvci.1999.0413.

[67] Ruiz M.E.: Combining image features, case descriptions and UMLS concepts to improve retrieval of medical images. In: *American Medical Informatics Association Annual Symposium Proceedings*, pp. 674–678, 2006.

[68] Shriram K., Priyadarsini P., Baskar A.: An intelligent system of content-based image retrieval for crime investigation, *International Journal of Advanced Intelligence Paradigms*, vol. 7(3-4), pp. 264–279, 2015. doi: 10.1504/IJAIP.2015.073707.

[69] Shyu C.R., Kak A., Brodley C.E., Broderick L.S.: Testing for human perceptual categories in a physician-in-the-loop CBIR system for medical imagery. In: *Workshop on Content-Based Access of Image and Video Libraries*, pp. 102–108, IEEE, 1999. doi: 10.1109/IVL.1999.781132.

[70] Singh S., Batra S.: An efficient bi-layer content based image retrieval system, *Multimedia Tools and Applications*, vol. 79, pp. 17731–17759, 2020. doi: 10.1007/s11042-019-08401-7.

[71] Srivastava P., Khare A.: Integration of wavelet transform, Local Binary Patterns and moments for content-based image retrieval, *Journal of Visual Communication and Image Representation*, vol. 42, pp. 78–103, 2017. doi: 10.1016/j.jvcir.2016.11.008.

[72] Srivastava P., Khare A.: Content-based image retrieval using multiresolution speeded-up robust feature, *International Journal of Computational Vision and Robotics*, vol. 8(4), pp. 375–387, 2018. doi: 10.1504/IJCVR.2018.093967.

[73] Srivastava P., Khare A.: Utilizing multiscale local binary pattern for content-based image retrieval, *Multimedia Tools and Applications*, vol. 77, pp. 12377–12403, 2018. doi: 10.1007/s11042-017-4894-4.

[74] Subash Kumar T., Nagarajan V.: Local curve pattern for content-based image retrieval, *Pattern Analysis and Applications*, vol. 22(3), pp. 1233–1242, 2019. doi: 10.1007/s10044-018-0724-1.

[75] Sun J., Wu X.: Chain code distribution-based image retrieval. In: *Intelligent Information Hiding and Multimedia Signal Processing*, pp. 139–142, IEEE, 2006. doi: 10.1109/IIH-MSP.2006.264973.

[76] Talib A., Mahmuddin M., Husni H., George L.E.: A weighted dominant color descriptor for content-based image retrieval, *Journal of Visual Communication and Image Representation*, vol. 24(3), pp. 345–360, 2013. doi: 10.1016/j.jvcir.2013.01.007.

[77] Tan X., Triggs B.: Enhanced local texture feature sets for face recognition under difficult lighting conditions. In: *International Workshop on Analysis and Modeling of Faces and Gestures*, pp. 168–182, Springer, 2007. doi: 10.1007/978-3-540-75690-3_13.

[78] Tiwari A., Bansal V.: PATSEEK: Content Based Image Retrieval System for Patent Database. In: *International Council on English Braille*, pp. 1167–1171, 2004. https://aisel.aisnet.org/iceb2004/199.

[79] Tyagi V.: *Content-Based Image Retrieval Ideas Influences and Current Trends*, Springer Nature Singapore Pte Ltd, 2017. doi: 10.1007/978-981-10-6759-4.

[80] Tănase-Avătavului M.: *Shape decomposition and retrieval*, Ph.D. thesis, Utrecht University, 2005. http://dspace.library.uu.nl/handle/1874/1700.

[81] Umamaheswaran S., Lakshmanan R., Vinothkumar V., Arvind K., Nagarajan S.: New and robust composite micro structure descriptor (CMSD) for CBIR, *International Journal of Speech Technology*, vol. 23(2), pp. 243–249, 2019. doi: 10.1007/s10772-019-09663-0.

[82] Vipparthi S.K., Nagar S.K.: Color Directional Local Quinary Patterns for Content Based Indexing and Retrieval, *Human-centric Computing and Information Sciences*, vol. 4(1), pp. 1–13, 2014. doi: 10.1186/s13673-014-0006-x.

[83] Vrochidis S., Papadopoulos S., Moumtzidou A., Sidiropoulos P., Pianta E., Kompatsiaris I.: Towards content-based patent image retrieval: A framework perspective, *World Patent Information*, vol. 32(2), pp. 94–106, 2010. doi: 10.1016/j.wpi.2009.05.010.

[84] Wang J.Z., Li J., Wiederhold G.: SIMPLIcity: Semantics-sensitive Integrated Matching for Picture LIbraries, *IEEE Transactions on Pattern Analysis and Machine Intelligence,*, vol. 23(9), pp. 947–963, 2001. doi: 10.1109/34.955109.

[85] Wong K.M., Po L.M., Cheung K.W.: Dominant color structure descriptor for image retrieval. In: *International Conference on Image Processing*, p. 365–368, 2006. doi: 10.1109/ICIP.2007.4379597.

[86] Xia Y., Wan S., Jin P., Yue L.: Multi-Scale Local Spatial Binary Patterns for Content-Based Image Retrieval. In: *International Conference on Active Media Technology*, pp. 423–432, Springer, 2013. doi: 10.1007/978-3-319-02750-0_45.

[87] Xie G., Guo B., Huang Z., Zheng Y., Yan Y.: Combination of Dominant Color Descriptor and Hu Moments in Consistent Zone for Content Based Image Retrieval, *IEEE Access*, vol. 8, pp. 146284–146299, 2020. doi: 10.1109/ACCESS.2020.3015285.

[88] Xu D., Yan S., Tao D., Lin S., Zhang H.J.: Marginal Fisher Analysis and Its Variants for Human Gait Recognition and Content-Based Image Retrieval, *IEEE Transactions on Image Processing*, vol. 16(11), pp. 2811–2821, 2007. doi: 10.1109/TIP.2007.906769.

[89] Yu J., Qin Z., Wan T., Zhang X.: Feature integration analysis of bag-of-features model for image retrieval, *Neurocomputing*, vol. 120, pp. 355–364, 2013. doi: 10.1016/j.neucom.2012.08.061.

[90] Yuan B.H., Liu G.H.: Image retrieval based on gradient-structures histogram, *Neural Computing and Applications*, vol. 32, pp. 11717–11727, 2020. doi: 10.1007/s00521-019-04657-0.

[91] Zhang D., Lu G.: A comparative study on shape retrieval using Fourier descriptors with different shape signatures. In: *International Conference on Intelligent Multimedia and Distance Education*, pp. 1–9, 2001.

[92] Zhang L., Hu Y., Li M., Ma W., Zhang H.: Efficient propagation for face annotation in family albums. In: *The 12th Annual International Conference on Multimedia*, pp. 716–723, ACM, 2004. doi: 10.1145/1027527.1027689.

[93] Zhou J., Liu X., Liu W., Gan J.: Image retrieval based on effective feature extraction and diffusion process, *Multimedia Tools and Applications*, vol. 78, pp. 6163–6190, 2019. doi: 10.1007/s11042-018-6192-1.

[94] Zhou J., Liu X., Xu T., Gan J., Liu W.: A new fusion approach for content based image retrieval with color histogram and local directional pattern, *International Journal of Machine Learning and Cybernetics*, vol. 9(4), pp. 677–689, 2016. doi: 10.1007/s13042-016-0597-9.

[95] Zhou W., Li H., Tian Q.: Recent Advance in Content-based Image Retrieval: A Literature Survey, *arXiv preprint arXiv:170606064*, 2017.

[96] Zhou X.S., Zillner S., Moeller M., Sintek M., Zhan Y., Krishnan A., Gupta A.: Semantics and CBIR: a medical imaging perspective. In: *The International Conference on Content-based Image and Video Retrieval*, pp. 571–580, ACM, 2008. doi: 10.1145/1386352.1386436.

[97] Zhu L., Jin H., Zheng R., Zhang Q., Xie X., Guo M.: Content-Based Design Patent Image Retrieval Using Structured Features and Multiple Feature Fusion. In: *6th International Conference on Image and Graphics*, pp. 969–974, IEEE, 2011. doi: 10.1109/ICIG.2011.121.

Affiliations

Sagar Chavda
Gujarat Technological University, Ahmedabad, India
Government Engineering College, Modasa, India
chavdasagar.m@gmail.com, ORCID ID: https://orcid.org/0000-0002-7478-6814

Mahesh Goyani
Gujarat Technological University, Ahmedabad, India
Government Engineering College, Modasa, India
mgoyani@gmail.com, ORCID ID: https://orcid.org/0000-0002-7269-1176

Received: 22.05.2020
Revised: 15.06.2021
Accepted: 13.09.2021