Expert Systems With Applications 232 (2023) 120774

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

A novel content-based image retrieval system with feature descriptor integration and accuracy noise reduction

Gabriel S. Vieira [a,b,*], Afonso U. Fonseca [b], Naiane M. Sousa [b], Juliana P. Felix [b], Fabrizzio Soares [b]

[a] Federal Institute Goiano/IFGoiano, Computer Vision Laboratory, Urutaí GO, Brazil
[b] Federal University of Goiás/UFG, Institute of Informatics, Goiânia GO, Brazil

ARTICLE INFO

Keywords:
Image retrieval
Image descriptor
Microstructures
Features fusion
Local binary pattern
Low-level features combination

ABSTRACT

Efficient algorithms, intelligent approaches, and intensive use of computers are crucial to extracting reliable information from large-scale data sets. It is the case, for example, in retrieving information from multimedia data. As different systems integrate individual electronic devices such as smartphones and cameras into storage, sharing, and social media platforms, the amount of multimedia data has increased dramatically in recent years. Therefore, content-based image retrieval (CBIR) is quite challenging. In this paper, we introduce a new image descriptors integration to represent the visual attributes of images. Furthermore, we noticed a pattern between different CBIR systems in which the first images retrieved are more likely to be assertive than those in the middle and final positions. Then, we investigated the interactions between the first images retrieved from CBIR systems to present a novel method for reducing accuracy noise. The performance of the proposed method is evaluated in different data sets (Corel-1K, Corel-5K, Corel-10K, GHIM-10K) and compared to related works. The experimental results demonstrate that the proposal consistently retrieves images with similar content. The code prepared by the authors is publicly available.

1. Introduction

Information can change the course of history, leverage business, create opportunities and open new perspectives to be explored. However, good information only exists if data is available to be used, investigated, and analyzed. Information can be limited and not very assertive when data are insufficient. Likewise, if there is much data, the excess can exceed human capacity to exploit it directly. Hence, requiring efficient algorithms, intelligent approaches, and intensive use of computers to extract reliable information.

At the current stage of technology, the existence of data is no longer a severe limitation. With digitization, sensing equipment, and the increase in storage capacity, persisting a diversity of data in large volumes is feasible. Hence, the real challenge is retrieving information from stored data, especially in multimedia (i.e., video and image data) (Ahmed, Afzal, Mufti, Mehmood, & Choi, 2020; Nazir & Nazir, 2018; Shikha, Gitanjali, & Kumar, 2020).

As different systems integrate individual electronic devices such as smartphones and cameras into storage, sharing, and social media platforms, the amount of multimedia data has increased dramatically in recent years. Although widely used, the conventional search through text-based image retrieval (TBIR) (Bibi et al., 2020; Wang, Wang, Liu, Ren, & Yuan, 2018) is insufficient to describe the diversity of content within images due to the lack of discriminative capability of written texts (Mehmood et al., 2018). Besides, the volume of data grows exponentially, making manual content annotations unfeasible.

Over the last few years, content-based image retrieval (CBIR) has been introduced to overcome the limitations of text-based content search. Henceforward, the human labor of filling in description tags could be replaced by image feature descriptors (Ahmed, Ummesafi, & Iqbal, 2019; Chu & Liu, 2020). Feature descriptors extract low-level features such as color, texture, shape, or spatial layout from images and encapsulate them into virtual image representations or vectors. Feature vectors are image signatures representing the images' visual attributes (Irtaza et al., 2018). Because of that, they are the basis of CBIR systems and are directly responsible for the quality of information retrieval (Mehmood et al., 2018; Niu, Zhao, Lin, & Zhang, 2020; Pradhan, Pal, & Banka, 2022)

A single descriptor can be used to describe images or combined with other image descriptors in fusion-based methods. When low-level image attributes are combined, distinct aspects of the images are represented through integrated feature vectors, increasing the representation of the image content (Vieira, Fonseca, & Soares, 2023). In this sense, pixel

* Corresponding author at: Federal Institute Goiano/IFGoiano, Computer Vision Laboratory, Urutaí GO, Brazil.
E-mail addresses: gabriel.vieira@ifgoiano.edu.br (G.S. Vieira), afonso@inf.ufg.br (A.U. Fonseca), naiane@inf.ufg.br (N.M. Sousa), jufelix16@gmail.com (J.P. Felix), fabrizzio@inf.ufg.br (F. Soares).

https://doi.org/10.1016/j.eswa.2023.120774
Received 6 September 2022; Received in revised form 22 March 2023; Accepted 7 June 2023
Available online 14 June 2023

intensity gives information about the color distribution of images, the color transition about the texture of images, and the object outline about the shape of image contents. Different combinations have been investigated and reported in comparative analyses, and those with multiple descriptors have achieved more consistent responses. The primary justification is that single feature extraction cannot describe complex images with only one observation point (Pradhan et al., 2022; Wei & Liu, 2020). Thus, as CBIR systems return one or more images associated with a query image, the number of wrong answers is decreased by considering integrated descriptors.

Although the descriptor combination strategies present better results than individual descriptors, the dimensionality of the feature vector can increase, demanding more processing capacity, storage, and execution time. Furthermore, information retrieval approaches can be compromised due to the high processing cost of similarity evaluation operations, hardware limitations, or excessively time-consuming outputs. As multimedia repositories are large databases, considerations regarding processing time are essential in designing CBIR systems. Thus, feature vector dimensionality and similarity evaluation mechanisms play a vital role in the modeling and constructing of information retrieval systems (Liu, Li, Zhang, & Xu, 2011).

This paper presents an effective and innovative CBIR to address these issues. In the proposal, color intensity values are organized into histograms (cl-MSD descriptor), image contrast is encoded through local variance directional responses (LDiPv descriptor), and local binary patterns are used to represent the spatial arrangement of color in image regions (LBP descriptor). Also, a pairwise distance between two sets of observations known as L_1 was integrated into the model, proving time efficient and assertive in retrieved results.

In this sense, we present a new fusion-based method and a similarity measure suitable for integrating the selected descriptors. We show that the proposed combination of descriptors increases assertiveness, enhancing information retrieval by presenting more assertive responses concerning individual descriptors. Furthermore, we present a new strategy that changes the positioning of retrieved images corroborating to reduce the accuracy error. We argue that the accuracy error in CBIR systems grows as the number of retrieved images increases. As expected, the retrieved images occupying the top positions are close to the query images in terms of content. When these images also become query images, other images can be retrieved where the first positions can be used to increase the number of potential candidates to assume the first positions of the initial query image. We encoded this note and compared the results with and without using our accuracy noise reduction approach.

In the analysis of the proposed method, we verified that integrating the cl-MSD, LBP, and LDiPv descriptors increases the percentage of correct answers by at least 11% when compared to the individual descriptors. Furthermore, we verified that our accuracy noise reduction strategy adds a 4 to 6% assertiveness boost to image retrieval. Thus, our attention is focused on the extraction and representation of image features and the search for assertive answers in terms of similarity between a query image and other related ones. Besides that, our approach aligns with conventional CBIR strategies in which correspondence assessment is performed individually between data set samples rather than learning models and object classification. In this way, it does not involve data annotation and training steps.

The major contributions of this work are as follows:

- a new fusion-based method that combines the cl-MSD, LBP, and LDiPv descriptors.
- a novel accuracy noise reduction approach based on clustering the initial responses of the matching image process.
- an efficient image retrieval method to deal with erroneous responses that occur when different image primitives generate similarities between feature vectors.
- a comparative analysis between the proposed method and state-of-the-art image retrieval methods using different data sets (Corel-1K, Corel-5K, Corel-10K, and GHIM-10K).

The remaining organization of this paper is given in the following way: related work is presented in Section 2, and the description of the proposed work has been given in Section 3 where the four main steps of the proposal, namely, image feature extraction, image similarity assessment, image retrieval and accuracy noise reduction, and replacement of duplicate entries are discussed. Experimental simulation and analysis have been given in Section 4. Finally, in Section 5, the conclusion with future trends has been given.

2. Related work

This section presents research related to developing content-based image retrieval systems. The section is divided into categories to emphasize design aspects of CBIR systems, such as feature vector dimensionality, similarity and performance evaluation, and image data set benchmarks. Furthermore, this section categorizes image retrieval solutions into local and global approaches, presents machine learning solutions and discusses the limitations of learning models, discusses accuracy noise in image retrieval, and presents the main contributions of the proposal.

2.1. Feature vector dimensionality

One of the first concerns about the CBIR system's design is the feature vectors' dimensionality. Higher dimensional vectors can bring details that better represent the visual content of the images while requiring more storage space and longer search time. On the other hand, lower dimensionality reduces the use of computational resources and operating time at the cost of reducing the power of discrimination. In this case, it is essential to prepare an adequate balance between retrieval accuracy, storage space, and retrieval speed (Liu et al., 2011). Chu and Liu (2020) proposed a CBIR method based on a multi-integration feature model encompassing image color and edge orientation information in a feature vector of size 102. Hua, Liu, and Song (2019) designed a color volume histogram to describe colors, textures, shapes, and spatial features in images, represented by a feature vector with dimensionality equaling 104. Dawood et al. (2019) established correlations between color, texture orientation, and intensity features to identify microstructures in images encoded in a feature vector of size 88. Raza et al. (2018) developed a feature descriptor with size 242 based on the correlation among color, orientation, and intensity information. Later, Raza, Nawaz, Dawood, and Dawood (2019) introduced a new method with a 172-dimensional feature vector. Likewise, Ahmed (2012), Dhall, Asthana, Goecke, and Gedeon (2011), Mohammad and Ali (2011) and Bashar, Khan, Ahmed, and Kabir (2014) proposed feature vectors of size 256. Others proposed descriptors with larger dimensions, such as Lei, Ahonen, Pietikäinen, and Li (2011) and Tan and Triggs (2010), that presented vectors of size 512.

2.2. Local and global approaches

There are several proposals for CBIR systems in the literature, each one with its particularities. Some capture low-level features through key points or salient regions, while others consider the entire image when building descriptors. These categories, named local and global approaches, capture relevant information for evaluating the similarity between a query image and data stored in a database. While global features represent the entire image (i.e., spatial information, texture, color, shape), local features are defined as the key points or parts of images, such as corners, blobs, and edges (Hameed, Abdulhussain, & Mahmmod, 2021). The method proposed by Pradhan et al. (2022) is

an example of a local CBIR. The authors divided the data set images into object and non-object regions using image saliency to group the image pixels based on the different intensity values. Then, the local discriminatory visual concentration was used as objects of interest (foreground), while the other regions were used as non-object regions (background). In contrast, Niu et al. (2020) considered the entire image and applied convolution operations to address the relationship between shape and texture features and between color features and image texture. In addition, they proposed a scheme to update the results from the images initially retrieved.

2.3. Similarity evaluation

A typical step in CBIR systems is the similarity assessment. In this step, the feature vectors are compared to list the images from the data set closely related to the query images. Different similarity measures have been investigated. In Nazir and Nazir (2018), the similarity among the images was computed through Manhattan distance. Sharif et al. (2019) presented an image-matching score based on histogram intersection. In Raja, Kumar, and Mahmood (2020), the similarity distance was estimated between the query image and stored image with similarity metrics like Manhattan, Euclidean, Chebyshev, Hamming, and Jaccard distances. Likewise, Liu et al. (2011) presented different similarity metrics to evaluate their proposal showing that L_1 (or city block) distance is very suitable for large-scale image data sets. Other researchers proposed their metrics, such as Wang et al. (2018), which presented the dominant granule structure similarity, and Zeng, Huang, Wang, and Kang (2016), which prepared a metric based on color-spatial histograms similarity.

2.4. Performance evaluation

In performance evaluation, the most used metrics are recall and precision. The first one determines the performance of the methods considering the total number of images in a specific database group. The second one measures the capability of the methods to retrieve relevant information given a predefined number of expected responses or retrieved images (NR). The NR value varies among related works, Ali et al. (2016), Irtaza et al. (2018) and Ahmed et al. (2019) considered the retrieval of twenty images ($NR = 20$). Chen, Ding, Li, Wang, Wang, and Deng (2014) worked with the return of ten, twenty, and forty images ($NR = \{10, 20, 40\}$). Sathiamoorthy and Natarajan (2020) made comparisons between different methods using the retrieval of ten images ($NR = 10$) and Shakarami and Tarrah (2020) worked with five and ten retrievals ($NR = \{5, 10\}$). Furthermore, some works considered each image within the data set as a query image to perform the experimental evaluations (Liu & Wei, 2020; Verma & Raman, 2018; Wang et al., 2018). Meanwhile, other studies randomly selected samples from the databases used, such as Liu, Yang, and Li (2015), Pavithra and Sharmila (2018). In this sense, the first approach allows a fairer comparative analysis as the images used for the search are known in advance. The second one allows only an approximation since the selected images may vary between works.

2.5. Image data set benchmarks

Although no benchmarking standard exists for models and test data sets for CBIR systems (Chu & Liu, 2020; Liu et al., 2011), some data sets have been used more frequently. Some are used in the more general evaluation and comparative analysis while others are used to address more specific questions, such as responses to different image textures and object detection (Pradhan et al., 2022; Verma & Raman, 2018). Among the various data sets available, the Corel series is one of the most used databases, which contains many images with different semantic contents and groups (Alzu'bi, Amira, & Ramzan, 2015). For this data set, there are some versions like Corel-1k, or Corel-1000 (Wang, Li, & Wiederhold, 2001), with 1,000 images; and Corel-10k and its shortened versions Corel-5k (Latif et al., 2019), Corel-2k (Ali et al., 2016) and Corel-1.5k (Baig et al., 2020; Hameed et al., 2021). However, the Corel-10k and the others that originate from it require special attention because three data sets are available for the Corel-10k version, whose sample images may vary. Li and Wang (2003) prepared one of these versions, another one was organized by Tao, Li, and Maybank (2007), and the third one by Liu et al. (2011), which is the most used version. In addition to the Corel family data sets, the GHIM-10k (Liu et al., 2015), COIL-100 (Nene, Nayar, Murase, et al., 1996), Holidays (Jegou, Douze, & Schmid, 2008), Oxford (Philbin, Chum, Isard, Sivic, & Zisserman, 2007) data sets, among others, are also well known.

2.6. Machine learning models and limitations

In addition to conventional image retrieval methods, many learning-based methods have recently been proposed. Ali et al. (2016) proposed the construction of histograms by integrating SIFT and SURF descriptors followed by classification using a support vector machine (SVM). Sharif et al. (2019) developed a fusion of the SIFT and BRISK feature descriptors followed by model training using SVM classification. Irtaza et al. (2018) introduced a genetic classifier learning method that combines the SVM and ANN learning bases for image retrieval. Ahmed et al. (2021) presented a method that combines texture, shape, and color features with the involvement of VGG-19 and GoogLeNet architectures to address the image retrieval task. Besides, Kanwal, Tehseen Ahmad, Khan, Alhusaini, and Jing (2021) designed a method that applies local features in combination with GoogLeNet, VGG-19, and ResNet-50.

Although learning-based methods achieve impressive results, they are limited by the existence of annotated databases, which is challenging in large web repositories. Besides, like other strategies, they suffer from the lack of semantic information (Shikha et al., 2020), i.e., while they deal with low-level features, they fail to describe high-level semantic concepts (Liu & Yang, 2013). Deep learning models have tried to reduce this semantic gap with a hierarchy of layers closer to human attention, showing superior performance to conventional image retrieval methods and shallow machine learning approaches (Ahmed et al., 2021; Kanwal et al., 2021; Maji & Bose, 2021; Tarawneh, Celik, Hassanat, & Chetverikov, 2020). However, there are some drawbacks, such as they require annotated data for supervised training and demand large amounts of data (Niu et al., 2020; Tian, Jiao, Liu, & Zhang, 2014). Also, they are not efficient for extrapolating or predicting unexpected scenarios and unknown data sets, and the training process of these networks is time-consuming (Chu & Liu, 2020; Wei & Liu, 2020).

2.7. Accuracy noise in image retrieval

The intuition about our method is based on the fact that CBIR systems present successful results when they only deliver the first images retrieved. As more recovered images are expected, the error progressively increases, as noted in different works. For example, in Niu et al. (2020), it is observed that the amount of images expected to be retrieved is decisive in obtaining better assertiveness. In their method, each new unit of the recovered image implies a loss in precision in the order of 2%. Therefore, the error progressively increased as a function of the number of images retrieved.

In other related works, this same pattern can be noted. In Pradhan et al. (2022), the assertiveness for five retrieved images is around ten percentage points above the retrieval of twenty images. In Shikha et al. (2020), the increase in the number of retrieved images impacted the accuracy value in different data sets, dropping almost 30% from ten retrieved images to fifty. In the same way, Verma and Raman (2018) showed that the number of retrieved images influences the degree of assertiveness, i.e., as more outputs are expected, fewer hits are achieved.

In this sense, the best results are obtained with the first images retrieved. When similarity assessment is applied and results are ordered, the images in the first positions are more likely to be assertive than those in the intermediate and final positions. This proportionality relationship has not been investigated and is a target of our study.

2.8. Main contributions

Considering the previous works, we present a new image retrieval approach based on integrating descriptors and adjusting outputs obtained by CBIR systems. The proposal integrates image descriptors to encode global image features like color and texture into a single solution. Thus, we introduce the Color Micro-Structure Descriptor(cl-MSD) (Liu et al., 2011), which is combined with the Local Binary Pattern (LBP) and Local Directional Pattern Variance (LDiPv) descriptors (Kabir, Jabid, & Chae, 2010; Ojala, Pietikainen, & Maenpaa, 2002). The resulting vector of the proposed combination encapsulates the image information in a descriptor of only 187 features. As the dimensionality is smaller than many individual image descriptors, our approach becomes competitive in computational processing and the space required to store the feature vectors. In addition, integrating the descriptors under study presents superior accuracy than those obtained with individual descriptors and related works.

We also present a novel strategy to increase assertiveness in image retrieval systems named accuracy noise reduction (ANR). Our starting point is given by the fact that CBIR systems present outputs with more hits in the first top images. When the number of required images increases, a drop in precision is observed, and the CBIR systems progressively present incorrect responses. In this sense, we argue that retrieved images occupying the first positions can be used to direct and suggest images of greater verisimilitude to them through their interactions. We encode the relationship between the top images so that each is used to retrieve other top images. This procedure is recursively executed, producing image subsets or image clusters. Then, the images that best meet the similarity criteria are grouped and returned by the accuracy noise reduction strategy. While our descriptor combination presents an assertiveness gain between 11 and 32% compared to individual descriptors, our accuracy noise reduction strategy improves the results in image retrieval regardless of the image descriptor used. For example, in the integration of the image descriptors cl-MSD, LBP, and LDiPv, an increase of 4 to 6% can be observed.

3. Proposed method

This section presents the four steps of the proposed method in which the image features are extracted, the similarities between the images are computed, the retrieved images are updated, and duplicate images are replaced. Initially, a query image is converted to the HSV color space to extract micro-structures from pixel intensities (cl-MSD descriptor). Furthermore, the features of the input image in its original RGB format are extracted from local intensity variations (LDiPv descriptor) and local binary patterns (LBP descriptor). Then, the results are grouped into a single feature vector which represents the primitive elements of the image. This process is also applied to all images in the data set.

In the second step, the feature vector that represents the query image is compared with the other images of a data set using the L_1 distance to evaluate the similarity between the query image and the data set. Results are ordered regarding how close the data set images are to the query image. Then, an initial cluster with the expected responses is obtained.

The initial cluster is updated in the third step, considering the images in the first positions. The retrieved images are used as queries, and the search process is re-instantiated to obtain images related to them and form secondary clusters. The first and second images of the initial cluster are found in the secondary clusters. Their positions are used to rank the secondary clusters in order of proximity to the initial cluster. Then, the first two positions of each secondary cluster are concatenated to form a cluster with the top images. In the fourth step, duplicate entries or repeated images are replaced, and the final result is obtained. Based on this strategy, the accuracy noise in the initial cluster is reduced so that there is a gain in assertiveness. Fig. 1 presents the workflow of the proposed method.

3.1. Image feature extraction

This study uses local-feature descriptors to represent images in feature vector form. The first is based on applying color and edge orientation micro-structures as proposed by Liu et al. (2011). Although we follow the design of the micro-structure descriptor (MSD), we prepared a version of this descriptor that uses only color information. Therefore, we named this descriptor cl-MSD. The second descriptor, local binary pattern (LBP), encodes the relationship between the central pixel and its neighbors to summarize local structures through neighborhood relations (Ojala et al., 2002). Finally, the third descriptor, local directional pattern variance (LDiPv), encodes the contrast information to characterize both spatial structure and density difference between the adjacent areas of each micro-pattern (Kabir et al., 2010).

3.1.1. Color micro-structure descriptor (cl-MSD)

As in the original MSD descriptor (Liu et al., 2011), in the color micro-structure descriptor (cl-MSD), the input image in the RGB color space is converted to HSV. Hence, the H, S, and V channels are uniformly quantized following Eq(s). (1)–(3).

$$\forall i \in \{1,2,\dots,b^k\}, \quad \mathbf{Q}^k_{rc} = \begin{cases} i, & \text{if } c^k_{rc} > l_i \wedge c^k_{rc} <= u_i \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

$$\forall i \in \{1,2,\dots,b^k\}, \quad U_i = \sum_{i=1}^{b^k} u_{i-1} + max(\mathbf{C}^k)/b^k \tag{2}$$

$$\forall i \in \{1,2,\dots,b^k\}, \quad L_i = u_i - max(\mathbf{C}^k)/b^k \tag{3}$$

where $c_{rc} \in \mathbf{C} = \{H, S, V\}$ is a pixel value in the position r (row) and c (column) of the HSV image. $b \in B$ is a bin value used to quantize an image channel, k is an index number for image channels and bin values, i.e., $k = \{1,2,3\}$. max(.) is a function that returns the maximum value in an image channel. $max(\mathbf{C}^k)/b^k$ is a constant used to compute the lower ($l \in L$) and upper ($u \in U$) limits of quantization. $\mathbf{Q}$ is the resulting image after quantization.

According to the previous equations, the HSV image is quantized, and the three color components ($\mathbf{Q}^1$, $\mathbf{Q}^2$, $\mathbf{Q}^3$) can be combined to form a typical two-dimensional image which is obtained following Eq. (4).

$$\forall r \in \{1,\dots,m\}, \forall c \in \{1,\dots,n\}, \quad \mathbf{I}_{rc} = b^2 b^3 q^1_{rc} + b^3 q^2_{rc} + q^3_{rc} \tag{4}$$

where $q \in \mathbf{Q}$ is a quantized pixel in the position r and c. m and n are the numbers of rows and columns of images $\mathbf{Q}$. $\mathbf{I}$ is the resulting image after combining the three quantized channels.

Then, each position of the $\mathbf{I}$ image becomes a central pixel used to check if its neighbors have the same value. If they have the same value, they are grouped by a sum operation whose result represents the color micro-structures of the image. For a central pixel T_x and its eight neighboring pixels $T_n = (n = 1,2,\dots,8)$, the color micro-structures can be computed as presented in Eq(s). (5) and (6).

$$M_{(T_x)} = M_{(T_x)} + \sum_{n=0}^{p-1} F_1(T_n - T_x) \tag{5}$$

$$F_1(I) = \begin{cases} 1, & \text{if } I = 0 \\ 0, & \text{otherwise} \end{cases} \tag{6}$$

where p is the number of neighboring pixels and M is the resulting vector that encodes the micro-structures. T_x is the central pixel of a

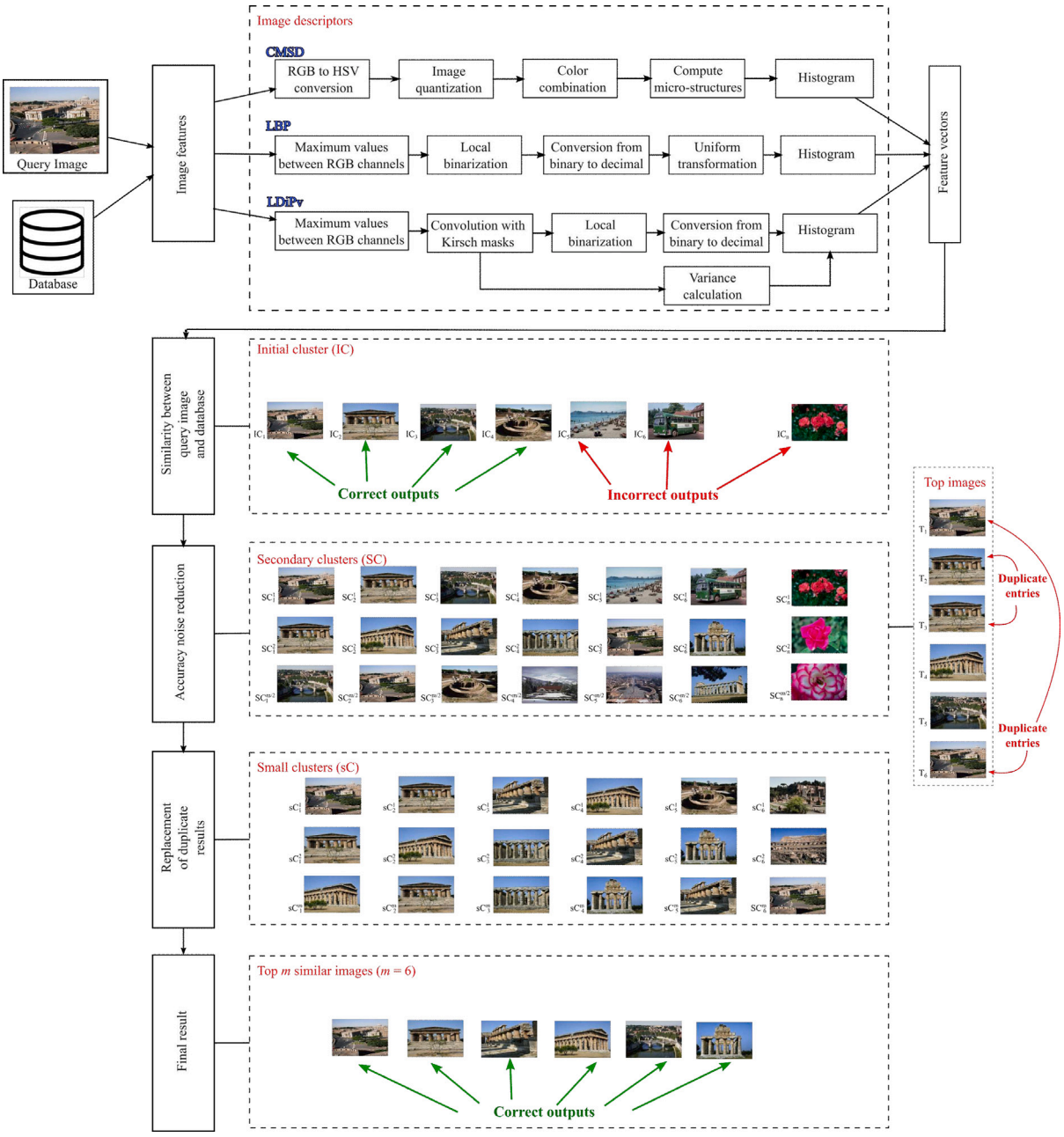


Fig. 1. Workflow of the proposed method.

sliding window, and since it is in the range $T_x \in [1, (b^1 \times b^2 \times b^3)]$, it is also an index used to update M.

Besides, the number of repetitions in $\mathbf{I}$ is counted to measure the frequency of the quantized values according to Eq(s). (7) and (8).

$$H_{(\tau)} = \sum_{r=1}^{m} \sum_{c=1}^{n} F_2(\mathbf{I}_{rc}, \tau) \tag{7}$$

$$F_2(a, b) = \begin{cases} 1, & \text{if } a = b \\ 0, & \text{otherwise} \end{cases} \tag{8}$$

where $\tau \in \{1, (b^1 \times b^2 \times b^3)\}$ is a value in the quantization range.

In this sense, while M encodes local patterns in the image, H encapsulates global patterns. We use a neighborhood of eight pixels to group these patterns to calculate micro-structures following Eq. (9).

$$\forall i \in \tau, \quad \text{cl-MSD}_i = \frac{M_i}{8H_i} \tag{9}$$

Besides that, the quantization bins are 8, 3, and 3 for each HSV channel ($B = \{8, 3, 3\}$). Then, with this parameterization, we obtain a 72-dimensional vector corresponding to the multiplication of the bin values ($b^1 \times b^2 \times b^3 = 8 \times 3 \times 3 = 72$). The quantization process was exhaustively investigated in Liu et al. (2011), Raza et al. (2018), and Niu et al. (2020). Thus, the values we used to agree with these

works as they are a good trade-off between feature vector size and execution time.

3.1.2. Local binary pattern (LBP)

The local binary pattern descriptor (LBP) encodes the relationship between the central pixel and its neighbors to present pattern features in the local patches (Ojala et al., 2002). To compute the LBP, we consider the maximum values among the three RGB channels to generate an output image with only one channel, as described in Eq. (10).

$$\forall r \in \{1,\dots,m\}, \forall c \in \{1,\dots,n\}, \quad \mathbf{T}_{rc} = max(x_{rc}, y_{rc}, z_{rc}) \tag{10}$$

where $x \in \mathbf{R}$, $y \in \mathbf{G}$, and $z \in \mathbf{B}$ are pixels in the coordinates r and c of the three channels of a RGB image, respectively. $\mathbf{T}$ is the resulting image, and $max(.)$ is a function that returns the maximum value among input arguments.

Then, considering a sliding window, the difference between the central pixel and its neighbors is computed. A unit-step function is used to obtain a binarized result for each patch. If the neighboring pixel has a value greater than or equal to the central pixel, then the value 1 signs the relation; otherwise, the value 0 is applied. Subsequently, the binary numbers are multiplied by some weights, which are added to represent the local binary pattern of a central pixel. Eq(s). (11) and (12) systematize these steps.

$$\mathbf{I}_{rc} = \sum_{n=0}^{p-1} F_3(T_n - T_x) \times 2^n \tag{11}$$

$$F_3(I) = \begin{cases} 1, & \text{if } I >= 0 \\ 0, & \text{otherwise} \end{cases} \tag{12}$$

where $\mathbf{I}_{rc}$ is the resulting LBP image, and r and c are row and column indexes, i.e., the central pixel coordinates. p is the number of neighboring pixels of the central pixel T_x while T_n represents a neighboring pixel. $F_3(.)$ is the unit-step function used in the LBP descriptor. Following the given formulation, the values assigned to $\mathbf{I}$ range from 0 to 255.

According to Ojala et al. (2002), there are a limited number of transitions or discontinuities in the circular presentation of the LBP pattern. Thus, the image $\mathbf{I}$ with a maximum pixel value of 256 can be converted to an image with a maximum pixel value of 59 by considering the uniform appearance of the local binary pattern. When this transformation is applied, we get the image $\mathbf{I}'$, and the final LBP histogram is obtained following Eq. (13).

$$LBP_{(\tau)} = \sum_{r=1}^{m} \sum_{c=1}^{n} F_2(\mathbf{I}'_{rc}, \tau) \tag{13}$$

where m and n are the number of rows and columns of the transformed LBP image $\mathbf{I}'$. τ is a pattern number which is in the range $\tau \in [1, 59]$ and function F_2 is presented in Eq. (8). In this way, we encapsulate the LBP features in a 59-dimensional vector.

3.1.3. Local directional pattern variance (LDiPv)

LDiPV was also designed to work with 1-channel images. Thus, we convert the original RGB images considering the maximum values between their channels (Eq. (10)). As in LBP, the LDiPv descriptor considers neighborhood regions for each central pixel. In contrast, the input image is first convolved with eight Kirsch masks resulting in eight different images as presented in Eq. (14).

$$\mathbf{I}^d = \mathbf{C} * M^d \tag{14}$$

where M is one of the eight Kirsch masks ($d \in [1,\dots,8]$), $\mathbf{C}$ is the input image, and $\mathbf{I}$ are the resulting images.

Then, the absolute value of each coordinate of $\mathbf{I}$ is computed as in Eq. (15) where r and c are the row and column position of $i \in \mathbf{I}$, m and n correspond to the image dimension ($m \times n$), and d is an index to the eight convoluted images.

$$\forall r \in \{1,\dots,m\}, \forall c \in \{1,\dots,n\}, \quad \hat{\mathbf{I}}^d_{rc} = |i^d_{rc}| \tag{15}$$

In the next step, the values of the eight images $\hat{\mathbf{I}}$ in the same position (r,c) are used to prepare binary codewords. In Eq. (16), eight binary images are obtained considering the k-th most significant directional response. When $k = 3$, the three maximum values in $[\hat{i}^1_{rc}, \dots, \hat{i}^8_{rc}]$ are set to 1, and the remaining bits are set to 0. In Eq. (17), M holds those k maximum values, and ψ is set with the minimum value among M. $F_3(.)$ is defined in Eq. (12).

$$\mathbf{B}^d_{rc} = \sum_{d=1}^{8} F_3(\hat{i}^d_{rc} - \psi) \tag{16}$$

$$\psi = min([M^d_1, M^d_2, M^d_3]) \tag{17}$$

A binary codeword corresponds to the eight values of $\mathbf{B}$ in the coordinates r and c, i.e., $w = \{b^8_{rc}, \dots, b^1_{rc}\}$ where $b \in \mathbf{B}$. Note that the superscript index is in reverse order from 8 to 1. Then, the binary codeword w is converted to a decimal number. Since the number of 1 bit in w is 3 (due to $k = 3$), there are only 56 possible decimal values for the w conversion. In Eq. (18), binary to decimal number conversions are performed by the f function, and the results are set to $\mathbf{D}$.

$$\forall r \in \{1,\dots,m\}, \forall c \in \{1,\dots,n\} \quad \mathbf{D}_{rc} = f(w_{rc}) \tag{18}$$

Besides that, the variance among the eight edge responses in $\hat{\mathbf{I}}$ is calculated according to Eq. (19) where $\bar{i}$ is the average value of the eight responses of $\hat{\mathbf{I}}$ in position (r, c).

$$\mathbf{V}_{rc} = \frac{1}{8} \sum_{d=1}^{8} (\hat{i}^d_{rc} - \bar{i}_{rc}) \tag{19}$$

After the previous steps, the histogram of $\mathbf{D}$ is computed by replacing the frequency count value with the variance in $\mathbf{V}$. With the application of the Eq(s). (20) and (21), the LDiPv features are encapsulated in a 56-dimensional vector.

$$LDiPv_{(\tau)} = \sum_{r=1}^{m} \sum_{c=1}^{n} F_4(\mathbf{D}_{rc}, \tau) \tag{20}$$

$$F_4(a, b) = \begin{cases} \mathbf{V}_{rc}, & \text{if } a = b \\ 0, & \text{otherwise} \end{cases} \tag{21}$$

where τ is a pattern number which is in the range $\tau \in [1, 56]$.

3.2. Image similarity assessment

In image retrieval, the images are ordered according to the similarity between a query image and database images. In our similarity evaluation, we apply the L_1 distance to measure the difference between the visual attributes of the images. The L_1 distance does not need square and square root operations, so it can reduce the computational cost when applied to large-scale data sets. In this sense, considering the two M-dimensional feature vectors, which are designated by Q = {cl-MSD, LBP, LDiPv} and T = {cl-MSD, LBP, LDiPv}, the similarity between them is calculated as follows in Eq. (22):

$$D_1(Q,T) = \sum_{i=1}^{M_1} |q^1_i - t^1_i|, \quad D_2(Q,T) = \sum_{i=1}^{M_2} |q^2_i - t^2_i|,$$
$$D_3(Q,T) = \sum_{i=1}^{M_3} |q^3_i - t^3_i| \tag{22}$$

where $q \in Q$ is a feature vector instance that represents a query image, $t \in T$ is a data set image instance, and M corresponds to the size of each of the three descriptors, i.e., $M = \{72, 59, 56\}$. The similarity evaluation between the cl-MSD, LBP, and LDiPv vectors is performed separately, whose results point to the distances D_1, D_2, and D_3.

When comparing the query image with all the images in the data set, we obtain three distance measurements for each one of them, resulting in $D^{(z)}_1$, $D^{(z)}_2$, and $D^{(z)}_3$, where z is in the range of 1 to the number

of images in the data set. Therefore, a normalization step is applied considering the similar results of the query image with the entire data set according to Eq. (23).

$$\hat{D}_1^{(z)} = \frac{D_1^{(z)} - \rho_1}{\varrho_1 - \rho_1}, \quad \hat{D}_2^{(z)} = \frac{D_2^{(z)} - \rho_2}{\varrho_2 - \rho_2}, \quad \hat{D}_3^{(z)} = \frac{D_3^{(z)} - \rho_3}{\varrho_3 - \rho_3} \tag{23}$$

where $\rho_{(.)}$ is the minimum value between all $D_{(.)}$ entries and $\varrho_{(.)}$ is the maximum distance value between a query image and the other images in the data set.

After that, the distance measurements are grouped by their average as shown in Eq. (24).

$$\bar{D}^{(z)} = \frac{1}{3}\sum_{i=1}^{3} \hat{D}_i^{(z)} \tag{24}$$

Finally, the results are sorted in ascending order, and the images that produce the smallest distance are assigned to the query image.

3.3. Image retrieval and accuracy noise reduction

The image retrieval process starts with constructing a two-dimensional array that stores the image data set's features. In this step, the n images of a data set D are processed and their feature vectors are obtained (Section 3.1). Then, a feature space $\mathbf{F}$ of size $n \times k$ is prepared where n is the number of images in the data set, and k is the length of the feature vector ($k = 187$, considering the image descriptors cl-MSD, LBP, and LDiPv). Algorithm 1 describes the steps to build the feature space.

Algorithm 1 – Feature space construction.

Input: Image data set D_n.

Output: A two-dimensional array that contains the feature vector of each image in the data set.

1. Select all data set images $d \in D_n$ one by one and construct the feature vectors (cl-MSD, LBP, and LDiPv).
2. Construct a two-dimensional array $\mathbf{F}$ where its rows are indexes to each entry of the image data set, and its columns represent the image features.
3. Return the two-dimensional array $\mathbf{F}$.

In the next step, the feature vector of a query image is compared to the image representations of the data set images using the L_1 distance. Then, the results are ordered to form an initial cluster (IC) with the expected responses (Section 3.2). The detailed steps of the initial image retrieval process are explained in Algorithm 2.

Algorithm 2 – Initial image retrieval.

Input: RGB query image and data set image representation ($\mathbf{F}$).

Output: Initial cluster IC with the expected responses.

1. Generate the input query image feature vectors: cl-MSD, LBP, and LDiPv.
2. Compute L_1 distance between query image feature vector fv and the entire feature vectors fv_n of the feature space $\mathbf{F}$ one by one.
3. Sort the index values of the calculated distance values in ascending order.
4. Return the index values according to similarity order to form the initial cluster IC with the expected responses.

With the initial response from the previous step, the positioning of the retrieved images is updated considering the entries with the highest similarity to the query image. For a number m of images to be retrieved ($NR = m$), the first $m/2$ entries from the initial cluster (IC) are used in building secondary clusters (SC). Values smaller than $m/2$ may limit the output with retrieved images below the required number. On the other hand, values above $m/2$ are acceptable but may reduce assertiveness because, as discussed in Section 2.7, retrieved images that move away from the first position may be incorrect, which leads to a progressive increase in error. Thus, we use $m/2$ as a trade-off between the number of images required for retrieval and those that enhance assertiveness. In future code versions, we intend to handle this issue dynamically.

The process of constructing secondary clusters is also applied to the entries in the secondary clusters' initial positions ($m/2$). Then, after obtaining all the secondary clusters, the position of the first two entries of the initial cluster is found in the secondary clusters to measure the similarity between the initial cluster (IC) and the secondary clusters (SC). After that, the secondary clusters are ordered considering their proximity to the initial cluster. Then, the first two entries of the secondary clusters are concatenated until the number of expected responses is reached. In this step, the number $m/2$ guarantees that the amount of images retrieved follows the number of responses required. As the concatenation is prepared with the first two entries of the $m/2$ secondary clusters, the final cluster will have the same size as required, i.e., equal to m. Algorithm 3 details the steps to retrieve the top m images.

Algorithm 3 – Accuracy noise reduction.

Input: Initial cluster IC.

Output: Top m similar images from data set D_n.

1. For each of the first $m/2$ entries of IC, compute their similarity (Algorithm 2) to the D_n data set and get SC_1, SC_2,..., $SC_{m/2}$ secondary clusters.
2. Select the first $m/2$ entries from each secondary cluster and concatenate them to form a single cluster C.
3. Remove duplicate entries in C.
4. Repeat step 1 for all entries in C to increase the number of secondary clusters.
5. When all entries in C were processed find where the first and second entries of IC are in the secondary clusters indexed by C and sum the two results.
6. Order the results obtained in step 5 and rearrange the C entries accordingly.
7. Select the first two entries from the secondary clusters considering the new arrangement of C obtained in step 6.
8. Return the first m (where, $m \leq n$) index values to retrieve the top m images T from the data set D_n.

3.4. Replacement of duplicate entries

The accuracy noise reduction strategy may present duplicate images after the final arrangement. As the secondary clusters come from the interaction between a query image and the first positions of the initial cluster, redundancy may occur due to the similarities between the top images. It is reasonable to consider this circumstance since the similarity between images is a two-way path, i.e., if image B is in the set of images retrieved from image A, then A may be in the set of images retrieved from B.

To deal with this issue, we treat the replacement of duplicate entries by constructing small clusters. In this approach, small clusters are prepared from the output obtained with the accuracy noise reduction strategy. After obtaining the small clusters, the repeated images are replaced, and a new arrangement of retrieval images is presented. The detailed steps of this process are explained in Algorithm 4. As our

proposal is modular, we also intend to investigate other strategies for replacing repeated entries in future work.

Algorithm 4 – Replacement of duplicate entries.

Input: Top images T.

Output: Final top m similar images.

1. For each entry of T find the top m images (Algorithm 3) and get small clusters $sC_1, sC_2, ..., sC_m$.
2. Concatenate the small clusters to form a single cluster C.
3. Remove duplicate entries in C.
4. Find duplicate entries in T.
5. Replace duplicate entries in T with entries in C that are not part of T.

4. Results and discussion

We conducted experiments on four databases commonly used in evaluating image-based retrieval methods: Corel-1K (Wang et al., 2001), Corel-5K (Liu et al., 2011), Corel-10K (Liu et al., 2011), and GHIM-10K (Liu et al., 2015). These data sets were selected because they are frequently used in related work, which allows us to compare our proposal with other methods for image retrieval. Besides that, as explained by Murala, Maheshwari, and Balasubramanian (2012), these data sets meet all the requirements to evaluate CBIR systems due to their large size and heterogeneous content ranging from animals to outdoor sports to natural images.

The performance of the proposed method is measured with precision and recall metrics, where every database image was used as a query image. Besides, all experimental simulations were performed on Matlab 2014a installed on a notebook with Core i7-9750H (2.6 GHz; 12 MB cache) and 16 GB of RAM. The code prepared by the authors is publicly available.[1]

4.1. Performance evaluation metrics

Precision and recall are the two most used measures in verifying the performance of CBIR systems. Precision measures the number of correctly identified images against the expected number of images retrieved. Recall checks the ratio of correctly returned images to the total number of images in a given category.

The mathematical definition for precision (P) and recall (R) is given in Eq. (25) as follows:

$$p_{(i,j)} = \frac{NS_{(i,j)}}{NR}, \quad r_{(i,j)} = \frac{NS_{(i,j)}}{ND} \tag{25}$$

where NS stands for the number of retrieved similar images considering a query image indexed by j in the category i, NR stands for the number of expected responses, and ND is the total number of images in the data set that are similar to the query image. NR and ND are prefix numbers defined by the user and data set specifications, respectively.

As a data set has different categories, we use C_i to represent the number of images in each. Therefore, precision and recall average for the ith category is calculated according to Eq. (26):

$$P_{(i)} = \frac{\sum_{j=1}^{C_i} p_{(i,j)}}{C_i}, \quad R_{(i)} = \frac{\sum_{j=1}^{C_i} r_{(i,j)}}{C_i} \tag{26}$$

Based on that, the average retrieval precision (ARP) and the average retrieval recall (ARR) are defined in Eq. (27) where NC is the number of categories in a given data set.

$$ARP = \frac{\sum_{i=1}^{NC} P(i)}{NC}, \quad ARR = \frac{\sum_{i=1}^{NC} R(i)}{NC} \tag{27}$$

[1] https://github.com/gabrieldgf4/cbir-anr

4.2. Retrieval results

We present a retrieval image method by considering the combination of three image descriptors and a strategy to update the retrieval system responses to reduce noise in accuracy. We refer to "cl-MSD+LBP+LDiPv" to represent the three descriptors and "cl-MSD+LBP+LDiPv+ANR" to indicate their use with the accuracy noise reduction approach (Section 3.3). When we use the reserved word "OUR", it means that we are using the complete approach, i.e., the three descriptors and the response update strategy.

4.2.1. Corel-1K data set

The Corel-1K data set (Wang et al., 2001) contains 1,000 images equally distributed in 10 categories: Africans, Beaches, Buildings, Buses, Dinosaurs, Elephants, Flowers, Horses, Mountains, and Food. The images in this data set have dimensions equaling to 256 × 384 or 384 × 256 pixels.

Fig. 2 presents the ARPs and ARRs generated by the image descriptor combination. When these three descriptors are combined, the results are leveraged, reaching 80.62% of precision. However, with the application of the accuracy noise reduction approach ("ANR"), assertiveness significantly improved to 84.38%. Complementary, Fig. 3 shows the ARP-ARR curves corresponding to these four methods with the number of retrieved images (NR) varying from 1 to 12.

Table 1 lists the results of different methods for the Corel-1K data set categories. The maximum values of ARP and ARR are highlighted in boldface as the overall average of the most prominent method. Compared to the methods listed, our method performed the best results in the "Elephants", "Flowers", and "Mountains" categories. Furthermore, our method achieved the best ARP and ARR among the ten categories with 84.38% precision and 10.13% recall.

Fig. 4 presents the top 12 images retrieved from the Corel-1K data set considering a query image which is also one of the results (first image from the upper left corner). Fig. 4(a) shows the results achieved with the "cl-MSD+LBP+LDiPv" method while Fig. 4(b) shows the result of method "cl-MSD+LBP+LDiPv+ANR". Incorrect answers are highlighted with red dashed boxes. The first method obtained 83.3% precision because two invalid responses were presented. Our accuracy noise reduction approach automatically fine-tuned responses to the query image, achieving 100% precision.

Fig. 5 presents a case where updating images does not obtain maximum precision. However, the update approach still achieves a more assertive final answer, jumping from 66.6% to 83.3% precision.

4.2.2. Corel-5K data set

The Corel-5k data set (Liu et al., 2011) contains 5,000 images distributed in 50 different categories each with 100 images. The images in this data set have dimensions equaling to 192 × 128 or 128 × 192 pixels.

Fig. 6 presents the ARPs and ARRs generated by the proposed image descriptors in combination with NR equaling 12. Fig. 7 shows the ARP-ARR curves corresponding to the image descriptors varying the number of images retrieved from 1 to 12. It is observed that when combining the three image descriptors, the results are improved in comparison with the isolated descriptors. The accuracy noise reduction approach achieves better results by standing out with the best performance.

Table 2 lists results from different CBIR methods on the Corel-5K data set. The proposed method obtains the best overall average precision (ARP) and average recall (ARR).

Fig. 8 shows from a query image (top-left image), the top 12 retrieved images from the Corel-5K data set. The retrieved similar images include the query image itself. In Fig. 8(a), the combination of the three descriptors achieved 100% precision. Likewise, the accuracy noise reduction approach kept the accuracy at the same level, but with some image position changes and other retrieved images (Fig. 8(b)).

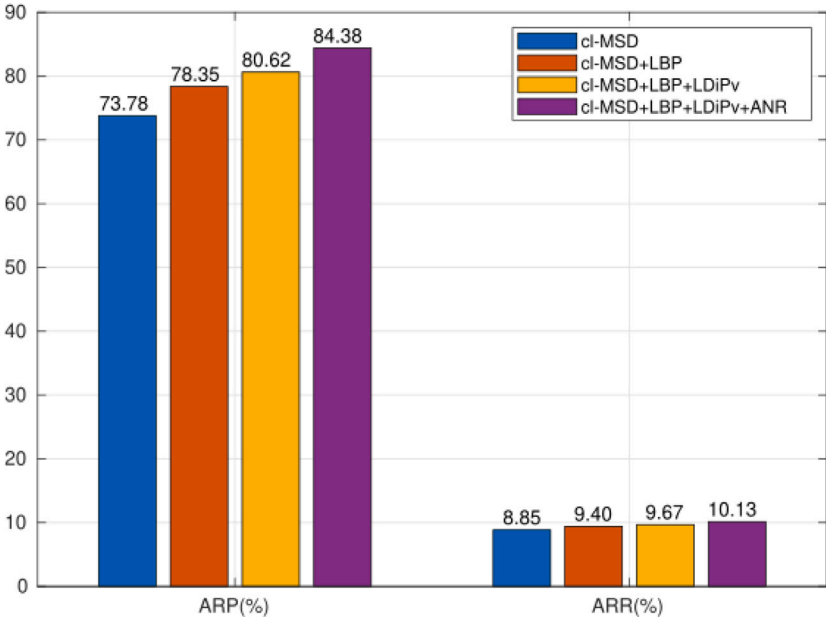


Fig. 2. Image descriptors performance on Corel-1K data set with $NR = 12$.

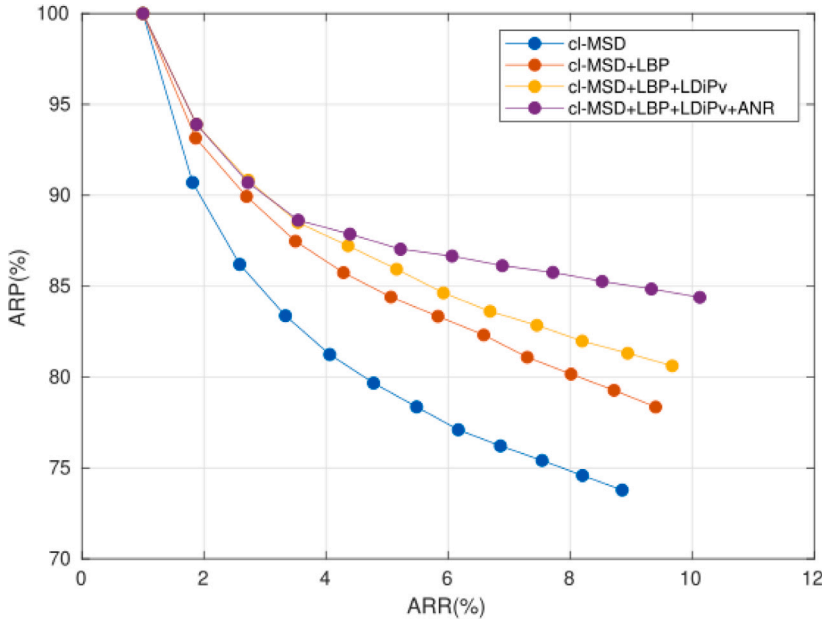


Fig. 3. ARP and ARR curves of the proposed descriptors on Corel-1K data set with $NR \in [1, 12]$.

4.2.3. Corel-10k data set

The Corel-10K data set (Liu et al., 2011) contains 100 images in each of its 100 categories. The total number of images in this data set is 10,000, which comprises the 50 categories of the Corel-5K data set and 50 categories with different attributes. The size of each image is either 192 × 128 or 128 × 192 pixels.

Fig. 9 presents the ARPs and ARRs generated by "cl-MSD", "cl-MSD+LBP", "cl-MSD+LBP+LDiPv", and "cl-MSD+LBP+LDiPv+ANR" method with NR equaling to 12. Fig. 10 shows the ARP-ARR curves corresponding to these four methods. As noted in previous experiments, the combined image descriptors and accuracy noise reduction approach significantly improve accuracy and recall results. While "cl-MSD" achieved 46.12% precision, the combination of the descriptors resulted in 57.00% precision. Then, using the accuracy noise reduction approach ("ANR"), 63.30% precision was obtained.

Table 3 lists some related works and presents their results obtained in the Corel-10K data set. In comparison, our method performed the best precision and recall result considering a NR equaling 12.

Fig. 11 presents the responses obtained by combining the three image descriptors and applying them to our accuracy noise reduction approach. In this example, incorrectly retrieved images are replaced with images of the same category as the query image, going from 50% precision to 100%.

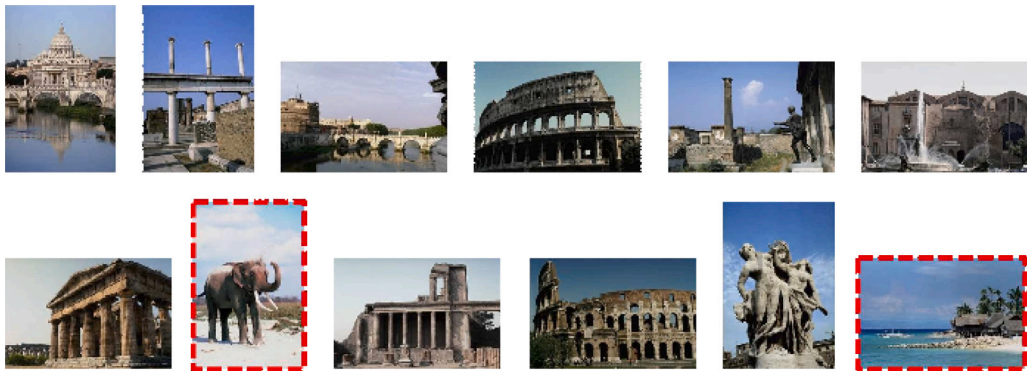
4.2.4. GHIM-10k data set

The GHIM-10K data set (Liu et al., 2015) contains 20 categories with 500 images in each one. Therefore, it has a total of 10,000 images that exhibits a size of either 300 × 400 or 400 × 300 pixels.

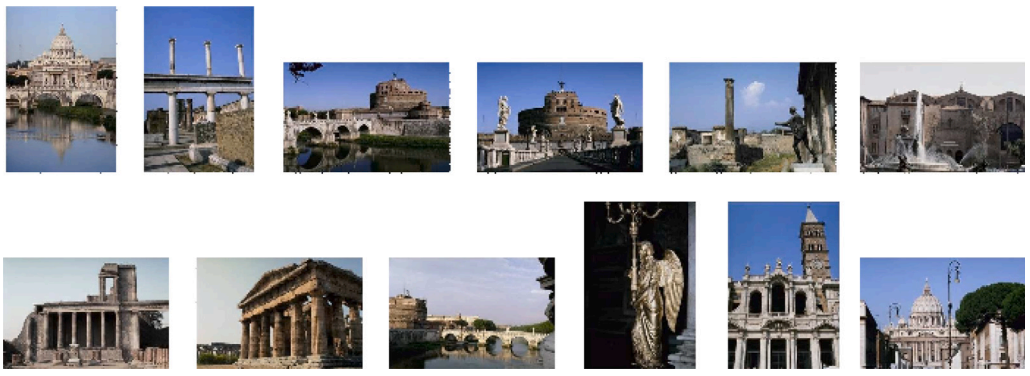
Fig. 12 presents the ARPs and ARRs generated by the descriptors investigated in this study ("cl-MSD", "LBP", "LDiPv") and the proposed

Table 1
Comparison of different image retrieval methods on the Corel-1K data set. The value of NR is 12. The bold values indicate the best results.

Category	Performance	Liu et al. (2011)	Feng, Wu, Liu, and Zhang (2015)	Raza et al. (2018)	Dawood et al. (2019)	Niu et al. (2020)	OUR
Africans	ARP	83.33	87.50	**91.66**	86.66	86.42	82.83
	ARR	10.00	10.50	**11.00**	10.40	9.50	9.94
Beaches	ARP	43.33	**68.33**	54.58	42.08	50.25	59.83
	ARR	5.20	**8.20**	6.55	5.05	6.03	7.18
Buildings	ARP	63.33	61.67	78.75	**81.66**	77.83	77.75
	ARR	7.60	7.40	9.45	**9.80**	9.34	9.33
Buses	ARP	76.67	80.00	86.25	81.66	**98.08**	96.75
	ARR	9.20	9.60	10.35	9.80	**11.77**	11.6
Dinosaurs	ARP	**100.00**	**100.00**	**100.00**	**100.00**	99.17	99.08
	ARR	**12.00**	**12.00**	**12.00**	**12.00**	11.90	11.89
Elephants	ARP	65.00	67.50	65.83	72.08	74.50	**84.17**
	ARR	7.80	8.10	7.90	8.65	8.94	**10.10**
Flowers	ARP	86.67	88.33	95.41	84.16	95.92	**99.00**
	ARR	10.40	10.60	11.45	10.10	11.51	**11.88**
Horses	ARP	97.50	**100.00**	93.33	94.16	95.92	95.08
	ARR	11.70	**12.00**	11.20	11.30	11.51	11.41
Mountains	ARP	29.17	55.00	56.25	50.00	53.25	**58.50**
	ARR	3.50	6.60	6.75	6.00	6.39	**7.02**
Food	ARP	76.67	74.17	85.83	**92.91**	91.08	90.83
	ARR	9.20	8.90	10.30	**11.15**	10.93	10.90
Average	ARP	72.17	72.85	80.79	78.54	82.24	**84.38**
	ARR	8.66	9.39	9.69	9.42	9.87	**10.13**



(a) Retrieved images using "cl-MSD+LBP+LDiPv" descriptor.



(b) Retrieved images using "cl-MSD+LBP+LDiPv" descriptor with the "ANR"approach.

Fig. 4. A query image (top-left corner of (a) and (b)) from the Corel-1K data set and its 12 top similar images. Similar images include the query image itself. Incorrect responses are marked with red dashed boxes.

accuracy noise reduction approach ("ANR") with NR equaling to 12. Fig. 13 shows the ARP-ARR curves corresponding to these methods with NR varying from 1 to 12. As noted, when the image descriptors are used alone, they achieve inferior results than their combination. In addition, the accuracy noise reduction strategy improves results substantially.

Table 4 lists the precision and recall results of different methods related to our proposal. In comparing these methods with ours, it can

(a) Retrieved images using “cl-MSD+LBP+LDiPv” descriptor.

(b) Retrieved images using “cl-MSD+LBP+LDiPv” descriptor with the “ANR”approach.

Fig. 5. A query image (top-left corner of (a) and (b)) from the Corel-1K data set and its 12 top similar images. Similar images include the query image itself. Incorrect responses are marked with red dashed boxes.

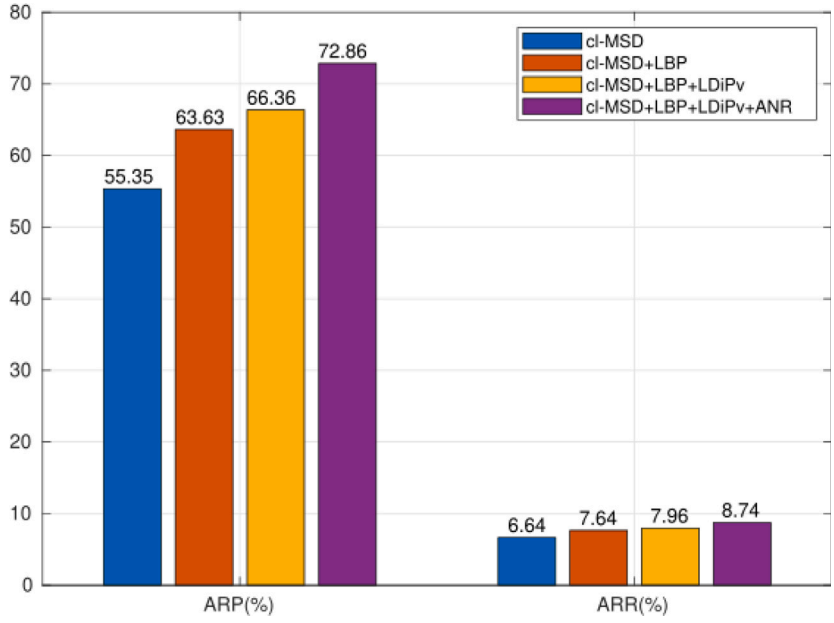


Fig. 6. Image descriptors performance on Corel-5K data set with $NR = 12$.

be noted that the proposed method has the best performance among them.

Fig. 14 shows the visual results for a query image (upper left side). As in the other examples, the query image is also part of the results, which can be seen in the first position among the top 12 images. The three descriptors combined achieved 83.33% precision with two incorrect responses. The update of the answers through the proposed method reached 91.67% precision with only one wrong answer.

Table 2
Comparison of different image retrieval methods on the Corel-5K data set. The value of NR is 12. The bold values indicate the best results.

Performance	Hua et al. (2019)	Dawood et al. (2019)	Chu and Liu (2020)	Liu and Wei (2020)	Wei and Liu (2020)	Niu et al. (2020)	OUR
ARP	60.13	63.14	60.16	63.50	66.91	67.93	**72.86**
ARR	7.21	7.54	7.21	7.62	8.03	8.15	**8.74**

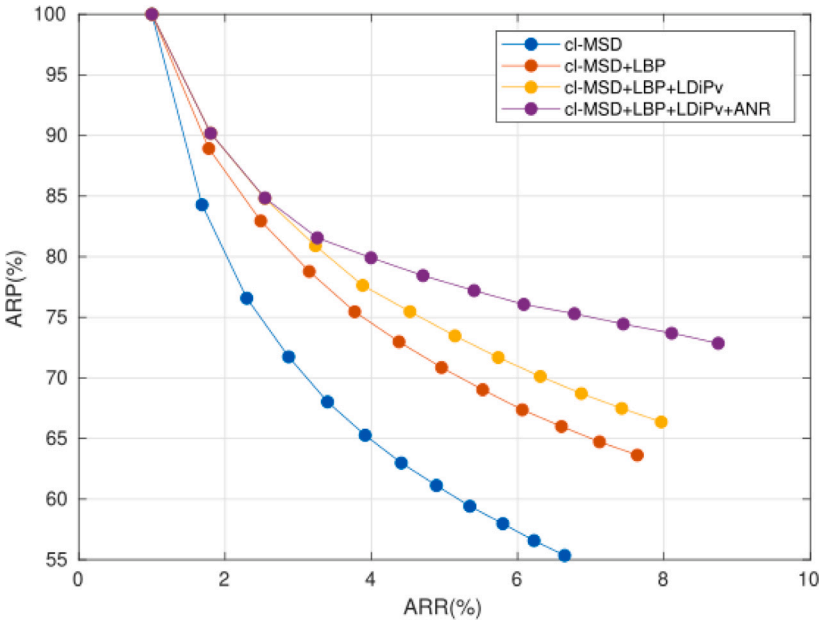


Fig. 7. ARP and ARR curves of the proposed descriptors on Corel-5K data set with $NR \in [1, 12]$.

(a) Retrieved images using "cl-MSD+LBP+LDiPv" descriptor.

(b) Retrieved images using "cl-MSD+LBP+LDiPv" descriptor with the "ANR"approach.

Fig. 8. A query image (top-left corner of (a) and (b)) from the Corel-5K data set and its 12 top similar images. Similar images include the query image itself.

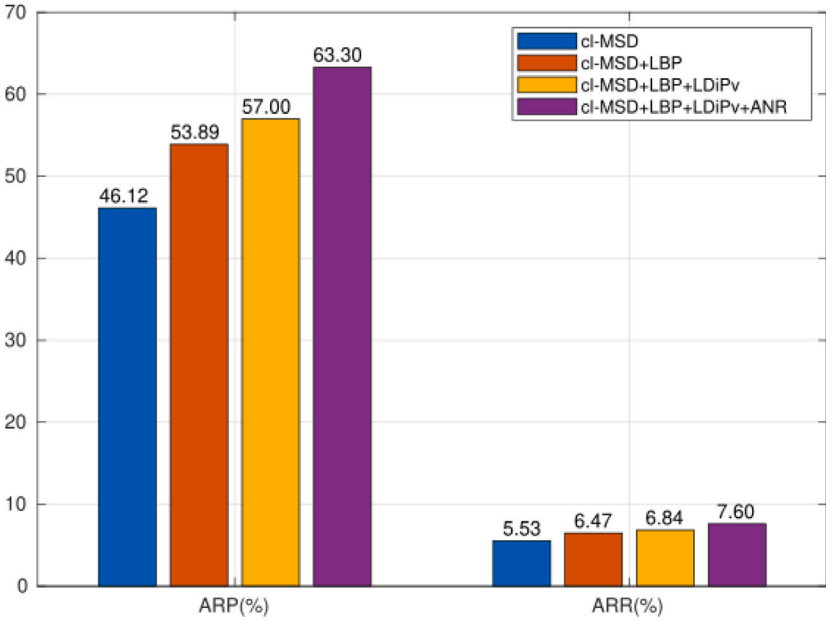


Fig. 9. Image descriptors performance on Corel-10K data set with $NR = 12$.

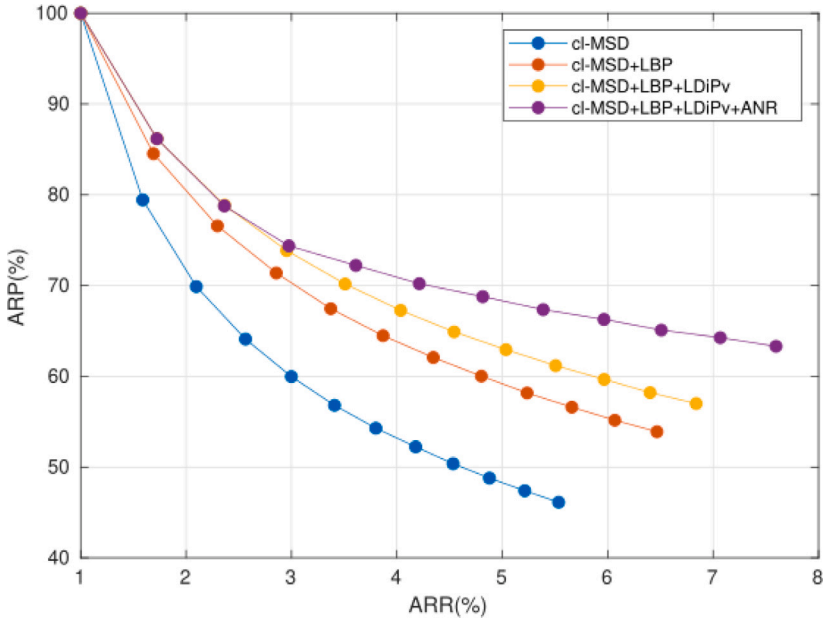


Fig. 10. ARP and ARR curves of the proposed descriptors on Corel-10K data set with $NR \in [1, 12]$.

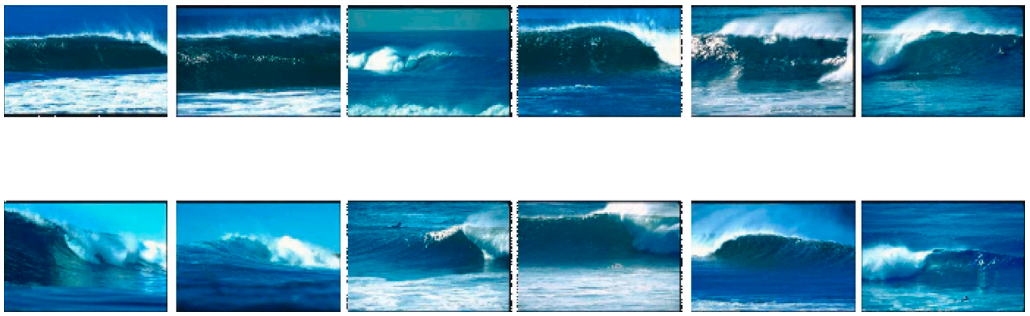
Table 3
Comparison of different image retrieval methods on the Corel-10K data set. The value of NR is 12. The bold values indicate the best results.

Performance	Hua et al. (2019)	Dawood et al. (2019)	Chu and Liu (2020)	Liu and Wei (2020)	Wei and Liu (2020)	Niu et al. (2020)	OUR
ARP	48.58	50.2	52.96	53.19	56.88	58.52	**63.31**
ARR	5.83	6.03	6.36	6.38	6.83	7.02	**7.60**

Table 4
Comparison of different image retrieval methods on the GHIM-10K data set. The value of NR is 12. The bold values indicate the best results.

Performance	Liu (2015)	Liu et al. (2015)	Liu (2016)	Zhou, Liu, Liu, and Gan (2019)	Chu and Liu (2020)	Niu et al. (2020)	OUR
ARP	57.51	61.16	56.68	67.10	56.48	68.50	**71.23**
ARR	1.38	1.47	1.36	1.61	1.36	1.65	**1.71**

(a) Retrieved images using “cl-MSD+LBP+LDiPv” descriptor.



(b) Retrieved images using “cl-MSD+LBP+LDiPv” descriptor with the “ANR”approach.

Fig. 11. A query image (top-left corner of (a) and (b)) from the Corel-10K data set and its 12 top similar images. Similar images include the query image itself. Incorrect responses are marked with red dashed boxes.

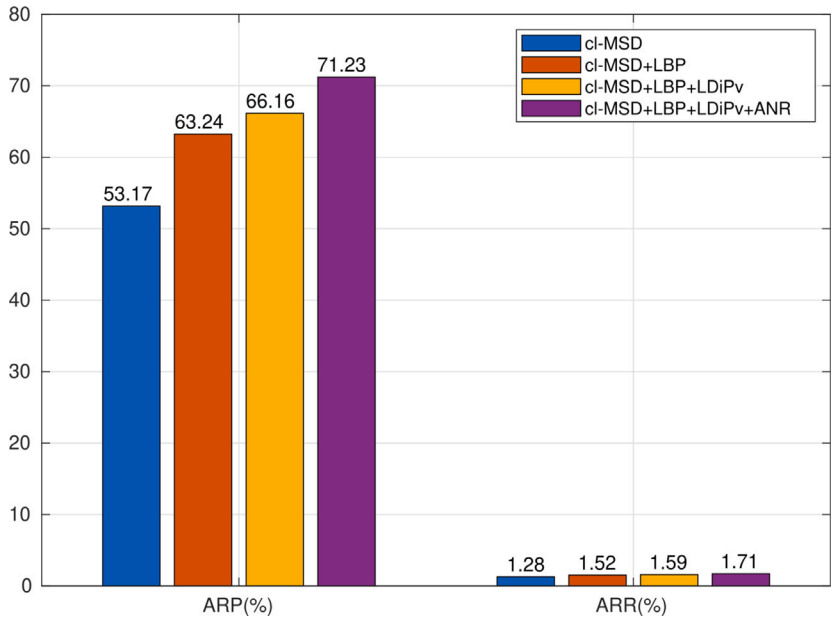


Fig. 12. Image descriptors performance on GHIM-10K data set with $NR = 12$.

4.3. Comparison with learning approaches

Table 5 shows the result of the accuracy and recall metrics for machine learning models and the proposed method results considering the top 20 similar images ($NR = 20$). As can be seen, the methods that use training steps achieve better results than our method in Corel-1K and Corel-10K databases. However, in the Corel-5K database, our method performed better than (Sharif et al., 2019).

Although the related works present a better performance in some cases, it is worth mentioning that machine learning methods have training steps that can be time-consuming. Moreover, supervised learning models require annotated data which can be challenging for large-scale multimedia repositories. In contrast, our method does not involve training steps and can be performed directly without categorizing the database.

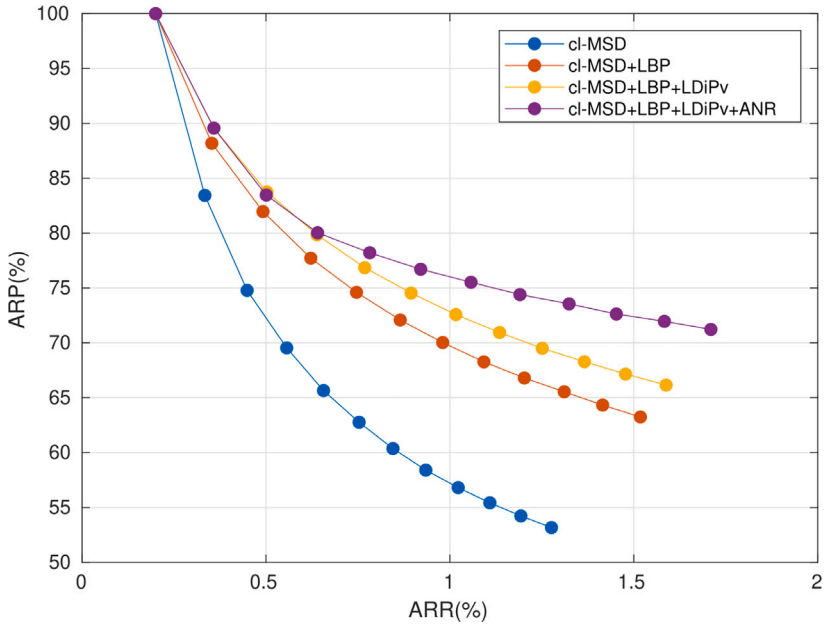


Fig. 13. ARP and ARR curves of the proposed descriptors on GHIM-10K data set with $NR \in [1, 12]$.

(a) Retrieved images using "cl-MSD+LBP+LDiPv" descriptor.

(b) Retrieved images using "cl-MSD+LBP+LDiPv" descriptor with the "ANR"approach.

Fig. 14. A query image (top-left corner of (a) and (b)) from the GHIM-10K data set and its 12 top similar images. Similar images include the query image itself. Incorrect responses are marked with red dashed boxes.

Another critical limitation of machine learning approaches is the difficulty of models in dealing with unseen scenarios. In this case, a learning model is designed for each database to obtain better results and may even require changes in network parameters. Furthermore, the training steps of machine learning models are time-consuming and may require suitable computers to process multiple image transformation operations. Our proposal proved consistent in image retrieval working in different databases without using training steps. In addition to the generalization capability, our proposal is efficient regarding data processing time, as explained in Section 4.4.

4.4. Time performance analysis

The temporal performance of the proposed CBIR system was analyzed in terms of CPU time taken by its processes. In the proposed CBIR system, four steps, presented in Algorithms 1 to 4, have been executed to perform the image retrieval. The time was measured for

Table 5
Comparative analysis with machine learning proposals for CBIR systems. The value of NR is 20. The bold values indicate the best results.

Data set	Performance	Method					
		Irtaza et al. (2018)	Ahmed et al. (2019)	Sharif et al. (2019)	Ahmed et al. (2020)	Baig et al. (2020)	OUR
Corel-1K	ARP	84.7	83.5	84.3	80.0	**86.4**	79.6
	ARR	16.90	12.3	16.87	–	**17.28**	15.93
Corel-5K	ARP	–	–	57.3	–	–	**67.1**
	ARR	–	–	11.47	–	–	**13.43**
Corel-10K	ARP	–	56.7	–	**65.0**	–	56.7
	ARR	–	–	–	**16.0**	–	11.35

Table 6
Average CPU time and Standard Deviation (in seconds) by each task in the image retrieval process considering different image data sets.

	Corel-1K (s)	Corel-5K (s)	Corel-10K (s)	GHIM-10K (s)
Algorithm 1	1.7175 ± 0.1642	0.4548 ± 0.0615	0.4269 ± 0.0336	2.0767 ± 0.0519
Algorithm 2	0.0988 ± 0.0086	0.5252 ± 0.0167	1.0245 ± 0.0230	1.0146 ± 0.0250
Algorithm 3	0.0004 ± 0.0004	0.0034 ± 0.0010	0.0073 ± 0.0021	0.0077 ± 0.0020
Algorithm 4	0.0003 ± 0.0005	0.0015 ± 0.0006	0.0031 ± 0.0012	0.0031 ± 0.0012
Total	**1.8170 ± 0.1737**	**0.9849 ± 0.0798**	**1.4618 ± 0.0599**	**3.1020 ± 0.0818**

Table 7
Comparison between distance measures using average retrieval precision (ARP) and average running time (in seconds).

	Corel-1K		Corel-5K		Corel-10K		GHIM-10K	
	ARP	Time	ARP	Time	ARP	Time	ARP	Time
Hamming	36.91	0.1885	17.30	0.9795	13.58	1.9893	21.70	1.7321
Chebychev	55.38	0.1955	30.73	0.9940	23.36	3.0880	31.18	1.7711
Cosine	54.47	0.2517	35.18	1.2519	26.79	2.5582	35.34	2.2980
Euclidean	58.35	0.1903	34.71	1.0601	27.15	1.9701	36.46	1.7802
Spearman	68.98	0.3464	50.21	1.7086	40.09	3.6337	49.88	3.0789
L_1	**84.38**	**0.0988**	**72.86**	**0.5252**	**63.31**	**1.0245**	**71.23**	**1.0146**

each processing step to show the computation time involved in the (i) Feature space construction, (ii) Initial image retrieval, (iii) Accuracy noise reduction, and (iv) Replacement of duplicate entries. Table 6 presents the average CPU time (in seconds) and standard deviation of each task in the image retrieval process performed on all four image data sets.

As shown in Table 6, the construction of feature vectors consumed most of the processing time. Once completed, the similarity evaluation, accuracy noise reduction, and duplicate replacement operations were performed in close to zero seconds. In evaluating the total processing time, it is observed that the processing time was below 1.8 s for the Corel1K and Corel10K databases. Furthermore, the average processing time for the Corel5K database was less than 1 s. On the other hand, the GHIM10K database required just over 3 s to complete the tasks. Considering GHIM10K has larger images than the other data sets, the time required to compute the feature vectors (Algorithm 1) was also proportionally higher.

4.5. Distance measures

In this section, we evaluate the performance of our image retrieval method, considering different distance measures. The distance metric numerically indicates the similarity between images, which is a crucial part of CBIR systems. We compared six distance metrics in this evaluation: Hamming, Chebychev, Cosine, Euclidean, Spearman, and L_1. Experiments were performed on the Corel1K, Corel5K, Corel10K, and GHIM10k with the number of retrieved images equal to 12. In addition to computing the average retrieval precision (ARP), the average running time of each distance metric is also measured.

Table 7 shows the comparative results of different distance metrics applied to our content-based image retrieval method. As can be seen, the L_1 distance showed the best accuracy results in all databases. Although Spearman distance presented the second-best result, it was far below the L_1 distance between 16 and 23 percentage points. Among the compared distance metrics, the Hamming distance presented the least assertive result for our fusion-based method, followed by the Chebychev and Cosine distances. Furthermore, the L_1 distance was processed faster than the other distances, practically half the time spent by the other metrics.

4.6. Comparison with individual descriptors

This section evaluates the performance of the proposed system using other image descriptors. We performed the experiments on the Corel1K, Corel5K, Corel10K, and GHIM10K databases and measured the average retrieval precision (ARP) and the average retrieval recall (ARR) by considering the performance of individual descriptors and applying our accuracy noise reduction strategy. This experiment compares ten image descriptors, namely Gradient Directional Pattern (GDP) (Ahmed, 2012), Gradient Local Ternary Pattern (GLTeP) (Islam, 2013), Local Arc Pattern (LAP) (Islam & Auwatanamo, 2014), Local Directional Number Pattern (LDN) (Rivera, Castillo, & Chae, 2012), Local Frequency Descriptor (LFD) (Lei et al., 2011), Local Monotonic Pattern (LMP) (Mohammad & Ali, 2011), Local Phase Quantization (LPQ) (Dhall et al., 2011), Local Ternary Pattern (LTeP) (Tan & Triggs, 2010), Median Binary Pattern (MBP) (Bashar et al., 2014), and Pyramid of Histogram of Oriented Gradients (PHOG) (Bosch, Zisserman, & Munoz, 2007). We use the code prepared by Turan and Lam (2018), which contains the implementation of the selected descriptors.

Table 8 presents the accuracy results for different image descriptors. The LPQ obtained the best results in the three Corel databases by comparing the individual descriptors. In contrast, LTeP performed best in the GHIM10K database, followed by the GLTP, LMP, and MBP descriptors. Furthermore, it is possible to observe that the accuracy noise reduction (ANR) strategy improved the results from 1 to 4%. For example, the initial result of LTeP on Corel1K was 64% accurate. After applying the ANR, the descriptor achieved 68% accuracy. In the Table, it is still possible to follow the results of our fusion-based method. As

Table 8
Comparative evaluation of image descriptors using mean values of precision and recall (ARP and ARR) and feature vector dimensionality ($dim = (.)$).

	Corel-1K		Corel-5K		Corel-10K		GHIM-10K	
	ARP	ARR	ARP	ARR	ARP	ARR	ARP	ARR
GDP (dim=256)	59.34	7.12	31.91	3.83	25.58	3.07	41.02	0.98
GDP+ANR	**62.47**	**7.50**	**33.77**	**4.05**	**26.71**	**3.21**	**42.11**	**1.01**
GLTeP (dim=512)	63.93	7.67	40.89	4.91	34.02	4.08	48.81	1.17
GLTeP+ANR	**66.35**	**7.96**	**44.01**	**5.28**	**36.58**	**4.39**	**50.47**	**1.21**
LAP (dim=272)	60.44	7.25	36.65	4.40	29.94	3.59	40.36	0.97
LAP+ANR	**62.29**	**7.47**	**38.58**	**4.63**	**31.85**	**3.82**	**41.55**	**1.00**
LDN (dim=56)	63.55	7.63	39.35	4.72	32.52	3.90	42.65	1.02
LDN+ANR	**64.87**	**7.88**	**41.38**	**4.97**	**34.53**	**4.14**	**44.12**	**1.06**
LFD (dim=512)	65.71	7.88	36.08	4.33	29.82	3.58	44.77	1.07
LFD+ANR	**68.59**	**8.23**	**37.95**	**4.55**	**31.49**	**3.78**	**45.75**	**1.10**
LMP (dim=256)	65.58	7.87	39.68	4.76	32.63	3.91	46.72	1.12
LMP+ANR	**67.82**	**8.14**	**42.06**	**5.05**	**34.55**	**4.15**	**48.25**	**1.16**
LPQ (dim=256)	69.08	8.29	45.08	5.41	36.89	4.43	45.88	1.10
LPQ+ANR	**72.08**	**8.65**	**48.25**	**5.79**	**39.46**	**4.73**	**47.71**	**1.15**
LTeP (dim=512)	64.58	7.75	44.04	5.28	36.47	4.38	49.46	1.19
LTeP+ANR	**68.00**	**8.16**	**47.67**	**5.72**	**39.79**	**4.78**	**51.33**	**1.23**
MBP (dim=256)	63.64	7.64	40.07	4.81	33.33	4.00	46.20	1.11
MBP+ANR	**66.43**	**7.97**	**42.56**	**5.11**	**35.52**	**4.26**	**47.88**	**1.15**
PHOG (dim=168)	48.42	5.81	36.60	4.39	29.24	3.51	42.06	1.01
PHOG+ANR	**52.62**	**6.31**	**40.56**	**4.87**	**32.11**	**3.85**	**44.52**	**1.07**
cl-MSD+LBP+LDiPv (dim=187)	80.62	9.67	66.36	7.96	57.00	6.84	66.16	1.59
cl-MSD+LBP+LDiPv+ANR	**84.38**	**10.13**	**72.86**	**8.74**	**63.30**	**7.60**	**71.23**	**1.71**

shown, the combination of descriptors is superior to any individual descriptors. Also, the vector's dimensionality (dim = 187) is smaller than most of the other descriptors. Therefore, it is competitive as it requires little storage space and can be processed with few computational resources.

4.7. Limitation and future work

The proposed method has some limitations, which can be noted in eventual circumstances. The first occurs when the query image has visual attributes resembling multiple image categories. In this case, the method may return incorrect responses, indicating images of a different category than expected due to the image similarities. The second occurs when the similarity evaluation returns only images whose categorization differs from the expected class. In this case, the accuracy noise reduction strategy is compromised because the correct category cannot be recovered. Finally, the third occurs when the constructed image clusters have the same images in the first positions. In this case, our replacement of duplicate entries algorithm may not adequately replace the multiple duplicate entries.

The presented limitations pave the way for constructing new approaches in future work. To deal with these issues, we intend to investigate other image descriptor combinations to propose other fusion-based methods, explore local approaches using image saliency techniques, use the findings of this study with machine learning techniques, update the proposed algorithms to dynamically deal with the top images parameter used by the accuracy noise reduction strategy, and develop other strategies to handle duplicate entries caused by our accuracy improvement process.

5. Conclusions

This paper presents a novel descriptor integration to represent image signatures. With the combination of the cl-MSD, LBP, and LDiPv descriptors, the visual attributes of the images were encapsulated in a 187-dimensional vector, which is a very reasonable size for large-scale image data sets. Furthermore, we explore the interactions between the top first images retrieved from CBIR systems to present a new accuracy noise reduction method. We evaluated the proposal considering four different data sets. The results showed consistency in dealing with erroneous responses when different image primitives generate similarity between feature vectors. Thus, compared with other image retrieval strategies, the proposed method is in line with the latest content-based image retrieval approaches. For future work, we would like to investigate this approach using other image descriptors and apply this method in combination with machine learning algorithms.

CRediT authorship contribution statement

Gabriel S. Vieira: Conceptualization, Methodology, Software, Writing – original draft. **Afonso U. Fonseca:** Methodology, Validation. **Naiane M. Sousa:** Visualization. **Juliana P. Felix:** Investigation. **Fabrizzio Soares:** Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

The authors would like to thank the Universidade Federal de Goiás (Brazil), Instituto Federal Goiano (Brazil), and CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brazil) [CAPES Finance Code #001] for partially supporting this research work. We also thank Dr. Guang-Hai Liu for sharing the Corel-10K data set with us.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.eswa.2023.120774. We have added to the article presentation the computer program code developed by the authors for content-based image retrieval. The program was thoroughly investigated in this study and is publicly available online. In Vieira et al. (2023) is presented more details of the implementation.

References

Ahmed, F. (2012). Gradient directional pattern: a robust feature descriptor for facial expression recognition. *Electronics Letters*, *48*(19), 1203–1204.

Ahmed, K. T., Afzal, H., Mufti, M. R., Mehmood, A., & Choi, G. S. (2020). Deep image sensing and retrieval using suppression, scale spacing and division, interpolation and spatial color coordinates with bag of words for large and complex datasets. *IEEE Access*, *8*, 90351–90379.

Ahmed, K. T., Jaffar, S., Hussain, M. G., Fareed, S., Mehmood, A., & Choi, G. S. (2021). Maximum response deep learning using Markov, retinal & primitive patch binding with GoogLeNet & VGG-19 for large image retrieval. *IEEE Access*, *9*, 41934–41957.

Ahmed, K. T., Ummesafi, S., & Iqbal, A. (2019). Content based image retrieval using image features information fusion. *Information Fusion*, *51*, 76–99.

Ali, N., Bajwa, K. B., Sablatnig, R., Chatzichristofis, S. A., Iqbal, Z., Rashid, M., et al. (2016). A novel image retrieval based on visual words integration of SIFT and SURF. *PLoS One*, *11*(6), Article e0157428.

Alzu'bi, A., Amira, A., & Ramzan, N. (2015). Semantic content-based image retrieval: A comprehensive study. *Journal of Visual Communication and Image Representation*, *32*, 20–54.

Baig, F., Mehmood, Z., Rashid, M., Javid, M. A., Rehman, A., Saba, T., et al. (2020). Boosting the performance of the BoVW model using SURF–cohog-based sparse features with relevance feedback for CBIR. *Iranian Journal of Science and Technology, Transactions of Electrical Engineering*, *44*(1), 99–118.

Bashar, F., Khan, A., Ahmed, F., & Kabir, M. H. (2014). Robust facial expression recognition based on median ternary pattern (MTP). In *2013 international conference on electrical information and communication technology* (pp. 1–5). IEEE.

Bibi, R., Mehmood, Z., Yousaf, R. M., Saba, T., Sardaraz, M., & Rehman, A. (2020). Query-by-visual-search: multimodal framework for content-based image retrieval. *Journal of Ambient Intelligence and Humanized Computing*, *11*(11), 5629–5648.

Bosch, A., Zisserman, A., & Munoz, X. (2007). Representing shape with a spatial pyramid kernel. In *Proceedings of the 6th ACM international conference on image and video retrieval* (pp. 401–408).

Chen, Q., Ding, Y., Li, H., Wang, X., Wang, J., & Deng, X. (2014). A novel multi-feature fusion and sparse coding-based framework for image retrieval. In *2014 IEEE international conference on systems, man, and cybernetics* (pp. 2391–2396). IEEE.

Chu, K., & Liu, G.-H. (2020). Image retrieval based on a multi-integration features model. *Mathematical Problems in Engineering*, *2020*.

Dawood, H., Alkinani, M. H., Raza, A., Dawood, H., Mehboob, R., & Shabbir, S. (2019). Correlated microstructure descriptor for image retrieval. *IEEE Access*, *7*, 55206–55228.

Dhall, A., Asthana, A., Goecke, R., & Gedeon, T. (2011). Emotion recognition using PHOG and LPQ features. In *2011 IEEE international conference on automatic face & gesture recognition* (pp. 878–883). IEEE.

Feng, L., Wu, J., Liu, S., & Zhang, H. (2015). Global correlation descriptor: a novel image representation for image retrieval. *Journal of Visual Communication and Image Representation*, *33*, 104–114.

Hameed, I. M., Abdulhussain, S. H., & Mahmmod, B. M. (2021). Content-based image retrieval: A review of recent trends. *Cogent Engineering*, *8*(1), Article 1927469.

Hua, J.-Z., Liu, G.-H., & Song, S.-X. (2019). Content-based image retrieval using color volume histograms. *International Journal of Pattern Recognition and Artificial Intelligence*, *33*(11), Article 1940010.

Irtaza, A., Adnan, S. M., Ahmed, K. T., Jaffar, A., Khan, A., Javed, A., et al. (2018). An ensemble based evolutionary approach to the class imbalance problem with applications in CBIR. *Applied Sciences*, *8*(4), 495.

Islam, M. S. (2013). Gender classification using gradient direction pattern. *Science International (Lahore)*, *25*(4), 797–799.

Islam, M. S., & Auwatanamo, S. (2014). Facial expression recognition using local arc pattern. *Trends in Applied Sciences Research*, *9*(2), 113.

Jegou, H., Douze, M., & Schmid, C. (2008). Hamming embedding and weak geometric consistency for large scale image search. In *European conference on computer vision* (pp. 304–317). Springer.

Kabir, M. H., Jabid, T., & Chae, O. (2010). A local directional pattern variance (LDPv) based face descriptor for human facial expression recognition. In *2010 7th IEEE international conference on advanced video and signal based surveillance* (pp. 526–532). IEEE.

Kanwal, K., Tehseen Ahmad, K., Khan, R., Alhusaini, N., & Jing, L. (2021). Deep learning using isotroping, laplacing, eigenvalues interpolative binding, and convolved determinants with normed mapping for large-scale image retrieval. *Sensors*, *21*(4), 1139.

Latif, A., Rasheed, A., Sajid, U., Ahmed, J., Ali, N., Ratyal, N. I., et al. (2019). Content-based image retrieval and feature extraction: a comprehensive review. *Mathematical Problems in Engineering*, *2019*.

Lei, Z., Ahonen, T., Pietikäinen, M., & Li, S. Z. (2011). Local frequency descriptor for low-resolution face recognition. In *2011 IEEE international conference on automatic face & gesture recognition* (pp. 161–166). IEEE.

Li, J., & Wang, J. Z. (2003). Automatic linguistic indexing of pictures by a statistical modeling approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *25*(9), 1075–1088.

Liu, G.-H. (2015). Content-based image retrieval based on visual attention and the conditional probability. In *International conference on chemical, material and food engineering* (pp. 843–847). Atlantis Press.

Liu, G.-H. (2016). Content-based image retrieval based on cauchy density function histogram. In *2016 12th international conference on natural computation, fuzzy systems and knowledge discovery (ICNC-FSKD)* (pp. 506–510). IEEE.

Liu, G.-H., Li, Z.-Y., Zhang, L., & Xu, Y. (2011). Image retrieval based on micro-structure descriptor. *Pattern Recognition*, *44*(9), 2123–2133.

Liu, G.-H., & Wei, Z. (2020). Image retrieval using the fused perceptual color histogram. *Computational Intelligence and Neuroscience*, *2020*.

Liu, G.-H., & Yang, J.-Y. (2013). Content-based image retrieval using color difference histogram. *Pattern Recognition*, *46*(1), 188–198.

Liu, G.-H., Yang, J.-Y., & Li, Z. (2015). Content-based image retrieval using computational visual attention model. *Pattern Recognition*, *48*(8), 2554–2566.

Maji, S., & Bose, S. (2021). CBIR using features derived by deep learning. *ACM/IMS Transactions on Data Science (TDS)*, *2*(3), 1–24.

Mehmood, Z., Gul, N., Altaf, M., Mahmood, T., Saba, T., Rehman, A., et al. (2018). Scene search based on the adapted triangular regions and soft clustering to improve the effectiveness of the visual-bag-of-words model. *EURASIP Journal on Image and Video Processing*, *2018*(1), 1–16.

Mohammad, T., & Ali, M. L. (2011). Robust facial expression recognition based on local monotonic pattern (LMP). In *14th international conference on computer and information technology (ICCIT 2011)* (pp. 572–576). IEEE.

Murala, S., Maheshwari, R., & Balasubramanian, R. (2012). Directional local extrema patterns: a new descriptor for content based image retrieval. *International Journal of Multimedia Information Retrieval*, *1*(3), 191–203.

Nazir, A., & Nazir, K. (2018). An efficient image retrieval based on fusion of low-level visual features. arXiv preprint arXiv:1811.12695.

Nene, S. A., Nayar, S. K., Murase, H., et al. (1996). Columbia object image library (coil-100).

Niu, D., Zhao, X., Lin, X., & Zhang, C. (2020). A novel image retrieval method based on multi-features fusion. *Signal Processing: Image Communication*, [ISSN: 0923-5965] *87*, Article 115911. http://dx.doi.org/10.1016/j.image.2020.115911.

Ojala, T., Pietikainen, M., & Maenpaa, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *24*(7), 971–987.

Pavithra, L., & Sharmila, T. S. (2018). An efficient framework for image retrieval using color, texture and edge features. *Computers & Electrical Engineering*, *70*, 580–593.

Philbin, J., Chum, O., Isard, M., Sivic, J., & Zisserman, A. (2007). Object retrieval with large vocabularies and fast spatial matching. In *2007 IEEE conference on computer vision and pattern recognition* (pp. 1–8). IEEE.

Pradhan, J., Pal, A. K., & Banka, H. (2022). A CBIR system based on saliency driven local image features and multi orientation texture features. *Journal of Visual Communication and Image Representation*, [ISSN: 1047-3203] *83*, Article 103396. http://dx.doi.org/10.1016/j.jvcir.2021.103396.

Raja, R., Kumar, S., & Mahmood, M. R. (2020). Color object detection based image retrieval using ROI segmentation with multi-feature method. *Wireless Personal Communications*, *112*(1), 169–192.

Raza, A., Dawood, H., Dawood, H., Shabbir, S., Mehboob, R., & Banjar, A. (2018). Correlated primary visual texton histogram features for content base image retrieval. *IEEE Access*, *6*, 46595–46616.

Raza, A., Nawaz, T., Dawood, H., & Dawood, H. (2019). Square texton histogram features for image retrieval. *Multimedia Tools and Applications*, *78*(3), 2719–2746.

Rivera, A. R., Castillo, J. R., & Chae, O. O. (2012). Local directional number pattern for face analysis: Face and expression recognition. *IEEE Transactions on Image Processing*, *22*(5), 1740–1752.

Sathiamoorthy, S., & Natarajan, M. (2020). An efficient content based image retrieval using enhanced multi-trend structure descriptor. *SN Applied Sciences*, *2*(2), 1–20.

Shakarami, A., & Tarrah, H. (2020). An efficient image descriptor for image classification and CBIR. *Optik*, *214*, Article 164833.

Sharif, U., Mehmood, Z., Mahmood, T., Javid, M. A., Rehman, A., & Saba, T. (2019). Scene analysis and search using local features and support vector machine for effective content-based image retrieval. *Artificial Intelligence Review*, *52*(2), 901–925.

Shikha, B., Gitanjali, P., & Kumar, D. P. (2020). An extreme learning machine-relevance feedback framework for enhancing the accuracy of a hybrid image retrieval system.. *International Journal of Interactive Multimedia & Artificial Intelligence*, *6*(2).

Tan, X., & Triggs, B. (2010). Enhanced local texture feature sets for face recognition under difficult lighting conditions. *IEEE Transactions on Image Processing*, *19*(6), 1635–1650.

Tao, D., Li, X., & Maybank, S. J. (2007). Negative samples analysis in relevance feedback. *IEEE Transactions on Knowledge and Data Engineering*, *19*(4), 568–580.

Tarawneh, A. S., Celik, C., Hassanat, A. B., & Chetverikov, D. (2020). Detailed investigation of deep features with sparse representation and dimensionality reduction in cbir: A comparative study. *Intelligent Data Analysis*, *24*(1), 47–68.

Tian, X., Jiao, L., Liu, X., & Zhang, X. (2014). Feature integration of EODH and color-SIFT: Application to image retrieval based on codebook. *Signal Processing: Image Communication*, *29*(4), 530–545.

Turan, C., & Lam, K.-M. (2018). Histogram-based local descriptors for facial expression recognition (FER): A comprehensive study. *Journal of Visual Communication and Image Representation*, *55*, 331–341.

Verma, M., & Raman, B. (2018). Local neighborhood difference pattern: A new feature descriptor for natural and texture image retrieval. *Multimedia Tools and Applications*, *77*(10), 11843–11866.

Vieira, G. S., Fonseca, A. U., & Soares, F. (2023). CBIR-ANR: A content-based image retrieval with accuracy noise reduction. *Software Impacts*, Article 100486.

Wang, J. Z., Li, J., & Wiederhold, G. (2001). Simplicity: Semantics-sensitive integrated matching for picture libraries. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *23*(9), 947–963.

Wang, J., Wang, L., Liu, X., Ren, Y., & Yuan, Y. (2018). Color-based image retrieval using proximity space theory. *Algorithms*, *11*(8), 115.

Wei, Z., & Liu, G.-H. (2020). Image retrieval using the intensity variation descriptor. *Mathematical Problems in Engineering*, *2020*.

Zeng, S., Huang, R., Wang, H., & Kang, Z. (2016). Image retrieval using spatiograms of colors quantized by gaussian mixture models. *Neurocomputing*, *171*, 673–684.

Zhou, J., Liu, X., Liu, W., & Gan, J. (2019). Image retrieval based on effective feature extraction and diffusion process. *Multimedia Tools and Applications*, *78*(5), 6163–6190.