

Article

Integration of Discrete Wavelet Transform, DBSCAN, and Classifiers for Efficient Content Based Image Retrieval

Muhammad Junaid Khalid ^{1,*}, Muhammad Irfan ², Tariq Ali ^{3,*}, Muqaddas Gull ⁴, Umar Draz ^{5,*}, Adam Glowacz ⁶, Maciej Sulowicz ^{7,*}, Arkadiusz Dziechciarz ⁷, Fahad Salem AlKahtani ² and Shafiq Hussain ⁵

¹ College of Engineering & Technology, University of Sargodha, Punjab 40100, Pakistan

² College of Engineering, Electrical Engineering Department, Najran University, Najran 61441, Saudi Arabia; irfan16.uetian@gmail.com (M.I.); fsalkahtani@nu.edu.sa (F.S.A.)

³ Computer Science Department, COMSATS University Islamabad, Sahiwal Campus, Sahiwal 57000, Pakistan

⁴ Department of CS & IT, University of Lahore, Lahore Road, Sargodha 40100, Pakistan; muqaddas.gull@sgd.uol.edu.pk

⁵ Department of Computer Science, University of Sahiwal, Sahiwal 57000, Pakistan; drshafiq@uosahiwal.edu.pk

⁶ Department of Automatic Control and Robotics, Faculty of Electrical Engineering, Automatics, Computer Science and Biomedical Engineering, AGH University of Science and Technology, al. A. Mickiewicza 30, 30-059 Kraków, Poland; adglow@agh.edu.pl

⁷ Faculty of Electrical and Computer Engineering, Cracow University of Technology, Warszawska 24 Str., 31-155 Cracow, Poland; arkadiusz.dziechciarz@gmail.com

* Correspondence: junaid.khalid@uos.edu.pk (M.J.K.); tariqali@cuisahiwal.edu.pk (T.A.); sheikhumar520@gmail.com (U.D.); maciej.sulowicz@pk.edu.pl (M.S.)

Received: 3 October 2020; Accepted: 6 November 2020; Published: 9 November 2020



Abstract: In the domain of computer vision, the efficient representation of an image feature vector for the retrieval of images remains a significant problem. Extensive research has been undertaken on Content-Based Image Retrieval (CBIR) using various descriptors, and machine learning algorithms with certain descriptors have significantly improved the performance of these systems. In this proposed research, a new scheme for CBIR was implemented to address the semantic gap issue and to form an efficient feature vector. This technique was based on the histogram formation of query and dataset images. The auto-correlogram of the images was computed w.r.t RGB format, followed by a moment's extraction. To form efficient feature vectors, Discrete Wavelet Transform (DWT) in a multi-resolution framework was applied. A codebook was formed using a density-based clustering approach known as Density-Based Spatial Clustering of Applications with Noise (DBSCAN). The similarity index was computed using the Euclidean distance between the feature vector of the query image and the dataset images. Different classifiers, like Support Vector (SVM), K-Nearest Neighbor (KNN), and Decision Tree, were used for the classification of images. The set experiment was performed on three publicly available datasets, and the performance of the proposed framework was compared with another state of the proposed frameworks which have had a positive performance in terms of accuracy.

Keywords: content-based image retrieval (CBIR); discrete wavelet transform (DWT); features extraction; support vector machine (SVM); decision tree; performance evaluation

1. Introduction

In multimedia research, with the incessant evolution in facet and mid-level vision, the retrieval of the desired image efficiently and accurately from a dataset has been in the domain of Content-Based Image Retrieval (CBIR). The efficient representation of image features and cutting of the semantic gap remains a significant problem in Image Retrieval (IR) tasks. The proposed approach is the interfusion of image features in a multilayer exploration context for CBIR, that has been influenced by wavelets. In the present research, the query images Discrete Wavelet Transform (DWT) coefficients computation was performed. Clustering, prospectively implemented by Density-Based Spatial Clustering of Applications with Noise (DBSCAN) and the Support Vector (SVM) classifier, was used for the classification of images. The similarity index between the test image and the query image can potentially be performed through the Euclidean distance. Extensive experiments were conducted on Corel datasets, and accomplishments of the proposed methodology were measured using certain performance evaluation matrices, like precision and recall. The proposed approach was implemented using matrix laboratory (MATLAB).

CBIR systems search large databases for images that match an image query. CBIR systems automatically search, retrieve, or index image databases. With the evolution of digital technologies, enormous volumes of spatial information, data analysis, and medical images are instantly available in the multimedia form [1]. The image content searching is based on two approaches: CBIR and TBIR. In the CBIR organization, the user enters a search query, then low features (e.g., text, shape, and color) of images are excerpted using feature extractors or visual content descriptors. The similarity index between the image search query and the retrieved image is found through the distance that is measured through the analysis of search and query image feature vectors of the dataset images. There is a significant gain in the similarity minimization gap between the search query and extracted image via CBIR, which further enables researchers to explore areas like browsing, remote sensing, crime avoidance, and ancient research.

The image texture, shape, and color features play a key role in CBIR systems for the retrieval of images. Certain vital color features, such as histograms, correlograms, and Hue Saturation Value (HSV) color techniques are used in CBIR systems. The image shape features are excerpted using several image descriptors, such as the moment invariant, Fourier, chain code, and curvature-based. The descriptors describe the color, shape, and text, which are the elementary properties of an image. The image retrieval is performed by the system based on extracted features [2,3]. The merger of more than one feature is also used to improve the CBIR and facial expression system performance [4]. The extraction of the texture feature in the spatial domain can be performed through the Local Binary Pattern (LBP) [5], Completed Volume Local Binary Pattern (CVLBP) [6], Volume Local Binary Pattern (VLBP) [7], Noise Invariant Structure Pattern (NISP) [8], Local Ternary Pattern (LTP) [9], Scale Invariant Feature Transform (SIFT), and Speeded-up Robust Features (SURF), among others algorithm which are utilizing it in different case study [10].

1.1. Problem Statement

There are numerous diverse challenges in CBIR, including efficient feature representation for the images' visual contents, the denotation rift between low and high-level attributes and high-level image semantics, loss of spatial information, automatic image annotation, manual image annotation, and image ranking problems. In this research, efficient feature representation of images and reducing the semantic gap are addressed. In CBIR, image descriptors describe the fundamental properties, which are essentially color, text, and shape. Wavelets contain the time and frequency values, and the features enable DWT, which is more beneficial for CBIR. The significance of this approach is to detect the features of an image that are left undetected at one resolution level. Clustering is the approach of combining a set of entities, such that the entities in the alike cluster are more alike to one another than to the entities present in different clusters. The relationship between two data entities is echoed through their likeness or similarity. This approach is typically used to study and explore data mining

and utilizes machine learning, information retrieval, computer vision, analysis of images, pattern recognition, and so forth. DBSCAN, based on the density-based notion of the cluster, discovers clusters of arbitrary shape in spatial databases with noise. The core idea of the grouping of points w.r.t. high-density and outliers show that these points exist alone in low-density regions. The machine learning-based classifier SVM and Decision Tree give better results for the classification of images. SVM has been extensively utilized for classification and other various learning purposes and achieves notable accuracy in results with less computational cost [11]. The data points in the classes can be separated based on the chosen hyperplane [12]. The SVM classifier is constructed based on the kernel type, gamma, and c value [13]. There are various kernel functions available which can be used in the models of SVM. These include the linear kernel, polynomial kernel, sigmoid kernel, and radial basis function (RBF) kernel and some other work is done [14]. The SVM forms nonlinear margins by using the kernel trick [15]. The kernel trick process includes the dual problem, summarization of the inner product, and kernel function definition. Finally, the most suitable images are retrieved and placed, bestowing their similarity score. This technique is used for the retrieval of similar images from image datasets after sending the image search query. In our proposed approach, the similarity is computed using the Euclidean distance, due to its positive performance [16,17]. The distance is measured between the query image factor vector and the database image feature vector.

This research proposes a CBIR system to elucidate the efficient feature representation and semantic gap problems linked with CBIR. The purpose of this study is to explore the presented approach's effectiveness while representing image features and bridging the semantic gap between them. The set experiment performance was tested on a publicly available benchmark dataset, explicitly the Wang datasets [18,19]. The fulfillment of the presented approach was assessed using precision (P) and recall (R), by forming a confusion matrix. The P-value specifies the retrieved relevant outcomes; it deals with the accuracy of a system [20]. The R-value specifies relevant outcomes retrieved; it deals with the coverage of the appropriate information of a system along with fault tolerance same as in [21].

An image contains edges, shapes, textures, colors, and interest points; these features are the basis for the retrieval of visually similar images. The information is considered fundamental in image analysis applications like identification, classification, recognition, and restoring. The image features are divided into low and high-level features [22,23]. The CBIR system forms feature vectors based on the extracted features of an image. The feature vector database is maintained to check the similarity matching score [24]. The image features also include named descriptors, which are used to extract image features for the construction of feature vectors that may be local or global [25,26]. The local descriptors, like the LBP, Local Intensity Order Pattern (LIOP), LTP, SIFT, and SURF, identify the image features based on color, shapes, texture, and spatial details, while global descriptors may include the Histogram of Oriented Gradient (HOG), Co-HOG, and Moment. The color information presented in an image is crucial and extensively utilized in CBIR approaches. The color is the wavelength-dependent approach and is said to be a core feature of images as it cannot be altered based on the orientation and size of the object. It has been widely used for image retrieval and recognition of objects [1,27,28].

The texture information embedded in images illustrates the presence of visual patterns with uniform invariant properties. Texture features [29] play a significant role in similarity matching and image retrieval. In most cases, texture features are integrated with a color feature to improve the performance of image retrieving systems. Texture features are further categorized into four classes: Geometrical, statistical, model, and spectral-based [30]. In texture analysis there are various approaches, like LBP [31], LIOP [32], Gabor filtering [33], gray-level co-occurrence [34], and Markov random field [35]. The shape features are important for CBIR systems; based on shape information, humans can simply differentiate objects present in an image. These are categorized into two sections: Boundary-based and region-based. The basic shape features are eccentricity, gravity center, area, and orientation. Due to segmentation accuracy problems, it is difficult to utilize these features as compared to color and texture features. In Author's work [36], eccentricity and image orientation are used to excerpt shape information from images, and in Author's work [37], a gross region descriptor is

constructed that is based on second-order moments and area. Along with the color, texture, and shape features, spatial features may also be used. These provide spatial location information. The spatial features are identified in the image top and bottom regions [38,39].

1.2. Literature Review

A range of research has been presented on the performance enhancement of CBIR systems [40]. Alzu'bi et al. [39] proposed bilinear Convolution Neural Network (CNN)-based framing using dual CNN in parallel for image attributes extraction without prior knowledge of image content semantic meta-data. These attributes are excerpted using the activation of convolutional layers and are afterward squeezed for low-extent illustration using the root bilinear compact pooling approach. The approach of CRB-CNN (M) and CRB-CNN (16) has been tuned to perform CBIR tasks. Publicly available datasets are used for the experiments, then distance scores are measured by the Manhattan, Euclidean, and Cityblock approaches. The retrieval accuracy of the techniques was measured through Mean Average Precision (mAP). The results revealed CRB-CNN shows high efficiency in learning and pre-training of the architectural deep model while measuring the distance-vector Euclidean distance overtakes Manhattan and Cityblock. Finally, the training was performed without annotations, meta-data, and other content tags, endorsing the high capability of the approach and giving a high retrieval performance.

Srivastava et al. [31] proposed a combo feature-based multiresolution exploration strategy for the retrieval of an image via the merging of texture and shape features. This approach combined Legendre Moments and Local binary patterns at various resolutions of wavelet disintegration. The performance was measured through recall and precision, and the analysis showed the proposed approach outperformed other available methodologies. However, low retrieval accuracy occurred when using large datasets, and results validity could be further enhanced by using LDP and LTP. Cui et al. [41] described a unique merging method based on text and visual applicability learning that collimates textual appropriateness from image labels, then integrates both visual and textual relevance for CBIR. The labeling/tagging completion is implemented to impart the absent tags and corrects datasets images with noisy labels. The results verified the utility of the proposed approach, however, the authors intend to improve the computational power for labeling of compressed sensing and matrix completion, giving a potential way out for the representation schemes of semantic modeling from related labels.

Zhou et al. [42] investigated the potential of integrating SIFT and CNN features for effective and efficient IR. An embedding algorithm based on collaborative indexing was proposed for the exaltation of SIFT and CNN attribute index files. The proposed conjoint directory planting algorithm for the retrieval of images gained significant accuracy in retrieval, and with less memory cost in contrast to the present advanced techniques. Jin et al. [43] studied the distance metric/similarity indexing method, which depends on Cost-Sensitive Learning (CSL). The authors replaced the conventional kernel function with large margin distribution learning machines. The fusion of learning distribution machines (LDM) and CSL was applied to tasks of CBIR for enactment classification. The results showed the proposed model achieved a good classification performance, with the least misclassification cost. Tzelepi et al. [44] proposed a model reskilling approach for acquiring efficient convolutional representations for CBIR. A deep CNN model, through the beginnings of the operating max-pooling layers of convolution, was used to obtain feature representations and, consequently, the author modified and retrained the network to yield effective image descriptors that enhanced memory requisites and retrieval performance. The analysis results indicated that the proposed methodology achieved good results in terms of retrieval tasks when compared with other retrieval techniques based on CNN and traditional feature-centered techniques.

Rana et al. [1] compared their proposed work with the existing seven approaches through ascertaining statistical metrics across databases of five images, through feature extraction through a hybrid CBIR method, based on color moments, ranklet transformation, and moment invariants. The ensuing feature extraction process similarity quantification was appraised using X2 and Chi-Square,

and the Euclidean, Canberra, and Manhattan distances were analyzed. The analysis showed Euclidean achieved improved performance evaluation parameters (precision and recall) values. The CRM approach combined with three invariant features was then performed on five various datasets, attaining substantial precision values as compared to current approaches.

Pavithra et al. [45] proposed a hybrid framework for CBIR to address the retrieval system accuracy issues using image shallow attributes. The presented approach took relevant images from datasets through moment information based on color. Then, for the mining of image edge and texture attributes, LBP and Canny edge detectors were utilized. The tendered technique had good outcomes in terms of precision and recall as compared to modern existing techniques when applied to the publicly available dataset. To ensure data privacy in a cloud environment while retrieving an image, private information should be cyphered before being implanted in a cloud environment. The encryption process causes complexities in the retrieval of images, and Xu et al. [46] proposed a privacy-preserving CBIR approach based on orthogonal decomposition. The implementation of orthogonal transform results in the decomposition of the image into dual orthogonal element fields. Then, the extraction of features and encryption was performed distinctly, resulting in a cloud provider that can extract an image from the cyphered catalog directly while preserving information confidentiality.

Hussain et al. [47] proposed a MapReduced and Spark-based model system for quick image indexing and retrieval. They further integrated the K-Nearest Neighbors classifier for image retrieval. Alsmadi et al. [48] proposed another CBIR framework based upon a metaheuristic approach, with a novel neutrosophic clustering and DWT, to extract better image feature vectors, which achieved good results on Corel-1000 datasets in terms of precision recall. Garg et al. [49] proposed another CBIR framework with a core focus on efficient feature extraction with minimal multi-feature involvement. The GLCM and LBP descriptors with swarm particle optimization approaches can also be used for better image retrieval. The proposed framework also gives good results on the Corel-1000 dataset. Moreover, the validity of the results can further be tested on large datasets.

Multiple types of research have been undertaken, however, image ranking, image annotation, feature vector formation, and the semantic gap effect on retrieval accuracy remain open for further study. All images feature vectors are formed by using image descriptors, then these feature vectors are stored in a database. The distance measurements between the query and dataset image feature vectors are computed mainly using Cityblock, Manhattan, and Euclidean, among others [50,51]. The most suitable images are retrieved and placed based on their similarity scores. Statistical measures are implemented to check the validity and performance of the proposed work. Finally, the retrieval performance of the suggested approach has been measured concerning precision [20], recall, and the confusion matrix. The remaining part of the research is organized as follows: Section 2 outlines the research method, while the results and analysis are given in Section 3. Finally, Section 4 concludes the research.

2. Materials and Methods

The proposed approach consists of three stages. In the first stage, image histograms were formed, and after that RGB space, auto-correlograms were calculated. The mean and standard deviation moments were formed according to the color space. Then, at different decomposition levels, DWT coefficients were computed. In the next stage, a codebook was generated using DBSCAN, then a histogram of retrieved images was formed. The classifier was used for better classification of the images. The distance metric was formed, and similar images related to the query image were retrieved. Finally, a confusion matrix was generated to inspect the accuracy of the proposed CBIR technique. The schema chart of the proposed methodology is shown in Figure 1. The proposed approach is implemented using MATLAB.

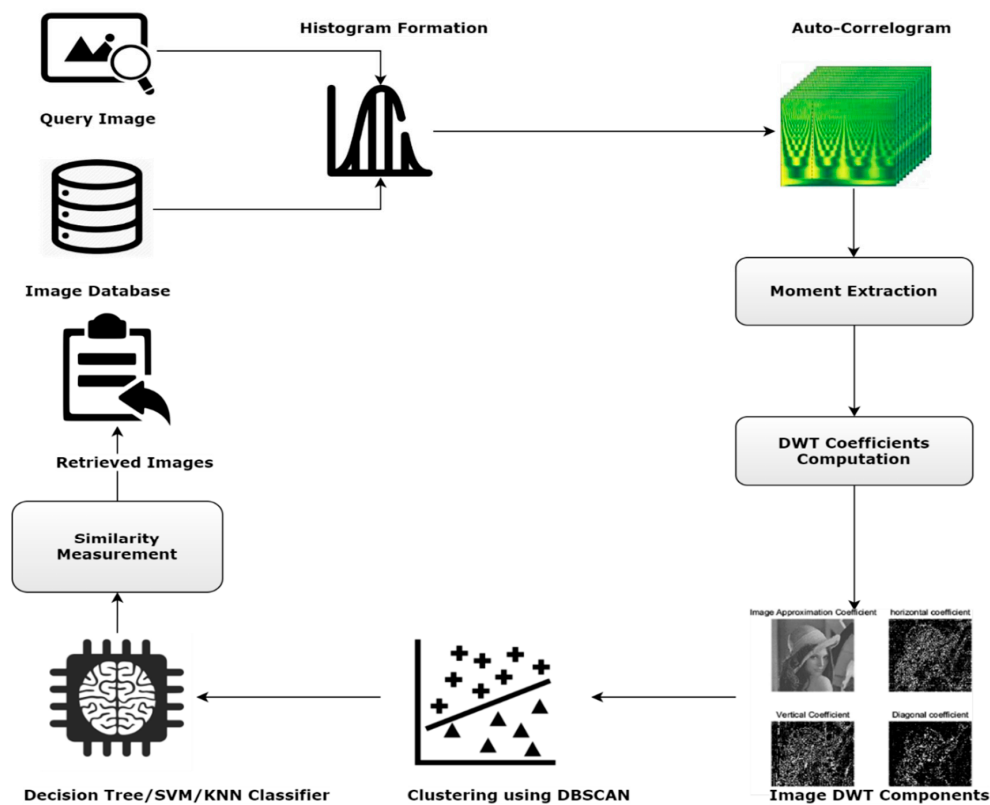


Figure 1. Proposed approach for Content-Based Image Retrieval (CBIR).

First of all, color histograms and auto-correlograms were formed from dataset and query images. The standard deviation and mean moments of the color space were computed. The 2D query image DWT coefficients were computed, and the 1D DWT transform can be enhanced for images with two dimensions. For two dimensional images, the ascending function was $\varphi(w,v)$ and there are three-dimensional wavelets, which were sensitive directionally and gauge variations of images horizontally $\psi^H(w,v)$, vertically $\psi^V(w,v)$, and diagonally $\psi^D(w,v)$. The proposed methodology features the wavelet approach due to its multi-resolution analysis feature. Once the feature descriptors were computed, the feature correspondences between different images can be automatically quantized under a similarity measure, such as the Euclidean distance. The grouping of similar images with similar features was done through DBSCAN. The goal of similarity measurement was to extract visually similar images to the query image from the database. Finally, a deep analysis was necessary to study various CBIR aspects of efficiency, the similarity score, and the proposed system performance.

2.1. Discrete Wavelet Transform

The image descriptors signify the correlative attributes of an image. In CBIR, image descriptors describe visual attributes of the content of an image. The descriptors describe the fundamental properties; essentially, color, text, and shape. The wavelets were well localized in time and frequency. Wavelets have various sizes and shapes, like Morlet, Daubechies, Coiflets, Biorthogonal, Mexican Hat, and Symlets as shown in Figure 2.

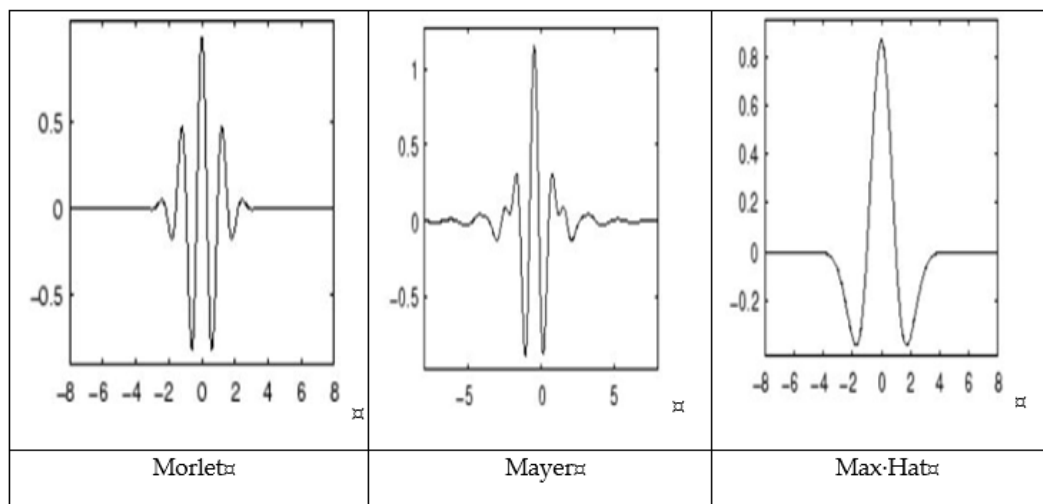


Figure 2. Wavelet types representation.

In this proposed research, the Gabor wavelet is used as a mother wavelet. Amid several bases of wavelet, Gabor provides the optimal resolution in terms of both the time and frequency domains, and the Gabor wavelet extracts optimal features compared to other bases [52]. There exist two variants of wavelets, called the Continuous Wavelet Transform (CWT) and DWT. DWT represents a lot of naturally occurring signals and images with fewer coefficients, which therefore consume less memory. This approach is useful when image datasets contain objects with high and low-level contrast, as well as the size of objects varying from small to large or vice-versa [41]. In DWT, the first-level decomposition of a given signal or image is passed through the High Pass Filter (HPF) and Low Pass Filter (LPF).

The Nyquist criterion was applied to sampling signals, and half of the samples were discarded. The low pass sub-bands were also named as approximate coefficients or approximate levels, and high pass sub-bands named as detail coefficients or the detail level.

The mathematical model of the expansion function of wavelet series $f(w)L_2(R)$ compared to the wavelet $\psi(w)$ and function for scaling (w) is illustrated by Equation (1).

$$f(x) = \sum_k C_{j_0}(k) \varphi_{j_0,k}(w) + \sum_{j=j_0} d_j(k) \psi_{j,k}(w) \quad (1)$$

where j_0 denotes arbitrary scale, $c_{j_0}(k)$ denotes approximation coefficients, and $d_j(k)$ denotes detailed coefficients. Equation (1) defines the growth in the wavelet series; it extracts a sequence of coefficients from a continuous variable function. The resultant coefficients are formed due to the growths of function consequences in order of numbers and are called DWT coefficients of function $f(x)$. In this scenario, the growth of Equation (2) becomes a DWT transform pair, where j_0 denotes the arbitrary preliminary measure integer index k , $f(w)$, and $\psi_{j,k}(w)$ and $\varphi_{j_0,k}(w)$ are the variables for the distance 1, 2, 3, ..., $M-1$ function.

$$W_\varphi(j_0, k) = \frac{1}{\sqrt{M}} \sum_x f(w) \psi_{j,k}(w) \quad (2)$$

$$W_\psi(j, k) = \frac{1}{\sqrt{M}} \sum_x f(w) \psi_{j,k}(w) \text{ for } j > j_0 \quad (3)$$

$$f(w) = \frac{1}{\sqrt{M}} \sum_w f(w) \varphi_{j_0,k}(w) + \frac{1}{\sqrt{M}} \sum_w f(w) \psi_{j,k}(w) \quad (4)$$

$$W_\varphi(j_0, m, n) = \frac{1}{\sqrt{MN}} \sum_{w=0}^{M-1} \sum_{v=0}^{N-1} f(w, v) \varphi_{j_0,m,n}(w, v) \quad (5)$$

where j_o denotes the arbitrary starting scale, and $W_{\psi}^i(j_o, m, n)$ denotes the coefficients' horizontal, vertical, and diagonal details. The 1-D DWT transform can be enhanced for images with two dimensions. For two dimensional images, the ascending function is $\varphi(w, v)$. There exist three-dimensional wavelets that are said to be sensitive directionally, and they measure variations for gray-scale images horizontally $\psi^H(w, v)$, vertically $\psi^V(w, v)$, and diagonally $\psi^D(w, v)$. The $f(x, y)$ function for DWT, if the images have $M \times N$ size, is as in Equations (5) and (6).

for $j \geq j_o$

$$W_{\psi}^i(j, m, n) = \frac{1}{\sqrt{MN}} \sum_{w=0}^{M-1} \sum_{v=0}^{N-1} f(w, v) \psi_{j,m,n}^i(w, v) \quad (6)$$

The DWT is centered on the coding mechanism of sub-band; it can further be utilized to relente a rapid reckoning of the wavelet transform. DWT fragments the information into two types of coefficient: The detail and approximation coefficients. The detail coefficient gives directional information and independently forms a feature vector. This feature vector is used for the retrieving of apparently related images. Another important feature of DWT is the examination of images at different frequencies and multiple resolutions. The analysis at different resolutions produces another detailed horizontal, vertical, and diagonal coefficient. The core advantage of this feature is while the feature may remain unidentified at one level, it may be identified at another level.

2.2. Dictionary Learning

The analysis of clusters remains problematic in data science. Scientists can utilize various clustering approaches for the comprehension of server collapse, genes grouping for a homogenous indication sequence, demand-supply gap, summarization of news, trend identification, and so forth. By definition, clustering is an approach to combining a set of entities, hence the entities in the alike cluster are more alike to one another than to the entities present in different clusters. The Density-Based Clustering algorithm is fundamental to the search for non-linear density-based structures. DBSCAN [53] is the most frequently used algorithm. It uses the concept of reachability connectivity based upon density. On the other hand, K-means is the simplest unsupervised learning technique that solves a known grouping problem. This algorithm is based on two parameters: ϵ and MinPts (N_p) [53]. This parameter defines the radius for the neighboring data point p . The criteria state that if the distance between these two points is less than or equivalent to " ϵ ", then these are pointed as neighbors. In the case of " ϵ " being too small, then a large portion of the data points will be measured as outliers. On the other hand, if the " ϵ " value is too large, then most data points will be merged into the same cluster. " ϵ " is calculated through the k-distance graph defined in Equation (7), while the MinPts (N_p) parameter denotes the minimum neighboring points within the defined " ϵ " radius. For a big dataset, a large value for this parameter should be specified. The minimum number of MinPts can be derived w.r.t. the number of dimensions (D) in the dataset. In general, this criterion for the minimum number of points should be $\text{MinPts} \geq D + 1$, or should be at least three.

$$N_{\epsilon(p)} = \{q \text{ in data set } D \mid \text{dist}(p, q) \leq \epsilon\} \quad (7)$$

DBSCAN holds three sorts of data points: Core, border, and outlier. The DBSCAN cluster contains the core and border points, and its description is as follows:

- The " P " dataset is given with the parameter radius (ϵ) and MinPts
- The cluster named C is a subset of dataset " P ", which satisfies two criteria:

(1). Maximality

$$\forall p, q \text{ if } p \in C \text{ and if } q \text{ is density reachable from } p, \text{ then also } q \in C$$

(2). Connectivity

$$\forall p, q \in C, p \text{ and } q \text{ are density connected}$$

2.3. Similarity Measurement

In similarity measurement score computation, first of all, the individual query image is instigated in the CBIR system. This similarity is indexed to fetch visually similar images from the image datasets concerning the query image. In our proposed approach, it is done by computing the Euclidean distance [16,17] between the query image factor vector and the database image feature vector. If the histogram of the query image coefficients are $h_Q = (h_{Q1}, h_{Q2}, h_{Q3}, \dots, h_{Qn})$ and the dataset image coefficients are $h_{DB} = (h_{DB1}, h_{DB2}, h_{DB3}, \dots, h_{DBn})$, the Euclidean distance is calculated using Equation (8). The values of the precision and recall are computed to draw a confusion matrix that portrays system performance.

$$D(h_Q, h_{DB}) = \sqrt{\sum_{i=1}^n (h_{Qi} - h_{DBi})^2} \quad (8)$$

2.4. Classification

Machine Learning (ML) combines with multiple statistical approaches in various ways, and many techniques have been developed to perform ML tasks, like SVM, Decision Tree, Naïve Bayes, Neural Networks, and Random Forest [11]. These approaches have been used to perform multiple tasks, including image retrieval, speech recognition, pattern recognition, texture classification, and biometric identification. SVM has been extensively utilized for classification and other various learning purposes and gives good accuracy in results with less computational cost [11]. The training data in SVM is represented by points in space that are dissected into classes, with an obvious gap between classes. The data points in the classes are separated based on the chosen hyperplane [11,12]. The main goal is to choose a hyperplane with a maximum margin between the data points of different classes. The data points on any side of the hyperplane are associated with different classes, and the hyperplane dimension is dependent on whether features exist. If two features are involved as input so the hyper plane is represented by drawing a single line, for three features as input, a two-dimensional plane is used. The representation of more than three features as input causes complexities for drawing them on a hyperplane. The data points are called support vectors, and are near the hyperplane and affect the hyperplane locus and orientation. The margins of the classifiers are maximized by involving these support vectors. The position of the hyperplane is affected by the deletion of any support-vector. Based on this support-vector, an SVM is formed.

The SVM approach works efficiently for multi-dimension systems. The output of a linear function is defined in SVM; if the output value is more than 1 it is identified as one, and for a -1 output value another class if defined. The value of -1 to 1 is used to define margins [12,54]. The SVM classifier is constructed based on the kernel type, gamma, and c value [13].

3. Results and Discussion

The experiment utilized various benchmark datasets, like Corel-1000, Corel-1500, and Corel-5K. The performance evaluation matrices were computed at various resolution levels of the selected dataset. The set experiment was performed using the proposed image retrieval technique to measure the retrieval performance of the system on the datasets. The size of each image in Corel-1000 is 256 by 384 or 384 by 256 [18], while it is 126 by 187 or 187 by 126 in Corel-1500 and Corel-5k.

The performance of the suggested approach was assessed concerning precision and recall. The experiment was performed after rescaling Corel-1000, Corel-1500, and Corel-5K images at 256 by 384 and 128 by 85, respectively. The images were transformed into a gray-scale scheme to increase computational performance. At the start of the experiment, the images' DWT coefficients were computed at multi-resolution for the efficient construction of feature vectors through Gabor wavelet

transform. The DWT produces three matrices that contain diagonal, vertical, and horizontal details. The DBSCAN was used to search for non-linear density-based structures. DBSCAN [54] is the most frequently used algorithm. It uses the concept of reachability connectivity, based upon density. The technique charts, in a simple way, to categorize a certain set of data for a certain number of groups. This approach clustered similar images related to a query image. Then, the similarity index was measured on these three coefficients using the Euclidean, Manhattan, and Cityblock distance measures. The similarity indexes of three coefficients sets were then combined to show similar images. The precision and recall values were computed for performance evaluation. The values were then used with the SVM, K-Nearest Neighbor (KNN), and Decision Tree classifier to show the accuracy of the proposed framework. The higher the value of these performance evaluation matrices, the better the system accuracy. The efficiency of the proposed framework was computed by using the SVM [54], KNN, and Decision Tree classifier.

The proposed framework was tested on Wang's datasets Corel-1000, Corel-1500, and Corel-5K. The framework retrieved twenty similar images to compute accuracy. The findings of the proposed framework showed that the Decision Tree framework more accurately classified the images, resulting in a better retrieval accuracy as compared to SVM and KNN. Circumspect data training of KNN is required and the distance measuring is critical, as this affects the accuracy of the classification. The further complexities with KNN are around data points that exist in low dimensional space, and higher dimensions create complexities while estimating nearest neighbors as compared to SVM. There are a lot of data points that exist in higher space; these are also responsible for the performance of the classifier. Along with this, the Decision Tree medium complexity classifier was also implemented with twenty splits to compare its significance over other classifiers. The maximum deviance reduction was used as a splitting criterion, and with ten maximum per node surrogates. The dataset images were classified using the Linear SVM, Quadratic SVM, Cubic SVM, Decision Tree, Weighted, and Cosine KNN. The proposed framework gives a good performance when classified with a Decision Tree with medium complexity. Decision Tree is non-parametric, and the input space is split into hyper-rectangles w.r.t. the set objective. The splitting nodes and tree pruning criterion are given in Equation (9). The observations that exist in node "t" is to class "s" as the mainstream class exists in node "t". The maximum deviance reduction was estimated using Equation (10). The deviance reduction splitting criterion was differentiable and easy to use for the optimization of numerical data, meaning it gives better accuracy as compared to other tested classifiers. The per true class results of the proposed methods with different classifier variants are shown in Figure 3.

$$O_{ts} = \frac{1}{N_t} \sum_{x_i \in R_t} R(J_i = s, t) \quad (9)$$

$$- \sum_{s=1}^S O_{ts} \log O_{ts} \quad (10)$$

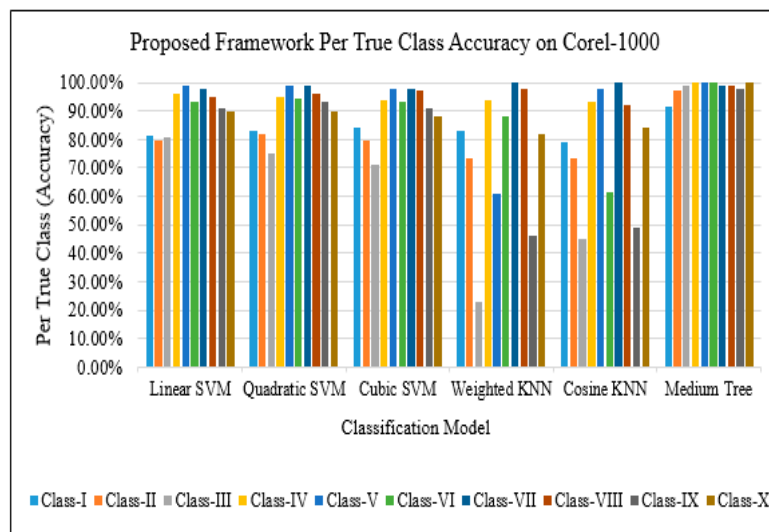


Figure 3. Proposed framework per True Class Accuracy on Corel-1000 classes.

The overall accuracy of each classifier on the Corel-1000, Corel-1500, and Corel-5k datasets is shown in Figure 4. Further, feature selection is very important in terms of predicting the accuracy of models with outliers. Hence, the accuracy of the framework is better with Decision Tree. The framework retrieved twenty of the best similar images to compute accuracy. The performance of the proposed framework was compared with another state of the proposed frameworks, tabulated in the appendices. We compared the performance of the proposed framework by considering the twenty best similar images of the proposed and other methods. The P-value specifies the retrieved relevant outcomes; it deals with the accuracy of a system [20]. The results in Table 1 show the results of the Corel-1000, Corel-1500, and Corel-5k datasets.

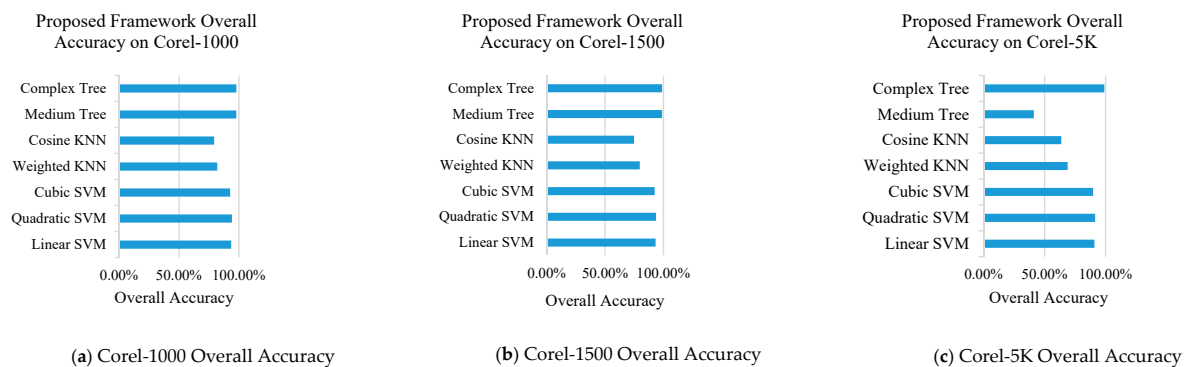


Figure 4. Proposed framework overall accuracy on Corel-1500.

Table 1. Performance comparison of the proposed approach on Corel-1000, Corel-1500, and Corel-5K as compared to other retrieval techniques in terms of accuracy (Average Precision (AP)).

Corel-1k	Lin et al. [55]	Irtaza et al. [56]	Wang et al. [57]	Walia et al. [58]	Walia et al. [59]	Pavithra et al. [45]	Proposed Approach
Class-I	55.5	53	80.5	41.25	73	81	91.6
Class-II	66	46	56	71	39.25	66	97.1
Class-III	53.5	59	48	46.75	46.25	78.75	99
Class-IV	84	73	70.5	59.25	82.5	96.25	100
Class-V	98.25	99.75	100	99.5	98	100	100
Class-VI	63.75	51	53.75	62	59.25	70.75	99
Class-VII	88.5	76.75	93	80.5	86	95.75	99
Class-VIII	87.25	70.25	89	68.75	89.75	98.75	99
Class-IX	48.75	62.5	52	69	41.75	67.75	98
Class-X	68.75	70.75	62.25	29.25	53.45	77.25	100
Overall AP	71.425%	66.20%	70.50%	62.725%	66.92%	83.225%	98.3%
Corel-1500 AP	-	-	-	-	-	-	98.9%
Corel-5K AP	58.753	53.238	60.58	49.67	56.72	68.605	98.8%

The results show that the proposed framework gives 98.3% average accuracy on Corel-1000, 98.9% average accuracy on Corel-1500, and 98.8% accuracy on Corel-5k. From ten classes of Corel-1000, the given framework accomplished better results in every class. Given the results in the appendices, and the Pavithra et al. [46] combined moments, LBP, and edge features, the proposed method's average accuracy is 83.225%. However, the proposed approach achieves a better performance as compared to other states of the art approaches.

4. Conclusions

In this paper, we aimed to further improve the performance of the proposed approach by conducting more discerning research with larger datasets and better feature vector formation. Therefore, vigorous customization of the training framework was contemplated, leading towards the essential construction of processing data at different training levels. Additionally, the primary approach for initial image description and classification can be enhanced using deep learning. In this research, a novel approach in which a feature vector was formed, then visually similar images retrieved based on the query and database image feature vector similarity index, was tested. The amalgamation of multiple image feature excerpts intermixed the benefits of multiresolution analysis, in contrast to a single feature. The significance of using DWT was the detection of features of an image that are left undetected at one resolution level. Further, the presented approach extracted shape features from a texture feature of images at multiresolution with DBSCAN clustering. As it gathers more similar images in groups, it does not require the specification of cluster numbers. Finally, different classifiers were tested to classify retrieved images and to check the accuracy of the system. The accuracy of the presented approach was tested by generating a confusion matrix and was compared in terms of average precision.

The results showed that the presented approach gives good results as compared to other states of the art approaches. However, the presented method may give low retrieval accuracy when used on bulky datasets like GHM-10K and Corel-10K. To improve the performance of the system, future work should focus on the use of segmentation for the extraction of objects from images and moments to form the feature vector. Further, deep learning algorithms such as deep CNN and transfer learning may be considered for better image classification, which helps to improve retrieval accuracy of the CBIR system.

Author Contributions: M.J.K. and T.A. proposed the research conceptualization and methodology. The technical and theoretical framework was prepared by T.A., M.G. The technical review and improvement were performed by M.I., U.D., and A.G. The overall technical support, guidance, source management, and project administration were done by M.S. and A.D. The editing and final proofread was done by F.S.A., S.H., A.D., and M.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by project NAWA to pay the APC of the journal.

Acknowledgments: The authors acknowledge the Ministry of Education and the Deanship of Scientific Research, Najran University, Kingdom of Saudi Arabia, under code number NU/ESCI/19/001.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Rana, S.P.; Dey, M.; Siarry, P. Boosting content based image retrieval performance through integration of parametric & nonparametric approaches. *J. Vis. Commun. Image Represent.* **2019**, *58*, 205–219.
2. Tyagi, V. Content-based image retrieval techniques: A review. In *Content-Based Image Retrieval*; Springer: Singapore, 2017; pp. 29–48.
3. Memon, M.H.; Li, J.; Memon, I.; Arain, Q.A. GEO matching regions: Multiple regions of interests using content based image retrieval based on relative locations. *Multimed. Tools Appl.* **2017**, *76*, 15377–15411. [\[CrossRef\]](#)
4. Ali, G.; Ali, A.; Ali, F.; Draz, U.; Majeed, F.; Yasin, S.; Haider, N. Artificial neural network based ensemble approach for multicultural facial expressions analysis. *IEEE Access.* **2020**, *8*, 134950–134963. [\[CrossRef\]](#)
5. Guo, Z.; Zhang, L.; Zhang, D. A completed modeling of local binary pattern operator for texture classification. *IEEE Trans. Image Process.* **2010**, *19*, 1657–1663. [\[PubMed\]](#)
6. Tiwari, D.; Tyagi, V. Dynamic Texture Recognition Based on Completed Volume Local Binary Pattern. *Multidimens. Syst. Signal Process.* **2016**, *27*, 563–575. [\[CrossRef\]](#)
7. Zhao, G.; Pietikäinen, M. Dynamic texture recognition using volume local binary patterns. In *Dynamical Vision*; Springer: Berlin/Heidelberg, Germany, 2006; pp. 165–177.
8. Shrivastava, N.; Tyagi, V. Noise-invariant structure pattern for image texture classification and retrieval. *Multimed. Tools Appl.* **2016**, *75*, 10887–10906. [\[CrossRef\]](#)
9. Agarwal, M.; Singhal, A.; Lall, B. Multi-channel local ternary pattern for content-based image retrieval. *Pattern Anal. Appl.* **2019**, *22*, 1585–1596. [\[CrossRef\]](#)
10. Ali, T.; Noureen, J.; Draz, U.; Shaf, A.; Yasin, S.; Ayaz, M. Participants Ranking Algorithm for Crowdsensing in Mobile Communication. *EAI Endorsed Trans. Scalable Inf. Syst.* **2018**, *5*, 154467. [\[CrossRef\]](#)
11. Hussain, A.; Draz, U.; Ali, T.; Tariq, S.; Irfan, M.; Glowacz, A.; Rahman, S. Waste Management and Prediction of Air Pollutants Using IoT and Machine Learning Approach. *Energies* **2020**, *13*, 3930. [\[CrossRef\]](#)
12. Burges, C.J. A tutorial on support vector machines for pattern recognition. *Data Min. Knowl. Discov.* **1998**, *2*, 121–167. [\[CrossRef\]](#)
13. Shawe-Taylor, J.; Sun, S. A review of optimization methodologies in support vector machines. *Neurocomputing* **2011**, *74*, 3609–3618. [\[CrossRef\]](#)
14. Qaisar, Z.H.; Irfan, M.; Ali, T.; Ahmad, A.; Ali, G.; Glowacz, A.; Glowacz, W.; Caesarendra, W.; Mashrafi, A.M.; Draz, U.; et al. Effective beamforming technique amid optimal value for wireless communication. *Electronics* **2020**, *9*, 1869. [\[CrossRef\]](#)
15. Schölkopf, B.; Burges, C.J.; Smola, A.J. *Advances in Kernel Methods: Support Vector Learning*; MIT Press: Cambridge, MA, USA, 1999.
16. Khosla, G.; Rajpal, N.; Singh, J. Evaluation of Euclidean and Manhattan metrics in content based image retrieval system. In Proceedings of the 2015 2nd International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, India, 11–13 March 2015; IEEE: Piscataway, NJ, USA, 2015.
17. Nalini, P.; Malleswari, B. An Empirical Study and Comparative Analysis of Content Based Image Retrieval (CBIR) Techniques with Various Similarity Measures. In Proceedings of the 3rd International Conference on Electrical, Electronics, Engineering Trends, Communication, Optimization and Sciences (EEECOS), Tadepalligudem, India, 1–2 June 2016.
18. Jia Li, J.Z.W. Corel-1000. Available online: <http://wang.ist.psu.edu/docs/related/> (accessed on 2 February 2020).
19. Li, J.; Wang, J.Z. Real-time Computerized Annotation of Pictures. *IEEE Trans. Pattern Anal. Mach. Intell.* **2008**, *30*, 985–1002.
20. Zhang, E.; Zhang, Y. Precision. In *Encyclopedia of Database Systems*; Liu, L., Özsu, M.T., Eds.; Springer: Boston, MA, USA, 2009; p. 2126.

21. Hussain, A.; Irfan, M.; Baloch, N.K.; Draz, U.; Ali, T.; Glowacz, A.; Antonino-Daviu, J. Savior: A Reliable Fault Resilient Router Architecture for Network-on-Chip. *Electronics* **2020**, *9*, 1783. [\[CrossRef\]](#)
22. Lowe, D.G. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* **2004**, *60*, 91–110. [\[CrossRef\]](#)
23. Gudivada, V.N.; Raghavan, V.V. Content based image retrieval systems. *Computer* **1995**, *28*, 18–22. [\[CrossRef\]](#)
24. Liu, Y.; Zhang, D.; Lu, G.; Ma, W.-Y. A survey of content-based image retrieval with high-level semantics. *Pattern Recognit.* **2007**, *40*, 262–282. [\[CrossRef\]](#)
25. Kokare, M.; Chatterji, B.; Biswas, P. Comparison of similarity metrics for texture image retrieval. In Proceedings of the Conference on Convergent Technologies for Asia-Pacific Region (TENCON 2003), Bangalore, India, 15–17 October 2003; IEEE: Piscataway, NJ, USA, 2003.
26. Graps, A. An introduction to wavelets. *IEEE Comput. Sci. Eng.* **1995**, *2*, 50–61. [\[CrossRef\]](#)
27. Ali, G.; Ali, T.; Irfan, M.; Draz, U.; Sohail, M.; Glowacz, A.; Martis, C. IoT Based Smart Parking System Using Deep Long Short Memory Network. *Electronics* **2020**, *9*, 1696. [\[CrossRef\]](#)
28. Ali, T.; Draz, U.; Yasin, S.; Noureen, J.; Shaf, A.; Ali, M. An Efficient Participant's Selection Algorithm for Crowdsensing. *Int. J. Adv. Comput. Sci. Appl.* **2018**, *9*, 399–404. [\[CrossRef\]](#)
29. Vassilieva, N.S. Content-based image retrieval methods. *Program. Comput. Softw.* **2009**, *35*, 158–180. [\[CrossRef\]](#)
30. Ojala, T.; Pietikäinen, M.; Mäenpää, T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans. Pattern Anal. Mach. Intel.* **2002**, *24*, 971–987. [\[CrossRef\]](#)
31. Srivastava, P.; Khare, A. Integration of wavelet transform, local binary patterns and moments for content-based image retrieval. *J. Vis. Commun. Image Represent.* **2017**, *42*, 78–103. [\[CrossRef\]](#)
32. Manjunath, B.S.; Ma, W.-Y. Texture features for browsing and retrieval of image data. *IEEE Trans. Pattern Anal. Mach. Intell.* **1996**, *18*, 837–842. [\[CrossRef\]](#)
33. Haralick, R.M.; Shanmugam, K. Textural features for image classification. *IEEE Trans. Syst. Man Cybern.* **1973**, SMC-3, 610–621. [\[CrossRef\]](#)
34. Cross, G.R.; Jain, A.K. Markov random field texture models. *IEEE Trans. Pattern Anal. Mach. Intell.* **1983**, PAMI-5, 25–39. [\[CrossRef\]](#)
35. Mezaris, V.; Kompatsiaris, I.; Strintzis, M.G. An ontology approach to object-based image retrieval. In Proceedings of the 2003 International Conference on Image Processing, Barcelona, Spain, 14–17 September 2003; IEEE: Piscataway, NJ, USA, 2003.
36. Town, C.; Sinclair, D. *Content Based Image Retrieval Using Semantic Visual Categories*; AT&T Laboratories Cambridge: Cambridge, UK, 2000.
37. Song, Y.; Wang, W.; Zhang, A. Automatic annotation and retrieval of images. *World Wide Web* **2003**, *6*, 209–231. [\[CrossRef\]](#)
38. Ahmed, K.T.; Naqvi, S.A.H.; Rehman, A.; Saba, T. Convolution, Approximation and Spatial Information Based Object and Color Signatures for Content Based Image Retrieval. In Proceedings of the 2019 International Conference on Computer and Information Sciences (ICCIS), Sakaka, Saudi Arabia, 3–4 April 2019; IEEE: Piscataway, NJ, USA, 2019.
39. Alzu'bi, A.; Amira, A.; Ramzan, N. Content-based image retrieval with compact deep convolutional features. *Neurocomputing* **2017**, *249*, 95–105. [\[CrossRef\]](#)
40. Shaf, A.; Ali, T.; Farooq, W.; Javaid, S.; Draz, U.; Yasin, S. Two Classes Classification Using Different Optimizers in Convolutional Neural Network. In Proceedings of the 2018 IEEE 21st International Multi-Topic Conference (INMIC), Karachi, Pakistan, 1–2 November 2018; pp. 1–6.
41. Cui, C.; Lin, P.; Nie, X.; Yin, Y.; Zhu, Q. Hybrid. textual-visual relevance learning for content-based image retrieval. *J. Vis. Commun. Image Represent.* **2017**, *48*, 367–374. [\[CrossRef\]](#)
42. Zhou, W.; Li, H.; Sun, J.; Tian, Q. Collaborative index embedding for image retrieval. *IEEE Trans. Pattern Anal. Mach. Intel.* **2018**, *40*, 1154–1166. [\[CrossRef\]](#)
43. Jin, C.; Jin, S.-W. Content-based image retrieval model based on cost sensitive learning. *J. Vis. Commun. Image Represent.* **2018**, *55*, 720–728. [\[CrossRef\]](#)
44. Tzelepi, M.; Tefas, A. Deep convolutional learning for content based image retrieval. *Neurocomputing* **2018**, *275*, 2467–2478. [\[CrossRef\]](#)
45. Pavithra, L.; Sharmila, T.S. An efficient framework for image retrieval using color, texture and edge features. *Comput. Electr. Eng.* **2018**, *70*, 580–593. [\[CrossRef\]](#)

46. Xu, Y.; Gong, J.; Xiong, L.; Xu, Z.; Wang, J.; Shi, Y.-S. A privacy-preserving content-based image retrieval method in cloud environment. *J. Vis. Commun. Image Represent.* **2017**, *43*, 164–172. [[CrossRef](#)]
47. Hussain, D.M.; Surendran, D. The efficient fast-response content-based image retrieval using spark and MapReduce model framework. *J. Ambient. Intell. Humaniz. Comput.* **2020**, *7*, 1–8.
48. Alsmadi, M.K. Content-Based Image Retrieval Using Color, Shape and Texture Descriptors and Features. *Arab. J. Sci. Eng.* **2020**, *45*, 3317–3330. [[CrossRef](#)]
49. Garg, M.; Dhiman, G. A novel content based image retrieval approach for classification using glcm features and texture fused lbp variants. *Neural. Comput. Appl.* **2020**. [[CrossRef](#)]
50. Cha, S.-H. Comprehensive survey on distance/similarity measures between probability density functions. *City* **2007**, *1*, 1.
51. Yue, J.; Li, Z.; Liu, L.; Fu, Z. Content-based image retrieval using color and texture fused features. *Math. Comput. Model.* **2011**, *54*, 1121–1127. [[CrossRef](#)]
52. Shen, L.; Bai, L. A review on Gabor wavelets for face recognition. *Pattern Anal. Appl.* **2006**, *9*, 273–292. [[CrossRef](#)]
53. Ester, M.; Kriegel, H.-P.; Sander, J.; Xu, X. A density-based algorithm for discovering clusters in large spatial databases with noise. In Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD), Portland, OR, USA, 2–4 August 1996.
54. Hastie, T.; Rosset, S.; Tibshirani, R.; Zhu, J. The entire regularization path for the support vector machine. *J. Mach. Learn. Res.* **2004**, *5*, 1391–1415.
55. Lin, C.-H.; Chen, R.-T.; Chan, Y.-K. A smart content-based image retrieval system based on color and texture feature. *Image Vis. Comput.* **2009**, *27*, 658–665. [[CrossRef](#)]
56. Irtaza, A.; Jaffar, A.; Aleisa, E.; Choi, T.-S. Embedding neural networks for semantic association in content based image retrieval. *Multimed. Tools Appl.* **2014**, *72*, 1911–1931. [[CrossRef](#)]
57. Wang, X.-Y.; Yu, Y.-J.; Yang, H.-Y. An effective image retrieval scheme using color, texture and shape features. *Comput. Stand. Interfaces* **2011**, *33*, 59–68. [[CrossRef](#)]
58. Walia, E.; Pal, A. Fusion framework for effective color image retrieval. *J. Vis. Commun. Image Represent.* **2014**, *25*, 1335–1348. [[CrossRef](#)]
59. Walia, E.; Vesal, S.; Pal, A. An effective and fast hybrid framework for color image retrieval. *Sens. Imaging* **2014**, *15*, 93. [[CrossRef](#)]

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).