Content-based image retrieval using hybrid k-means moth flame optimization algorithm

Annrose Joseph[1] · Edwin Selva Rex[2] · Seldev Christopher[3] · Jenifer Jose[4]

Abstract

Content-based image retrieval plays a key role in many domains and the volume of the image databases increases tremendously; it is very difficult to compare query image feature with every image in the dataset during the retrieval phase. Hence, search space and computational complexity increase which degrades the performance of recognition accuracy. This system investigates various search space reduction techniques, which partition or classify the image collection into a subset of related images. This study proposes an image clustering using the hybrid K-means moth flame optimization algorithm (KMFO). It enhances the performance of the K-means algorithm by assigning the optimum number of clusters and cluster centroids using the number of flames and flame values obtained in MFO. It uses color moments, HSV color histogram, color correlogram, GLCM, wavelet transform, dominant color, and region-based descriptors as feature vectors. The experiments are tested on Corel 1K dataset, and it shows competent results when compared with other retrieval techniques.

Keywords Image retrieval · Moth flame optimization · K-means clustering · Feature reduction · Search space reduction

Introduction

The world is moving towards the fourth technological revolution, where everything gets automated due to the integration of different technologies that blurs the boundary between physical, biological, and digital spheres. In this digital environment, there is an exponential growth in multimedia data from every domain, in which one-third of data are images. According to the International Data Corporation (IDC) report, the big data market will reach over US$125 billion by 2019, and the number of sensors will increase by 1 trillion in 2030 (Marjani et al. 2017)

As the number of image repositories in various domains such as medical, digital image archives, art galleries, geographic information systems, e-commerce, law enforcement, biometric identification, historical analysis, and so on increases dramatically. During 1970s, active research started to meet the challenges of image matching and retrieval system. Text-based image retrieval is a common technique in the early stages, in which images are matched based on textual descriptions such as labels, captions, keywords, semantic context, and so on (Yue et al. 2011). It does not capture the entire image content; different users may give different interpretations and annotations for the same image. It is also more subjective and incomplete. Therefore, for large image repositories, text-based image retrieval is not a standard and practical method. Hence, content-based image retrieval is the better alternative method that overcomes the problem of text-based image retrieval. Active research on content-based image retrieval had been started early 1990s. Deep learning is the

✉ Annrose Joseph
roseedjo@gmail.com

Edwin Selva Rex
cresrex@gmail.com

Seldev Christopher
cseldev@gmail.com

Jenifer Jose
jenifer@sxcce.edu.in

[1] Department of IT, St. Xavier's Catholic College of Engineering, Chunkankadai, Nagercoil, Tamil Nadu 629003, India

[2] Department of EEE, Vignana Bharathi Institute of Technology, Hyderabad, Telungana, India

[3] Department of CSE, St. Xavier's Catholic College of Engineering, Chunkankadai, Nagercoil, Tamil Nadu 629003, India

[4] Department of MCA, St. Xavier's Catholic College of Engineering, Chunkankadai, Nagercoil, Tamil Nadu 629003, India

hottest research area in computer vision and machine learning applications in the last decade. It mimics the human brain with high computing and processing power through multiple stages of transformation without hand-crafted features. Some of the its application include face recognition, image detection, voice recognition, video analysis, health care, smart cities, smart agriculture, smart grid energy usage analysis, business intelligence, natural language processing, and more. A convolution neural network is a stack of operations like convolution, pooling, and activation layers that recognize the visual pattern of images. It starts its milestone with the supervision of a deep convolution neural network model called AlexNet in 2012 (Krizhevsky et al. 2012) on ImageNet Large Scale Visual Recognition Challenge (ILSVRC) to classify images. Popular deep neural network models perform object detection, object localization, and image classification tasks on these challenges (Russakovsky et al. 2015). ZFNet is an extension of AlexNet with a small change in the filter size to avoid pixel loss. It uses a 7×7 filter size which is lesser than AlexNet, but it fails to reduce the computational cost. The inception model is also called as GoogleNet which provides a very deeper network of having 22 layers, but it reduces the computational complexity, the number of parameters, and memory usage by the use of the inception module with dimension reduction using 1×1 convolutions. Microsoft's research team developed Residual Network (ResNet) that overcomes the problem of vanishing/exploding gradient in extremely deep networks by the introduction of residual networks. A residual network is a stack of residual blocks by providing identical connections between layers. ResNeXt used a stack of blocks and then use the ResNet approach of residual blocks. It has a feature known as cardinality referred to the size of the set of transformations. Because of uniformity in topology, a fewer number of parameters were required for deeper networks. Progressive Neural Architecture Search (PNASNet) used a new learning structure for CNN using reinforcement learning and evolutionary algorithms which produced optimized results than the previous models. The taxonomy of the image retrieval system is shown in figure (Fig. 1). The first category shows the text-based retrieval, and it specifies the possible ways to annotate or describe the images. The next category content-based image retrieval describes the image by its features or content, and its types are specified in the figure. The last category is the hybrid approach, which combines both text and content to describe the images.

In content-based image retrieval, images get matched with the features or contents available in the images. QBIC (Faloutsos et al. 1994), Photobook (Pentland et al. 1996), Virage (Gupta and Jain 1997), VisualSEEK (Smith and Chang 1996), Netra (Ma and Manjunath 1997), and SIMPLIcity (Wang et al. 2001) are some of the commercial CBIR systems, and wide-ranging surveys of CBIR can be found in the journals (Liua et al. 2007; Datta et al.

2005). Improving the efficiency of the retrieval system, reducing the response time, semantic gap, and sensory gap are the primary objectives of every image retrieval system. A survey on the high-level semantic-based system to narrow down the semantic gap is presented in this article (Liua et al. 2007). Based on the type of features, user intervention, and computational intelligence, CBIR has been classified into low-level image retrieval, high or semantic level image retrieval, image retrieval using relevance feedback, and intelligent image retrieval system. For a large-scale image repository, the conventional method of matching during the query phase degrades the performance of the retrieval system. So, it is necessary to select the initial collection of the most relevant images from the large repository, and then, the query image features are matched with the selected subset of images. K-means algorithm is the most simple and common clustering algorithm which is an unsupervised machine learning technique that could be applied to select the initial subset of images.

The standard K-means algorithm was first proposed by Lloyd in 1957 at Bell Labs, and it is one of the most popular data mining clustering algorithms, because of its efficiency and simplicity. It increases intra-cluster similarity and decreases inter-cluster similarity using the sum of squared distances between two feature points. The generalized version of the K-means algorithm is presented in this article (Cheung 2003) to reduce the problems of the conventional K-means algorithm. Conventional K-means algorithm requires predetermination of cluster numbers, and it suffers due to the dead unit problem; which means incorrect initialization of cluster centers. K-means algorithm provides a faster convergence rate in local optima, but it fails to find the global optimum solution. To overcome the above drawbacks, moth flame optimizer is applied before K-means clustering (Mirjalili 2015). Moth flame optimizer was developed by Mirillaji in the year 2015; it produces a higher convergence rate in the global solution. Hence, to improve the convergence and trapping in local optima, this system reduces the search space by combining the K-means clustering algorithm with a bio-inspired algorithm called moth flame optimizer. MFO algorithm improves the initial random solutions and convergence to a better point in the search space. So, the initial seed values like the number of clusters and cluster centroids are assigned to the K-means algorithm from the moth flame algorithm. The performance efficiency is tested on WANG or COREL1K dataset and COIL dataset.

Related Work

The particle swarm optimization is combined with K-means clustering to reduce the search space by clustering the images are proposed in this article (Younus et al. 2015). K-means

Fig. 1 Taxonomy of image retrieval systems
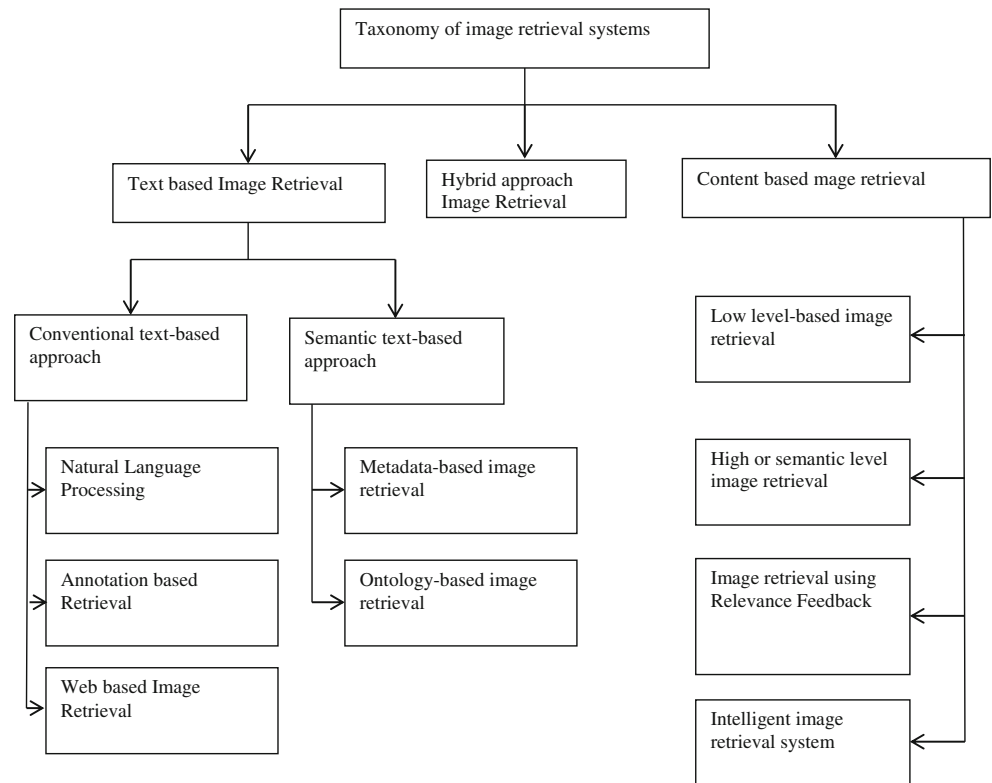
algorithm finds local optimal solution effectively, but it rarely catches the global optimal solution. This method uses particle swarm optimization initially to locate the cluster centroid optimally, and these centroids are given as a seed value to the K-means clustering. The experiment was conducted on the WANG dataset, and it proved to be better than the other
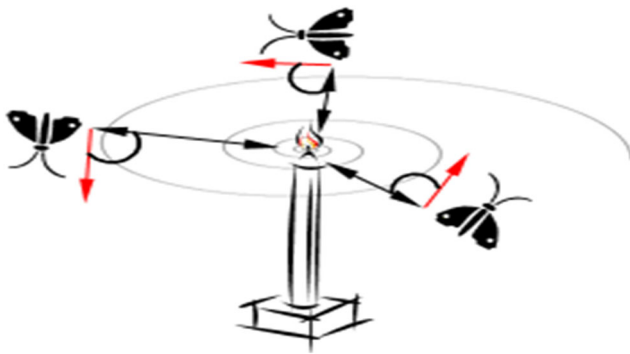
Fig. 2 Proposed CBIR architecture
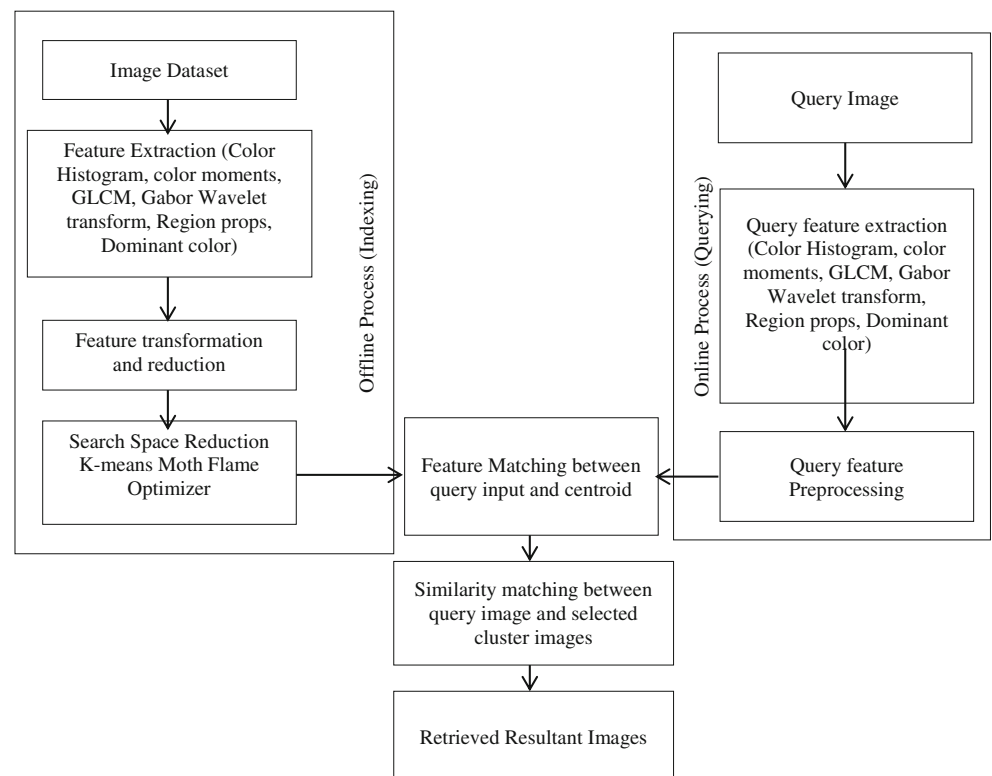
Fig. 3 Flying behavior of moth around the light flame

CBIR systems. The inter-class boundary problem is addressed in the feature space by replacing simple distance-based retrieval for texture databases (Dash et al. 2015). It reduces the searching time in class membership-based retrieval. The class membership and classification confidence-based retrieval (CM-CCR) seem to be computationally efficient than class membership-based retrieval and yield better retrieval performance than classification confidence-based retrieval. Texture-based image retrieval using two novel wavelet features is proposed in this article (Huang and Dai 2003). Energy distribution pattern strings are a fuzzy matching mechanism that acts as a filter, and the selected images are compared with the query images using composite sub-band gradient vectors.

The graph-theoretical cluster-based image retrieval using unsupervised learning could be embedded with any CBIR systems including relevance feedback. Most of the image classification or clustering algorithms are global, static, and independent of the query. It is a dynamic clustering algorithm because it captures the characteristics of the query image (Chen et al. 2005). An efficient framework for an image retrieval system based on a rule-based system is proposed in this system (ElAlami 2011a). The retrieval process is limited to the set of images matched to the rule with the query image. It requires rule generation and rule pruning. A survey of cluster techniques is proposed in (Xu and Wunsch 2005). Unsupervised clustering is used to select a subset of relevant images which is used to narrow down the feature search space. Images within a particular cluster tend to have high similarity, and it is dissimilar to other clusters. From their findings, none
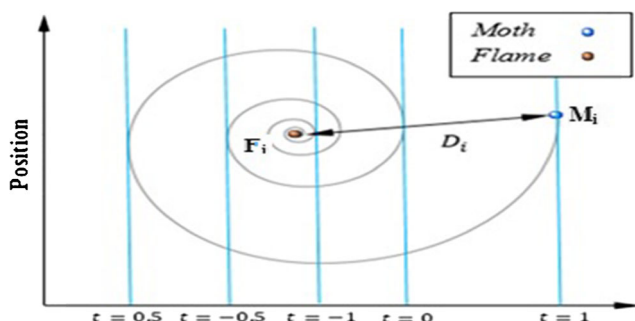

Fig. 4 A logarithmic spiral, space around a flame

of the cluster algorithms is universally accepted to provide more promising results for general dataset. Hence, cluster algorithms can be selected based on the domain-specific information with different proximity measures and a criterion function.

K-means clustering with B+ tree indexing is used in this system (Yildizer et al. 2012), when the querying phase relevant images are retrieved by matching the cluster centroid and the first three topmost closest clusters are targeted for similarity matching with the query image. It uses two parameters CG and CS to determine the distance range which leads to an increase in the computation complexity. Unlike static clustering, SQL-based query which is a dynamic way of selecting the closest images concerning the query image is used in this system (Annrose and Seldev 2016). It uses search space reduction using rule generation. Here, intra-normalization is applied to divide each feature element into five intervals. During querying, the feature interval of each query image feature element is determined, and it is combined using a Boolean operator, and a rule has been generated. This method misses out on some true positive images that lie in the border of the adjacent intervals. Hence, it affects the recognition rate. To overcome this issue, SQL-based range query is used to select the initial level of images (Annrose and CC, 2018). It selects the images that lie within the range which is closer to the query image features. Deep learning is the hottest research area in computer vision and machine learning applications in the last decade. It mimics the human brain with high computing and processing power through multiple stages of transformation without hand-crafted features. A convolution neural network is a stack of operations like convolution, pooling, and activation layers that recognize the visual pattern of images. It starts its milestone with the supervision of a deep convolution neural network model called AlexNet in 2012 on ImageNet Large Scale Visual Recognition Challenge (ILSVRC) to classify images. Deep learning specifically addresses feature representation and similarity matching related to CBIR tasks. The deep learning-based clustering techniques cluster the data points based on complex patterns rather than distance measures. Running K-means on representation vectors learned by deep autoencoder tend to give better results comparing to running K-means directly on the input vectors. This article (Wan et al., 2014) provides an empirical study on deep CNN for CBIR task. The following conclusions were drawn: pre-trained CNN models can be used for feature extraction that captures low-level features to high-level semantic information, and it was demonstrated that these features outperform traditional hand-crafted features, resulting in significant improvements in retrieval efficiency. Caron et al. (2019) proposed a deep clustering, which learns the hyper-parameters of neural network and cluster assignment. It is a novel end-to-end learning of convnets that works with K-means clustering. K-means clustering works onset of features generated by convNet. A

Fig. 5 Proposed K-means moth flame optimizer
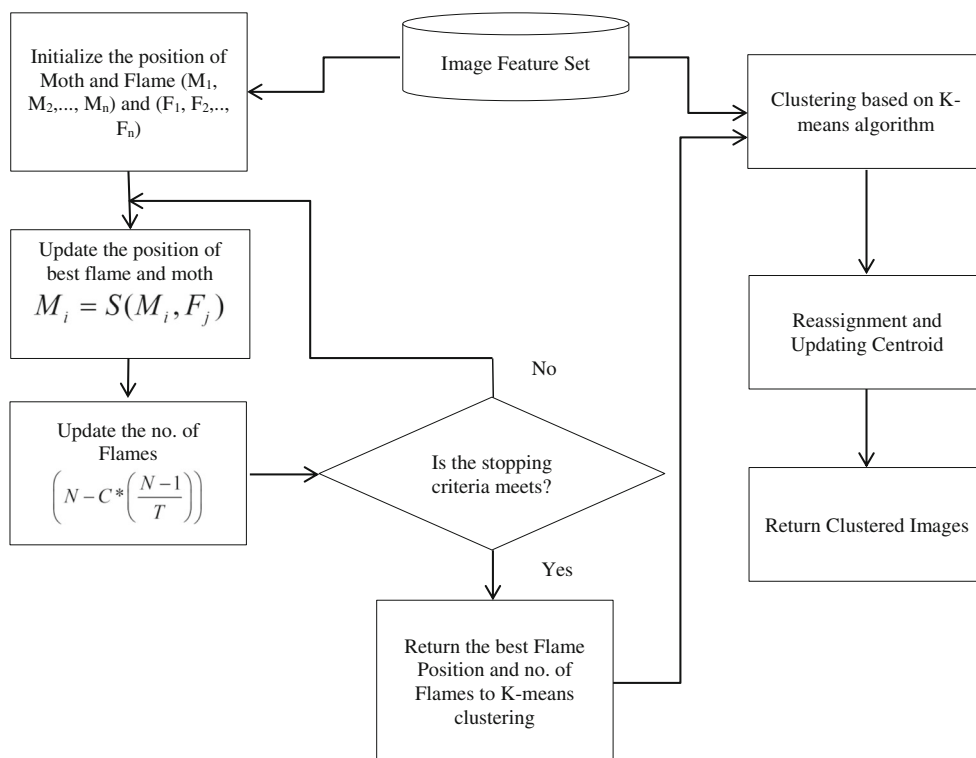
recurrent framework was proposed in this article (Yang et al. 2016) to iteratively learn convnet features and clusters within this model. It shows promising performance on small datasets but may be challenging to large-scale image datasets. Although all the deep neural network models provide the best results, it requires high computing GPUs; hence, it is mandated to upgrade the system requirements.

Proposed System

Content-based image retrieval is commonly used in a broad range of fields, from handheld devices to large-scale applications such as commerce, satellite image processing, and the medical industry. The proposed CBIR system shown in figure (Fig. 2) uses four primary stages; the first three stages are performed during the indexing or offline process, and then, the querying or retrieval phase is performed during the online process.

Indexing (offline process)

- Feature extraction
- Feature transformation and reduction
- A subset of image selection (KMFO clustering)

Querying phase (online process)

- Query feature extraction and pre-processing
- Query feature matching with cluster centroid
- Similarity matching with selected subset of images

The two phases in designing CBIR based image retrieval systems are:

Indexing phase

In this phase, the image information like color, texture, or shape is separated into features that are stored in an index data structure along with a corresponding link to the actual image. Indexing enhances the data access speed and improves the accuracy of the retrieval process, and hence, it is an important factor in image database systems. Indexing in content-based image retrieval systems is used to facilitate automatic identification and abstraction of the visual content of an image.

Image acquisition

This method uses a heterogeneous collection of broad domain image dataset called the COREL1K dataset. It is otherwise called as WANG dataset (Wang et al. 2001) which consists of 1000 images with 10 different categories.

Feature extraction

Feature extraction is the primary reduction process in an image retrieval system. It is the process of transforming images from pixel space to feature space (Otávio et al. 2012). This system extracts both global and local features by fusing low-level shape, color, and texture features. Gray-level co-occurrence matrix (GLCM) and Gabor wavelet transform are used

for extracting texture features. The dominant color descriptor, HSV color histogram, color correlogram, and color moments are extracted for color features, and shape features are extracted using region-based shape components to determine the largest connected component.

Gray-level co-occurrence matrix

It is a second derivative statistical analysis method, which specifies the intensity variation of pixel elements in different direction and orientation. It gives the distribution of common gray values frequently scattered throughout the image. It calculates the probability of co-occurrence of gray values in different positions. For an $n \times m$ image, a co-occurrence matrix GLCM is defined as:

$$glcm(x,y) = \sum_{p=1}^{n} \sum_{q=1}^{m} \begin{cases} 1, if\ A(p,q) = x\ and\ A(r,s) = y \\ 0, otherwise \end{cases} \quad (1)$$

where

$|p\text{-}r|=|q\text{-}s|=|p\text{-}q|=|r\text{-}s|=1$

x and y are the image intensity values

(p, q) and (r, s) are the adjacent spatial positions in an image A

The contrast, correlation, energy, and homogeneity are extracted using spatial distribution. The contrast gives the gray-level intensity difference between two adjacent pixels over the entire image.

$$Contrast = \sum_{x,y}^{n} |x{-}y|^2 glcm(x,y) \quad (2)$$

Inverse difference moment gives the local homogeneity, and it is the inverse of contrast.

$$Inverse\ Diff\ Moment = \sum_{x,y}^{n} \frac{1}{|x{-}y|^2} glcm(x,y)\ if\ x{\neq}y \quad (3)$$

Inverse Inversecorrelation specifies the measure of the interrelationship between adjacent pixel values over the whole image.

$$Correlation = \sum_{x,y}^{n} \frac{(x{-}\mu_x)(y{-}\mu_y) glcm(x,y)}{\sigma_x \sigma_y} \quad (4)$$

Energy is one for the constant image.

$$Energy = \sum_{x,y} glcm(x,y)^2 \quad (5)$$

The uniformity and closeness of pixels in an image are determined by the homogeneity.

$$homogeneity = \sum_{x,y} \frac{glcm(x,y)}{1+|x{-}y|} \quad (6)$$

HSV color histogram

It is the widely used feature that calculates the histogram of images in the HSV color model.

Region-based shape descriptor

Object shape features provide a boundless sign to object identity. The object is recognized and diagnosed easily from its shape. Two categories of shape features are contour or boundary-based and region-based descriptors (Su et al. 2010). In contour-based shape descriptors, peripheral information of object shapes is used instead of using interior shape details. In region-based methods, shape descriptors use information from both peripheral and interior regions of the shape. Region props are used to measure the properties of image regions, and it is a global feature extraction method. It determines the number of connected components in an image, and here, we extract the largest connected component. The following five shape features like area, centroid, perimeter, solidity, and circularity are extracted from the largest region.

Dominant color

The histogram of each color space is determined in the RGB color model. With the histogram values, the percentage of red, green, and blue color components are determined.

Color moments

The distribution of color is provided by color moments. Using this method, the first-, second-, and third-order color moments are determined using the following equation:

$$M_i = {}^1\!/_n \sum_{j=1}^{n} A(i,j) \quad (7)$$

Standard deviation is the second-order moment, and it is obtained by applying the square root of the variance, and it is given as:

$$SD_i = \sqrt{1/n \sum_{j=1}^{n} (A(i,j){-}M_i)^2} \quad (8)$$

Skewness specifies the third-order color moment, and it gives the shape, and it shows the asymmetric property of the color distribution; it is calculated as follows:

$$Sw_i = \sqrt[3]{1/n \sum_{j=1}^{n} (A(i,j){-}M_i)^3} \quad (9)$$

The color moment is an invariant color feature vector that could be applied to any size of images.

Wavelet transform

The multi-resolution analysis of an image is represented using a wavelet that provides signals in both space and frequency domains. In one-dimensional discrete wavelet transform, images are decomposed into high-frequency and low-frequency components (Arai and Rahmad 2012). Two-dimensional DWT decomposes an image into four components, LL retains the approximation details, HL gives the vertical edge detail, horizontal edge detail is retained in LH, and high-frequency values are provided by HH. Features are extracted by determining the first-order mean and second-order standard deviation of 2D DWT.

Color correlogram

Color correlogram includes the spatial correlation of colors; it can be used to describe the global distribution of local spatial correlation of colors. The image is quantized to 16 levels hence it generates 64 feature dimensions.

Feature Transformation and Selection

Feature transformation is a pre-processing step that could be applied before any data mining or machine learning techniques. This system uses multiple features and each feature has its domain or range of values. During similarity matching, a larger feature value dominates and assigns more weight than

small range values. Hence, feature transformation is necessary to convert all the feature elements with the same significance (Aksoy and Haralick 2001). This method uses intra-normalization where each feature element gets normalized independently.

$$FV_{i,j} = \frac{FV_{i,j} - Max_i}{Max_i - Min_i} \tag{10}$$

$FV_{i,j}$ Feature vector
Max_i Maximum value of each feature element
Min_i Minimum value of each feature element

The transformed features undergo a selection process to remove less significant features. Three types of feature selection are:

- Filter-based feature selection is independent of any classifier or learning algorithms.
- The wrapper-based selection method depends on learning algorithms.
- The hybrid method combines both the filter method and the wrapper method.

This paper uses a filter-based feature selection method which uses a ranking algorithm to sort out feature elements having a greater number of null values and less distinct values.

Algorithm 1 SQL-based feature reduction

```
Input   : Feature Vector (F), Threshold (th)
Output  : Reduced Feature Vector (FV)
Intialize
        R={}
        F={F1,F2,F3,…Fn} // Feature vector with n dimension
        th=(4N-N)/4  // N is the number of images
Repeat
        loop
        for each Fi€F
         d=select distinct(Fi) from F;      // No. of distinct values in each feature element
         e=select * from F where Fi==NULL;  // No. of NULL values in each feature element
         if d<th ||e<th
            R=RU{Fi}
         end if
        end loop
FV=F∩ R
return FV
```

Thus, features that have a large number of repeated values and NULL values are eliminated. Repeated values and NULL values in each feature element are selected, and if it is greater

than the threshold, then those features are removed; thereby, reduced feature set is obtained. In the COREL dataset, 158

feature elements are initially extracted, and the reduced feature set consists of 143 feature dimensions.

Search space reduction

When feature reduction techniques are used, they lose out on certain important features, causing retrieval efficiency to suffer. To improve the retrieval accuracy and to reduce the response time, it is necessary to reduce the search space. This system uses the K-means algorithm to cluster the images; hence, the query image features are compared to the cluster center initially, and the most relevant clustered group of images are further used for similarity matching. Hence, the NP-

hard problem will be reduced to a polynomial problem. K-means algorithm is an unsupervised hard clustering that is widely used for most applications. Conventional K-means algorithm requires pre-determination of cluster numbers, and it suffers due to the dead unit problem, which means incorrect initialization of cluster centers. Hence, this work combines the moth flame optimization algorithm with K-means clustering. Using the MFO algorithm optimum number of flames is determined, and these flame values are initialized as the cluster centers, and the number of flames is assigned to number of clusters (k) for the K-means algorithm. The following algorithm gives the procedure of K-means clustering algorithm.

Algorithm 2 K-means algorithm

```
Input    : Feature Vector (FV), No. of clusters (k)
Output   : Clustered Set, Cluster Centroid(μ)
Intialize
   k= {S₁,S₂,S₃,…,Sₖ}     // Randomly selected k samples from FV
   I= {I₁,I₂,I₃,…Iₘ}   // No. of images in the dataset
         FV=({FV₁₁, FV₁₂, …FV₁ₙ},{FV₂₁, FV₂₂, …FV₂ₙ}, …,FVₘ₁, FVₘ₂, …FVₘₙ})
Repeat until no change in k
   Reassign each FVi values to the nearest cluster centroid
   for each Sᵢ in k
      μᵢ= Sᵢ
      if | μᵢ - FVᵢⱼ|< |μᵢ₊₁-FVᵢⱼ|
          Cᵢ(i) = Cᵢ(i) U FVᵢ // Reassignment of images
      end if
   end for
   for i in 1 to k
```

$$\mu_i = \sum_{\vec{\mu} \in c} \vec{\mu} \quad \text{// Update the Cluster Centroid}$$

```
   end for
end loop
return μ₁ ,μ₂ ,…, μₖ
```

Moth flame optimizer

Seyedali Mirjalili in 2015 developed a moth flame optimization algorithm (MFO) which is a bio-inspired algorithm. The moth is the decorative insect that belongs to the butterfly family. It has a special navigation mechanism in which a moth flies by retaining a fixed angle concerning the moon position. In the artificial light, the moths fly spirally around it and drop towards the artificial flame and it is represented in figure (Fig. 3) .

The set of moths is represented in a matrix as follows, and it represents the image feature vector:

$$M = \begin{bmatrix} M_{1,1} & M_{1,2} & .. & M_{1,d} \\ M_{2,1} & M_{2,2} & .. & M_{2,d} \\ \vdots & \vdots & \vdots & \vdots \\ M_{n,1} & M_{n,2} & .. & M_{n,d} \end{bmatrix} \qquad (11)$$

where

n number of moth (no. of images)
d number of feature element (no. of feature dimensions)

Every moth has an array for storing the resultant fitness values:

$$OM = \begin{bmatrix} OM_1 \\ OM_2 \\ \vdots \\ OM_n \end{bmatrix} \tag{12}$$

where OM_i is the fitness value of ith moth M_i.

It uses flame as an important component. Flame matrix is also equal to the moth matrix and the set of flame is represented in a matrix as follows:

$$F = \begin{bmatrix} F_{1,1} & F_{1,2} & .. & F_{1,d} \\ F_{2,1} & F_{2,2} & .. & F_{2,d} \\ \vdots & \vdots & \vdots & \vdots \\ F_{n,1} & F_{n,2} & .. & F_{n,d} \end{bmatrix} \tag{13}$$

where

m number of flames
d number of
    variables

Every flame has an array for storing the resultant fitness values

$$OF = \begin{bmatrix} OF_1 \\ OF_2 \\ \vdots \\ OF_n \end{bmatrix} \tag{14}$$

Here, the moths and flames are both solutions. The key distinction between moths and flames is that moths are search agents that drive towards the best flame, while flames are moth's best positions. The mathematical model and behavior is specified in the following equation, where moth in each position is updated concerning a flame:

$$M_i = S(M_i, F_j) \tag{15}$$

where M_i is the ith moth, F_j is the jth flame, and S indicates spiral function.

The logarithmic spiral function is stated as follows:

$$S(M_i, F_j) = D_i.e^{bt}.\cos(2\prod t) + F_j \tag{16}$$

where D_i is the distance between an ith moth and j^{th} flame, the shape of the logarithmic spiral is represented by a constant b, and t is a random number that lies between -1 and 1 and it is illustrated in figure (Fig. 4).

The distance between an ith moth and the jth flame is calculated as follows:

$$D_i = |F_j - M_i| \tag{17}$$

Here, the number of flames can be decreased during the iterative process, and the number of flame can be calculated as:

$$\text{Flame no.} = \text{round}\left(N - C*\left(\frac{N-1}{T}\right)\right) \tag{18}$$

where C is represented as the current iteration number, the maximum number of flames is N, and maximum number of iterations is represented as T. Hence, decrease in the number of flames balances the exploration and exploitation of the search space.

K-means moth flame optimizer

The proposed work is illustrated in figure (Fig. 5) which combines moth flame algorithm with K-means clustering. Mirjalili's (2015) MFO algorithm is theoretically able to improve the initial random solutions and convergence to a better point in the search space. Using the moth flame algorithm number of clusters and cluster centroid is determined, and it is given as an initial seed value to the K-means algorithm.

Step 1: Moth value matrix and flame matrix are initialized by an image feature vector. In the context of clustering, a single flame position signifies the centroid of clusters.
       M= {M_1, M_2,…,M_n} and F= {F_1,F_2,…,F_n} // Initial Moth and Flame Values

Step 2: For each moth

  a) Calculate the Moth fitness based on clustering criteria args min| M_i - F_j |. The best position of moth and flame are updated.
  b) Calculate the number of flames using Equation (18)
  c) Repeat until the stopping criteria are satisfied (maximum no. of iteration)

Step 3: Apply K-means algorithm using the best flame positions and a number of flames obtained in MFO.

Step 4: Return the clustered images and the cluster centroids.

Retrieval phase

In this phase, the searching for the desired query image in the CBIR index is performed. Description of the properties of the image is done either by supplying a query image or denoting the image features. Generally, the collections of images are represented as a set of feature vectors. For the query input, the same set of features is extracted and processed using the feature transformation and selection technique. Then, the query image feature is compared with the cluster centroids obtained using the proposed KMFO algorithm. Then, the topmost matched cluster images are matched with the query image to generate the top most relevant images.

Fig. 6 **a** Sample input from bus category with 16 relevant images out of 20 retrieved. **b** Input from horse category with 19 relevant images out of 20 images retrieved
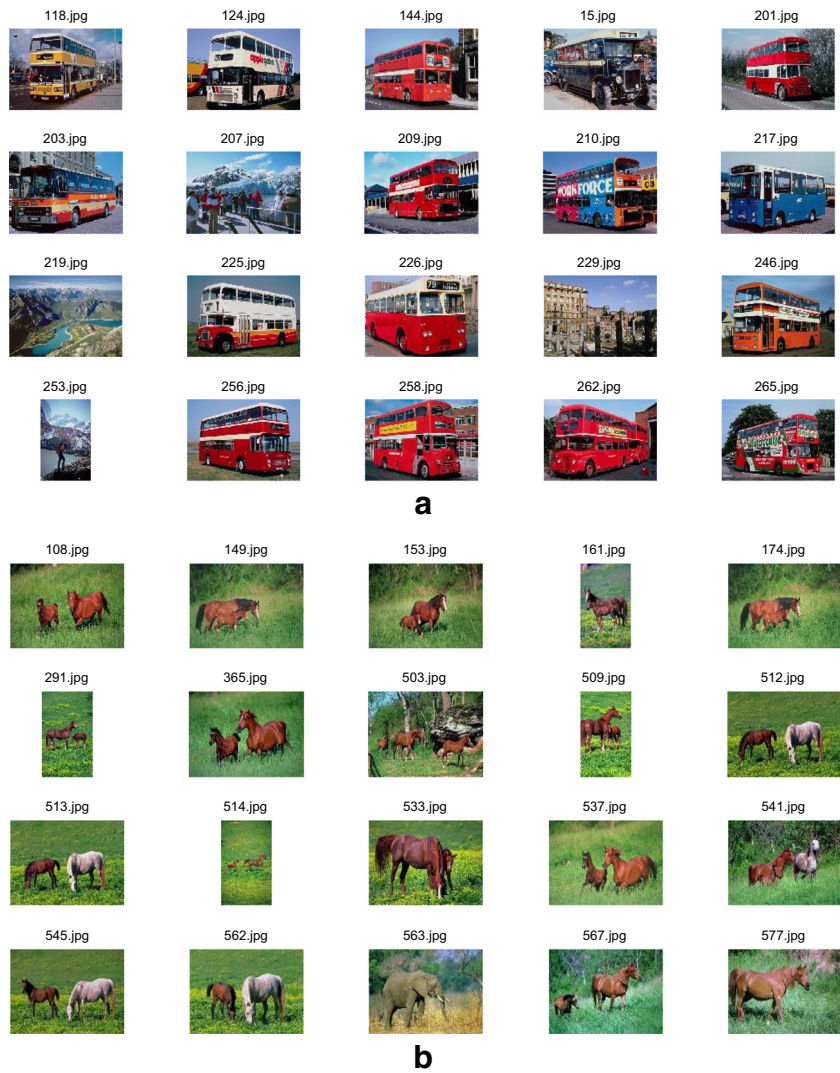
a

b

Table 1 Precision comparison of the different image retrieval system

Query	Huang and Dai 2003	Jhanwar et al. 2004	Rao et al. 2011	Chuen et al. 2009	Afifi and Ashour 2012	ElAlami, 2011b	Younus et al. 2015	Rule-based query	Range query	Proposed KMFO
Africans	0.424	0.4525	0.562	0.683	0.6	0.703	0.89	0.7	0.86	0.81
Beach	0.4455	0.3975	0.536	0.54	0.7	0.561	0.73	0.8	0.81	0.78
Architecture	0.4105	0.3735	0.61	0.562	0.8	0.571	0.701	0.4	0.5	0.80
Buses	0.8515	0.741	0.893	0.888	0.7	0.876	0.838	0.9	0.86	0.81
Dinosaurs	0.5865	0.9145	0.984	0.992	0.7	0.987	0.998	1	1	1
Elephants	0.4255	0.304	0.578	0.658	0.5	0.675	0.795	0.7	0.84	0.81
Flowers	0.8975	0.8515	0.899	0.891	0.7	0.914	0.925	1	1	0.89
Horses	0.589	0.568	0.78	0.803	0.5	0.834	0.871	0.7	1	0.95
Mountain	0.268	0.2925	0.512	0.522	0.5	0.536	0.55	0.4	0.56	0.68
Foods	0.4265	0.3695	0.694	0.733	0.4	0.741	0.755	0.6	0.76	0.6
Average	0.53245	0.52645	0.7048	0.7272	0.61	0.7398	0.8053	0.72	0.819	0.813

Fig. 7 Recall comparison of the different image retrieval system
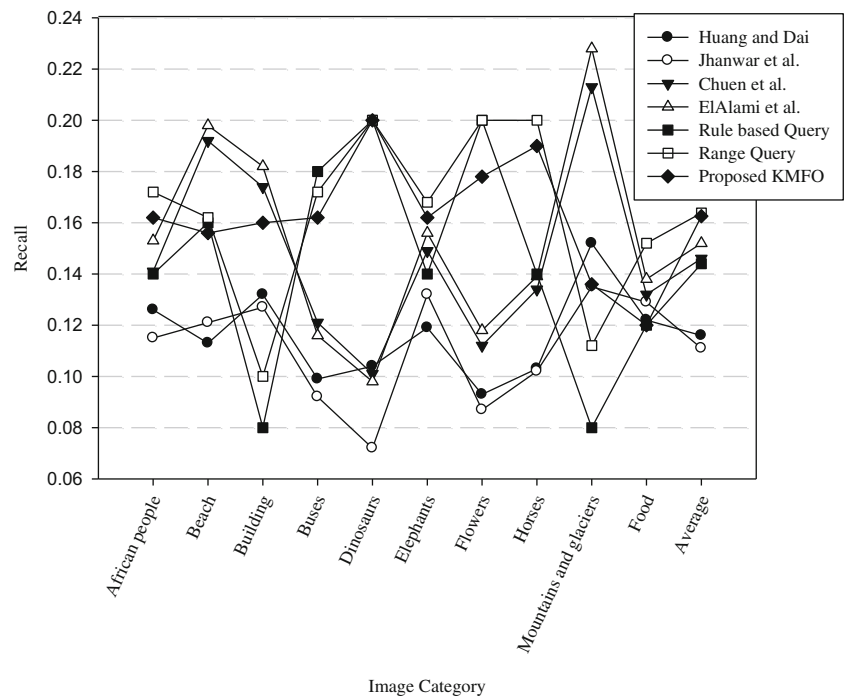
Experimental Results

This section presents the experimental results of the proposed method, and it is compared with other existing CBIR systems. In the proposed system, images are clustered in the offline process by combining moth flame optimization with K-means algorithm. During the online or querying phase, the query image is initially compared with the clustered centroid to identify the most relevant cluster set. The images belonging to the selected cluster are then compared with the query image to retrieve the most relevant images.

The performance measures used to evaluate the efficiency of the proposed system are precision, recall, and F-measure, mean average precision. Precision gives the ratio of no. relevant images retrieved (Nr) to the whole number of images retrieved (Rt). Recall specifies the ratio of number of relevant images retrieved (Nr) to the total relevant images present in the dataset (Nt). F-measure is the harmonic mean of precision and recall. The computation time of the query phase is determined using the response time. Precision and recall should be high to show high retrieval performance; hence, the joint precision-recall curve is used to characterize the performance of the image retrieval system.

Table 2 Precision, recall, and F-measure of K-mean and KMFO (COREL1K dataset)

Image category	Precision		Recall		F1-measure	
	KMFO	K-means	KMFO	K-means	KMFO	K-means
Africans	0.81	0.72	0.162	0.144	0.27	0.24
Beach	0.78	0.76	0.156	0.152	0.26	0.253333
Architecture	0.80	0.55	0.16	0.11	0.2666667	0.183333
Buses	0.81	0.79	0.162	0.158	0.27	0.263333
Dinosaurs	1	1	0.2	0.2	0.3333333	0.333333
Elephants	0.81	0.7	0.162	0.14	0.27	0.233333
Flowers	0.89	0.87	0.178	0.174	0.2966667	0.29
Horses	0.95	0.7	0.19	0.14	0.3166667	0.233333
Mountain	0.68	0.58	0.136	0.116	0.2266667	0.193333
Foods	0.6	0.56	0.12	0.112	0.2	0.186667
Average	0.813	0.723	0.162	0.1446	0.271	0.241

Fig. 8 Sample images in each category of the COIL dataset

Fig. 9 Comparison of retrieval time on COIL dataset

$$Precision = \frac{Nr}{Rt} = \frac{TP}{TP + FP} \qquad (19)$$

$$Recall = \frac{Nr}{Nt} = \frac{TP}{TP + FN} \qquad (20)$$

$$F-Measure = 2[(Precision*Recall)/(Precision + Recall)] \qquad (21)$$

The test is conducted on two different datasets: COREL and COIL datasets. More than half of the surveyed papers use the COREL dataset which consists of various content including animals, buildings, African people, and natural scenery. This system uses the COREL1K dataset, which is mostly accepted because of its heterogeneity and human-annotated ground truth images. The images are pre-classified into 10 different
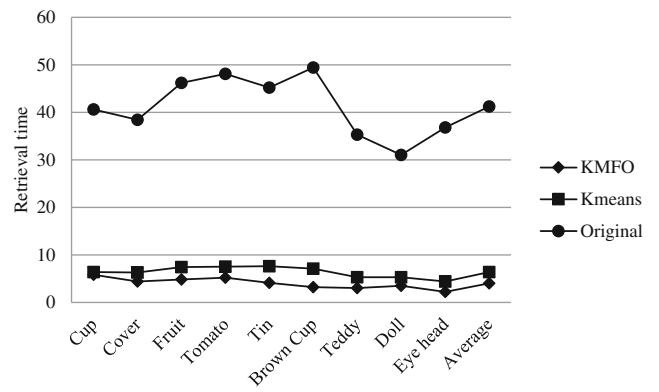
categories of size 100 images by domain experts. Initially, 158 image features are extracted in the feature extraction phase, and 15 features are removed which is having fewer number of distinct values and more null values. Then, it is grouped into different clusters using KMFO, and the query input is randomly selected from each category to test the average true positive images. The following figure specifies the top 20 images retrieved related to the sample query image. The first figure in each category denotes the query image. Figure 6a shows that four of the twenty images are unrelated to the query image, which results in 80 percent retrieval precision. The same set of tests is performed several times with different query images in the same category, and the average precision rate is 81% which is depicted in Table 1. Figure 6b shows that one image is unrelated to the query image, with a retrieval precision of 95% and an average precision rate of 95% for the horse class. Table 1 specifies the average precision comparison of the proposed KMFO with other image retrieval systems; it shows that the average precision of KMFO is better than all the other systems except range query-based image retrieval. Figure 7 specifies the comparison of recall with other retrieval systems on the corel1K dataset.

Table 3 Precision, recall, and F-measure of K-mean and KMFO (COIL dataset)

Input image category	Precision		Recall		F-measure	
	K-means	KMFO	K-means	KMFO	K-means	KMFO
Cup	0.54	0.72	0.19	N 0.25	0.28	0.37
Cover	1	1	0.35	0.35	0.52	0.52
Fruit	1	1	0.35	0.35	0.52	0.52
Tomato	1	1	0.35	0.35	0.52	0.52
Tin	1	1	0.35	0.35	0.52	0.52
BrownCup	0.26	0.38	0.09	0.13	0.13	0.19
Teddy	0.50	0.70	0.17	0.24	0.25	0.36
Doll	0.86	0.88	0.3	0.31	0.44	0.46
Eye Toy	1	1	0.35	0.35	0.52	0.52
Average	0.796	0.853	0.28	0.3	0.41	0.44

The next type of comparison is image retrieval with K-means and KMFO, and the performance measures like precision, recall, and F-measure are shown in Table 2. The results show that moth flame optimizer improves the performance of K-means to some extent.

COIL dataset (Nene et al., n.d.) is the next experimental image dataset which consists of 1440 images with 20 different categories and each category consisting of 72 images. It was captured using a motioned camera with the common black background, and the object was rotated 360 degrees placed on a turntable corresponds to 72 different positions. After feature extraction, search space is reduced by grouping similar images using the K-means algorithm with the moth flame optimizer. The results are compared with and without the moth flame optimizer. The sample image in each category is shown in Fig. 8.

The following Table 3 specifies the retrieval precision, recall, and F-measure of the COIL dataset using K-means and KMFO algorithm. The result shows that the performance of the K-means algorithm is improved slightly by initializing the seed value of K-means parameters using moth flame optimizer.

The retrieval time is compared with and without search space reduction. Figure 9 shows the execution time of the three methods, and KMFO is slightly less than K-means, and it is drastically improved compared with the original method that is without search space reduction.

Conclusion

Clustering algorithms have been applied to the feature space to reduce the searching time, thereby reducing the response time without compromising the retrieval accuracy. The proposed CBIR system presents the moth flame optimization algorithm with K-means clustering which overcomes the drawbacks of the conventional K-means clustering algorithm. Random selection of initial cluster centroids and number of clusters in the K-means algorithm leads to a dead unit problem. Hence, it is reduced by providing optimal value through MFO. The comparability of the proposed method with the other existing system demonstrates the usage of COREL and COIL dataset. The result proves that this system provides a satisfactory outcome, and it is slightly better than other methods. The future work of the study is to improve the retrieval accuracy by including feature dimension reduction and applying other bio-inspired optimization algorithms.

References

Afifi AJ, Ashour WM (2012) Content-based image retrieval using invariant color and texture features. In: Int Conf on Digital Image Computing Techniques and Applications (DICTA). IEEE, Fremantle WA, pp 1–6 http://hdl.handle.net/20.500.12358/24456

Aksoy S, Haralick RM (2001) Feature normalization and likelihood-based similarity measures for image retrieval. Pattern Recogn Lett: 563–582. https://doi.org/10.1016/S0167-8655(00)00112-4

Annrose J, Seldev CC (2016) Content based image retrieval using query based feature reduction with k-means cluster index. Asian J Res Soc Sci Humanit 6:852–872. https://doi.org/10.5958/2249-7315.2016.01334.4

Annrose J, CC CS (2018) An efficient image retrieval system with structured query based feature selection and filtering initial level relevant images using range query. Optik 157:1053–1064. https://doi.org/10.1016/j.ijleo.2017.11.179

Arai K, Rahmad C (2012) Wavelet based image retrieval method. Int J Adv Comput Sci Appl 3:6–11. https://doi.org/10.14569/IJACSA.2012.030402

Caron M, Bojanowski P, Joulin A, and Douze M (2019) Deep clustering for unsupervised learning of visual features. Computer Vision and Pattern Recognition, 1-30. https://arxiv.org/abs/1807.05520

Chen Y, James ZW, Krovetz R (2005) CLUE: cluster-based retrieval of images by unsupervised learning. IEEE Trans Image Process 14: 1187–1201. https://doi.org/10.1109/tip.2005.849770

Cheung Y (2003) K-means: a new generalized k-means clustering algorithm. Pattern Recogn Lett 24:2883–2893. https://doi.org/10.1016/S0167-8655(03)00146-6

Chuen L, Chen RT, Chan YK (2009) A smart content-based image retrieval system based on color and texture feature. Image Vis Comput 27:658–665. https://doi.org/10.1016/j.imavis.2008.07.004

Dash JK, Mukhopadhyay S, Gupta RD (2015) Content-based image retrieval using fuzzy class membership and rules based on classifier confidence. IET Image Process 9:836–848. https://doi.org/10.1049/iet-ipr.2014.0299

Datta R, Li J & Wang JZ (2005) Content-based image retrieval - approaches and trends of the new age. MIR '05 Proceedings of the 7th ACM SIGMM international workshop on Multimedia information retrieval: 253-262. https://doi.org/10.1145/1101826.1101866

ElAlami ME (2011a) A novel image retrieval model based on the most relevant features. Knowl-Based Syst 24:23–32. https://doi.org/10.1016/j.knosys.2010.06.001

ElAlami ME (2011b) Supporting image retrieval framework with rule base system. Knowl-Based Syst 24:331–340. https://doi.org/10.1016/j.knosys.2010.10.005

Faloutsos C, Barber R, Flickner M, Hafner J, Niblack W, Petkovic D, Equitz W (1994) Efficient and effective querying by image content. J Intell Inf Syst 3:231–262. https://doi.org/10.1007/BF00962238

Gupta A, Jain R (1997) Visual information retrieval. Commun ACM 40: 70–79. https://doi.org/10.1145/253769.253798

Huang PW, Dai SK (2003) Image retrieval by texture similarity. Pattern Recogn 36:665–679. https://doi.org/10.1016/S0031-3203(02)00083-3

Jhanwar N, Chaudhuri S, Seetharaman G, Zavidovique B (2004) Content based image retrieval using motif co-occurrence matrix. Image Vis

Comput 22:1211–1220. https://doi.org/10.1016/j.imavis.2004.03.026

Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. NIPS'12: Proceedings of the 25th International Conference on Neural Information Processing Systems 1:1097–1105. https://doi.org/10.5555/2999134.2999257

Liua Y, Zhanga D, Lua G, Mab W (2007) A survey of content-based image retrieval with high-level semantics. Pattern Recogn 40:262–282. https://doi.org/10.1016/j.patcog.2006.04.045

Ma WY, Manjunath B (1997) Netra: a toolbox for navigating large image databases. Proceedings of the IEEE International Conference on Image Processing 568–571. https://doi.org/10.1007/s005300050121

Marjani M, Nasaruddin F, Gani A, Karim A, Hashem IAT, Siddiqa A (2017) Big IoT data analytics: architecture, opportunities, and open research challenges. IEEE Access 5:5247–5261. https://doi.org/10.1109/ACCESS.2017.2689040

Mirjalili S (2015) Moth-flame optimization algorithm: a novel nature-inspired heuristic paradigm. Knowl-Based Syst 89:228–249. https://doi.org/10.1016/j.knosys.2015.07.006

Nene SA, Nayar SK, Murase H (n.d.) Columbia Object Image Library (COIL-100). Center for Research on Intelligent Systems at the Department of Computer Science, Columbia University

Otávio AB, Valle PE, Torre RS (2012) Comparative study of global color and texture descriptors for web image retrieval. J Vis Commun Image Represent 23:359–380. https://doi.org/10.1016/j.jvcir.2011.11.002

Pentland A, Picard RW, Scaroff S (1996) Photobook: content-based manipulation for image databases. Int J Comput Vis 18:233–254. https://doi.org/10.1007/BF00123143

Rao MB, Rao BP, Govardhan A (2011) CTDCIRS: content based image retrieval system based on dominant color and texture features. Int J Comput Appl 18(6):40–46. https://doi.org/10.5120/2285-2961

Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, Huang Z, Karpathy A, Khosla A, Bernstein M, Berg AC, Fei-Fei L (2015) ImageNet Large Scale Visual Recognition Challenge. IJCV https://arxiv.org/abs/1409.0575

Smith JR, Chang SF (1996) VisualSeek: a fully automatic content based query system. Proceedings of the Fourth ACM International Conference on Multimedia. 87–98. https://www.ee.columbia.edu/ln/dvmm/publications/96/smith96f.pdf

Su WT, Chen JC, Lien JJJ (2010) Region-based image retrieval system with heuristic pre-clustering relevance feedback. Expert Syst Appl 37:4984–4998. https://doi.org/10.1016/j.eswa.2009.12.015

Wan J, Wang D, Hoi SCH, Wu P, Zhu J, Zhang Y, Li J (2014) Deep learning for content-based image retrieval: a comprehensive study. Proc of the ACM Int Conf on Multimedia. https://doi.org/10.1145/2647868.2654948

Wang JZ, Li J, Wiederhold G (2001) SIMPLIcity: semantics-sensitive integrated matching for picture libraries. IEEE Trans Pattern Anal Mach Intell 23:947–963. https://doi.org/10.1109/34.955109

Xu R, Wunsch D (2005) Survey of clustering algorithms. IEEE Trans on Neural Netw 16(3):645–677. https://doi.org/10.1109/TNN.2005.845141

Yang J, Parikh D, Batra D (2016) Joint unsupervised learning of deep representations and image clusters. IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 1-19. https://arxiv.org/abs/1604.03628

Yildizer E, Balci AM, Jarada TN, Alhajj R (2012) Integrating wavelets with clustering and indexing for effective content-based image retrieval. Knowl-Based Syst 31:55–66. https://doi.org/10.1016/j.knosys.2012.01.013

Younus ZS, Mohamad D, Saba T, Alkawaz MH, Rehman A, Al-Rodhaan Z, Al-Dhelaan A (2015) Content-based image retrieval using PSO and k-means clustering algorithm. Arab J Geosci 8:6211–6224. https://doi.org/10.1007/s12517-014-1584-7

Yue J, Li Z, Liu L, Fu Z (2011) Content-based image retrieval using color and texture fused features. Math Comput Model 54:1121–1127. https://doi.org/10.1016/j.mcm.2010.11.044