



Clustering method and sine cosine algorithm for image segmentation

Lahbib Khrissi¹ · Nabil El Akkad^{1,2} · Hassan Satori¹ · Khalid Satori¹

Received: 19 February 2020 / Revised: 7 October 2020 / Accepted: 1 December 2020 / Published online: 3 January 2021
© The Author(s), under exclusive licence to Springer-Verlag GmbH, DE part of Springer Nature 2021

Abstract

This article presents a new image segmentation approach based on the principle of clustering optimized by the meta-heuristic algorithm namely: SCA (Algorithm Sinus Cosine). This algorithm uses a mathematical model based on trigonometric functions to solve optimization problems. Such an approach was developed to solve the drawbacks existing in classic clustering techniques such as the initialization of cluster centers and convergence towards the local optimum. In fact, to obtain an “optimal” cluster center and to improve the image segmentation quality, we propose this technique which begins with the generation of a random population. Then, we determine the number of clusters to exploit. Later, we formulate an objective function to maximize the interclass distance and minimize the intra-class distance. The resolution of this function gives the best overall solution used to update the rest of the population. The performances of the proposed approach are evaluated using a set of reference images and compared to several classic clustering methods, like k-means or fuzzy c-means and other meta-heuristic approaches, such as genetic algorithms and particle swarm optimization. The results obtained from the different methods are analyzed based on the best fitness values, PSNR, RMSE, SC, XB, PC, S, SC, CE and the computation time. The experimental results show that the proposed approach gives satisfactory results compared to the other methods.

Keywords Clustering · Classification · Image segmentation · Sine cosine algorithm · Metaheuristic

1 Introduction

Image processing is a discipline of computer science and applied mathematics that studies digital images and their transformations in order to improve their quality or extract information from them. It is used in several fields, namely 3D reconstruction [1], medicine [2], camera self-calibration [3], security [4, 5]. In an image processing system, the most important operation is image segmentation. It is an important task in any image processing and analysis process. This task is present in several applications of computer vision, such as medical imaging, video analysis, remote sensing, biology, etc. Today, there is no

universal method of image segmentation. Any technique is only effective for a given type of image, for a given type of application and in each IT context. Due to these constraints, the various image processing strategies that have been proposed have affirmed their inadequacies and their limits. It is therefore natural to explore new horizons and find new, more flexible and effective methods. Image segmentation techniques are generally based on the search for local discontinuities (edge detection) or the detection of areas of the image with homogeneous characteristics (region extraction). Therefore, Image segmentation aims to partition an image into several regions that do not overlap and are consistent with the requirements of different applications. It is always one of the most difficult tasks in understanding the image in image processing. In the literature, several approaches to image segmentation [6] have been proposed, but none of them is robust enough for many images. Image segmentation methods include clustering [7, 8], Split/Merge [9], Region Growth [10], Active Contour [11], Graph Cut [12], Genetic Algorithms [13], Markov Random Field [14], and others. Among these techniques, clustering (an unsupervised learning method)

✉ Lahbib Khrissi
lahbib.khrissi@usmba.ac.ma

¹ LISAC, Department of Computer Science, Faculty of Science, Dhar El Mahraz, Sidi Mohamed Ben Abdellah University, Fez, Morocco

² Laboratory of Engineering, Systems and Applications, ENSA of Fez, Sidi Mohamed Ben Abdellah University, Fez, Morocco

is one of the most commonly used image segmentation methods because of its efficiency and speed. The purpose of clustering is to divide a set into several clusters so that the members of the same cluster can be similar, and the elements of different clusters are different. There are two types of clustering: non-hierarchical clustering (partitioning) [15, 16], and Hierarchical clustering [17]. One of the most used non-hierarchical clustering algorithms is k-means. It is simple and faster to compute than is hierarchical clustering. It can also work for many variables [15]. The different types of segmentations available pose a major problem in the evaluation of image segmentation, which implies a fundamental conflict between generality and objectivity [18].

In order to improve the segmentation results, bio-inspired or nature-inspired methods were, then, used to solve the problem posed. In recent years, computer scientists have found in the natural world an inexhaustible source of inspiration for the design of new computer systems. It is a question of drawing from the behaviors of natural phenomena new approaches to solving difficult problems. These stochastic global optimization algorithms can be applied to any problem if it is formulated in the form of criteria optimization. These algorithms are inspired by analogies with physics (simulated annealing, annealing), biology (evolutionary algorithms) or ethology (ant colonies, particle swarms). There are several methods that use metaheuristics to improve the quality of image segmentation. For example, GHOSH et al. [19] proposed a method based on genetic algorithms (GA), Yang et al. [20] presented an algorithm based on the search for harmony (HS) and Vijay et al. [21], submitted a technique based on particle swarming (PSO) for image segmentation.

On the other hand, a recent metaheuristic which is called SCA has appeared. It is a new population-based optimization algorithm for solving optimization problems. In population-based optimization techniques usually start with a set of random solutions. This set of random solutions is evaluated several times using an objective function. These solutions are improved by following a set of rules, which in turn, are the ones that form the basis of this optimization technique. SCA is an optimization algorithm that uses a mathematical model based on the sine and cosine functions to carry out the exploitation and exploration of the research space in order to solve

optimization problems. SCA has been tested on many reference functions, showing better performance in terms of convergence speed, optimal solution search precision and stability compared to approaches like in [22].

In this article, a new image segmentation technique based on clustering using SCA is presented. We are particularly interested in the SCA algorithm which is extended to solve the clustering problem as an optimization problem. The aim is to use the advantages of the sine cosine algorithm such as the low number of parameters and lack of optimal local trapping to solve clustering problems, and our main goal is to consolidate untagged data using the SCA algorithm so that we can get better results with simple solutions and perform a full search compared to existing methods in the literature. The proposed automatic clustering method begins with the reception of the data set and then initializes the set of solutions. Later, we determine the number of clusters to operate and which represents the positions of the centroids. The next step is to calculate the fitness function of each solution to maximize the interclass distance and minimize the intra-class distance. Thereafter, the population is updated by using the SCA operators to carry out the operations of the exploitation and exploration of the research space, which gives them a great capacity to find the optimal solution. Then, the previous two steps are repeated until the termination conditions are met and return the best solution representing the best one to the clustering problem. The performance of the proposed algorithm was tested on different types of images and compared with several classic and metaheuristic clustering algorithms in order to have the best result of image segmentation by clustering using SCA.

The main contributions of this document can be summarized as follows:

- Proposal for a clustering image segmentation approach based on the SCA algorithm.
- Determination of both the number of clusters and the center of each cluster.
- Improvement of operating capacity by using SCA operators.
- Evaluation of the performance of the proposed method using a set of metric measures

- A comparison is provided between the proposed method and other classical clustering and meta-heuristic algorithms.

The rest of this article is organized as follows: In the second part, we discuss related work. The third part deals with the theoretical framework. The proposed clustering method by SCA is presented in the fourth part. The fifth part is devoted to the presentation of experimentations and the discussion of the results obtained, and the conclusion is addressed in the last part.

2 Related work

To date, there have been many works in the field of image segmentation using different methods depending on the field of application. Among these methods, we find unsupervised classification (clustering) as a data analysis and exploration technique which aims to group data into a set of classes or clusters, so that the objects of the same class have a strong similarity to each other and differ greatly from objects of other classes. The data clustering methods which have been very successful are: The k-means algorithm and the fuzzy c-means (FCM) algorithm, but the solution to these algorithms depend on the initial random state and always converge to the local optimum [23].

Due to the large amount of information and the complexity of the problems, classical optimization methods are unable to solve most optimization problems; that is why optimization metaheuristics and image segmentation are among the most active areas of research. In recent years, several works have created a bridge between these two fields of research. Indeed, the image segmentation problem is formalized as a single-criteria or multi-criteria combinatorial optimization problem, where the use of metaheuristics allows finding acceptable solutions within a reasonable time. Meanwhile, the clustering method is often seen as an optimization problem, consisting in finding a partition of objects that optimizes a given criterion [24]. Many researchers are also trying to produce new nature-inspired methods that are more effective than existing methods and show a better result in image segmentation. Among these are:

Pham et al. [25] used an improvement method for the image segmentation problem using particle swarm optimization (PSO) with a new objective function based on the core of an improved fuzzy entropy clustering algorithm with local spatial information, called PSO-KFECS. The proposed approach was evaluated on several benchmark test images, including synthetic and MRI images. Experimental results show that the proposed PSO-KFECS algorithm can work better than competing algorithms.

Benaichouche et al. [26] presented a technique for improving image segmentation using the fuzzy c-means clustering (FCM) algorithm. The proposed method, called IFCMS spatial blur enhanced c-means, has been evaluated on several test images, including synthetic and MRI images). This method is compared to the most widely used FCM algorithms in the literature. The results demonstrate the effectiveness of the ideas presented by the authors.

SINHA et al. [27] described a new clustering algorithm for distributed data sets, using a combination of the genetic algorithm (GA) with Mahalanobis distance and the k-means clustering algorithm. Experiments were conducted for several synthetic and real data sets to measure the efficiencies of the proposed algorithm. The results were compared with algorithms based on MapReduce, mrk-means, parallel k-means and GA scaling.

Maulik et al. [28] proposed a new approach of automatic fuzzy clustering based on a modified differential evolution that automatically changes the number of clusters and the appropriate partitioning from a data set. The assignment of points to different classes is based on a Xie-Beni index, where the Euclidean distance is considered. Indeed, the technique proposed by the authors to identify different land cover regions in remote sensing imagery shows efficiency compared to other existing techniques such as automatic clustering based on an improved differential evolution, fuzzy automatic clustering based on a classical differential evolution and fuzzy clustering based on a genetic algorithm.

Niknam et al. [29] presented a new hybrid scalable algorithm to solve the problem of nonlinear partitioning clustering. The proposed method is the combination of the algorithms FAPSO (Fuzzy Adaptive Particle Swarm Optimization), ACO (Ant Colony Optimization) and k-means,

called FAPSO-ACO-K, which can find a better cluster partitioning. The efficiency of the proposed algorithm is evaluated, and the results show that the performance of the proposed algorithm is better than other algorithms such as PSO, ACO, simulated annealing (SA)..., for the partial clustering problem.

Recently, Mirjalili et al. [22] have proposed a new population-based meta-heuristic optimization algorithm to solve optimization problems, called Sine Cosine Algorithm (SCA). The SCA has been used in a wide range of applications. For example, Li et al. [30] propose an improved sine cosine algorithm based on sample theft. The main idea is the following: Individuals who do not score progress will update their position using the variable Levy flight parameters in order to improve the ability to explore. This algorithm has shown good performance in complex nonlinear problem optimization. In the FS (features selection) domain, Hafez et al. [31] have proposed SCA as a feature selection system. SCA has been implemented and tested on 18 datasets. The experiments show that SCA is ahead of other research methods such as PSO and GA. In addition, Emary et al. [32] used the Lévy flight technique (LF) to improve the exploration capacity of an individual in SCA; he was able to approximate the satisfactory solutions and decrease the probability of getting stuck around local optima. Finally, Wang et al. [33] used SCA to develop an efficient wind speed forecasting system which simultaneously obtained high accuracy and high stability. With its unique optimization principles and theoretical advantages, the SCA algorithm has proven to be superior to PSO, genetic algorithm (GA) and other existing algorithms in most examples and has been applied in many practical problems. SCA has been recognized as being sufficiently competitive with other meta-heuristic methods well known in the literature. Based on the characteristics and advantages of the SCA algorithm, we are motivated to take advantage of their operators to propose an alternative method of clustering to find out a good quality of image segmentation.

3 Theoretical framework

3.1 Clustering problem formulation

Clustering problem is the search for distinct groups in the feature space. It is expected that these groups have different structures and that they can be clearly differentiated. The clustering task separates the data into number of partitions, which are in the N-dimensional feature space.

This process is achieved by minimizing the within-cluster variation (or intra-cluster distance), while maximizing the between-cluster variation (or inter-cluster distance) [34].

This problem can be formulated by assuming that the dataset consists of N_S samples $S = [s_1, s_2, \dots, s_{N_S}]$. Each sample has D features (i.e., $s_i = [s_{i1}, s_{i2}, \dots, s_{iD}]$). Then, any clustering algorithm aims to divide the data into K_{max} clusters (i.e., $C_1, C_2, \dots, C_{K_{max}}$) subject to [34]:

$$\bigcup_{k=1}^{K_{max}} C_k = S, C_k \neq \emptyset, \quad k = 1, 2, \dots, K_{max} \quad (1)$$

$$C_k \cap C_{k1} = \emptyset, k, k1 = 1, 2, \dots, K_{max}, \quad k \neq k1 \quad (2)$$

The value of K may or may not be known a priori. The partition matrix is represented as $U = [u_{kl}]$, $k = 1, 2, \dots, K_{max}$ and $l = 1, 2, \dots, N$, where u_{kl} is the membership of data point x_j to cluster C_k . For the hard partitioning of the dataset, the following condition must be satisfied [35].

$$u_{kl} = \begin{cases} 1 & \text{if } x_l \in C_k \\ 0 & \text{if } x_l \notin C_k \end{cases} \quad (3)$$

For the fuzzy partitioning of the dataset, the following condition must be satisfied.

$$0 < \sum_{l=1}^N u_{kl} < N \quad \forall k \in \{1, 2, \dots, k_{max}\} \quad (4)$$

3.2 The sine cosine algorithm (SCA)

The sine cosine algorithm (SCA) proposed by [22] could be a modern optimization strategy to solve optimization problems. The SCA makes some irregular candidate arrangements at the beginning and forces them to fluctuate outwards or towards the finest arrangement using a numerical demonstration based on sine and cosine capabilities. In addition, some arbitrary and versatile factors are coordinates in this algorithm to focus on investigation and misuse of space at distinct stages of optimization.

The algorithm of sine cosine is a stochastic algorithm [22] that contains several iterations. We update, for each iteration, the solutions according to the sine or cosine function by the following expressions:

$$X(i, j)_{t+1} = X(i, j)_t + r_1 * \sin(r_2) * |r_3 * P(j)_t - X(i, j)_t| \quad (5)$$

$$X(i, j)_{t+1} = X(i, j)_t + r_1 * \cos(r_2) * |r_3 * P(j)_t - X(i, j)_t| \quad (6)$$

Such as,

$X(i, j)_t$: the position of the current solution (i) in dimension (j) at the iteration (t).

$P(j)_t$: the position of the best individual in dimension (j) at the iteration (t),

$|\cdot|$: the absolute value of (e).

r_1, r_2 and r_3 are three random variables.

The two expressions (5) and (6) presented below give the following formula:

$$X(i, j)_{t+1} = \begin{cases} X(i, j)_t + r_1 * \sin(r_2) * |r_3 * P(j)_t - X(i, j)_t| & \text{if } r_4 < 0,5 \\ X(i, j)_t + r_1 * \cos(r_2) * |r_3 * P(j)_t - X(i, j)_t| & \text{if } r_4 \geq 0,5 \end{cases} \quad (7)$$

Such as $r_4 \in [0,1]$.

Expression (7) represents a circular search model [22].

In this model, the best solution is in the center of a circle and the search area surrounds this solution. This area is divided into sub-areas representing possible exploration areas for X_i . The parameters r_1, r_2, r_3 and r_4 are defined as follows:

- r_1 controls how X_i varies in these areas. In fact, if $r_1 > 1$ then X_i moves towards P. Otherwise, this point moves away from P. In addition, r_1 is used in the balancing of the exploration and exploitation phases, since it allows to be reduced by the expression (8).
- r_2 controls how far X_i moves along its direction according to r_1 .
- r_3 brings a random weight to the destination in order to stochastically emphasize ($r_3 > 1$) or de-emphasize ($r_3 < 1$) the destination effect in the distance definition.
- r_4 switches between the sine and cosine components of Eq. (7).

The random number r_1 determines the exploration phase (when $r_1 > 1$) or exploitation phase (when $r_1 < 1$). The value of which is adopted in the SCA is given by the formula (8):

$$r_1 = a - a * \frac{t}{T_{\max}} \quad (8)$$

where T_{\max} is the maximum number of iterations to be given as a stopping criterion of the algorithm. t is the current iteration and a is a constant.

The SCA algorithm is given in Algorithm 1.

Algorithm1: Sine Cosine Algorithm.

1. Initialize a set of random agents
 2. Evaluate the fitness of solutions using objective function.
 3. Select the best solution P
 4. Initialize r_1, r_2, r_3 and r_4
 5. Initialize the generation count $t = 0$
 6. **While** ($t < T_{\max}$)
 7. **For each** candidate solution **do**
 8. Update the solution using equations (7) and (8)
 9. **End for**
 10. Calculate the objective function value of update solution
 11. Update the best solution P
 12. Update the parameters r_1, r_2, r_3 and r_4
 13. $t = t + 1$
 14. **End while**
 15. Return the best solution P.
-

4 Proposed clustering method

The SCA has proven to be superior to some existing traditional optimization algorithms because of its unique optimization principle which is based on a mathematical model (sine and cosine functions) and because it is easy to implement. It is characterized by a good convergence speed, search precision and stability. Because of its benefits, it has been used in clustering to improve the quality of image segmentation. In the proposed approach, we on the one hand, explore and describe the applicability of the sine cosine algorithm to the development of the clustering technique and, on the other hand, we show that the SCA is capable of producing the optimal cluster centers. For clustering, data from the same cluster must have higher similarities while data from different clusters have higher dissimilarities [36].

In this part, we deal with and solve some image segmentation problems using our proposed clustering approach. This approach uses a clustering technique based on the SCA algorithm in order to find combinations of characteristics which maximize the inter-class distance and minimize the intra-class distance. The search space of each characteristic, represented in an individual dimension, and the scope of each dimension ranging from 0 to 1 are very huge and, therefore require an intelligent search method. Indeed, the aim of the latter is to find an optimal point in the search space which minimizes the fitness function that we formulated in expression (9).

The main steps of the algorithm proposed by our approach are presented in algorithm 2 below:

Algorithm 2: Algorithm of the proposed method

1. Initialize the problem and the parameters of the algorithm.
2. Evaluate the fitness of solutions using objective function.
 % Use of the fitness function described in equation (9) to evaluate the performance of the clustering algorithm.
3. Initialize the search agents (solutions) X_k ($k = 1, 2, 3, \dots, m$).
 % The search areas are initialized with randomly selected image pixels from the given pixel set.
4. Select the best solution X_{best} from the population of solutions.
5. Initialize the generation count $t = 0$
6. Initialize the max numbers of iteration T_{max}
7. **While** ($t < T_{max}$) **do**
 % Based in equation (9), the performance of search agents is calculated using the fitness function.
8. **for** each solution X_k **do**
9. Evaluate objective function $f(X_k)$
 % The position of the best search agent is updated using the fitness values.
10. **if** $f(X_k) < f(X_{best})$ **then**
11. $X_{best} = X_k$
12. **end if**
13. **end for**
 %The position of the search agents is updated using equation (8).
14. Update the parameter r_1
15. Updated the parameters (r_2, r_3 and r_4).
16. **for** each solution X_k **do**
17. update X_k
18. **End for**
19. $t = t + 1$
20. **end while**
21. Return the best solution X_{best} .

The calculation is complete when the stoppage criterion is met. Otherwise, steps 3 and 4 are repeated. Finally, the best search agent is selected and is considered the best cluster center of the set of points in the image.

First, the process of image segmentation according to the proposed approach, is started by the reception of the data set, and then initializes the set of solutions. Next, we determine the number of clusters to operate and which represents the positions of the centroids. The random distribution of individuals in the clusters is evaluated by a fitness function of each solution to maximize the interclass distance and minimize the infra-class distance. Then, the population is updated by using the SCA operators to carry out the operations of the exploitation and exploration of the research space, which gives them a great capacity to find the best overall solution. The process was repeated until convergence towards a global solution was

reached, which represents the solution to the clustering problem, or the stop conditions were fulfilled.

In general, to calculate similarities between data elements, the Euclidean distance calculation is applied. The calculation of the distance is done in the following way:

Given N objects, each element is first selected from one of the K clusters and then the sum squared of the Euclidean distances between each element and the center of the cluster is minimized and deducted from each allocated element.

Our proposed method uses SCA to find feature combinations that maximize classification performance and minimize the selected features. The search space for each agent represented in a single dimension and the range of each dimension from 0 to 1 is very large and requires an intelligent search method to find an optimal point in the search space that minimizes the specified physical entity.

In this paper, we formulate a non-linear cost function to solve the clustering problem. This function (9) is cited below. The minimization of this cost function, which is one of the main objectives of Clustering, leads to a greater similarity in each cluster and to the difference with others.

$$f(x_i, c_j) = \sum_{i=1}^N \sum_{j=1}^K \|x_i - c_{ij}\|^2 \quad (9)$$

where c_{ij} refers to the cluster center vector of the i th member of the j th cluster (c_{ij}).

N and K are the number of picture elements and the number of clusters, respectively.

The main objective of the proposed SCA-based clustering method is the search for suitable cluster centers with minimal clustering criteria.

Our proposed approach uses the sine cosine algorithm (SCA) to find combinations of characteristics that maximize the inter-class distance and minimize the intra-class distance. The search space with each characteristic represented in an individual dimension and the scope of each dimension ranging from 0 to 1 are very huge and, indeed, require an intelligent search method to find an optimal point in the search space which minimizes the fitness function given in expression (9). That is to say, our method follows the conventional search strategy similar to that of GA and PSO, as it initializes the process of optimization with a random set of research agents (individuals). Therefore, the search procedure of the proposed algorithm (algorithm 2) initially generates a random set of search agents whose aptitude is evaluated using a fitness function (9). Classification error rate and after a random initialization, the quality of the search agents is measured using the objective function. Upon evaluating the fitness function of the initial population, the best solutions that have been initiated so far will be considered the best, which will be used to evaluate the solutions to come. In addition, to highlight the exploitation of the search space, the values of the parameters r_1 , r_2 , r_3 and r_4 are updated at each iteration. Our technique achieves the expected solution by undergoing several iterations (generations). The ranges of sine and cosine functions will be updated as the iteration counter increases. Finally, by obtaining the best optimal solution for clustering, the algorithm stops satisfying the termination criteria.

The proposed algorithm is presented in the form of the flowchart in Fig. 1.

4.1 Time complexity analysis

The time complexity of an algorithm is a function that qualitatively describes the execution time of the algorithm. For the basic SCA algorithm (algorithm 1), the time complexity of lines 1 (initialization), 2 (evaluation

of the fitness function) and 3 (selection of the best solution) are $\mathcal{O}(m)$, $\mathcal{O}(N)$ (see formula (9), where K is a constant) and $\mathcal{O}(m)$, respectively. Moreover, the time complexity of the “for loop” of line 7 to 9 is $\mathcal{O}(m * N)$, the time complexity of lines 10 and 11 are $\mathcal{O}(N)$ and $\mathcal{O}(m)$, respectively. Thus, the time complexity of the “while loop” is $\mathcal{O}(T_{max} * \max\{m * N; N; m\}) = \mathcal{O}(T_{max} * m * N)$. Consequently, the time complexity of algorithm 1 is $\mathcal{O}(\max\{m; N; m; T_{max} * m * N\}) = \mathcal{O}(T_{max} * m * N)$.

In the algorithm of our approach (algorithm 2), the time complexity of the Construction of the initial population is $\mathcal{O}(N)$ and the time complexity of lines 2, 3 and 4 are $\mathcal{O}(N)$, $\mathcal{O}(m)$ and $\mathcal{O}(m)$, respectively. Moreover, the time complexity of line 9 is $\mathcal{O}(N)$ (see formula (9), where K is a constant). Then, the time complexity of the first “for loop” is $\mathcal{O}(m * N)$. Furthermore, the time complexity of the second “for loop” is $\mathcal{O}(m)$. Consequently, the time complexity of the “while loop” is $\mathcal{O}(T_{max} * \max\{m * N; m\}) = \mathcal{O}(T_{max} * m * N)$. Thus, the time complexity of our approach (i.e. algorithm 2) is $\mathcal{O}(\max\{N; m; m; T_{max} * m * N\}) = \mathcal{O}(T_{max} * m * N)$. We note that the time complexity of our approach has the same order of magnitude as that of the standard SCA algorithm. Additionally, we evaluate our approach in terms of the execution time in the experimentation part.

5 Experimental results

In order to validate our approach and show its performance and efficiency in the field of image segmentation, we have conducted several experiments on reference images. In addition, we compared our approach with four other clustering methods giving good results in the literature, namely: two metaheuristic algorithms which are the particle swarm optimization algorithm (PSO) [37] and the genetic algorithms (GA) [38], and two traditional algorithms for clustering which are the k-means algorithm and the c-fuzzy algorithm (FCM). The experiment is of carried out on the basis data from Berkeley Image Segmentation [39]. The parameters of these methods were determined empirically and their values are presented in Table 1. The results were calculated after running each algorithm 20 times. In addition, the population size was set at 30 [40], the maximum iteration was set at 200 for each method. The experiment was carried out on a Windows 10 64-bit platform on a 3rd generation Intel Core (TM) processor with 4 GB of RAM using MATLAB 2014b.

The obtained results from the different methods were analyzed on the basis of the best fitness values, PSNR, RMSE, SC, XB, PC, S, SC, CE. We carried out several tests by modifying the number of clusters k . To avoid putting all these different tests in this part, we will only present the results obtained using $k = 4$.

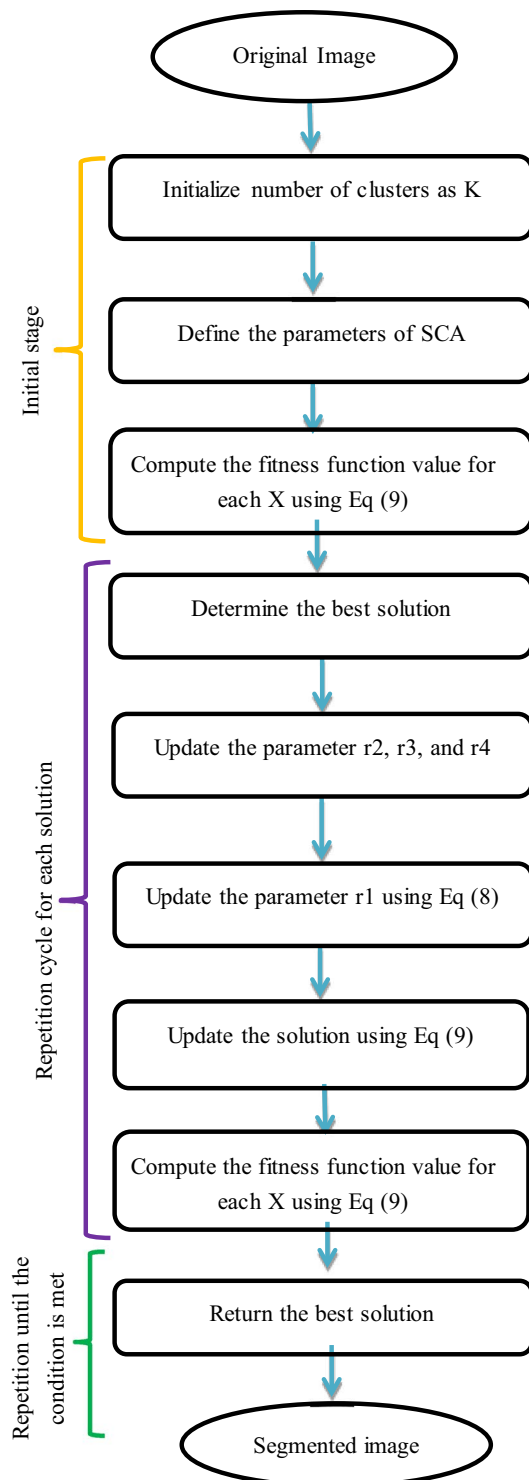


Fig. 1 The clustering flowchart of the proposed approach

Table 1 The parameters of the algorithm and their values

Algorithm	Parameter	Value
GA	Cross over percentage	0.8
	Mutation percentage	0.3
	Mutation rate	0.02
PSO	Maximum velocity (V_{max})	4.0
	Cognitive coefficient (C1)	1.49
	Cognitive coefficient (C2)	1.49
	The inertia weight	0.72
Proposed approach	a	2.0
	r_1, r_2, r_3 and r_4	Random numbers

To present the measures of the evaluation criteria of our proposed approach, we will focus in this section on clustering for $k = 4$.

Figure 2 shows that our approach outperforms other methods.

The quality of the images segmented by the different methods is analyzed using the RMSE measurement values and the signal-to-noise ratio (PSNR) values [41, 42].

The root mean square error is used as a standard performance measure of the output image. It gives the amount of output image that deviates from the input image, which is expressed as follows:

$$RMSE = \sqrt{\frac{\sum_0^{n_x-1} \sum_0^{n_y-1} [I_{in}(x, y) - I_{seg}(x, y)]^2}{n_x n_y}} \quad (10)$$

where n_x and n_y are the width and the height of the image respectively.

The signal-to-noise ratio represents the proportion between the maximum powers that can be reached and the parasite noise which influences the resemblance of the image. It is used to measure the quality of the output image formulated in the following formula:

$$PSNR = 20 \cdot \log_{10} \left[\frac{\max(I_{in}(x, y))}{RMSE} \right] \quad (11)$$

where $I_{in}(x, y)$ is the input image and $I_{seg}(x, y)$ is the segmented image. On the one hand, the lower value of RMSE means that the image is of good quality and, on the other hand, a smaller value of PSNR means that the image is of poor quality. The RMSE and PSNR values of the segmented images are calculated by the classical algorithms of k-means and FCM as well as by meta-heuristic algorithms (GA clustering, PSO clustering and SCA clustering). Based on the results (Figs. 3 and 4), we note that the RMSE values obtained by our approach are very small. On the other hand, the values of the PSNR obtained by our approach are very

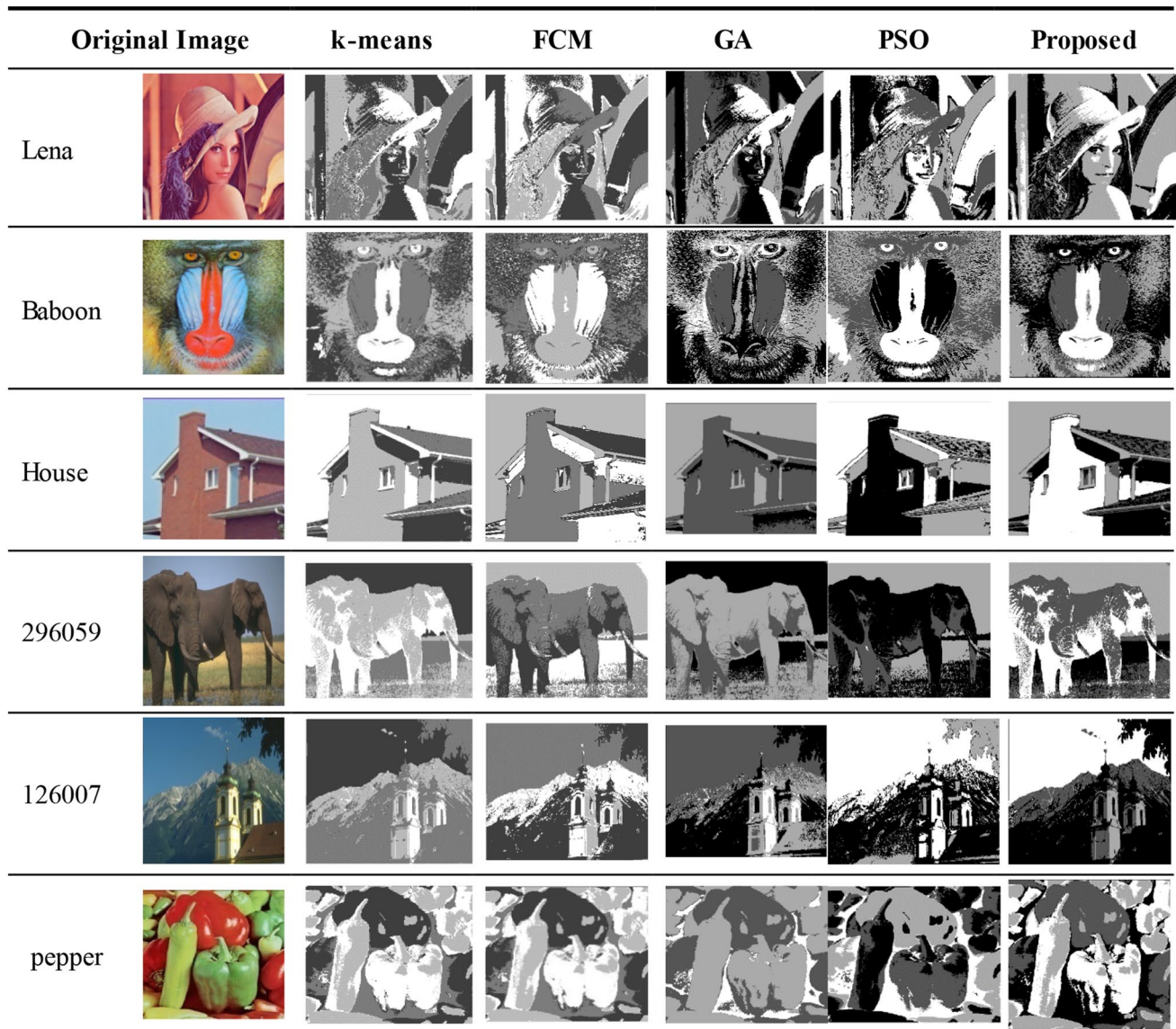


Fig. 2 Image segmentation results by the different clustering algorithms

large. Therefore, we can say that our approach gives good results, which shows that the image segmented by the proposed method is of a better quality.

Figure 3, below, shows the RMSE values obtained by the five methods as a function of the test images.

Figure 4, below, shows the PSNR values obtained by the five methods as a function of the test images.

From Fig. 4, we notice that the RMSE values obtained by our approach are low compared to those obtained by the other four methods. In addition, from Fig. 5, we can see that the PSNR values obtained by our approach are higher than those obtained by the other four methods. From the previous analysis of the result of the proposed approach, the sine cosine optimization algorithm is very useful in this problem and able to avoid the main

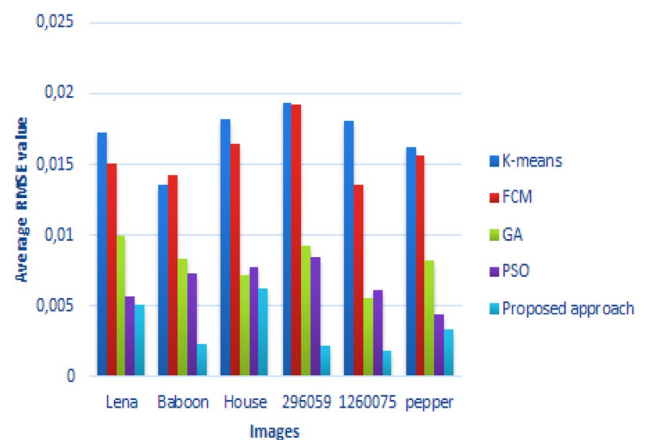


Fig. 3 RMSE values obtained by the five methods

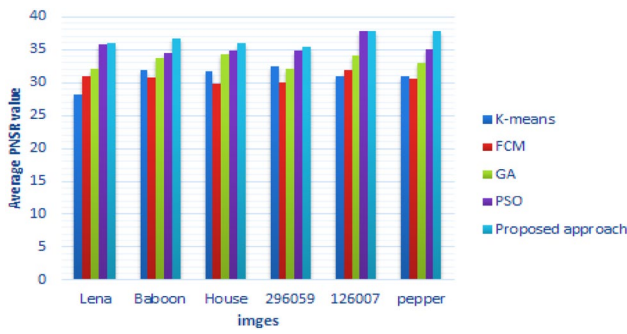


Fig. 4 PSNR values obtained by the five methods

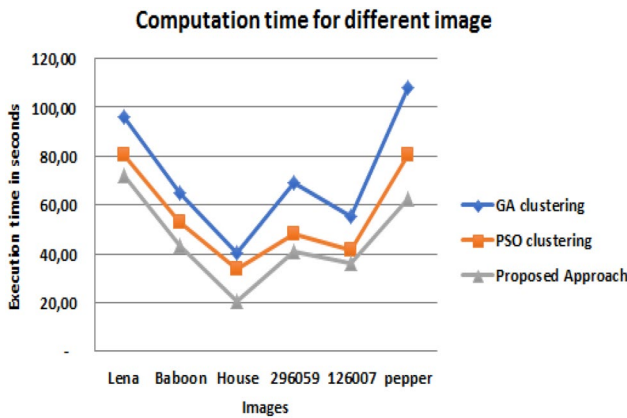


Fig. 5 Comparison of the execution time of different algorithms

disadvantage of traditional clustering methods such as k-means or FCM (local minima problem). We can, therefore, conclude that our approach gives satisfactory results compared to those obtained by the other methods. This means that the proposed approach gives better quality segmented images.

There are other quantitative assessment techniques (commonly known as the cluster validity method) to test the quality of the cluster. Internal quality compares different sets of clusters without any reference to external knowledge. A good clustering technique has a strong similarity within the cluster and a weak inter-cluster similarity. Cluster validation [43] is the technique for estimating the adequacy of a partition with the data structure. Many criteria have been developed to determine the validity of clusters, all with a common objective of finding clustering which results in compact, well-separated clusters. Moreover, validity indices are independent of clustering algorithms as explained in [44]. The partition that best matches the data is selected by running the clustering algorithm multiple times with a different K value for each run. In general, these approaches require running the selected clustering algorithm multiple times, and the clustering partitions that optimize the validity index are selected

as the best partition. In order to show the quality of our approach, we study the quality of the segmented image using the five measures SC, XB, CE, S, PC [45, 46], presented in the expressions (12), (13), (14), (15) and (16).

- The sub-area coefficient (SC) measures the ratio of the sum of the compactness and the separation of the clusters:

$$SC = \sum_{i=1}^c \frac{\sum_{k=1}^n (u_{ik})^m \|x_k - v_i\|^2}{n_i \sum_{j=1}^c \|v_j - v_i\|^2} \tag{12}$$

where n_i is the number of samples that belongs to clustering i .

- The Xie and Beni index (XB) determines the ratio of the total variation of the interior of a cluster to the separation of the clusters:

$$XB = \frac{\sum_{i=1}^c \sum_{k=1}^n (u_{i,k})^m \|x_k - v_i\|^2}{n \min_{ik} \|x_k - v_i\|^2} \tag{13}$$

- The classification entropy (CE) calculates the blurring of the cluster partition:

$$CE = -\frac{1}{n} \sum_{i=1}^c \sum_{k=1}^n u_{ik} \log u_{ik} \tag{14}$$

- The separation index (S) uses a minimum separation distance for the validity of the partition:

$$S = \frac{\sum_{i=1}^c \sum_{k=1}^n (u_{i,k})^m \|x_k - v_i\|^2}{n \min_{ik} \|v_k - v_i\|^2} \tag{15}$$

- The partition coefficient (PC) determines the amount of overlap between clusters:

$$PC = \frac{1}{n} \sum_{i=1}^c \sum_{k=1}^n (u_{ik})^2 \tag{16}$$

A clustering method is of better performance if the values of SC, XB and CE are low while the values of S and PC are high.

The metric values of the images presented in Fig. 2 are given in Tables 2, 3, 4, 5, 6 and 7.

Table 2 Comparison of the clustering methods on Lena image

	k-means	FCM	GA	PSO	Proposed
SC	1.0084	1.1032	0.0318	0.0299	0.0237
CE	0.8871	0.9006	0.5391	0.4297	0.4335
PC	0.4916	0.5318	0.7384	0.7586	0.7745
XB	665.9231	550.7512	527.5608	516.2461	516.3101
S	0.049	0.0482	0.00546	0.0585	0.1678

Table 3 Comparison of the clustering methods on Baboon image

	k-means	FCM	GA	PSO	Proposed
SC	0.9375	1.0667	0.0511	0.0260	0.0211
CE	0.7931	0.9885	0.4933	0.4314	0.3920
PC	0.6316	0.5388	0.6994	0.7764	0.8244
XB	3.8992e+03	3.8970e+03	3.7862e+03	2.8192e+03	2.0143e+03
S	0.0569	0.0651	0.0482	0.0479	1.3697

Table 4 Comparison of the clustering methods on House image

	k-means	FCM	GA	PSO	Proposed
SC	0.7932	0.8824	0.0330	0.0212	0.0212
CE	0.7712	0.6187	0.4221	0.3123	0.3210
PC	0.6642	0.6341	0.8643	0.8132	0.8610
XB	2.9982e+03	2.7962e+03	2.5982e+03	2.2382e+03	1.9752e+03
S	0.0567	0.0528	0.0545	0.0609	0.0698

Table 5 Comparison of the clustering methods on 296,059 image

	k-means	FCM	GA	PSO	Proposed
SC	0.7932	0.8824	0.0330	0.0212	0.0212
CE	0.7712	0.6187	0.4221	0.3123	0.3210
PC	0.642	0.6341	0.8643	0.8132	0.8610
XB	1.3182e+03	1.3197e+03	1.3154e+03	1.3128e+03	1.2682e+03
S	0.0576	0.0577	0.0576	0.0575	0.0578

Table 6 Comparison of the clustering methods on 126,007 image

	k-means	FCM	GA	PSO	Proposed
SC	0.0371	0.04756	0.0267	0.0212	0.0232
CE	0.4712	0.5187	0.4221	0.3923	0.3819
PC	0.7135	0.7241	0.7643	0.7332	0.7710
XB	413.4504	413.3804	543.9865	2.1982e+03	2.0982e+03
S	0.0561	0.0557	0.0601	0.0634	0.0709

The results of the different methods obtained from the five measurements SC, CE, XB, S and PC are presented in Tables 2, 3, 4, 5, 6 and 7, where the value set in bold in each line indicates the best result obtained among the five algorithms.

Tables 2, 3, 4, 5, 6 and 7 show the comparison between the proposed SCA-based clustering approach and the

techniques mentioned above in terms of cluster quality measures for the sets of test images. None of the algorithms can overwhelm the others in all metrics.

From the results of the first image Lena, we can see that the image obtained with our method is better than the others because its contour is sharper, especially the lip and eye areas. In addition, according to Table 2, the S and PC values of our proposed approach are slightly better than those of the other methods. But in terms of XB, clustering by PSO achieves the best clustering effect.

For the second Baboon image, our approach obtains the best values of the 5 metrics compared to the other algorithms depending on the definition of each metric. Thus, our approach obtains a good segmentation effect. Generally speaking, the image segmented by the proposed approach may be the best, because it shows obvious features, i.e. eyes and nostrils.

Table 7 Comparison of the clustering methods on pepper image

	k-means	FCM	GA	PSO	Proposed
SC	0.0381	0.0362	0.0301	0.0291	0.0243
CE	0.4712	0.4487	0.4221	0.4123	0.4010
PC	0.7042	0.7341	0.7643	0.7432	0.7660
XB	2.9982e+03	2.7962e+03	2.5982e+03	1.9752e+03	2.0211e+03
S	0.0469	0.0528	0.0555	0.0599	0.0698

From the results of the House image, we could see the obvious edges on the image of our method, and its XB value is much smaller than other methods. In terms of PCs, clustering by GA gets the best result.

On images 296,059 and 126,007, our algorithm shows better performance than other versions of clustering, and its images embody sharp edges and significant features.

For the last pepper image, the segmentation effect by our approach is quite obvious. Furthermore, according to Table 7, the S and PC values of our proposed approach are better than those of the other approaches. Therefore, our algorithm shows a preferable segmentation effect on most of the test images.

Generally, clustering by FCM and k-means has good PC values but has bad values of the other four metric values. Its XB values are very large while its S values are very small, which indicates that its partition quality is poor. GA clustering achieves a good XB value but displays ordinary results for the other four measurements. PSO clustering achieves relatively balanced results in terms of five measures, and the overall performance of SCA clustering is better in terms of five measures.

Altogether clustering based on SCA combines global migration and local migration to better balance exploration and exploitation than other versions, and thus obtain clustering results that better balance intra-class polymerization and inter-class difference thanks to its principle. This finds the sine and cosine functions, when the value of the function $r1 \cdot \sin(r2)$ or $r1 \cdot \cos(r2)$ is between -1 and 1 , the algorithm performs local operation, when the value of the function $r1 \cdot \sin(r2)$ or $r1 \cdot \cos(r2)$ is less than -1 or greater than 1 , the algorithm performs a global exploration.

With its unique optimization principles and theoretical advantages, clustering by the SCA algorithm has proven to be superior to PSO, GA and other existing algorithms in most of the examples and has been applied to many practical problems, like clustering. However, it still has some drawbacks, such as the low precision of the solutions. The reason is that when the SCA algorithm updates the individual, the next generation location of the individual is only expanded around its current optimal location. With the decrease in the diversity of the population in the last period of the algorithm, the overall search capacity deteriorates. But, with an increase in the number of iterations and random solutions, it is likely that the solution converges towards the global optimum.

In summary, our proposed approach efficiently segments the test images and obtains clear results, which shows that using the SCA algorithm to perform grouping is beneficial in improving the precision and accuracy of image segmentation. However, a disadvantage of SCA clustering is that its PC values are generally small. This indicates that the overlap between the resulting clusters is low. We can see that the

other clustering algorithms cannot significantly improve the PC values, mainly due to the intrinsically similar migration operators they use. Overall, as we can notice, our proposed approach obtains the best PC, S values and the worst values of XB, SC and CE. In particular, the SC and CE values of the proposed closure are close to zero, which indicates that the cohesion within the clusters is very high. In summary, the experimental results reveal that SCA-based clustering gives better values in terms of cluster quality measures.

In terms of computational costs, the literature shows that individual algorithms (i.e. local search) like k-means and FCM are characterized by their speed of convergence compared to population-based algorithms. Despite this advantage, they often fall into the trap of poor local solutions and give unsatisfactory results. For this reason, we compare below the execution time of our proposed approach with methods which are based on the population such as GA and PSO.

The execution time of each algorithm varies depending on the content of the image and the number of clusters that are we going choose. When the number of clusters increases, the execution time increases. In this article, the parameters chosen in the experiments for the selected algorithms are shown in Table 1. We run the algorithm 10 times for each image to calculate the average computational time. Figure 5 shows the comparison, in terms of average execution time, between the algorithms selected for each image.

From Fig. 5, we notice that the execution time of our proposed approach is smaller compared to the other methods (clustering by GA and clustering by PSO). In addition, when the image contains more details, the difference is significant. That means the difference in computational time between the selected algorithms is varied according to test images. Indeed, the execution time strongly depends on the content of the image.

Experimental results given in Fig. 5 above demonstrate that the difference rate of the execution time varies from one image to another. For example, for the Lena, Baboon and Pepper images, we observe that our approach allows us to save an average execution time of 25%, 34% and 42% respectively, compared to the clustering algorithm by GA, and an average execution time of 10%, 19% and 22% respectively, compared to the clustering algorithm by PSO. Overall, our proposed approach allows us to save an average execution time of 37% compared to the clustering algorithm by GA and an average execution time of 20% compared to the clustering algorithm by PSO on the images chosen. We also find that clustering by genetic algorithms cost more time than other clustering algorithms and this comes down to their complexity in time and search mechanism. In addition, qualitatively, we can say that our method is more stringent. Indeed, in Fig. 2, we can observe that our approach gives more details on the segmented image.

The experiments on the different reference images show a high performance of our approach not only to produce a reliable and good segmentation result, but also to achieve image segmentation in short execution time.

6 Conclusion

In this article, we addressed the problem of unsupervised classification of images using a new meta-heuristic optimization method. This method is based on sine and cosine mathematical functions to perform the exploitation and exploration of the search space. Besides, we have well established the segmentation of images by algorithm-based clustering (SCA), taking into consideration that the number of clusters is known in advance. In our approach, a set of individuals from a population is generated first. Each member of this set is a vector of length k with identifiers. Which refer to the centers of each of the k clusters and structure the final clustering solution. The proposed approach is compared with a set of methods based on several evaluation criteria.

The statistical analysis of experimental results on different images demonstrates a high performance of our approach, in terms of efficiency, reliability and rapidity.

References

1. El Akkad N, El Hazzat S, Saaidi A, Satori K (2016) Reconstruction of 3D scenes by camera self-calibration and using genetic algorithms. *3D Research* 7(1):6
2. Chen X, Williams BM, Vallabhaneni SR, Czanner G, Williams R, Zheng Y (2019) Learning active contour models for medical image segmentation. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp 11632–11640
3. El Hazzat S, Merras M, El Akkad N, Saaidi A, Satori K (2018) 3D reconstruction system based on incremental structure from motion using a camera with varying parameters. *Vis Comput* 34(10):1443–1460
4. Es-Sabry M, El Akkad N, Merras M, Saaidi A, Satori K (2020) A new image encryption algorithm using random numbers generation of two matrices and bit-shift operators. *Soft Comput* 24(5):3829–3848
5. Li C, Li S, Asim M, Nunez J, Alvarez G, Chen G (2009) On the security defects of an image encryption scheme. *Image Vis Comput* 27(9):1371–1381
6. Pont-Tuset J, Arbelaez P, Barron JT, Marques F, Malik J (2016) Multiscale combinatorial grouping for image segmentation and object proposal generation. *IEEE Trans Pattern Anal Mach Intell* 39(1):128–140
7. Jiang Y, Zhao K, Xia K, Xue J, Zhou L, Ding Y, Qian P (2019) A novel distributed multitask fuzzy clustering algorithm for automatic MR brain image segmentation. *J Med Syst* 43(5):118
8. Khrissi L, Akkad NE, Satori H, Satori K (2019) Color image segmentation based on hybridization between Canny and k-means. In: *2019 7th mediterranean congress of telecommunications (CMT)*. IEEE, pp 1–4
9. Aliniya Z, Mirroshandel SA (2019) A novel combinatorial merge-split approach for automatic clustering using imperialist competitive algorithm. *Expert Syst Appl* 117:243–266
10. Javed A, Kim YC, Khoo MC, Ward SLD, Nayak KS (2015) Dynamic 3-D MR visualization and detection of upper airway obstruction during sleep using region-growing segmentation. *IEEE Trans Biomed Eng* 63(2):431–437
11. Fang L, Qiu T, Zhao H, Lv F (2019) A hybrid active contour model based on global and local information for medical image segmentation. *Multidimens Syst Signal Process* 30(2):689–703
12. Mahapatra D (2017) Semi-supervised learning and graph cuts for consensus based medical image segmentation. *Pattern Recognit* 63:700–709
13. Khrissi L, El Akkad N, Satori H, Satori K (2020) Image segmentation based on k-means and genetic algorithms. In: *Embedded systems and artificial intelligence*. Springer, Singapore, pp 489–497
14. Tirandaz Z, Akbarizadeh G, Kaabi H (2020) PolSAR image segmentation based on feature extraction and data compression using Weighted Neighborhood Filter Bank and Hidden Markov random field-expectation maximization. *Measurement* 153:107432
15. Dhanachandra N, Manglem K, Chanu YJ (2015) Image segmentation using K-means clustering algorithm and subtractive clustering algorithm. *Procedia Comput Sci* 54:764–771
16. Ahmad A, Hashmi S (2016) K-Harmonic means type clustering algorithm for mixed datasets. *Appl Soft Comput* 48:39–49
17. Contreras P, Murtagh F (2010) Fast hierarchical clustering from the Baire distance. In: *Classification as a tool for research*. Springer, Berlin, Heidelberg, pp 235–243
18. Kaur J, Agrawal S, Vig R (2012) A methodology for the performance analysis of cluster based image segmentation. *Int J Eng Res Appl* 2(2):664–667
19. Ghosh P, Mitchell M, Tanyi JA, Hung AY (2016) Incorporating priors for medical image segmentation using a genetic algorithm. *Neurocomputing* 195:181–194
20. Yang Z, Shufan Y, Li G, Weifeng D (2016) Segmentation of MRI brain images with an improved harmony searching algorithm. *BioMed Research International*, Cairo
21. Vijay V, Kavitha AR, Rebecca SR (2016) Automated brain tumor segmentation and detection in MRI using enhanced Darwinian particle swarm optimization (EDPSO). *Procedia Comput Sci* 92:475–480
22. Mirjalili S (2016) SCA: a sine cosine algorithm for solving optimization problems. *Knowl-Based Syst* 96:120–133
23. Jain AK, Dubes RC (1988) *Algorithms for clustering data*. Prentice-Hall, Inc., Upper Saddle River
24. Gonzalez TF (1985) Clustering to minimize the maximum inter-cluster distance. *Theor Comput Sci* 38:293–306
25. Pham, TX, Siarry P, Oulhadj H (2017) Image clustering using improved particle swarm optimization. In: *International conference on industrial networks and intelligent systems*. Springer, Cham, pp 359–373
26. Benaichouche AN, Oulhadj H, Siarry P (2013) Improved spatial fuzzy c-means clustering for image segmentation using PSO initialization, Mahalanobis distance and post-segmentation correction. *Digit Signal Process* 23(5):1390–1400
27. Sinha A, Jana PK (2018) A hybrid MapReduce-based k-means clustering using genetic algorithm for distributed datasets. *J Supercomput* 74(4):1562–1579
28. Maulik U, Saha I (2010) Automatic fuzzy clustering using modified differential evolution for image classification. *IEEE Trans Geosci Remote Sens* 48(9):3503–3510
29. Niknam T, Amiri B (2010) An efficient hybrid approach based on PSO, ACO and k-means for cluster analysis. *Appl Soft Comput* 10(1):183–197
30. Li N, Li G, Deng Z (2017) An improved sine cosine algorithm based on levy flight. In: *Ninth international conference on digital*

- image processing (ICDIP 2017), Vol 10420. International Society for Optics and Photonics, p 104204R
31. Hafez AI, Zawbaa HM, Emary E, Hassanien AE (2016) Sine cosine optimization algorithm for feature selection. In: 2016 international symposium on innovations in intelligent systems and applications (INISTA). IEEE, pp 1–5
 32. Emary E, Zawbaa HM (2019) Feature selection via Lévy Antlion optimization. *Pattern Anal Appl* 22(3):857–876
 33. Wang J, Yang W, Du P et al (2018) A novel hybrid forecasting system of wind speed based on a newly developed multi-objective sine cosine algorithm. *Energy Convers Manag* 163:134–150
 34. Jain AK, Murty MN, Flynn PJ (1999) Data clustering: a review. *ACM comput Surv (CSUR)* 31(3):264–323
 35. Xu R, Wunsch D (2009) Data visualization and high-dimensional data clustering
 36. Han J, Pei J, Kamber M (2011) *Data mining: concepts and techniques*. Elsevier, Amsterdam
 37. Omran MG, Engelbrecht AP, Salman A (2004) Image classification using particle swarm optimization. In: *Recent advances in simulated evolution and learning*, pp 347–365
 38. Jaiswal V, Sharma V, Varma S (2019) An implementation of novel genetic based clustering algorithm for color image segmentation. *Telkonnika* 17(2):1461–1467
 39. Martin D, Fowlkes C, Tal D, Malik J (2001) A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In: *Proceedings eighth IEEE international conference on computer vision, ICCV 2001, vol 2*. IEEE, pp 416–423
 40. Du WB, Ying W, Yan G (2016) The impact of population structure on particle swarm optimization: a network science perspective. In: *International conference on swarm intelligence*. Springer, Cham, pp 341–349
 41. Liu Z, Laganière R (2007) Phase congruence measurement for image similarity assessment. *Pattern Recognit Lett* 28(1):166–172
 42. Chai T, Draxler RR (2014) Root mean square error (RMSE) or mean absolute error (MAE)? Arguments against avoiding RMSE in the literature. *Geosci Model Dev* 7(3):1247–1250
 43. Campello RJ, Hruschka ER (2006) A fuzzy extension of the silhouette width criterion for cluster analysis. *Fuzzy Sets Syst* 157(21):2858–2875
 44. Pakhira MK, Bandyopadhyay S, Maulik U (2005) A study of some fuzzy cluster validity indices, genetic clustering and application to pixel classification. *Fuzzy Sets Syst* 155(2):191–214
 45. Balasko B, Abonyi J, Feil B (2005) *Fuzzy clustering and data analysis toolbox*. Department of Process Engineering, University of Veszprem, Veszprem
 46. Bezdek JC (1981) Objective function clustering. In: *Pattern recognition with fuzzy objective function algorithms*. Springer, Boston, MA, pp 43–93

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.