Automatic lung CT clustering using MKM Algorithm Optimized by PSO

SaeidFazli, Mitrajafari

Abstract—Automatic detection systems in medical sciences improve the accuracy of the diagnosis and reduce the time of analysis. This field in lung disease referred to computer aided diagnosis (CAD) system. Lung nodule detection is the most important task of these systems. CAD systems used the combination of image processing techniques in order to detect the lung nodules. Lung CT image clustering is one of the most important steps in CAT systems. The accurate clustering method reduces the complex following steps and reduces the rate of fault. In this paper we propose an improved MKM clustering method which resolves the weakness of the previous methods and provides an accurate segmentation. In conventional concept of moving, the members in cluster with highest fitness are forced to move to the cluster with the lowest fitness. In the enhanced version of the Moving K-Means algorithm the members move in a defined range. We use particle swarm optimization (PSO) algorithm to optimize the range of the moving radius. Our proposed optimization algorithm with PSO clusters the CT images into the homogenous segment and resolves the problem of the dead center and the trapped center. Simulation results show the effectiveness of the proposed algorithm. The resultant of moving radius range shows that the proposed algorithm fulfilled the relationship condition with optimum value of R=1.8998 $e^{-4}$and (R)´=5.6045 e^{-4}.

Index Terms—Computer aided diagnosis (CAD), K-Means, clustering, optimization, PSO.

1 Introduction

Recently, image processing and machine learning techniques are widely used in different fields in medical sciences. In addition, in clinical practice automatic systems based on these techniques are improve the quality and quantity of earlier detection of the abnormalities. This field in pulmonary diseases referred to the Computer aided diagnosis (CAD) systems. Time factor is one the most important factor that increases the chance of remission in most diseases especially in different types of cancers. Lung cancer is now the leading cause of cancer death in the world. Therefore many studies investigate to improve the Computer aided diagnosis (CAD) systems. Lung nodule detection is the most important goal of these systems. The accuracy of the diagnosis systems depend on how it can analyze and interpret the images. Automatic analysis systems can significantly reduce the chance of Indiscrimination. Conventional CAD systems consist of the following steps in order to detect lung nodules. These steps are preprocessing, segmentation, feature extraction and classification for nodule detection. Segmentation step is the most important steps in CAD systems. An accurate segmentation can efficiently reduce the positive and negative faults in following steps and also reduce the intermediate complex computations. Segmentation is often a necessary step in computer analysis. In a lung CT image there are many different internal organs that must be considered in lung segmentation. Diagnosis nodules from airways, the vessels and the lung lobes pose many challenges [1]. Therefore, a method which we used to segment the lung nodules has a great effect to reduce the time of diagnosis and the rate of failure. Among the algorithms which segment an image based on similarity or having a particular feature like thresholding, region growing, edge detection and clustering methods [2], clustering methods are used in various filed in medical image segmentation. Among these clustering algorithms which based on minimizing the formal objective functions, many studies have been performed to the K-Means clustering algorithm. Despite the advantage of this clustering method, this method has some weaknesses. In order to overcome these weaknesses modified versions of conventional K-Means proposed. Moving K-Means method proposed to overcome the problem of dead center and the problem of local minima. The way that the MKM method applied to move the cluster members isn't appropriate then the adaptive versions such as AMKM, EMKM proposed. In our proposed method we use particle swarm optimization (PSO) algorithm in order to define the optimum radius range in which the members move to the other cluster. This radius variable considers as PSO cost parameter which we try to optimize it and the condition in which the best segmentation occur consider as a cost function.

This paper is organized as follows. In section 2 we explain the Moving K-Means (MKM) clustering algorithm. In section 3 the Particle Swarm Optimization (PSO) is discussed this followed in section 4 with set the cost parameters R and R´ of MKM algorithm with the PSO and in section 5 we have experimental results and in section 6 is consist of the discussion of this paper.

• *SaeidFazli, Research Institute of Modern Biological TechniquesinUniversity of zanjan, Zanjan, Iran. fazli@znu.ac.ir*
• *MitraJafari, M.Sc. Student of Electronics Eng,Electrical Engineering Department, University of zanjan, Zanjan, Iran.mitra_jafariy@yahoo.com*

2 Moving K-Means Segmentation Method

Image segmentation is a process of classifying an image in the groups with the similar patterns [3], [4]. Segmentation of medical images is done in order to assist radiologists to have a simplified representation of an image that is easier to analyze and understand. Moreover an accurate segmentation can significantly reduce the complexity and execution time of the

ISSN 2229-5518

processes and calculations in next steps and also increase the accuracy and Sensitivity of the diagnosis systems. In order to achieve the stated goals, we present a method that resolves problems of previous proposed conventional methods.

Many segmentation algorithms are proposed in these years in order to represent a better method that could classify similar patterns in to the appropriate homogenous groups [5]. Among them, k-means algorithm is the most popular segmentation algorithm. Low execution time and easy implementation of the K-Means algorithm, proposed it as an efficient algorithm in image segmentation. K-Means is a center-based algorithm which begins with initializing centers value and is followed by iteratively updating centers value until the desired minima variance is reached. The conventional K-Means have some important disadvantages. The algorithm sensitivity to the initial center, cause a poor segmentation without appropriate center initialization. And the other problems are trapped center which occurs when a cluster center is trapped at local minima which caused no further updates in next iterations. With this event, the cluster center remains in its initial place and its members lose their chance to update and move to their proper cluster. In other case the dead center could occur. This happened when a cluster center have no members.

To reduce these two problems, the Moving K-Means (MKM) was proposed. The idea of this modified method is that the cluster with the highest fitness forced its members to move into the cluster with the lowest fitness [6]. With this work, the fitness function in clusters becomes more univocal and update operation is performed on a cluster with the lowest fitness. Despite this, moving data to inappropriate cluster was not the problem's solution and couldn't produce a good segmentation. In the proposed method in Fasahat et al paper Ref.9, Adaptive Moving K-Means (AMKM), the members of the cluster with the highest fitness or highest variance transfer to its nearest neighboring cluster instead of transfer to the cluster with lowest fitness (variance). Also fuzzy methods proposed to resolve the above mentioned problems. The fuzzy concept of MKM means (FMKM) and the adaptive version of this method (AFMKM) presented in [6]. These methods couldn't be successful to resolve the problems. The latest improved version of MKM presented in Ref. [7] and Ref. [5] named respectively enhanced moving K−means (EMKM) and Optimized K-means (OKM). In the EMKM algorithm, in order to overcome the weakness of MKM the authors transfer members in a definition range. Despite this, range definition method didn't have a mathematical concept. In these facts both these methods couldn't really be successful to resolve the mentioned problems and also due to lack of precise and specific criteria for data transmission sometimes result disorganization in clusters. We use particle swarm optimization in order to determine the optimum range of the movement radius. The radius is modeled as an optimization problem and PSO estimate the best value of these parameters.

3 Particle Swarm Optimization

The particle swarm optimization is an optimization algorithm motivated by a social behavior of birds or fishes. This algorithm was first proposed in 1995 by Kennedy and Eberhart and contains a set of solutions developed to achieve an appropriate solution for multidimensional space problems [6]. The various versions of PSO algorithm is proposed based on continuity, fuzziness, velocity, convergence rate, topology, mobility, accommodation, particle distribution function and etc.[7].

In the basic concept of the particle swarm algorithm, the swarm consists of individual particles that are also a member of a society which moves in the space from one place to another place and searching for optimum solution. Each individual particle has memory of its best previous place and velocity and holds its memory until it reaches the best point. In the connected swarm all particles consider themselves as a part of a society and share their individual information in the search space. Therefore each particle has the information of the best position of other particles in every time. This information consists of any particle's best position and velocity. The goal of the PSO is to reach the global optimum of the fitness function in the search space. The PSO algorithm used the concept of local best and global best concurrently [7]. In the local best concept the population divided into the smaller groups in which the particles just share their information with their neighbors in the group. In global best concept all particles share their information in the connected swarm and any particle can know the best position and velocity of other particles. . The consequential of these directions is the vector that established the new place of the particle. The best solution is achieved when the fitness function which is the area under the curve ROC, is maximized.

Consider matrix X as a random arbitrary society. In this matrix each row is corresponding with each particle and all the particles in the Swarm are positioned in the n-dimensional space and the particles dimension defined whit the number of objective function parameters which is the column of matrixX. In our proposed method we used the PSO algorithm in order to estimate the cost parameters of ensemble classifier LogitBoost. To achieve this goal each particle was considered as a random result of the problem. In each repetition of the algorithm the local best of each particle and the global best where saved and the new velocity and new position of the particles computed in Eq. (1) and Eq. (2).

$$\mathrm{V_i(t) = wV_i(t-1) + c_1 Rand_1\left(P_{i,best} - X_i(t)\right) + c_2 Rand_2\left(P_{g,best} - X_i(t)\right)} \qquad (1)$$

$$X_i(t) = X_i(t-1) + V_i(t) \qquad (2)$$

4 Proposed Method

The proposed algorithm has membership function similar to the conventional K-Means algorithm. Consider an image with N×M size. K-Means clustering aims to partition the n observations into k sets [7]. Assign each observation to the cluster whose mean yields the least within-cluster sum of squares. This means all data will be assigned to nearest center based on Euclidean distance. Then calculate means to be the centroids of the observations in the new clusters. The new position of each center calculated based upon:

ISSN 2229-5518

$$C_j = \frac{1}{n_j}\sum_{i\in C_j} V_i \quad (3)$$

Where, n_j is the number of members in the $j-$thcluster. The fitness function for each cluster calculated according to the:

$$f(C_j) = \sum_{i\in C_j}(\|v_i - C_j\|)^2; \quad (4)$$

MKM algorithm for a good clustering and removing the problem of trapped centers proposed a relationship within the centers. Suppose C_s is a cluster center with the smallest fitness and C_l is a cluster center with the largest fitness. The relationship among the centers must fulfill the following condition:

$$f(C_s) \geq \alpha_a f(C_l) \quad (5)$$

Where, α_a is the constant initial with α_0 and the value of α_0 is designed in range $0 < \alpha_0 < \frac{1}{3}$. If Eq (5) is not fulfilled in conventional MKM, C_s get its member from C_l in order to come back to the active region. The members in cluster C_l which has the value lower than C_l are moving toward the smallest fitness cluster C_s and other members remain in C_l cluster. Both C_l and C_s will be updated as bellow:

$$C_l = \frac{1}{n_l}\sum_{i\in C_l} V_i \quad (6)$$

$$C_s = \frac{1}{n_s}\sum_{i\in C_s} V_i \quad (7)$$

Where, n_l , n_s are the new cluster members of C_l and C_s respectively. In order to the function (4), in MKM algorithm, if members of the cluster are more similar to the cluster center, the Euclidean distance becomes smaller and consequently the variance and fitness function are close to their minimum. This event appears in smallest fitness clusterC_s. Hence when we forced the members of the cluster C_l to move to the clusterC_s, the average variance of this cluster become larger. But forcing the migration from inappropriate cluster to the C_s caused clutter in the image and the poor segmentation appeared. In a adaptive method of MKM means Adaptive Moving K-Means (AMKM), the members of $C_l$which fulfill the $v_i < C_l$ Condition are moved to the nearest cluster instead of being forced to become the members of C_s and the cluster C_s obtain its member from the nearest cluster instead of the cluster C_l. If we consider C_n as the center of the nearest cluster to the C_s . The members fulfilled the condition $v_i > C_s$moved to the clusterC_s. Although this method resolve the problem of poor segmentation but the problem of trapped center that is far from other centers is already not resolved. Also the method we applied for member migration isn't appropriate. The members with the values less thanC_l are moved to the nearest cluster although they may be more similar to the C_l than the remained members. This caused the variance become greater. To avoid these phenomena, Enhanced Moving K-Means (EMKM) algorithm proposed two methods. First in the cluster C_l the members which is not fulfilled the range $\frac{1}{2C_l}$ should moved to the nearest cluster and the members within this range maintain inside the cluster. In the nearest cluster to the clusterC_s , if the members are outside the range $\frac{1}{2C_n}$ will be moved to the clusterC_s. And in EMKM-2, the members in the border of C_l and C_j moved to the cluster C_j and in other hand the members in the border of C_s and C_j moved to the clusterC_s [9].

As we can see EMKM didn't employ a mathematical method for determining the range that the members will be moved according to it. In our proposed method, according to the variance function, we define the R radius of the cluster C_s and cluster C_l in which the transmission should be employed. For avoiding of the trapped center and in order to come back to the active region, our proposed method defines the optimum range of the radius by using the PSO algorithm. In our method radius R is the range of transition in the nearest cluster ofC_s. Variance of the members in this radius is named δ_R. If we move members of adjacent cluster in the obtained range R, the variance of the cluster C_s will increase with the appropriate value. Hence with this way we can normalized the variance of the cluster C_s and the cluster C_s will come back to the active region and can update again. Also in cluster with highest variance value C_l, for reducing its variance value, the members of cluster C_l moved to its nearest cluster in the range $\acute{R}$. The $\acute{R}$ must be chose from the farther members because how much the members grey level similar to the cluster center the value of the cluster's variance becomes smaller. The members which place inside of $\acute{R}$moved to the nearest cluster. With this way we reduce the variance of cluster C_l and normalized its fitness in comparison with other clusters. The R and $\acute{R}$ are define as the PSO variable cost parameters and we define two different ranges for these variables. Each particle searches in two different spaces. Consider R is the radius in which the members in the nearest cluster of C_s which is named C_n, should move to the cluster C_s first space is from the nearest member of the smallest cluster C_s to the farthest member until reaches to the size of the radius R and consider $\acute{R}$ is the range in which the member of cluster C_l are moved to its nearest cluster. The second space is from of the farthest member toward the center of the cluster until reaches the size of $\acute{R}$. The random variable are chose from these two ranges. PSO estimate the best value of these parameters. The new variances are defined as bellow:

$$\delta_{C_s} + \delta_R = \delta_{C_{s\,new}} \quad (8)$$
$$\delta_{C_l} - \delta_{\acute{R}} = \delta_{C_{l\,new}} \quad (9)$$

Updating the cluster centers and the membership process will continue repeatedly until $\alpha_a$reached to 3.

$$\alpha_a = \alpha_a - \frac{\alpha_a}{K} \quad (10)$$

$$f(C_s) \geq \alpha_b f(C_l) \quad (11)$$

The$\alpha_a$andα_b are repeatedly updated until the following condition met.

$$\alpha_a = \alpha_0$$
$$\alpha_b = \alpha_a - \frac{\alpha_b}{K}$$

We use this condition as a PSO cost function and the members movement continue until this condition is full filled.

The cost function is defined as bellow:

$$F(i) = \delta_{C_{s\,new}} - \alpha_b \times \delta_{C_{l\,new}} \quad (12)$$

ISSN 2229-5518

In Fig.1 the defined radius and the proposed member's movement concept is shown.

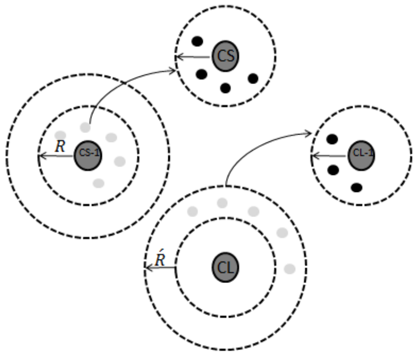


Fig. 1. The resulting movement radius from proposed method.

The PSO define the optimum value of R and Ŕ with minimizing the cost function value when the $F(\imath)' > 0$.

The sudo-code of proposed algorithm has been shown as bellow:

Input: fitness function, Number of particles, Number of parameters per particle, velocity coefficient, weight coefficient, damp ratio, parameter range

Output: Best input that optimize F(i) criteria

1- Fetch data, initialize population of PSO
   2- For j=1 to maximum iteration
      a. Compute the new random value, calculate weight coefficient 'w'
      b. Compute new velocity vector from (1)
      c. Compute new position from (2)
      d. If new computed position is define out of the range, set it to boundary value
      e. Fitness evaluation with new defined radius R and Ŕ from:

$$F(i) = \delta_{C_{s\,new}} - \alpha_b \times \delta_{C_{l\,new}} \&$$

$$F(i) > 0$$

      f. If new criteria is better than previous value THEN
         i. Replace better value as global best
         ii. Replace better value with local best
      g. End if

loop

The parameters of PSO algorithm are shown in Table I.

TABLE I. PARAMETERS OF PSO ALGORITHM

Parameter	Variable	Value
Weight coefficient	*w*	0.5
Local velocity coefficient	*c1*	2
Global velocity coefficient	*c2*	2
No. of Particles	*NP*	10
Max. Iteration	*Iter*	100
Damp Ratio	*DR*	0.95
Lower Bound	*LB*	0
Upper Bound	*UB*	100

5 EXPERIMENTAL RESULT

In this paper in order to evaluate our proposed segmentation method we use a data set of 100 lung CT images in DICOM format from hospital clinic data. The data set choose from the Lung Image Database Consortium (LIDC). These images consist of 220 real nodules. We use whole this data set in order to evaluate our proposed segmentation method. The resultants obtain from our proposed algorithm shows the accuracy of our proposed method. The fitness function repeated for 100 times.Fig.2 shows the convergence diagram for lung CT data set and in Fig 3 shows Resultant segmented images with the optimized MKM method with PSO. The best optimum value of the movement radius computed with the PSO algorithm as its cost parameters. In TABLE II the resultant value of R and Ŕ is shown.

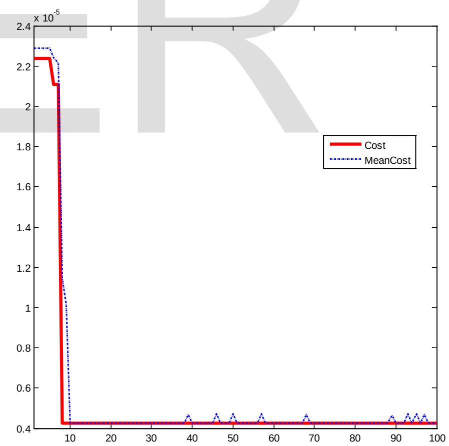
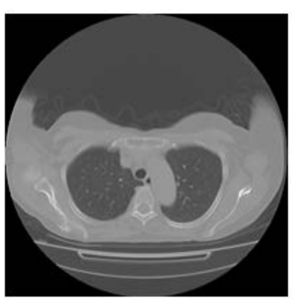


Fig. 2. Related convergence diagram for lung CT data



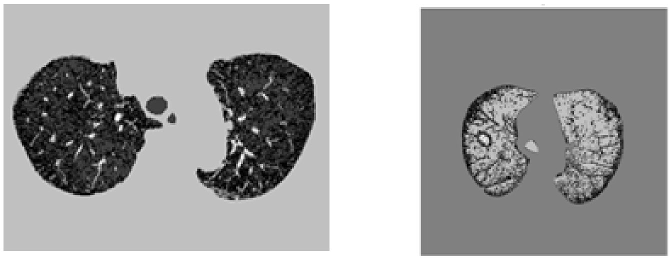


Fig. 3. Resultant segmented images with optimized MKM by PSO proposed algorithm .

TABLE II. TABLE STYLES

Resultant from PSO	***Radius R and $\acute{R}$***	***Old variance***	***New variance***
Cluster C_s	0.0018998	1.348e-07	1.7601e-06
Cluster C_l	0.0056045	5.2603e-06	2.8284e-06

6 DISCUSSION

In this paper the proposed algorithm is modified the conventional MKM algorithm by using the Particle Swarm Optimization algorithm. In our proposed algorithm the defined range of movement with PSO cause better segmentation and resolve the problem of trapped center and dead center with normalizing the value of fitness function (variance) in clusters C_s and C_l. The proposed method is result an accurate segmentation which is caused the accurate image analysis.

REFERENCES

[1] I. Sluimer, M. Prokop and B. van Ginneken"Towards automated segmentation of the pathological lung in CT," *Transactions on Medical Imaging* 24 (8), 2005.

[2] S.N. Sulaiman and N.A.M. Isa, "Adaptive fuzzy−K−means clustering algorithm for image segmentation", *IEEE T.Consum. Electr*. 56, 2661–2668 (2010).

[3] Linda G. Shapiro and George C. Stockman"*Computer Vision*,", pp 279-325, New Jersey, Prentice-Hall, ISBN 0-13-030796-3.

[4] Barghout, Lauren, and Lawrence W. Lee. "Perceptual information processing system." Paravue Inc. 10. U.S. Patent Application 10/618,543, filed July 11, 2003.

[5] F.U. Siddiqui and N.A.M. Isa, "Optimized K-means (OKM) clustering algorithm for image segmentation," *OPTO−ELECTRONICS REVIEW* 20(3), 216–225.

[6] N.A.M. Isa, S.A. Salamah, and U.K. Ngah, "Adaptive fuzzy moving K−means clustering algorithm for image segmentation"*IEEE T. Consum*. Electr. 55, 2145–2153 (2010).

[7] F.U. Siddiqui and N.A.M. Isa, "Enhanced moving K−means(EMKM) algorithm for image segmentation,"*IEEE T.Consum. Electr.* 57, 833–841 (2011).

[8] Kennedy, J., Eberhart, R., *"Particle Swarm Optimization"*, *IEEE Int. Conf. on Neural Networks*, Piscataway, NJ, Vol. 4, pp. 1942-1948, 1995.

[9] Sedighizadeh, D., Masehian, E., "Particle Swarm Optimization Methods, Taxonomy and Applications", *Int. J. Computer Theory and Engineering (IJCTE)*, Vol. 1, No. 5, pp. 486-502, 2009.