An approach for segmentation of medical images using pillar K-means algorithm

M.Pavani#1, Prof. S.Balaji*2

[#] *Department of Electronics and Computers Engineering, K.L. University Vaddeswaram, Vijayawada, India.*

[*] *Department of Electronics and Computers Engineering, K.L. University Vaddeswaram, Vijayawada, India.*

ABSTRACT

This paper presents an approach for image segmentation using pillar K-Means algorithm. In this paper the segmentation process includes a mechanism for clustering the elements of high resolution images. By using this process we can improve precision and reduce computational time. The system applies K-means clustering to image segmentation after optimized by pillar algorithm. The pillar algorithm considers that pillars placement should be located as far as possible from each other. The pillars placement is located far from each other to withstand against the pressure distribution of a roof, as identical to number of centroids among the data distribution. This algorithm is able to optimize the K-means clustering for image segmentation in terms of precision and computational time. By calculating the accumulated distance metric between each data point and all previous centroids it designates the initial centroids position and then it selects the data points which have maximum distance as new initial centroids. According to accumulated distance metric all the initial centroids are distributed in his algorithm. This paper evaluates by using an existing approach for image segmentation. But here we use medical images for segmentation. The experimental results clarify that this approach improves the segmentation quality in terms of precision and computational time.

Keywords - Image segmentation, K-means clustering, Pillar algorithm.

I. INTRODUCTION

In computer vision, image segmentation[2] is the process of partitioning a digital image into multiple segments (sets of pixels, also known as super pixels). The goal of segmentation is to simplify and/or change the representation of an image into something that is more meaningful and easier to analyze.[1] Image segmentation is typically used to locate objects and boundaries (lines, curves, etc.) in images. More precisely, image segmentation is the process of assigning a label to every pixel in an image such that pixels with the same label share certain visual characteristics.

Applications:

Some of the practical applications of image segmentation are:

1. Content-based image retrieval
2. Machine vision
3. Medical imaging[2]
   a. Locate tumors and other pathologies
   b. Measure tissue volumes
   c. Diagnosis, study of anatomical structure
4. Object detection
   a. Pedestrian detection
   b. Face detection
   c. Brake light detection
   d. Locate objects in satellite images(roads, forests, crops etc)
5. Recognition Tasks
   a. Face recognition
   b. Fingerprint recognition
   c. Iris recognition
6. Traffic control systems
7. Video surveillance

Several general-purpose algorithms and techniques have been developed for image segmentation. To be useful, these techniques must typically be combined with a domain's specific knowledge in order to effectively solve the domain's segmentation problems

II. RELATED WORK

2.1 Clustering methods:

The K-means algorithm is an iterative technique that is used to partition an image into K clusters. The basic algorithm is:

1. Pick K cluster centers, either randomly or based on some heuristic

2. Assign each pixel in the image to the cluster that minimizes the distance between the pixel and the cluster center
3. Re-compute the cluster centers by averaging all of the pixels in the cluster
4. Repeat steps 2 and 3 until convergence is attained (e.g. no pixels change clusters)

In this case, distance is the squared or absolute difference between a pixel and a cluster center. The difference is typically based on pixel color, intensity, texture, and location, or a weighted combination of these factors. K can be selected manually, randomly, or by a heuristic. This algorithm is guaranteed to converge, but it may not return the optimal solution. The quality of the solution depends on the initial set of clusters and the value of K.

2.2 K-Means clustering algorithm

An algorithm[3] for partitioning (or clustering) N data points into K disjoint subsets S_j containing N_j data points so as to minimize the sum-of-squares criterion

$$j = \sum_{j=1}^{k} \sum_{n \in s_j} \left| x_n - \mu_j \right|^2 \qquad (1)$$

where x_n is a vector representing the nth data point and μ_j is the geometric centroid of the data points in S_j. In general, the algorithm does not achieve a global minimum of J over the assignments. In fact, since the algorithm uses discrete assignment rather than a set of continuous parameters, the "minimum" it reaches cannot even be properly called a local minimum. Despite these limitations, the algorithm is used fairly frequently as a result of its ease of implementation.

The algorithm consists of a simple re-estimation procedure as follows. Initially, the data points are assigned at random to the K sets. For step 1, the centroid is computed for each set. In step 2, every point is assigned to the cluster whose centroid is closest to that point. These two steps are alternated until a stopping criterion is met, i.e., when there is no further change in the assignment of the data points.

2.3 K-Means algorithm

Let us talk about the K-means algorithm[4]. The algorithm accepts two inputs. The data itself, and "k", the number of clusters. We will talk about the implications of specifying "k" later. The output is k clusters with input data partitioned among them.

The aim of K-means (or clustering) is this : We want to group the items into k clusters such that all items in same cluster are as similar to each other as possible. And items not in same cluster are as different as possible. We use the distance measures to calculate similarity and dissimilarity. One of the important concept in K-means is that of centroid. Each cluster has a centroid. You can consider it as the point that is most representative of the cluster. Equivalently, centroid is point that is the "center" of a cluster.

2.4 Algorithm

1. Randomly choose k items and make them as initial centroids.
2. For each point, find the nearest centroid and assign the point to the cluster associated with the nearest centroid.
3. Update the centroid of each cluster based on the items in that cluster. Typically, the new centroid will be the average of all points in the cluster.
4. Repeats steps 2 and 3, till no point switches clusters.

As you can see, the algorithm is extremely simple. After some iterations, we will get k-clusters within which each points are similar. It is quite obvious that the algorithm should work.

2.5 Medical imaging

Medical imaging[21] is the technique and process used to create images of the human body (or parts and function thereof) for clinical purposes (medical procedures seeking to reveal, diagnose, or examine disease) or medical science (including the study of normalanatomy and physiology).

Measurement and recording techniques which are not primarily designed to produce images, such as electroencephalography (EEG), magnetoencephalo graphy (MEG),electrocardiography (EKG), and others, but which produce data susceptible to be represented as maps (i.e., containing positional information), can be seen as forms of medical imaging.

III. PROPOSED METHOD

3.1 The Basic Theory Of K-Means Clustering

This section briefly explains the basic theory of K-means clustering[1]. Let A={ai | i=1,…,f} be attributes of f-dimensional vectors and X={xi | i=1,…,N} be each data of A. The K-means clustering separates X into k partitions called clusters S={si | i=1,…,k} where M ∈X is Mi={mij | j=1,…, n(si)} as members of si, where n(si) is number of members for si. Each cluster has cluster center of C={ci | i=1,…,k}. K-means clustering algorithm can be described as follows [6]:

1. Initiate its algorithm by generating random starting points of initial centroids C.
2. Calculate the distance d between X to cluster center C. Euclidean distance is commonly used to express the distance.
3. Separate xi for i=1..N into S in which it has minimum d(xi,C).
4. Determine the new cluster centers ci for i=1..k defined as:

$$c_i = \frac{1}{n_i}\sum_{j=1}^{n(s_i)} m_{ij} \in s_i \qquad (2)$$

5. Go back to step 2 until all centroids are convergent.

The centroids can be said converged if their positions do not change in the iteration. It also may stop in the t iteration with a threshold ε [15] if those positions have been updated by the distance below ε:

$$\left|\frac{c^r - c^{r-1}}{c^r}\right| < \in \qquad (3)$$

3.2 An approach for Image Segmentation

The image segmentation is important to unify contiguous colors in the color vector space into representative colors [5]. It can improve significantly performance of the information extraction, such as color, shape, texture, and structure. This section describes our approach for image \ segmentation using our proposed Pillar algorithm [6] to optimize K-means clustering. The image segmentation system pre-proceeds three steps[1]: noise removal, color space transformation and dataset normalization. First, the image is enhanced by applying adaptive noise removal filtering. Then, our system provides a function to convert RGB of an image into HSL and CIELAB color systems. Because of different ranges of data in HSL and CIELAB, we apply the data normalization. Then, the system clusters the image for segmentation by applying K-means clustering after optimized by Pillar algorithm. Fig. 1 shows the computational steps of our approach for image segmentation.

A. Noise Removal

An adaptive noise removal[1] filtering using the Wiener filter is applied for noise removal of images. The Wiener filter can be considered as one of the most fundamental noise reduction approaches and widely used for solution for image restoration problems [7,8]. In our system, we use 3x3 neighborhoods of filtering size.

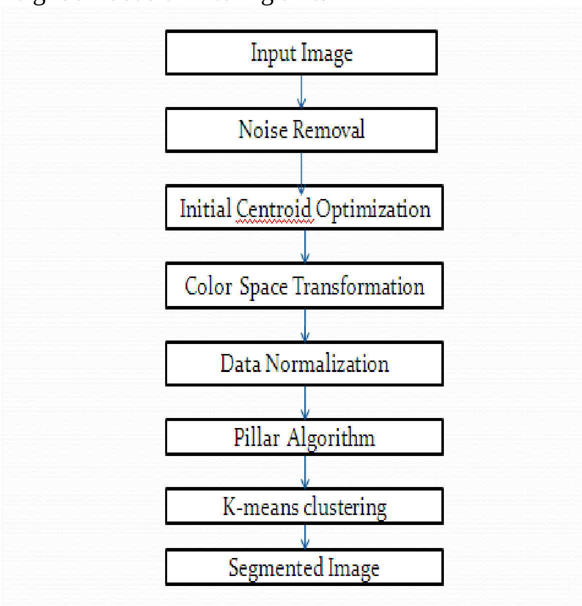


Fig. 1 Computational steps of our approach for image segmentation

B. Color Space Transformation

Our image segmentation system pre-proceeds the image by transforming the color space[1] from RGB to HSL and CIELAB color systems. HSL is well-known as an improved color space of HSV because it represents brightness much better than saturation. Beside, since the hue component in the HSL color space integrates all chromatic information, it is more powerful and successful for segmentation of color images than the primary colors [9]. The CIELAB color system has the advantage of being approximately perceptually uniform, and it is better than the RGB color system based on the assumption of three statistically independent color attributes [10]. The CIELAB color space is also widely-used for image restoration and segmentation [11][12][13][14]. Considering the advantages of each color system of HSL and CIELAB, in our system we utilize both of them as hybrid color systems for image segmentation.

C. Data Normalization

Because of different ranges of data points in HSL and CIELAB color spaces, we need to normalize the datasets. In our system, Softmax algorithm [15][16] is used for the data normalization[1]. The Softmax can reach softly toward its maximum and minimum value, but never getting there. The transformation using Softmax is more or less linear in the middle range, and has a smooth nonlinearity at both ends. The output range is between 0 and 1. A function in principle used to obtain the needed S-curve is the logistic function [17]:

$$f(x_i) = \frac{1}{1+e^{-x_i}} \qquad (4)$$

The logistic function produces the needed S-curve but not over the needed range of values, and there is also no way to select the range of linear response. In order to resolve this problem, {x} should be first transformed linearly to vary around the mean x in the following way:

$$x_i = \frac{x_i - \bar{x}}{\lambda(\sigma_x / 2\pi)} \qquad (5)$$

where:

x is the mean value of variable x

x σ is the standard deviation of variable x

λ is the linear response measured in standard deviation. It describes in terms of how many normally distributed standard deviations of the variables are to have a linear response. In our case, we set λ=10 in order to make smoother for normalizing the datasets.

D. Image Segmentation using Pillar Algorithm

The system uses the real size of the image in order to perform high quality of the image segmentation. It causes high-resolution image data points to be clustered. Therefore we use the K-means algorithm[1] for clustering image data considering that its ability to cluster huge data, and also outliers, quickly and efficiently [18][6]. However, Because of initial starting points generated randomly, K-means algorithm is difficult to reach global optimum, but only to one of local minima [19] which it will lead to incorrect clustering results [20]. Barakbah and Helen [18] performed that the error ratio of K-means is more than 60% for well-separated datasets. To avoid this phenomenon, we use our previous work regarding initial clusters optimization for K-means using Pillar algorithm [6]. The Pillar algorithm is very robust and superior for initial centroids optimization for K-means by positioning all centroids far separately among them in the data distribution.

This algorithm is inspired by the thought process of determining a set of pillars' locations in order to make a stable house or building. Fig. 2 illustrates the locating of two, three, and four pillars, in order to withstand the pressure distributions of several different roof structures composed of discrete points. It is inspiring that by distributing the pillars as far as possible from each other within the pressure distribution of a roof, the pillars can withstand the roof's pressure and stabilize a house or building. It considers the pillars which should be located as far as possible from each other to withstand against the pressure distribution of a roof, as number of centroids among the gravity weight of data distribution in the vector space. Therefore, this algorithm designates positions of initial centroids in the farthest accumulated distance between them in the data distribution.

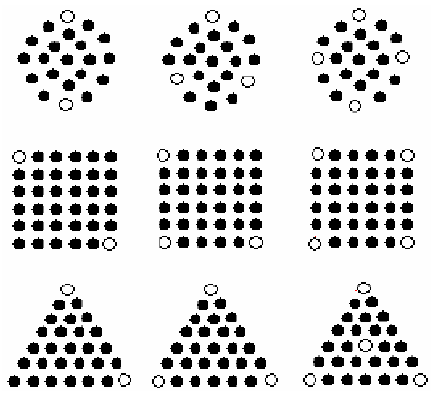


Fig. 2 Illustration of locating a set of pillars (white points) [1] withstanding against different pressure distribution of roofs.

The Pillar algorithm[1] is described as follows. Let X={xi | i=1,…,n} be data, k be number of clusters, C={ci | i=1,…,k} be initial centroids, SX ⊆ X be identification for X which are already selected in the sequence of process, DM={xi |i=1,…,n} be accumulated distance metric, D={xi | i=1,…,n} be distance metric for each iteration, and m be the grand mean of X. The following execution steps[1] of the proposed algorithm are described as:

1. Set C=∅ , SX=∅ , and DM=[]
2. Calculate D ・ dis(X,m)
3. Set number of neighbors nmin = α . n / k
4. Assign dmax ・ argmax(D)
5. Set neighborhood boundary nbdis = β . dmax
6. Set i=1 as counter to determine the i-th initial centroid

7. DM = DM + D
8. Select ж □ xargmax(DM) as the candidate for i-th initial centroids
9. SX=SX ∪ж
10. Set D as the distance metric between X to ж.
11. Set no▪ number of data points fulfilling D ≤ nbdis
12. Assign DM(ж)=0
13. If no < nmin, go to step 8
14. Assign D(SX)=0
15. C = C ∪ж
16. i = i + 1
17. If i ≤ k, go back to step 7
18. Finish in which C is the solution as optimized initial centroids.

However, the computation time may take long time if we apply the Pillar algorithm directly for all elements of high resolution image data points. In order to solve this problem, we reduce the image size to 5%, and then we apply the Pillar algorithm. After getting the optimized initial centroids as shown in Fig. 3, we apply clustering using the K-means algorithm and then obtain the position of final centroids. We use these final centroids as the initial centroids for the real size of the image as shown in Figure 1, and then apply the image data point clustering using K-means. This mechanism is able to improve segmentation results and make faster computation for the image segmentation.

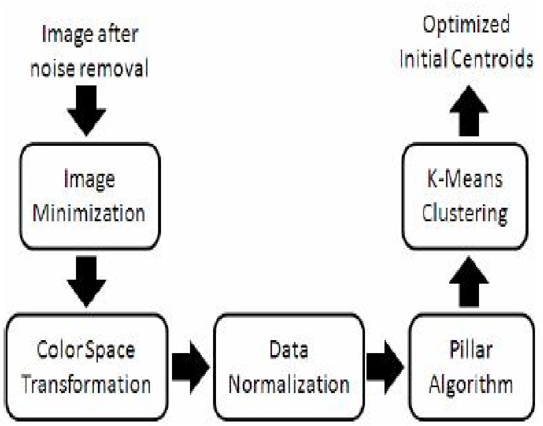
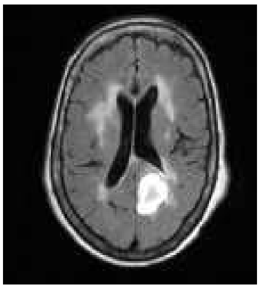


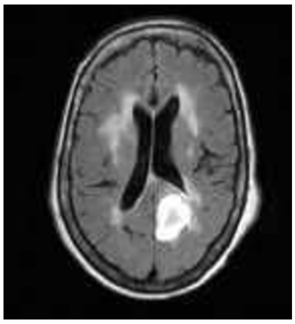
Fig. 3 Initial centroid optimization of K-means clustering for image segmentation.[1]

IV. EXPERIMENTAL RESULTS

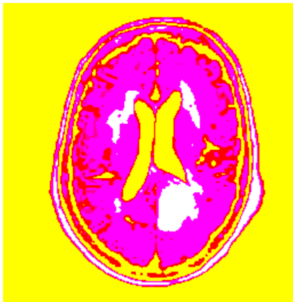
ORIGINAL IMAGE



AFTER MINIMIZATION IMAGE



COLOR SPACE TRANSLATED IMAGE



Segmented image with PK-means clustering

with K=51.4665 50.2431 49.1106

Truly segmented image using PILLAR K-means

CONCLUSION

In this paper we have presented a technique for segmentation of medical images using pillar K-means algorithm to find particular object. Here in results we find the tumor in the brain.

ACKNOWLEDGEMENTS

The authors would like to thank everyone, who ever remained a great source of help and inspirations in this humble presentation. The authors would like to thank K.L. University management for providing necessary facilities to carry out this work

REFERENCES

[1] A New Approach for Image Segmentation using Pillar-Kmeans Algorihm, *World Academy of Science*, Engineering and Technology 59 2009, Ali Ridho Barakbah and Yasushi Kiyoki

[2] http://en.wikipedia.org/wiki/Image_segmentation

[3] Bishop, C. M. Neural Networks for Pattern Recognition. Oxford, England: Oxford University Press, 1995, http://mathworld.wolfram.com/K-Means Clustering Algorithm.html

[4] http://saravananthirumuruganathan.wordpress.com/2010/01/27/k-means-clustering-algorithm/

[5] A.R. Barakbah, Y. Kiyoki, "An Image Database Retrieval System with 3D Color Vector Quantization and Cluster-based Shape and Structure Features", *The 19th European-Japanese Conference on Information Modelling and Knowledge Bases*, Maribor, 2009.

[6] A.R. Barakbah, Y. Kiyoki, "A Pillar Algorithm for K-Means Optimization by Distance Maximization for Initial Centroid Designation", *IEEE Symposium on Computational Intelligence and Data Mining (CIDM), Nashville-Tennessee, 2009.*

[7] A. Murli, L. D'Amore, V.D. Simone, "The Wiener Filter and Regularization Methods for Image Restoration Problems", Proc. *The 10th International Conference on Image Analysis and Processing, pp. 394-399, 1999*

[8] J. Chen, J. Benesty, Y.A. Huang, S. Doclo, "New Insights Into the Noise Reduction Wiener Filter", *IEEE Transactions on Audio, Speech, and Language Processing*, Vol. 14, No. 4, 2006.

[9] B. Wei, Y. Liu, Y. Pan, "Using Hybrid Knowledge Engineering and Image Processing in Color Virtual Restoration of Ancient Murals", *IEEE Transactions on Knowledge and Data Engineering*, Vol. 15, No. 5, 2003.

[10] B. Mičušík, A. Hanbury, "Automatic Image Segmentation by Positioning a Seed*", *ECCV 2006, Part II, LNCS 3952,* Springer Berlin/Heidelberg, pp. 468-480, 2006.

[11] Y.Gdalyahu, D. Weinshall, M. Wermen, "Self-Organization in Vision: Stochastic clustering for Image Segmentation, Perceptual Grouping, and Image database Organization", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 23, No. 12, pp. 1053-1074, 2001.

[12] H.M. Lotfy, A.S. Elmaghraby, "CoIRS: Cluster-oriented Image Retrieval System", *Proc. 16th IEEE International Conference on Tools with Artificial Intelligence (ICTAI), pp. 224-231, 2004.*

[13] N.J. Kwak, D.J. Kwon, Y.G. Kim, J.H. Ahn, "Color image segmentation using edge and adaptive threshold value based on the image characteristics", *Proc. International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS), pp. 255-258. 2004.*

[14] L. Jin, D. Li, "A Switching vector median based on the CIELAB color space for color image restoration", Signal Processing, Vol. 87, pp.1345-1354, 2007.

[15] T. Adani, H. Ni, B. Wang, "Partial likelihood for estimation of multiclass posterior probabilities", *Proc. the IEEE International Conference on Acoustics, Speech, and Signal Processing, Vol. 2, pp. 1053-1056, 1999.*

[16] K.L. Priddy, P.E. Keller, Artificial Neural Networks, pp. 16-17, *SPIE Publications*, 2005.

[17] A.R. Barakbah, Y. Kiyoki, "An Image Database Retrieval System with 3D Color Vector Quantization and Cluster-based Shape and Structure Features", *The 19th European-Japanese Conference on Information Modelling and Knowledge Bases, Maribor, 2009.*

[18] A.R. Barakbah, A. Helen, "Optimized K-means: an algorithm of initial centroids optimization for K-means", *Proc. Seminar on Soft Computing, Intelligent System, and Information Technology (SIIT), Surabaya, 2005.*

[19] B. Kövesi, J.M. Boucher, S. Saoudi, "Stochastic K-means algorithm for vector quantization", *Pattern Recognition Letters, Vol. 22, pp. 603-610,2001.*

[20] Y.M. Cheung, "k*-Means: A new generalized k-means clustering algorithm", *Pattern Recognition Letters, Vol. 24, pp. 2883-2893, 2003.*

[21] http://en.wikipedia.org/wiki/Medical_imaging