An approach of image segmentation technique using Improved K-means Clustering Algorithm

Monika Nagpal[2], Ms. Naveen Kumari[2]
[1]*PURCITM Mohali, Punjab, India*
[2] *PURCITM Mohali, Punjab, India*

***Abstract*—** Image segmentation is the classification of an image into different groups. Many researchers have been done in the area of image segmentation using clustering. There are different methods and one of the most popular methods is K-means clustering algorithm. Improved K-means clustering algorithm is an unsupervised algorithm and it is used to segment the interest area from the background. But before applying Improved K-means algorithm, first partial stretching enhancement is applied to the image to improve the quality of the image. Subtractive clustering method is data clustering method where it generates the centroid based on the potential value of the data points. So subtractive cluster is used to generate the initial centers and these centers are used in K-means algorithm for the segmentation of image.

Keywords— *Improved K-means Clustering, Median filter; Partial contrast stretching, Image segmentation.*

I. INTRODUCTION

Images square measure thought of mutually of the foremost vital medium of conveyance info. Understanding pictures and extracting the knowledge from them such the knowledge are often used for alternative tasks is a crucial side of Machine learning. Associate in Nursing example of identical would be the utilization of pictures for navigation of robots. Alternative applications like extracting malign tissues from body scans etc type integral a part of diagnosis.[1] One in every of the primary steps in direction of understanding pictures is to section them and verify completely different objects in them. There square measure completely different techniques for image segmentation like threshold based mostly, edge based, cluster based mostly, neural network based. From the various technique one in every of the foremost economical strategies is that the cluster methodology. Once more there square measure differing types of cluster: K-means clustering, Fuzzy C-means cluster, mountain cluster methodology and subtractive cluster methodology. [1]

One of most used cluster formula is K-means cluster. It's easy and computationally quicker than the ranked cluster. And it may also work for big range of variable. However it produces completely different cluster result for various range of range of cluster. therefore it's needed to initialize the correct range of range of cluster, k2. Again, it's needed to initialize the k range of center of mass. Different worth of initial center of mass would result different cluster.[2] Therefore choice of correct initial center of mass is additionally a crucial task. these days image segmentation becomes one in every of vital tool in medical space wherever it's accustomed extract or region of interest from the background.[2] Therefore medical pictures square measure segmental victimisation completely different technique and method outputs square measure used for the more analysis in medical. however medical pictures within their raw type square measure delineated by the arrays of numbers in the laptop, with the quantity indicating the values of relevant physical quantities that show distinction between differing types of body components. process and analysis of medical pictures square measure helpful in reworking raw pictures into a quantitative symbolic type, in extracting pregnant qualitative info to assist diagnosing and in desegregation complementary information from multiple imaging modalities.[2] And one in every of the elemental issues in medical analysis is that the image segmentation that identifies the boundaries of objects like organs or abnormal region in pictures. Results from the segmentation build it doable for form analysis, detection volume modification, and creating an explicit radiation treatment plant.[2] Associate in Nursing example of image segmentation is given below in Fig. 1

Fig1. Image Segmentation [1]

II. PARTIAL CONTRAST STRETCHING

Medical images which have been used for the analysis may have their own weakness such as blurred or low contrast. So a contrast enhancement technique such as Partial Spatial Stretching (PCS) is used to improve the image quality and contrast of the image[8]. It is done by stretching and compression process. By applying this technique, the pixel range of lower threshold value and upper threshold value will be mapped to a new pixel range and stretched linearly to a

wide range of pixels within new lower stretching value, and the remaining pixels will experience compression.[1]

III. MEDIAN FILTER

Median filtering is used as a noise removal in order to obtain a noise free image. After segmentation is done, the segmented image may still present some unwanted regions or noise. So to make the image a good and better quality, the median filter is applied to the segmented image. We can use different neighbourhood of n × n. But generally neighbourhood of n = 7 is used because large neighbourhoods produce more severe smoothing.[1]

IV.IMPROVED K-means CLUSTERING ALGORITHM[6]

Input: data set x contains n data points; the number of cluster is k.
Output: k clusters of meet the criterion function convergence.
Program process:
Step 1. Initialize the cluster center.
Step 2. Assigned the n data points from data set X to the closet cluster.
Step 3. Adjust each cluster center K by the formula (3).
Step 4. Calculate the distance of various data objects from each cluster center by formula (4),
and redistribute the n data points to corresponding cluster.
Step 5. Adjust each cluster center K by the formula (3).
Step 6. Calculate the criterion function E using formula (1), to determine whether the convergence, if convergence, then continue; otherwise, jump to Step 4.
Usually, the K-means algorithm criterion function adopts square error criterion, be defined as:

$$\sum_{j=1}^{k} \quad \sum^{} \quad \text{..... (1)}$$

In which, E is total square error of all the objects in the data cluster, xi bellows to data object set, mi is mean value of cluster C_i (x and m are both multi-dimensional). The function of this criterion is to make the generated cluster be as compacted and independent as possible.
The cluster center. Differences from the traditional clustering adjustment, the improved clustering algorithm add the weight of data point to the cluster center. Data points near the center of the cluster weights, on the contrary, the value of data points away from the cluster center is less weight. The formula of cluster center defined as follow:

$$\mathbf{k=} \quad \frac{djh}{D}xj1 + \frac{dj(h-1)}{D}xj2 + \cdots + \frac{dj2}{D}xj(h-2) + \frac{dj2}{D} \quad \textbf{... (2)}$$

where j represents the j_{th} cluster, h is the number of data points in the cluster, d_{jh} represents the distance between the h_{th} data point which belongs to cluster c and cluster center.

The Euclidean distance between data points and the cluster center. The distance between data point and the cluster center determine the cluster which data point belongs to, the formula of Euclidean distance is defined as follows:

$$\mathbf{d_{ji} = (1-\sigma_j / \sigma)d_{ji}} \text{ ..... } \textbf{(3)}$$

where j represents the jth cluster cj , i represents the ith data point xi, dji is the Euclidean distance between data point xi and the cluster center cj , i represents the squares error of the cluster cj , is the squares error sum of the K clusters c.

select the data point with maximum potential as the first cluster centre. Let us consider x1 and p1 as first cluster centre and its corresponding potential respectively. Then revise the potential of each data point by using the formula given below.

$$\mathbf{P_n = P_n - P_{1e}}^{(-4xn - x2/r\wedge 2_b)} \text{ ..... } \textbf{(4)}$$

r_b is the hyper sphere penalty radius in data space and it is a positive constant. Here an amount of potential is subtracted from each data point as a function of distance from the first cluster center. So the data points near the first cluster center will have greatly reduced potential, and therefore it have less chance to select as next cluster center. After calculating the revise potential of each data points, find the next highest potential as the next cluster center. So these processes continue until a sufficient number of cluster centre are obtained.
Consider a collection of n data points: $X = \{x_1, x_2, x_3 \ldots x_n\}$. Then each point is considered as a potential cluster center. The potential of data point's x_n is defined as:

$$\mathbf{P_n =} \qquad \text{..... (5)}$$

where r_a is hyper sphere cluster radius in data space and it is a positive constant which is used to define the neighbourhood. The symbol ||.|| denotes the Euclidean distance. So the measure of the potential for the data point is a measure of function of distance to all other data points.

V. IMAGE SEGMENTATION TECHNIQUES

The analysis on Image segmentation for several years has been a high degree of attention. Thousands of various segmentation techniques square measure gift within the literature, however there's not one methodology which might be thought-about smart for various pictures, all strategies don't seem to be equally smart for a specific sort of image [2]. Thus, rule development for one category of image might not perpetually be applied to different category of pictures. Hence, there square measure several difficult problems like development of a unified approach to image segmentation which might be applied to all or any sort of pictures, even the choice of associate degree applicable technique for a particular sort of image may be a tough downside.[2]

A. Segmentation supported Edge Detection:
This methodology makes an attempt to resolve image segmentation by police work the sides or pixels between totally different regions that have fast transition in intensity square measure extracted [2] and coupled to make closed object boundaries. The result's a binary image supported theory there square measure 2 main edge primarily {based} segmentation methodology's- grey bar graph and gradient based method. [2]

B.Thresholding Method:
Image segmentation by thresholding may be a easy however powerful approach for segmenting pictures having light-weight objects on dark background [7]. Thresholding technique is predicated on image space regions i.e. on characteristics of image. Thresholding operation convert a structure image into a binary image i.e., it select a correct threshold T, to divide image pixels into many regions and separate objects from background. Any pel (x, y) is taken into account as a section of object if its intensity is larger than or up to threshold worth i.e., f(x, y) ≥T, else pel belong to background. In native thresholding, multiple thresholds square measure wont to complete uneven illumination. Threshold choice is usually done interactively but, it's attainable to derive automatic threshold choice algorithms. [7]

C. Region primarily based Segmentation Methods:
Compared to edge detection methodology, segmentation algorithms based on region square measure comparatively easy and a lot of proof against noise. Edge primarily {based} strategies partition a picture supported fast changes in intensity close to edges whereas region based strategies, partition a picture into regions that square measure similar in keeping with a collection of predefined criteria.[4]

D. Segmentation strategies supported PDE (Partial Differential Equation):
Using a PDE primarily based methodology & finding the PDE equation by a numerical theme one will section the image. Image segmentation supported PDEs is principally applied by active contour model or snakes. This methodology was initial introduced by Kassetal in 1987. Kass developed this methodology to seek out acquainted objects in presence of noise and different ambiguities.[4] The central plan of snake is reworking a segmentation downside into a PDE framework. That is, the evolution of a given curve, surface or image is handled by PDEs and therefore the answer of those PDEs is what we glance forward to numerous strategies for image segmentation square measure - snake, level set and Mumford-shah model.[4]

E. Segmentation supported Artificial Neural Network:
Neural Network primarily based segmentation is completely totally different from typical segmentation algorithms. In this, a picture is first of all mapped into a Neural Network. wherever each vegetative cell stands for a pel, so image segmentation downside is reborn into energy step-down downside. The neural network was trained with coaching sample set so as to work out the affiliation and weights between nodes. Then the new pictures were segmental with trained neural network, for instance, we will extract image edges by mistreatment dynamic equations that direct the state of each vegetative cell towards minimum energy outlined by neural network. Neural network segmentation includes 2 vital steps feature extraction and image segmentation supported neural network.[6] Feature extraction is extremely crucial because it determines input file of neural network, first of all some options square measure extracted from the photographs, such they become appropriate for segmentation so they were the input of the neural network. All of the chosen options compose of extremely non-linear feature house of cluster boundary. [6]

F. Segmentation supported bunch
Clustering is associate degree unattended learning task, wherever one has to establish a finite set of classes called clusters to classify pixels. bunch use no coaching stages rather train themselves mistreatment accessible information. bunch is principally used once categories square measure famed earlier. A similarity criteria is outlined between pixels, so similar pixels square measure sorted along to make clusters. [7] The grouping of pixels into clusters is predicated on the principle of maximizing the intra category similarity and maximizing the put down category similarity. the standard of a bunch result depends on each the similarity live employed by the tactic and its implementation. bunch algorithms square measure classified as onerous bunch, k- suggests that bunch, fuzzy bunch, etc.[7]

VI.APPLICATIONS OF IMAGE SEGMENTATION

Image segmentation is mainly used to locate objects or object boundary, lines etc in an image so it can be used in applications which involve a particular kind of object recognition such as:[8]

- Face Recognition
- Fingerprint Recognition
- Locating objects in satellite images
- Traffic control systems
- Brake light detection
- Machine vision
- Agricultural imaging – crop disease detection.
- Medical imaging
- Locate tumors and other pathologies
- Measure tissue volumes
- Computer-guided surgery Diagnosis
- Treatment planning
- Study of anatomical structure

VII. PROPOSED ALGORITHM

II. Load the image to be segmented.
III. Apply partial contrast stretching. Initialize cluster centre.
IV. Use equation (5) to calculate the potential for every pixel value of the image.

V. Find maximum potential in step 3 and set that point be first center cluster and its corresponding potential as maximum potential.

VI. Use equation (2) to update the potential value of other remaining pixels based on the first cluster center.

VII. Again find the maximum potential in the step 4 and let that point be second point.

VIII. Adjust each cluster center K by the formula (3).

IX. Calculate the distance of various data objects from each cluster center by formula (4), and redistribute the n data points to corresponding cluster.

X. Calculate the criterion function E using formula (1), to determine whether the convergence, if convergence, then continue; otherwise, jump to Step 4.

XI. Assign the pixel with minimum distance with respect to centroid to its respective cluster of the centroid.

XII. Repeat the steps 10–12, until it satisfies the tolerance or error value.

XIII. Reshape the cluster into image.

XIV. Median filter is applied to the segmented image to remove any unwanted noise or region.

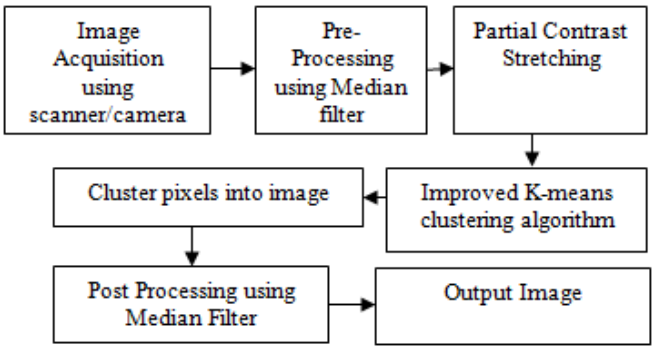
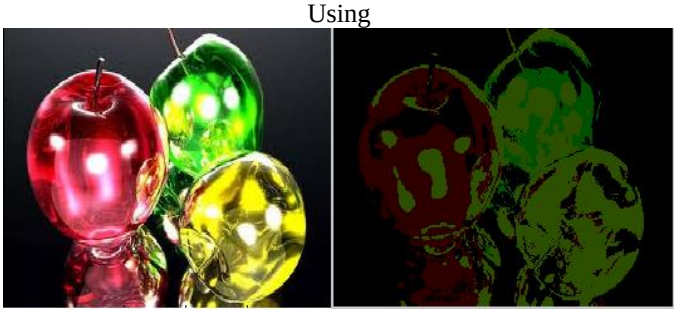


Fig 2: Block Diagram of Proposed Algorithm

Using



(a) (b)

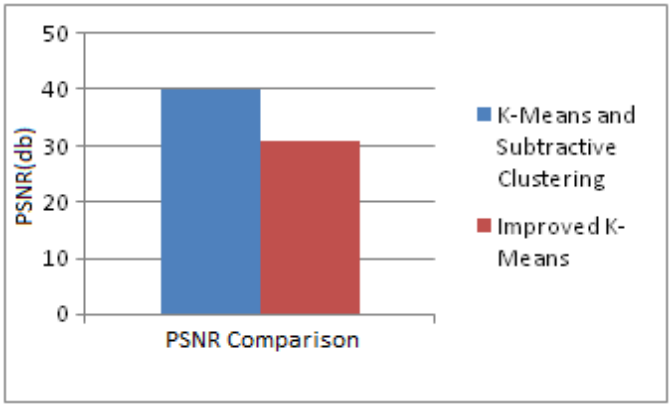
(c) (d)

Fig 3: (a),(c) Original Image; (b),(d) Proposed Algorithm

VIII RESULTS

For implementing improved K-mean approach for image segmentation firstly we take RCB image, and then convert it into processed image and segmented the objects in image and make different clusters.

PSNR: The peak to signal noise ratio is the proportion between maximum attainable powers and the corrupting noise that influence likeness of image. It is used to measure the quality of the output image. [1]



$$PSNR = 10.\log10\ (MAX2l / MSE) \quad \ldots\ldots\ldots\ldots\ldots(6)$$

Fig 4: PSNR

As shown in fig 3 the PSNR in the proposed approach i.e. improved K-mean based segmentation is 31dB where as in case of subtractive clustering and k-means it is approx 40 dB.

MSE: In statistics, the mean squared error (MSE) or mean squared deviation (MSD) of an estimator measures the average of the squares of the errors or deviations, that is, the difference between the estimator and what is estimated. MSE is a risk function, corresponding to the expected value of the squared error loss or quadratic loss. The difference occurs because of randomness or because the estimator doesn't account for information that could produce a more accurate estimate.

$$MSE = 1/mn \sum\_(i=0)^(m-1)\sum\_(j=0)^(n-1)[I(i,j)-K(i,j)\ ]^2 \quad \ldots\ldots\ldots(7)$$

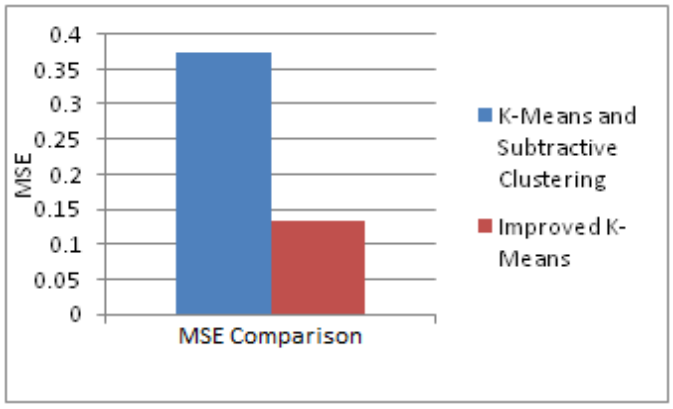


Fig 5: MSE

As shown in fig 3 the PSNR in the proposed approach i.e. improved K-means based segmentation is 0.132 where as in case of subtractive clustering and K-means it is approx 0.374.

RMSE: Root Mean Square Error has been used as a standard performance measurement of the output image. It gives how much output image is deviated from the input image.[1]

$$RMSE=\sqrt{MSE} \quad (8)$$

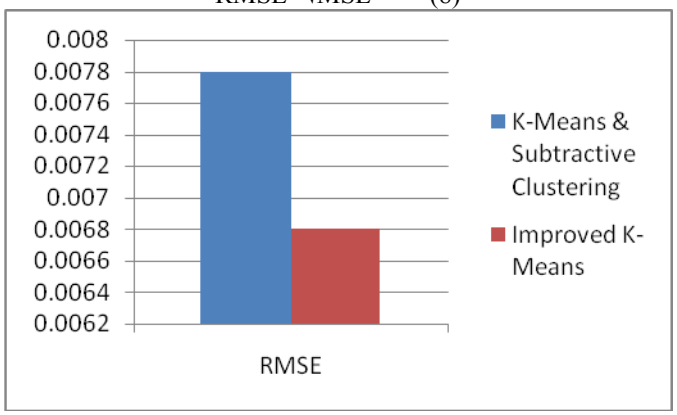


Fig 6: RMSE

IX.CONCLUSION

Image segmentation is very essential and critical to image processing and pattern recognition. This survey provides a summary of color image segmentation techniques available now. Basically color segmentation approaches are based on monochrome segmentation approaches operating in different color spaces. All of the existing color image segmentation approaches are by nature ad hoc. An image segmentation problem is basically one of psychophysical perception and it is essential to supplement any mathematical solutions by a priori knowledge about the picture knowledge. Most gray level image segmentation techniques could be extended to color image such as histogram thresholding clustering region growing edge detection and fuzzy based approaches. They can be directly applied to each component of a color space then the results can be combined in some way to obtain the final segmentation result. However one of the problems is how to employ the color information as a whole for each pixel. When color is projected onto three components the color information is so scattered that the color image becomes simply a multispectral image and the color information that human can perceive is lost.

X. REFERENCES

[1]. Nameirakpam Dhanachandra, Khumanthem Manglem, Yambem Jina Chanu, "Image Segmentation using K-means Clustering Algorithm and Subtractive Clustering Algorithm", Eleventh International Multi-Conference on Information Processing, Vol: 54, 2015, pp:764-771

[2]. Eman Abdel-Maksoud, Mohammed Elmogy, Rashid Al-Awadi, "Brain tumor segmentation based on a hybrid clustering technique", Egyptian Informatics Journal, Vol: 16, Issue: 1, March 2015, pp: 71-81

[3]. Zhensong Chena, Zhiquan Qib, Fan Mengac, Limeng Cuic,Yong Shi," Image Segmentation via Improving Clustering Algorithms with Density and Distance" Information Technology and Quantitative Management ,Vol:55,2015 ,pp:1015-1022

[4]. Fumitaka Hosotani, Yuya Inuzuka, Masaya Hasegawa, Shigeki Hirobayashi, Tadanobu Misawa, "Image Denoising With Edge-Preserving and Segmentation Based on Mask NHA", IEEE Transactions on Image Processing, Vol. 24, Issue. 12, December 2015, pp: 6025-6033 Mashiat Fatma, Jaya Sharma, "Leukemia Image Segmentation using K-Means Clustering and HSI Color Image Segmentation", International Journal of Computer Applications, Vol: 94, Issue: 12, 2014, pp:6-9

[5]. Kalpana Shrivastava, Neelesh Gupta, Neetu Sharma, "Medical Image Segmentation using Modified K Means Clustering", International Journal of Computer Applications, Vol: 103, Issue: 16, 2014, pp:12-16

[6]. Chunfei Zhang, Zhiyi Fang, "An Improved K-means Clustering Algorithm", Journal of Information & Computational Science, Vol: 10, Issue: 1, 2013, pp: 193-199

[7]. Salima Nebti," Bio-Inspired Algorithms for Color Image Segmentation", International Journal of Computer Applications, Vol :73,Issue :18, 2013 pp:887-975

[8]. Peter Karasev, Ivan Kolesov, Karl Fritscher, Patricio Vela, Phillip Mitchell, Allen Tannenbaum, "Interactive Medical Image Segmentation Using PDE Control of Active Contours", IEEE TRANSACTIONS ON MEDICAL IMAGING, VOL. 32, NO. 11, NOVEMBER 2013, pp: 2127-2139

[9]. Youguo Li, Haiyan Wu, "A Clustering Method Based on K-Means Algorithm", International Conference on Solid State Devices and Materials Science, Vol: 25, 2012, pp: 1104-1109

Monika Nagpal
MTECH
PURCITM
monikanagpal89@gmail.com