International Journal of Hybrid Information Technology
Vol.9, No.3 (2016), pp. 399-406
http://dx.doi.org/10.14257/ijhit.2016.9.3.37

EISCR: Efficient Image Segmentation using Cluster Representatives for Carcinoma Images

Mustafa Sabah

AL-Mansour University College,muustafa_bayat@yahoo.com

Abstract

Medical care requires intensively Image segmentation which a significant tool used to extract origin of interest from the background. Different segmentation techniques are deployed in medical images leading to an essential developments in both diagnosis and detection process. Such development has managed to assist specialists and doctors specified in the medical care system to diagnose the patients accurately.

This study will propose a more developed methodology via incorporating cluster representatives and second derivative filter technique on Carcinoma image. The conventional segmentation algorithm is widely deployed in medical care image, in spite of its advantages; being able to produce a complete division of the image, but it contains major drawbacks represented by the over-segmentations and sensitivity to false edges.

This study will expose these major problems of the conventional algorithm and fixing these problems via deploying Cluster Representatives; that uses a set of pixels as representatives that represent the clusters and producing advanced results. The other process deployed is a second derivative filter applied to the segmented image resulting of the demolition of unwanted region from the image.

Nevertheless, results of this study has proved that the number of partitions in medical care images are much fewer when comparing partitions produced by deploying the traditional algorithm deployed in medical care images. Such result will assist experts to easily diagnose health problems in a more accurate measure.

***Keywords**: Medical care, Image segmentation, second derivative filter, Cluster Representatives*

1. Introduction

Cancer can be a class associated with diseases that cause death among humans and there are various kinds of melanoma. Tongue cancer is one of the raise melanoma forms, and the tumor needs to be identified for further examination. Discovery associated with cancerous growth has an essential function for the management of virtually any melanoma. Image segmentation can be a basic process within in a decision oriented application [1]. The task connected with image segmentation is to partition an image into a number of non-overlapping places together with homogeneous features. There are several segmentation techniques that is broadly utilized, including histogram based procedures, edge-based procedures, artificial neural network based segmentation procedures, region-based methods (region splitting, growing, and merging), and clustering methods (Fuzzy C-means clustering, K-means clustering and Mean Shift) [2]. There are various difficult issues to image segmentation such as growth of one approach that can be used on different types involving image along with purposes.

Even, the selection proper way of a particular kind for a particular kind of image is a difficult problem. Therefore, there is no universal accepted method for image segmentation. Therefore, it remains a challenging problem in image processing and computer vision fields [3]. Clustering method is one of the most popular algorithms in image segmentation area. It is an approach of classifying patterns or data into categories, which is on the basis of the samples in the same group have the higher similarity than the

ISSN: 1738-9968 IJHIT
Copyright ⓒ 2016 SERSC

ones in different groups[3][4]. Although, some clustering algorithms have drawbacks, like K-means clustering that it doesn't guarantee continuous areas. In this paper, we present EISCR algorithm as a robust clustering algorithm for image segmentation due to its features as follows: 1) Dealing with a large number of pixels. 2) Suitable to handle strange shape clusters. 3) Working well with different clusters sizes. The suggested approach shows superior results as compared with k-means algorithm. The rest of the paper organized as follows section 2 shows some of the related works, k-means algorithm is presented in section 3, section 4 presents the Laplacian of Gaussian filter, Difference of Gaussians (DOG) filter is presented in section 5, the proposed algorithm is presented in section 6, results is presented in section 7, section 8 presents the conclusions.

2. Related Work

There were several works done in the area associated with image segmentation by employing various procedures. Many tend to be carried out according to various application associated with image segmentation.

Qixiang Ye [5] recommends a technique for segmentation problem through integrating the spatial connectivity as well as color feature in the pixels. Considering that an image is usually deemed the dataset where each and every pixel includes a spatial place along with a color importance, color image segmentation can be had through clustering these types of pixels directly into various multiple coherent spatial connectivity as well as color. To find out the spatial connectivity of the pixels, density-based clustering must be used, which often an effective is actually clustering procedure used in information exploration for acquiring spatial directories.

Mahesh Yambal [6] provides the most up-to-date questionnaire connected with diverse technological know-how utilized in health-related image segmentation using Fuzzy C Means (FCM). The conventional fuzzy c-means algorithm is definitely an effective clustering algorithm that may be utilized in health-related image segmentation. In order to replace the study connected with image segmentation the particular questionnaire provides conducted. This techniques used by this kind of questionnaire are Head Tumor Detection Making use of Segmentation According to Hierarchical Self Coordinating Guide and Fuzzy C Means Approach along with Histogram Centroid Initialization intended for Head Structure Segmentation.

Dibya Jyoti Bora [7] presents an innovative methodology designed for clustering based segmentation. At this point, this paper give significance on color space then pick l*a*b* for this assignment. The well-known clustering procedure K-means is used, nevertheless as its performance is reliant on picking a suitable distance measure (cosine distance measure). Formerly the segmented image is filtered through sobel filter. Finally some criteria are estimated to observe the performance.

Shu-Hung Leung [8] recommends a novel dissimilarity measure that incorporates the color dissimilarity in addition to the spatial distance in relations of an elliptic shape function. Since the attendance of the elliptic shape function, the novel measure is able to distinguish the pixels having similar color information but placed in dissimilar areas. An innovative iterative procedure for the resolve of the membership as well as centroid for each class is consequent, which is exposed to provide good variation among the lip area in addition to the non-lip area.

B.M.Nagarajan [9] recommends an instinctive segmentation technique established on unsupervised segmentation done on Ultrasound (US) images established from the radiologist. These imaging is extensively used in clinical analysis besides image-guided interferences, but then again undergoes from poor quality. One of the furthermost significant problems in image processing as well as investigation is segmentation.

 Copyright ⓒ 2016 SERSC

3. K-Means Clustering Algorithm

This algorithm represents as an unsupervised clustering procedure that categorizes the input data points hooked on multiple modules constructed on their essential distance from each other. The iterative k-means clustering process was first suggested by MacQueen . This procedure intentions at separating the data set, containing of ∂ expression arrangements {x1,…, x∂} in an n-dimensional space, into k separate clusters {C_i}i=lk, such that the appearance patterns in every cluster are further related to each other than to the appearance patterns in other clusters. There are two common separated clustering policies: square-error in addition to mixture modelling. The sum of the squared Euclidian distances among the models in a cluster besides the cluster center is termed as within-cluster distinction that is used as a standard in K-means clustering. This technique is an iterative procedure that reduces the sum of distances among every object besides its cluster centroid [10].

This technique is an iterative in addition to unsupervised that produces an exact number of non-hierarchical clusters. The procedure accepts the image features then attempts to discover usual grouping between these [11]. Completely the data points are randomly assigned which are similar in number as that of clusters around centroid $\mu \forall i = 1 \ldots k$

$$V = \sum_{i=1}^{k} \sum_{x_j \in S_i} \left(x_j - \mu_i\right)^2 \quad (1)$$

The above function is impartial, where there are k clusters Si,i = 1,2,…,k and μI is the mean point moreover named as centroid, of all the points xj ∈ Si

4. Laplacian of Gaussian filter (LoG)

This filter suggested by Marr and Hildreth is a well-known edge detector. The input image was convolved through this mask. Then, zero-crossings of the filtered image were detected as the edges. It has two additional advantages. Firstly, the LoG filter is an orientation-independent operator. Secondly, the detected edges form closed curves [12].

The Laplacian is a 2-D isotropic measure of the second spatial derivative of an image [9]. The Laplacian of an image highlights regions of rapid intensity change and is often used for edge detection. The Laplacian L(x,y) of an image with pixel intensity values I(x,y) is given as follows:

$$L(x,y) = \frac{\partial^2 I}{\partial x^2} + \frac{\partial^2 I}{\partial y^2} \quad (2)$$

Because it is approximating a second derivative measurement on image, it is very sensitive to noise. To reduce its sensitivity to noise, the Laplacian is often applied to an image that first has been smoothed by a Gaussian smoothing filter. We call this combined filter Laplacian of Gaussian filter. The 2-D LoG (Laplacian of Gaussian) function centered on zero and with Gaussian standard deviation σ has the form:

$$LoG(x,y) = -\frac{1}{\pi\sigma^4}\left[1 - \frac{x^2+y^2}{2\sigma^2}\right] e^{\frac{x^2+y^2}{2\sigma^2}} \quad (3)$$

Where, σ is the Gaussian standard deviation. Since the image is represented as a set of discrete pixels, we have to find a discrete convolution kernel that can approximate the Laplacian operator. Set different values of the Gaussianσ, we can get different LoG operators.

The Laplacian is a 2-D isotropic measure of the 2nd spatial derivative of an image. The Laplacian of an image highlights regions of rapid intensity change and is therefore often used for edge detection. The Laplacian is often applied to an image that has first been smoothed with something approximating a Gaussian smoothing filter in order to reduce

Copyright ⓒ 2016 SERSC

its sensitivity to noise. The operator normally takes a single gray level image as input and produces another gray level image as output [13].

5. Difference of Gaussians (DOG)

This filter does edge detection using the so-called "Difference of Gaussians" algorithm, which works by performing two different Gaussian blurs on the image, with a different blurring radius for each, and subtracting them to yield the result. This algorithm is very broadly used in artificial vision (maybe in biological vision as well), and is pretty fast because there are very efficient methods for doing Gaussian blurs. The most important parameters are the blurring radii for the two Gaussian blurs. It is probably easiest to set them using the preview, but it may help to know that increasing the smaller radius tends to give thicker-appearing edges, and decreasing the larger radius tends to increase the "threshold" for recognizing something as an edge [14].

This module is a filter that identifies edges. The DOG filter is similar to the LOG filters in that it is a two stage edge detection process. The DOG performs edge detection by performing a Gaussian blur on an image at a specified theta (also known as sigma or standard deviation). The resulting image is a blurred version of the source image. The module then performs another blur with a sharper theta that blurs the image less than previously. The final image is then calculated by replacing each pixel with the difference between the two blurred images and detecting when the values cross zero, i.e. negative becomes positive and vice versa. The resulting zero crossings will be focused at edges or areas of pixels that have some variation in their surrounding neighborhood [15].

A majority of the edge enhancement algorithms commonly employed in digital image processing often produce the unwanted side effect of increasing random noise in the image. Because it removes high-frequency spatial detail that can include random noise, the difference of Gaussians algorithm is useful for enhancing edges in noisy digital images. This interactive tutorial explores application of the difference of Gaussians algorithm to images captured in the microscope.

As an image enhancement procedure, the difference of Gaussians can be employed to growth the visibility of edges in addition to other detail present in a digital image. A wide diversity of alternative edge sharpening filters activate through enhancing high frequency detail, but because random noise also has a high spatial frequency, many of these sharpening filters tend to enhance noise, an undesirable artifact. The difference of Gaussians algorithm removes high frequency detail that often includes random noise, rendering this approach one of the most suitable for processing images with a high degree of noise. A major drawback to application of the difference of Gaussians algorithm is an inherent reduction in overall image contrast created via the operation [16].

6. Proposed Algorithm: Cluster formation using EISCR Algorithm

One of the essential and efficient algorithms in data mining is "Clustering Using Representatives", which is considered as a basic part of the hierarchical clustering algorithm. This algorithm possess the ability to partition sets of records into several groups adopting the similarity functions in order to produce similar record in the same group.

This study has proposed EISCR algorithm, the reason behind such proposal is; (1) the ability to deal with a large number of pixels that are randomly distributed, (2) the suitability to handle strange shape clusters, (3) working well with different clusters sizes.

The procedure of the proposed EISCR algorithm is represented by choosing a small sample of data (pixels in medical images) and clustering such data in the main memory. The second step utilized, is taking a fixed small set of pixels from each cluster and considering them as representative points. The representative points will be separated from one another, and moved in the 0.2-0.7 fraction towards the centroid of the cluster.

 Copyright ⓒ 2016 SERSC

Then, the merging process occurred between two clusters will be resulted if the following condition will be occurred; first; when they possess a pair of representative points; and secondly; when each one of the representative points of each cluster is sufficiently close.

The process of cluster merging is repeated till the distance between each neighbor cluster cannot be closer. When comparing points brought from secondary storage with representative points in the main memory, such comparison is to define accurately to which cluster these point belong.

For instance, in cancer images it is important to determine the ideal numbers of clusters, as the optimal numbers of clusters has to be as large as possible according to merge threshold value

The proposed algorithm contains the following steps:

- For an image with N pixels, EISCR takes a small number of pixels from the image, and each pixel represents a cluster.
- If the distance between clusters is found the shortest distance among any pair of pixels, then two clusters must be merged.
- Select a small set of pixels from each cluster with constant number R_p pixels to represent a cluster representative pixel. The pixels selected must be as far as possible from one another (lying exclusively in the boundary of the cluster). A right cluster must be found for the values of R_p, which is equal or greater than 10, the latter figure represents the expected threshold value. If the value of R_p is less than 10, then the representative pixels will fail to preserve the real geometry of the cluster.
- Merge two clusters if they have a pair of representative pixels, one from each cluster that is sufficiently close. This merging step is repeated until there are no more sufficiently close clusters.
- Each pixel Pi is compared with the clusters representative pixels and assigns to the cluster of the representative pixel that is closest to Pi.
- Repeat the previous steps, until it satisfies the tolerance or error value.
- Reshape the cluster into image.
- Post processing: Second derivative filters are applied to the segmented image to remove any unwanted noise or region.

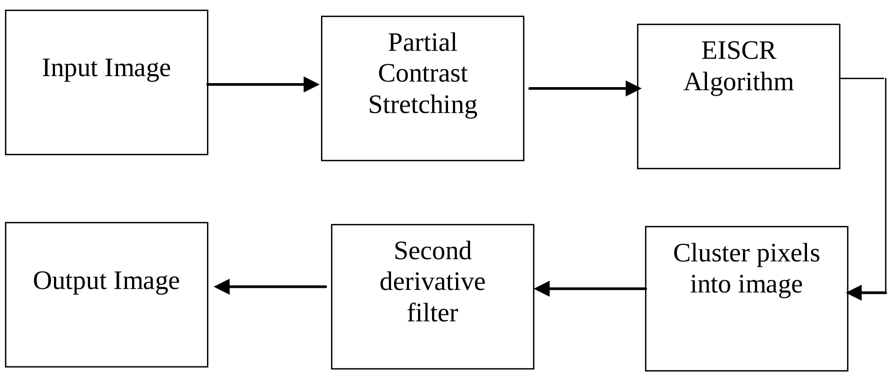

Figure 1. Block Diagram of the Proposed Algorithm

7. Results

The study has experienced two types of medical images for testing the proposed algorithm. Matlab is used to implement the proposed algorithm. When comparing the result of k-means algorithm with the proposed algorithm as shown in Figure. 2 it has resulted the following:

First sample tongue image as shown in figure 2-a; when deploying the k-means algorithm, the number chosen clusters is(k = 3), when deploying the EISCR algorithm the number of clusters starting with is7 clusters, two clusters are merged resulting 5 clusters to remain.

Second sample tongue, the number of clusters selected are (k = 6). In EISCR algorithm, 6 clusters were chosen to start with, and 3 clusters were merged to result 3 clusters to remain. From figure 2, we can notice that clusters resulted by applying EISCR algorithm are more accurate and has better segmentation results compared to the classical k-means algorithm. Because there are flexibility in the proposed algorithm that does not exist in the classical algorithm, the flexibility is represented by the movable number of clusters

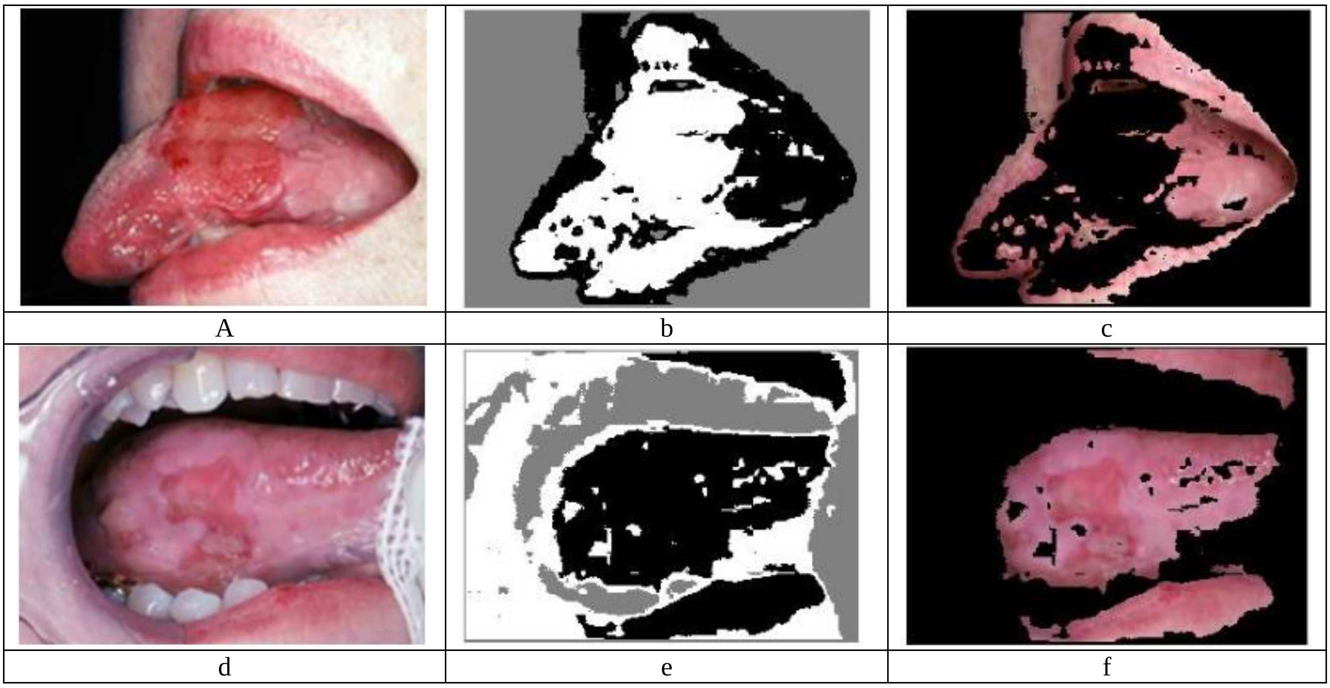

Figure 2. (a),(d) Original Images (b),(e) K-means Algorithm (c),(f) Proposed Algorithm

The quality of the segmented image is analyzed using two criteria to measure the value , the first criteria is nominated as (Root Mean Square Error (RMSE)) while the second criteria is called as(Peak to Signal Noise Ratio (PSNR)) the two criteria means the following:

1. Root Mean Square Error: is a standard performance measurement used at the output image. It provides the size of output image deviated from the input image.

$$RMSE = \sqrt{\frac{1}{n_x n_y} \frac{\sum_0^{n_x -1} \sum_0^{n_y -1} [(r(x,y))]^2}{\sum_0^{n_x-1} \sum_0^{n_y-1} [r(x,y)-t(x,y)]^2}} \qquad (4)$$

 Copyright ⓒ 2016 SERSC

2. Peak to Signal Noise Ratio: is the fraction between maximum attainable powers in addition to the corrupting noise that influence likeness of image. So, it is used to measure the quality of the output image.

$$PSNR = 10.log_1 0 \left[\frac{max(r(x,y))^2}{\sqrt{\frac{1}{n_x n_y} \frac{\sum_0^{n_x -1} \sum_0^{n_y -1} [(r(x,y))]^2}{\sum_0^{n_x-1} \sum_0^{n_y-1} [r(x,y)-t(x,y)]^2}}} \right] \quad (5)$$

In order to implement the above criteria's, there is some assumption: the *r (x, y) represents* the input image, and *t (x, y)* is the segmented image. The smaller value of RMSE means the image is in decent quality, while smaller value of PSNR means image of deprived quality. The resulted value of these testes of the segmented image is given below in Table 1.

The RMSE besides PSNR value are calculated in both algorithms, the results issued through the proposed algorithm shows must better besides efficient results from the typical algorithm. The test has shown that the value of RMSE is declining while the value of the PSNR is increasing. It can be concluded that the output image resulted from the proposed algorithm are better quality.

Table 1. RMSE and PSNR values

Image	RMSE Proposed algorithm	PSNR Proposed algorithm	RMSE K-means algorithm	PSNR K-means algorithm
Tongue Sample 1	0.0015	37.14	0.0071	33.46
Tongue Sample 2	0.0014	38.21	0.0061	32.65

8. Conclusion

This study segmented the tongue sample1, and tongue sample2 images through using EISCR clustering algorithm, using pixels as cluster representatives to generate the clusters and find the cluster centroid. At the same time partial contrast stretching is used to improve the quality of original image. Laplacian of Gaussian filter besides Difference of Gaussians (DOG) are used to improve segmented image. And the final segmented result is compare with *k*-means clustering algorithm and we can conclude that the proposed clustering algorithm has better segmentation.

Finally RMSE besides PSNR are checked and observed that they have small and large value respective, which are the condition for good image segmentation quality. And comparison for RMSE and PSNR are done for proposed method and classical *K*-means algorithm and it is found that the proposed method have better performance result. In the upcoming, we can improve the quality of the output image more through using the morphological operation then get better performance measurement.

References

[1] K. Hammouda, "A Comparative study of Data Clustering technique. Department of System Design Engineering", University of Waterloo, Canada.
[2] K. Arai and A. R. Barakbah, "Heirarchical K-means: An algorithm for Centroid initialization for K-means, Saga University", **(2007)**.
[3] P. Purohit and R. Joshi, "A New Efficient Approach towards k-means Clustering Algorithm", In International Journal of Computer Applications, (0975-8887), vol. 65, no. 11, **(2013)**.
[4] A. S. Abdul, M. Y. Masor and Z. Mohamed, "Colour Image Segmentation Approach for Detection of Malaria Parasiter using Various Colour Models and k-Means Clustering", In WSEAS Transaction on Biology and Biomedecine, vol. 10, **(2013)**.

Copyright ⓒ 2016 SERSC

[5] Q. Ye, W. Gao and W. Zeng, "color image segmentation using density –based clusreting", 0-7803-7663-3, IEEE III - 345 ICASSP, **(2003)**.
[6] M. Yambal, H. Gupta," Image Segmentation using Fuzzy C Means Clustering: A survey",International Journal of Advanced Research in Computer and Communication Engineering, vol. 2, no. 7, **(2013)**.
[7] D. Jyoti Bora, A. K. Gupta, "A Novel Approach Towards Clustering Based Image Segmentation", International Journal of Emerging Science and Engineering (IJESE), vol. 2, no. 11, **(2014)**.
[8] S.-H. Leung, S.-L. Wang, and W.-H. Lau, "Lip Image Segmentation Using Fuzzy Clustering Incorporating an Elliptic Shape Function", IEEE Transactions on Image Processing, vol. 13, no. 1, **(2004)**.
[9] B. M. Nagarajan, P. G. Prem and B. P. Senthil, "Image Segmentation Using Unsupervised Techniques", International Journal of Innovative Research in Science, Engineering and Technology, vol. 3, no. 3, **(2014)**.
[10] A. Jose, S. Ravi and M. Sambath, "Brain Tumor Segmentation using K-means Clustering and Fuzzy C-means Algorithm and its Area Calculation", International Journal of Innovative Research in Computer and Communication Engineering, vol. 2, issue 2, **(2014)**.
[11] V. Jumb, M. Sohani, A. Shrivas, Color image segmentation using k-means clustering and otsus adaptive thresholding, Int. J. Innov., Technol. Explor. Eng 3, no. 9, **(2014)**, pp. 72–76.
[12] J.Kaur, S. Agarwal and R. Vig, "A Methodology for the Performance Analysis of Cluster Based Image", In International Journal of Engineering Research and Application, vol. 2, **(2012)**.
[13] J Acharya, Gadhiya S, Raviya., "Segmentation techniques for image analysis: a review", Int J Comput Sci Manage Res2013; vol. 2, no. 4, pp. 1218–21.
[14] D Naik, Shah P., "A review on image segmentation clustering algorithms", Int J Comput Sci Inform Technol 2014; vol. 5, no. 3, pp. 3289–93.
[15] M. Bhat, "Digital Image Processing". International Journal of Science & Technology Research, vol. 3, no. 1, **(2014)**.
[16] Y.Yang, S. Hallman, D. Ramanan, C. C. Fowlkes. "Layered object Models for Image Segmentation", **(2010)**.

 Copyright ⓒ 2016 SERSC