COMPARATIVE STUDY OF WATERSHED ALGORITHM AND K-MEANS CLUSTERING TECHNIQUES FOR BRAIN TUMOR SEGEMENTATION

Mr.Chekuri Mahesh[1]., Zeba Afreen Zama[2]., Merugu Manjusha[3]

1 Assistant Professor, Department of ECE., Malla Reddy College of Engineering for Women., Maisammaguda., Medchal., TS, India (✉ chekurimahesh364@gmail.com)

2, 3 B.Tech ECE, (19RG1A04H5, 19RG1A04F7),

Malla Reddy College of Engineering for Women., Maisammaguda., Medchal., TS, India

Abstract:

We provide a comprehensive analysis of the many methods described for detecting tumours in MRI scans of the brain. Here, we'll look at three distinct methods and discuss their relative merits. K-Means and Fuzzy C-means clustering are two examples of such methods. The watershed segmentation approach is utilised once the tumour has been mapped. Information that isn't necessary may be removed from a database using the clustering process. In this investigation, we evaluate the similarities and differences between two popular cluster analysis techniques, K-Means and Fuzzy C-Means. The Fuzzy C-Means method is a soft segmentation approach used for clustering, in contrast to the target function employed by K-Means. MATLAB 2013a is used to test the procedures, and the results are described.

INTRODUCTION

Image processing enhances the accuracy of analysis by enhancing picture quality and clarity. Robotics and automated product inspection are only a few of the many fields that may benefit from satellite and other spacecraft-based remote sensing, picture transmission and storage, medical analysis of radar, sonar, and acoustic data, and more. Image segmentation is a crucial image processing method with wide-ranging applications in the medical, industrial, and other fields. Extracting elements of interest from photos requires first dividing the images into segments based on the characteristics of the underlying tissue. Early diagnosis of human brain tumours by magnetic resonance imaging has become the gold standard. Manually sifting through MRI scans for brain tumours is a laborious and time-consuming process. An automatic segmentation approach is proposed to reduce operator fatigue and boost trust in the results. In recent years, researchers have paid particular attention to the use of magnetic resonance imaging

(MRI) to segment brain tumours. Brain tumour identification is crucial for pinpointing the precise location and size of the tumour. As a result of its high protein concentration, the fluid released by brain tumour cells is both thick and luminous. Many strategies have been developed for analysing tumour cells. Several writers [3, 4, 8, 9] have suggested watershed segmentation as a method for spotting brain tumours. K-Means and Fuzzy C-

Means are two further methods for grouping data that have been developed [5, 6]. When applied to cancer cells, these methods are very effective. Next, we'll take a closer look at some concrete cases where these methods were put to use.

TECHNIQUES

Originally introduced by Diggable and Languorous, the watershed transform [1] [2] has since been refined by Beucher and Languorous for use in picture segmentation. Among geographical data splitting methods is the watershed transform. Envision a landscape that is submerged entirely in a lake, save for openings cut out at all of the local minima. Dams are constructed at the points where water from various catchment basins would otherwise mix as a consequence of these regional minimums. The procedure ends after the water level reaches the highest position. For this reason, the terrain is divided into basins by artificial barriers known as watershed lines.

The watershed segmentation method [11] can detect watershed boundaries in any given picture by cycling through a number of recursive flooding phases (XXY).

Notations:

a) $C(M_i)$ denotes the coordinates of the points in the catchment basin associated with regional minimum M_i (i=1, 2,…R).

b) The minimum and maximum gray levels of $g(x,y)$ are denoted as min and max.

c) $T[n]$ denotes the set of coordinates (s,t) for which $g(s,t) < n$.

d) Topography is flooded in integer flood increments from $n-min+1$ to $n-max+1$.

e) $C_n(M_i)$ denotes values of $C(M_i)$ flooded at stage n.

$$C_n(M_i) = C(M_i) \cap T[n] \qquad (1)$$

f) $C[n]$ denotes the union of the flooded catchment basin portions at stage 'n'

g) $C[max + 1]$ denotes the union of all catchment basins.

$$C[n] = \bigcup_{i=1}^{R} C_n(M_i) \qquad (2)$$

$$C[max+1] = \bigcup_{i=1}^{R} C(M_i) \qquad (3)$$

2) Initialization: Let $C[min+1]=T[min+1]$.

3) Algorithm:

a) At each step n, assume $C[n-1]$ has been constructed. The goal is to obtain $C[n]$ from $C[n-1]$.

b) $Q[n]$ denotes the set of connected components in $T[n]$. For each $q \in Q[n]$, there are three possibilities:

c) If $q \cap C[n-1]$ is empty.

d) A new minimum is encountered.

e) q is incorporated into $C[n-1]$ to form $C[n]$

f) $q \cap C[n-1]$ contains one connected component of $C[n-1]$.

g) q is incorporated into $C[n-1]$ to form $C[n]$

h) $q \cap C[n-1]$ contains more than one connected components of $C[n-1]$

i) A ridge separating two or more catchment basins has been encountered

j) A dam has to be built within q to prevent overflow between the catchment basins

k) Repeat the procedure until n=max+1

The problem of over segmentation arises when dividing up watersheds. Resolving this problem successfully using a marker-based technique has been shown. We utilise the regional maxima points of the distance converted blood cell masks as markers to address the problem of over segmentation and improve the algorithm's performance in the presence of ambiguity due to overlapping cells.

The watershed algorithm's ability to detect cancer cells has been bolstered by the addition of other processing techniques. The whole investigation is shown in Figure 1.

K-Means Clustering

In the realm of clustering techniques, the K means approach may be categorised as an unsupervised method. The desired number of clusters, represented by "k," is supplied into the procedure by the user.
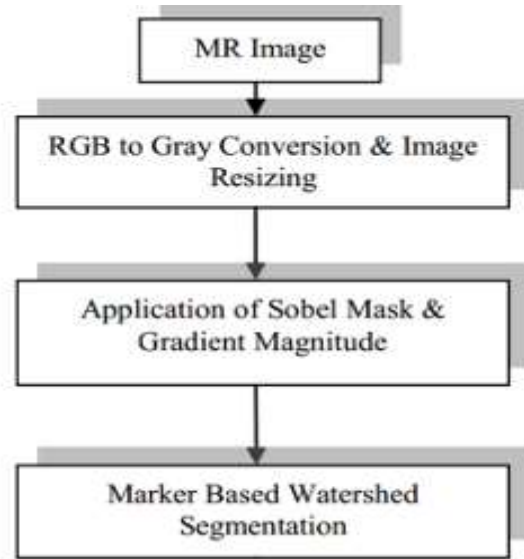

Figure 1. Block diagram of tumour cell segmentation using watershed algorithm.

There is a lot of reiteration. Convergence does take place; however, it happens only to a local minimum. The dataset is divided into an equal part, and each piece is assigned to the cluster that contains the average of the other n pieces. It's a way to get close to an NP-hard combinatorial optimization issue.

K means set-up is as follows:

1) $x_1, …, x_N$ are data points or vectors of observations

2) Each observation (vector xi) will be assigned to one and only one cluster

3) $C(i)$ denotes cluster number for the ith observation

4) Dissimilarity measure: Euclidean distance metric

5) K-means minimizes within-cluster point scatter:

$$W(C) = \frac{1}{2} \sum_{k=1}^{K} \sum_{C(i)=k} \sum_{C(j)=k} \|x_i - x_j\|^2 = \sum_{k=1}^{K} N_k \sum_{C(i)=k} \|x_i - m_k\|^2 \qquad (4)$$

where m_k is the mean vector of the kth cluster and N_k is the number of observations in k^{th} cluster.

K means algorithm can be concluded with the following steps:

1) For a given cluster assignment C of the data points, compute the cluster means m_k:

$$m_k = \frac{\sum_{i:C(i)=k} x_i}{N_k}, k = 1,...,K \qquad (5)$$

2) For a current set of cluster means, assign each observation as:

$$C(i) = \arg\min_{1 \le k \le K} \|x_i - m_k\|^2, i = 1,...,N \qquad (6)$$

3) Iterate above two steps until convergence.

Fuzzy C-Means Clustering:

Each data point in fuzzy clustering has only a weak association to a specific cluster, similar to the way it is in fuzzy logic. For this reason, nodes farther from the cluster's centre may sometimes contribute less. Insufficiencies of C-Means Clustering Due to the fact that each data point belongs to many clusters to varying degrees, clustering is a more forgiving variant of K-means. Some examples of such a clustering approach are the fuzzy c-means algorithm and fuzzy C-partitioning. It is common practise to partition dataset X using various fuzzy Cs.

$X = \{x_1, x_2, ..., x_n\}$: a set of given data.
$\rho = \{A_1, A_2, ..., A_c\}$: fuzzy c – partition of X.
c : the number of classes.
$A_1, ..., A_c$: fuzzy subsets of X
ρ satisfies two properties :

$$(1) \sum_{i=1}^{c} A_i(x_k) = 1$$

$$(2) \; 0 < \sum_{k=1}^{n} A_i(x_k) < n$$

The Fuzzy C-Means Algorithm may be used if m is a real number between 1 and 5, and if C is set to a modest positive value that ensures there are enough clusters to reliably separate the data. You may see the steps done below:

1) Let $t-0$. Select an initial fuzzy pseudo partition $\rho^{(o)}$.

2) Calculate the 'c' clustering centers $v_1^{(t)}, ..., v_2^{(t)}$ for $\rho^{(t)}$ and chosen value of m.

$$v_i^t = \frac{\sum_{k=1}^{n} [A_i^{(t)}(x_k)]^m x_k}{\sum_{k=1}^{n} [A_i^{(t)}(x_k)]^m} \qquad (7)$$

3) Update $\rho^{(t+1)}$

a) For each $x_{(k)} \in X$, if $\|x_k - v_i^{(t)}\|^2 > 0$ for all $i \in N_c$

b) $A_i^{(t+1)}(x_k) = [\sum_{j=1}^{c} (\frac{\|x_k - v_i^{(t)}\|^2}{\|x_k - v_j^{(t)}\|^2})^{\frac{1}{m-1}}]^{-1}$ (8)

4) If $\|x^k - v_i^{(t)}\|^2 - 0$ for some $I \in I \subseteq N_c$, then define $A_i^{(t+1)}(x_k)$ for $i \in I$ by any nonnegative real numbers satisfying

$$\sum_{i \in I} A_i^{(t+1)}(x_k) = 1 \qquad (9)$$

and define $A_i^{(t)}(x_k)$ for $i \in N_c - I$.

5) Compare $\rho^{(t+1)}$ and $\rho^{(t)}$.

If $|\rho^{(t+1)} - \rho^{(t)}| \le \epsilon$, then stop; otherwise, increase t by one and return to step 2.

RESULTS AND DISCUSSION

Here, we provide information gathered from MATLAB® simulations of the aforementioned three methods. The original picture that will be sliced is seen in Figure 2. Figure 3 displays the results of applying each method to this picture in an effort to separate tumour cells. To simulate a broad variety of cluster sizes, many clustering techniques were used (one to eight). In contrast, when the cluster size was set to 6, both approaches were equally effective.
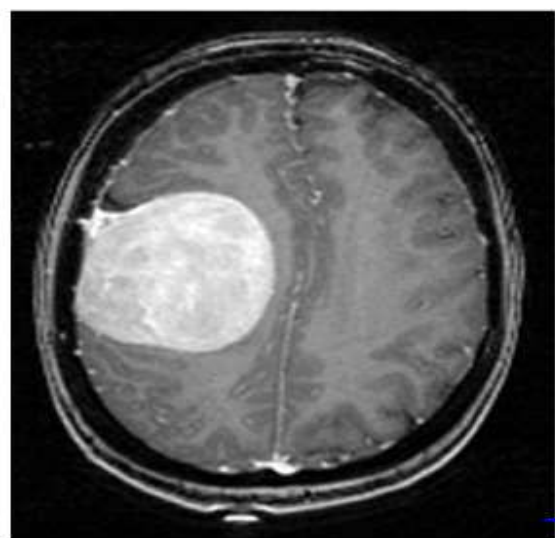
Figure 2. MR Image of brain with tumour.

Figure 3 demonstrates that all three algorithms are able to distinguish between tumour cells and healthy ones. Nonetheless, there may be localised regions of incorrect detection when using clustering. Nevertheless, morphological methods provide the efficient elimination of such localised regions. Instead, watershed segmentation may reliably isolate tumour cells with almost no false positives. However, when looking at how long each code takes to calculate, Fuzzy Means comes out on the bottom, whereas K-Means comes out on top. You can see how the different choices stack up in terms of time complexity in Table I.
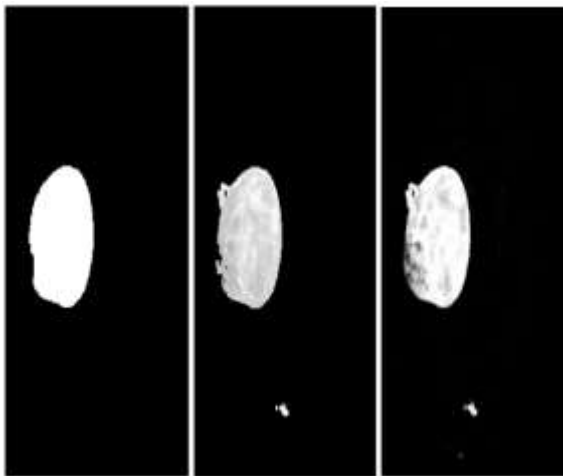
Figure 3. Segmented tumor cell. a) by watershed method. b) by KMeans algorithm. c) by Fuzzy C

TABLE I. C OMPLEXITY C OMPARISON O F A LGORITHMSMeans algorithm.

Algorithm	Computational Time
Watershed Segmentation	1.531461 seconds
K-Means Clustering	0.641206 seconds
Fuzzy C-Means Algorithm	8.368616 s seconds

CONCLUSION

Here, three methods are explored for using MRI scans to distinguish between tumour cells and healthy brain tissue. Methods and MATLAB® code examples are provided. The papers also give the outcomes of each algorithm. Finally, everyone agrees that all three strategies provide respectable outcomes and are therefore deserving of more study. Comparing simulation times, K-Means clustering was shown to be the most efficient method, followed by watershed segmentation and the FCM algorithm. Further, the watershed method successfully isolated the cancer cell whereas the clustering techniques had trouble avoiding false positives. Specialized methods, such as morphological processes, are required in this context.

REFERENCES

1. *L. Vincent, P. Soille, "Watersheds in digital spaces: an efficient algorithm based on immersion simulations," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.13, no.6, pp.583-598, Jun 1991.*
2. *J.B.T.M. Roerdink, A. Meijster, "The watershed transform: definitions, algorithms and parallelization strategies," Fundamenta Informaticae, vol. 41, pp. 187-228, Jan 2000.*
3. *R. J. Lapeer, A. C. Tan, R. V. Aldridge, "A combined approach to 3D medical image segmentation using marker-based watersheds and active contours: the active watershed method," Medical Image Understanding and Analysis (MIUA 2002), 22-23 July 2002, Portsmouth, UK.*
4. *C. Yu, Y. Li , "A watershed method for MR renography segmentation," International Conference on Biomedical Engineering and Biotechnology (iCBEB), pp.700-703, 28-30 May 2012.*
5. *Pan Lin ,Yong Yang, Chong-Xun Zheng, Jian-Wen Gu, "An Efficient Automatic Framework for Segmentation of MRI Brain Image", Proceedings of the Fourth International Conference on Computer and Information Technology(CIT'04), IEEE, 2004.*
6. *Abdel-Halim Elamy, Maidong Hu. Mining "Brain Tumors & Their Growth rates". 872-875 IEEE Image Processing Society, 2007.*
7. *Pratik P. Singhai, Siddharth A. Ladhake, "Brain Tumor Detection Using Marker Based Watershed Segmentation from Digital MR Images", International Journal of Innovative Technology and Exploring Engineering, vol. 2, issue 5, pp. 201-204, April 2013.*
8. *Vrji, K.S.A.; Jayakumari, J., "Automatic detection of brain tumor based on magnetic resonance image using CAD System with watershed segmentation," International Conference on Signal Processing, Communication, Computing and Networking Technologies (ICSCCN), pp.145,150, 21-22 July 2011.*
9. *Selvakumar, J.; Lakshmi, A.; Arivoli, T., "Brain tumor segmentation and its area calculation in brain MR images using K-mean clustering and Fuzzy C-mean algorithm," International Conference on Advances in Engineering, Science and Management (ICAESM), pp.186,190, 30-31 March 2012*
10. *R. C. Gonzalez and R. E. Woods, Digital Image Processing, 2nd ed., Upper Saddle River, NJ, Prentice Hall, 2002, pp. 541-553, 639-647.*