Fast and Automated Segmentation of Medical Image using Subtracting Clustering Fuzzy C-mean Algorithm

A. Himanshu Patel[1] ,**B. Asim Banerjee**[2]
[1] Biomedical Engineering, Ganpat University, Mehsana, Gujarat, India
[2] Electrical Engineering, DAIICT, Gandhinagar, Gujarat, India

Abstract – *Medical image processing and analysis covers a broad number of potential areas, the image segmentation is one of the key step in image processing and analysis. Texture is one of the important characteristics used in identifying objects OR regions of interest in an image. This paper deals with fast extraction of textural features of medical image using Sum and Difference Histogram Probabilities. The use of these extracted features with Fuzzy c-mean algorithm technique makes the segmentation process automated. The entire approach is unsupervised.*

Keywords: *Image segmentation, texture, histogram, clustering, fuzzy c-mean.*

1 Introduction

Medical image analysis has involved variety of directions with the aim to assist in diagnosis and clinical studies.

Image texture is an important diagnostic feature and is inherent to the human visual system. The human ability to distinguish subtle variation in texture in an object is limited. Computer based texture analysis is, therefore, a powerful diagnostic tool and has proved itself to be useful in the characterization of lesions in variety of anatomical sites from a range of imaging modalities. Fig. 1 shows the steps of texture (image) analysis system. The dotted lines outline the components of the texture feature extraction algorithms.

Sum and Difference Histogram probabilities are used to derive the textural features (UNSER, 1986) [1]. This method use vector in the form of Sum and Difference Histogram for purpose of generating co-occurrence texture features. The texture properties are derived by applying 1) first order and 2) second order statistics on derived Sum and Difference Histogram probabilities.

The data set is prepared from statistical texture features derived using Sum and Difference Histogram probabilities and presented to easy to implement, unsupervised fuzzy c-mean with subtracting clustering for the segmentation of images. Fuzzy c-mean algorithm with subtractive clustering technique automatically determines the number of cluster center in the image to be segmented and makes the segmentation process automatic.

In the present work the authors cascaded the texture properties and standard fuzzy c-mean algorithm with subtractive clustering technique (WEN-WAN LTU, 2003) [7]. The algorithm was applied to selected MRI, CT and ultrasound medical images and results obtained are promising.

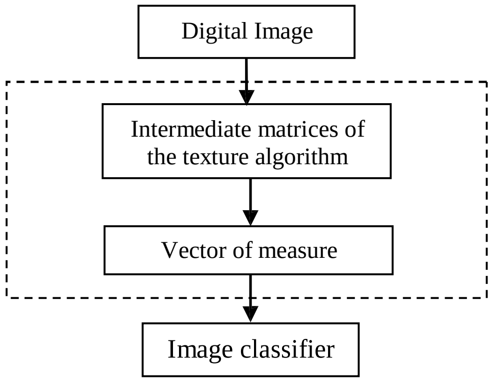


Fig. 1 steps of texture (image) analysis system.

2 Texture Feature Extraction

2.1 Texture Description

Image texture is the recognition of image regions using textural properties, and is a function of the spatial variation in pixel intensities (gray values).Texture is the most important visual cue in identifying the types of homogenous regions. Texture classification is the identification of the homogeneous regions, where each uniform textured region is identified with the texture class it belongs to. Texture segmentation is to obtain the boundary map separating the different textured regions in an image.

2.2 Texture Analysis Approaches

There are three principle approaches used in image processing to describe the texture of a region are 1) Statistical 2) Syntactic and 3) Spectral. In case of statistical approach texture is defined by a set of statistically extracted features represented as vector in multidimensional feature space (HARALICK, 1979) [2]. Medical images have high degree of variability and highly random texture. Statistical approach is particularly useful for random textures and author used approach to derive textural feature called statistical textural features.

2.3 Statistical Approach

The statistical features could be based on 1st, 2nd or higher order statistics of gray level of an image (R. M. HARALICK and K. SHANMUGAN, 1973) [4].

In 1st order statistical features mean original gray values are used to derive statistical features. One of the simplest approaches for describing texture is to use statistical moments of the gray-level histogram of an image or region. The statistical moments are calculated from the first order gray-level histogram. Let z be a random variable denoting gray levels and let $p(z_i)$, i = 0, 1, 2,…., L– 1, be the corresponding histogram, where L is the number of distinct gray levels. The nth moment of z about the mean is given by

$$\mu_n(z) = \sum_{i=0}^{L-1} (z_i - m)^n p(z_i) \qquad (1)$$

The various 1st order statistical features computed by statistical moment and used in this paper are tabulated in table 1.

TABLE 1
1st ORDER STATISTICAL FEATURES

Moment	Expression	Measure of Texture
Mean	$m = \sum_{i=0}^{L-1} z_i p(z_i)$	A measure of average intensity
Standard deviation	$\sigma = \sqrt{\mu_2(z)} = \sqrt{\sigma^2}$	A measure of average contrast.

Using 1st order statistical features hidden information is not always available. In order to improve the segmentation result second order statistical features are combined with 1st order features. Sum and Difference Histogram probability matrix is used to derive 2nd order statistical feature in this paper, that characterizes the spatial interrelationships of the gray tones in an image. The values of the sum and difference histogram probability matrix elements present relative frequencies with which two neighboring pixels separated by distance d appear on the image, where one of them has gray level i and other j. Such matrix is symmetric and also a function of the angular relationship between two neighboring pixels.

The non-normalized sum and difference associated with the relative displacement ($d1$, $d2$), are defined as

$$s_{k,l} = y_{k,l} + y_{k+d1,\, l+d2}$$
$$d_{k,l} = y_{k,l} - y_{k+d1,\, l+d2} \qquad (2)$$

Thus, the dynamic range of the sum and difference is generates twice the range of the original image. Now, the sum and difference histograms with parameters ($d1$, $d2$) over the domain D are defined in a manner similar to the dependence matrix definition

$$hs(i; d1, d2) = hs(i) \qquad (3)$$
$$hd(j; d1, d2) = hd(j) \qquad (4)$$

With, the total number of elements

$$N = \sum_i hs(i) = \sum_j hd(j)\,, \qquad (5)$$

the normalized sum and difference histograms probabilities are

$$\hat{P}s(i) = hs(i) / N; \quad (i = 2,\ldots\ldots\ldots,2Ng) \qquad (6)$$
$$\hat{P}d(i) = hd(j) / N; \quad (j = Ng+1,\ldots,Ng-1) \qquad (7)$$

Where, Ng is quantized gray level.

Each feature measure is obtained for user defined distances (d = 1, 2, 3….) and 4 angles (θ =0º, 45º, 90º, 135º). For each distance, the average across the 4 angles is done to provide "rotation invariance" to the technique.

Table 2 shows the various second order statistical features derived from the normalized sum and difference histograms probabilities shown by equation 6 and equation 7.

TABLE 2
2nd ORDER STATISTICAL FEATURES DERIVED FROM SUM AND DIFFERENCE HISTOGRAM PROBABILITIES

Texture Feature	Sum & Difference Histogram
Mean	$f1 = 0.5 \sum_i i.\hat{P}s(i)$
Variance	$f2 = 0.5 \{\sum_i (i-2\mu)^2.\hat{P}s(i) + \sum_j j^2\, \hat{P}d(j)\}$
Energy	$f3 = \sum_i \hat{P}s(i)^2.\sum_j \hat{P}d(i)^2$
Correlation	$f4 = 0.5 \{\sum_i (i-2\mu)^2.\hat{P}s(i) - \sum_j j^2.\hat{P}d(j)\}$
Entropy	$f5 = -\sum_i \hat{P}s(i).\log(\hat{P}s(i)) - \sum_j \hat{P}d(j).\log(\hat{P}d(j))$
Homogeneity	$f7 = \sum_j 1/1+j^2.\, \hat{P}d(j)$

3 Methodology and Algorithm

The scheme and methodology for sum and difference histogram probabilities calculation and feature extraction is shown in Fig. 2.

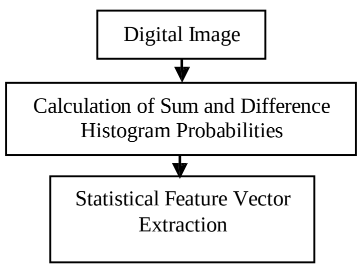


Fig. 2 Scheme for statistical textural feature vector extraction

In current approach use following steps to derive statistical texture feature using sum and difference histogram.

Step-1 Read the digital image and determine size of image.

Step-2 Using linear processing divides the whole digital image into desired size block.
Step-3 Enter the distance (d) over which sum and difference is calculated.
Step-4 Read the each block of the image and determine 1st order statistical features.
Step-5 Do padding on each block to calculate sum and difference over desired distance.
Step-6 Determine sum (S1,S2,S3,S4) and difference (D1,D2,D3,D4) of each image block in step-6 for angle $0^{\circ}, 45^{\circ}, 90^{\circ}, 135^{\circ}$.
Step-7 Calculate sum and difference probabilities for each sum (S1, S2, S3, S4) and difference (D1, D2, D3, D4) obtained in step-7.
Step-8 Apply second order statistics on normalized sum and difference probabilities obtained in step-8 to obtain 2nd order statistical feature.
Step-9 Until last block of the image reached, repeat step-5 to step-9.
Step-10 Combine 1st and 2nd order statistical features.
Step-11 Do averaging on combined features to obtained rotation invariance statistical feature vector.

4 Image Segmentation

4.1 Segmentation

Various medical imaging modalities such as X-ray, CT, MRI, PET and ultrasound are indispensable for the precise analysis of various medical pathologies. Computer power and medical scanner data with appropriate segmentation technique is necessary to extract the necessary boundaries, surface and segmented volumes in the spatial and temporal domain. This technique of data extraction is segmentation. 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.

5 Technique for Image Segmentation

Several general-purpose algorithms and techniques have been developed for image segmentation. Since there is no general solution to the image segmentation problem.

In this paper the authors used clustering technique to segment the image because 1) Easy to implement 2) It is unsupervised learning technique and 3) It permits to combine information from a set of different images.

6 FCM with Subtractive Clustering

FCM is an iterative, unsupervised clustering algorithm and has been applied widely to medical image segmentation (SUPATRA SAHAPHONG, 2007) [6], and regarded as one of the most promising method. Fuzzy c-mean is straightforward implementation, fairly robust behavior and has the ability of uncertainty data modeling. It is very sensitive to its initial cluster number value when we use fuzzy c-means (FCM) for fuzzy clustering. It also requests to give the number of clusters before we use it and wrong value of it does not give optimal solution. To obtain optimum segmentation result and to make segmentation process automatic in this paper the authors used Subtractive Clustering technique. Subtractive clustering is a fast, one-pass algorithm for estimating the number of clusters and the cluster centers in a set of data. The cluster estimates obtained by Subtractive Clustering technique is used to initialize iterative optimization-based FCM algorithm.

FCM is based on minimizing an objective function, with respect to fuzzy membership U, and set of cluster centroids *V:*

$$J_{FCM}(U, V : X) = \sum_{i=1}^{C}\sum_{k=1}^{n}(u_{ik})^m (\|x_k - v_i\|)^2$$

In the above equation, $X=\{x_1,x_2,\ldots xi\ldots x_N\}$ is a p x n data matrix, where p represents the dimension of each x_i, feature vector, and n represents the number of feature vectors. C is the number of clusters. $U_{ik} \in U\,(pxnxC)$, is the membership function of vector x*i* to the i-th cluster, which satisfies: $u_{ik} \in [0,1]$ and $\sum_{i=1}^{C} uik = 1, \forall k$. Where matrix $U = [\,u_{ik}\,]C \times n$, while set $V = \{v_i\}_{i=1}^{C}, v_i \in R^q$ is the prototype of i-th cluster *m* =>1 is a weighting exponent which determines the degree of fuzziness of FCM. ‖X*k*−V*i*‖ is an inner product induced norm on ***to*** measure the distance from X*k* to V*i*. The higher the membership, the higher the possibility that the pixel belongs to the cluster. The optimal partition is accessed via minimizing approximately the sum of intra-cluster squared errors.

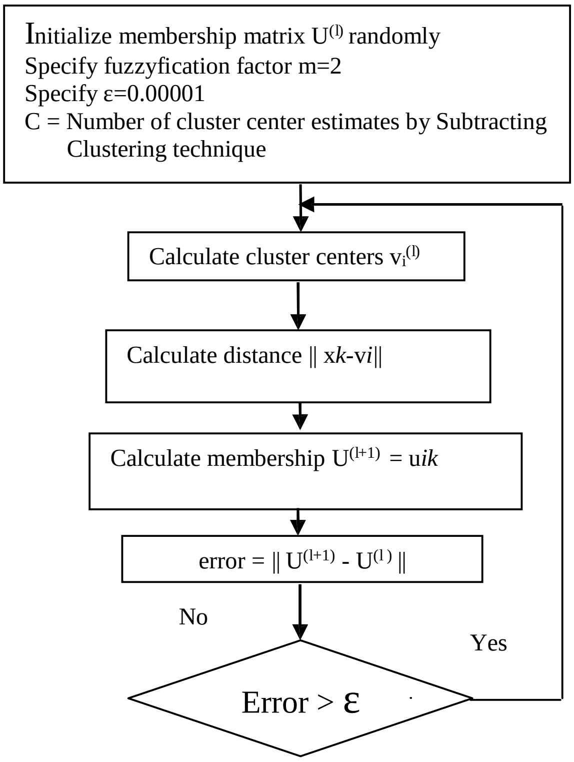
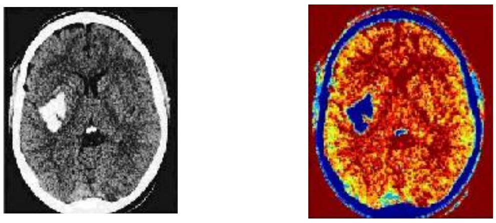


Fig. 3 Flow chart of FCM Algorithm with Subtractive Clustering Technique

7 Experimental Results

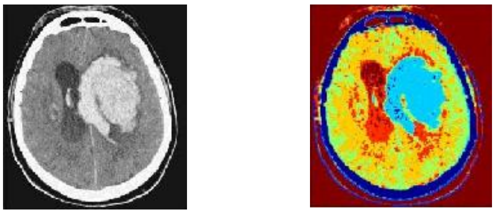
Algorithm is tested for various biomedical images of modalities like CT, MRI and ultrasound.

Original Images Segmented Images

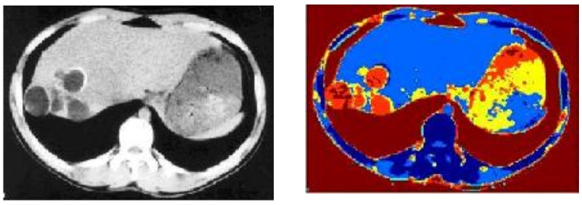


(a) CT image of Right ganglionic Hematoma

Original Images Segmented Images



(b) CT image of Intracerebral hemorrhage



(c) MRI image of Multiple hydatid cysts of the liver

Fig. 4 Original images & segmented images using proposed algorithm.

TABLE 3
ORIGINAL PIXEL VALUES AND DERIVED STATISTICAL TEXTURAL FEATURES

Original Pixel Value							Statistical Textural Features							
241	249	247	240	242	245	242	Mean	Variance	Energy	Correlation	Contrast	Homogeneity	Mean	SD
250	247	246	246	246	245	250	54.87	4403.7	0.014137	3725.9	3.6667	0.22404	247.33	2.0000
250	247	246	249	250	250	246	54.87	4402.7	0.014127	3924.6	2.3511	0.22770	247.22	2.1667
248	248	243	236	240	244	254	54.69	4437.0	0.012095	3898.0	6.9444	0.19097	246.00	2.8284
247	244	241	229	223	220	238	54.54	4828.8	0.019624	3876.2	5.7500	0.18758	245.44	2.351
246	246	246	243	241	240	242	54.31	4312.5	0.013903	3845.7	4.3333	0.18825	245.22	2.0480
248	245	241	245	247	248	249								
Right ganglionic Hematoma														
176	215	222	197	210	219	216	Mean	Variance	Energy	Correlation	Contrast	Homogeneity	Mean	SD
209	222	231	193	175	194	225	43.95	3462.1	0.011088	3083.0	52.500	0.051424	215.33	16.031
216	223	224	186	173	191	224	49.08	3553.5	0.011164	3152.4	29.556	0.054633	221.56	07.6992
229	228	212	206	200	206	227	48.90	3521.5	0.011736	3096.1	39.472	0.079871	222.56	06.8577
227	230	214	214	210	216	232	49.18	3544.4	0.010102	3144.1	17.250	0.079461	223.89	06.9722
221	231	223	213	206	216	238	49.79	3652.0	0.010898	3236.3	44.806	0.053849	225.44	09.1808
211	232	240	228	192	208	244								
Intracerebral hemorrhage														
093	117	114	098	094	112	125	Mean	Variance	Energy	Correlation	Contrast	Homogeneity	Mean	SD
099	112	101	090	084	098	114	20.30	735.24	0.010212	522.62	45.889	0.060800	101.56	10.979
092	095	097	102	097	096	101	21.58	721.73	0.010326	718.89	70.500	0.060412	102.22	9.5757
089	088	091	093	100	101	104	20.55	739.44	0.010555	557.21	41.000	0.060607	56.333	7.2801
096	102	105	096	108	121	111	21.27	669.42	0.011355	593.90	90.278	0.049225	55.333	5.5678
109	112	110	097	108	121	113	21.41	703.38	0.012612	588.52	43.833	0.051609	100.56	9.1394
112	115	122	112	111	110	111								
Multiple hydatid cysts of the liver														

Fig. 4 (a), (b) and (c) shows the original images and its segmented counterparts using the proposed algorithm. As shown in fig. 4 the FCM with clustering technique gives better segmentation then other *state -of- art* clustering techniques. Table 3 shows the part of the pixel values of the original images and statistical textural features calculated from it.

TABLE 4
TIME COMPARISION BETWEEN SUM AND DIFFERENCE HISTOGRAM AND OTHER METHOD FOR STATISTICAL FEAURE CALCULATION

Image	Other Method	S & D Histogram
Right ganglionic Hematoma	2447.06 Sec.	436.65 Sec.
Intracerebral Hemorrhage	2927.58 Sec.	477.44 Sec.
Multiple hydatid cysts of the liver	2857.90 Sec.	471.60 Sec.

Table 4 shows that the Sum and Difference Histogram is fast to calculate the statistical textural feature.

8 Conclusions

The algorithm presented by the authors confirms the use of sum and difference histogram probabilities for fast statistical textural feature extraction and medical image segmentation by FCM using subtractive clustering technique. The results obtained by authors for CT and MRI images of brain Hematoma, brain hemorrhage and multiple cysts in liver are promising. It can be applied to all the three popular medical imaging modalities of CT, MRI and ultrasound types for better visualization of diagnostic features.

9 References

[1] M.Unser, “Sum and difference Histograms for Texture Classification,” IEEE Transaction. on PAMI, Vol.8, pp.118-125, Jan.1986.

[2] R. M. Haralick, “Statistical and structural approaches to texture”, *Proc. of IEEE* , Vol.67, No.5, pp.786-804, May 1979.

[3] D.A. Clausi and M.E. Jernigan, “A fast method to determine cooccurrence texture features,” *IEEE Trans. Geosci. Remote Sensing*, vol. 36, pp. 298-300, Feb. 1998.

[4] R. M. Haralick, K. Shanmugan, and I. Dinstein, “Textural features for image classification,” IEEE Trans. on Systems, Man and Cybernet, Vol.3, pp. 610-621, Nov. 1973.

[5] Mounir Sayadi, Lotfi Tlig and Farhat Fnaiech, “A New Texture Segmentation Method Based on the Fuzzy C-Mean Algorithm and Statistical Features,” Applied Mathematical Sciences, Vol. 1, no. 60, pp. 2999 – 3007, 2007.

[6] Supatra Sahaphong, Nualsawat Hiransakolwong, " Unsupervised Image Segmentation Using Automated Fuzzy c-Means," IEEE, Seventh International Conference on Computer and Information Technology, 2007.

[7] Wen-Wan Ltu, Chun-Jing Xiao, Bao-Wen Wang, Yan Shi and Shu-Fen Fang, " Study On Combining Subtractive Clustering With Fuzzy C-Means Clustering." IEEE, Proceedings of the Second International Conference on Machine Learning and Cybernetics, Wan, 2-5 November 2003.

Prof. H. A. Patel was born in Visnagar, Gujarat, in 1977. He received the B.Tech degree in Instrumentation and Control from S. S. Engg. College, Bhavnagar in 1999. He received M.Tech Degree in Biomedical Engineering from IT-BHU, Varanasi in 2008. He has fifteen years of teaching experience and presently working as a Associate Professor at U. V. Patel College of Engineering, Ganpat University. Mehsana, Gujarat.

Dr. Asim Banerjee did his graduation in Electronics from the Pune University and went on to do his Masters in Electronic Instrumentation from NIT, Warangal. He joined IIT, Bombay as a research engineer in the DRDO sponsored project and then went on to do his PhD from IIT, Bombay in the area of Bio-medical Image processing. Presently working as a Professor at DAIICT, Gandhinagar, Gujarat.