A SEQUENTIAL K-MEANS CLUSTERING FOR MAMMOGRAM SEGMENTATION

[1]K. KARTEEKA PAVAN, [2]CH. SRNIVASA RAO

[1,2]RVR & JC College of Engineering, Chowdavaram, Guntur-522019, India
E-mail: [1]karteeka@yahoo.com, [2]chereddy_sriny@yahoo.com

Abstract— K-means is the simple, efficient clustering technique from past 50 years. Application of k-means is an important characteristic in many applications including medical image segmentation. One of the drawbacks in k-means algorithm is it does not use the spatial information of image space in the clustering process. This paper proposes a simple algorithm that combine intensity and texture based clustering in sequence using k-means to determine regions of interest in mammograms. Experiments are conducted on each image of MIAS database. The results demonstrated the accuracy and efficiency of the algorithm in identifying the masses of mammograms.

Index Terms- -means, intensity, Region of interest, segmentation, texture.

I. INTRODUCTION

Mammography is became the most reliable method that help radiologist in early detection of abnormalities and treatment planning1. Several Computer aided detection methods are available in literature for the detection and classification of mammograms2-7. In all such methods image segmentation is an important issue. Segmentation plays an important role in wide variety of applications including remote sensing and medical imaging. Image segmentation is the process of grouping the pixels of image space into homogenous regions, with respect to specific characteristics. Real world applications may involve multiple characteristics. Mammogram segmentation can be posed as one of such problems, involves multiple characteristics.

Clustering methods are one of the most commonly used techniques for image segmentation8. K-means is the simple, efficient partitional clustering and one of the top 10 clustering algorithms from past 50 years 9 -11. Sahiner et al used k-means for mass segmentation 12-13. Performance of the clustering algorithms is improved by involving multiple features. Li et al incorporated spatial information using adaptive thresholding 14-15. The Fuzzy C-Means (FCM) is a soft clustering algorithm in which each element is associated to each cluster using a fuzzy membership 16. Velthuizen, Chen and Lee used FCM with different objectives to find homogeneous regions with respect to grey-level values 17-18.

Texture analysis is also one of the methods widely used in image segmentation and classification. Texture describes the spatial distribution of colour, orientation and intensity in an image. Textural features can be classified into Statistical approaches, filtering methods, structural approaches and probability models. Textural features are used for image classification, information retrieval, segmentation etc. 19-22. Texture segmentation is a process of partitioning an image into similar regions on the basis of texture features. K-means is used to estimate the parameters in Fisher Linear Discriminant for texture segmentation 23. A combination of contour and texture analysis is used for image segmentation 24. A method is proposed for segmentation using colour fuzzy texture descriptor for colour texture images 25. A disadvantage of watershed algorithm is overcome using morphological dilation to integrate intensity and texture gradients 26. Little work can be observed in the texture segmentation using co-occurrence matrices 27. This paper proposes a sequential k-means clustering algorithm to determine the region of interest in mammogram by combining intensity and textual features of image space. Textural feature, homogeneity, is extracted by constructing GLCM for each pixel with the associated neighbour pixels of size 5x5 window. The experimental results are conducted on each image of MIAS database. The results are very prominent and the algorithm determined the masses accurately.

II. METHODOLOGY

Image segmentation is a process of partitioning an image into homogeneous groups with respect to a specific characteristics i.e., intensity, textual information, shape etc., Image segmentation can be define as follows. Segmentation of an image I into a set of K regions Rk,

k=1,2,.....K, such that

1. $\bigcup_{k=1}^{k} R_K = I$ with $R_i \bigcap R_j = \Phi, i \neq j$
2. $H(R_k) =$ true $\forall k$
3. $H(R_i \bigcup R_j) =$ false $\forall R_i$ and R_j adjacent

Where I is the image and H is the predictive of homogeneity. The proposed segmentation algorithm for mammograms contains 2 phases in sequence using k-means. Clustering the image with respect to pixel intensity is the first phase. Clustering the brightest group (find out in the first phase) pixels, by extracting texture feature for each pixel, is the second phase. Output the brightest group in the second as the result. The proposed algorithm presented in the following figure Fig.1.

.

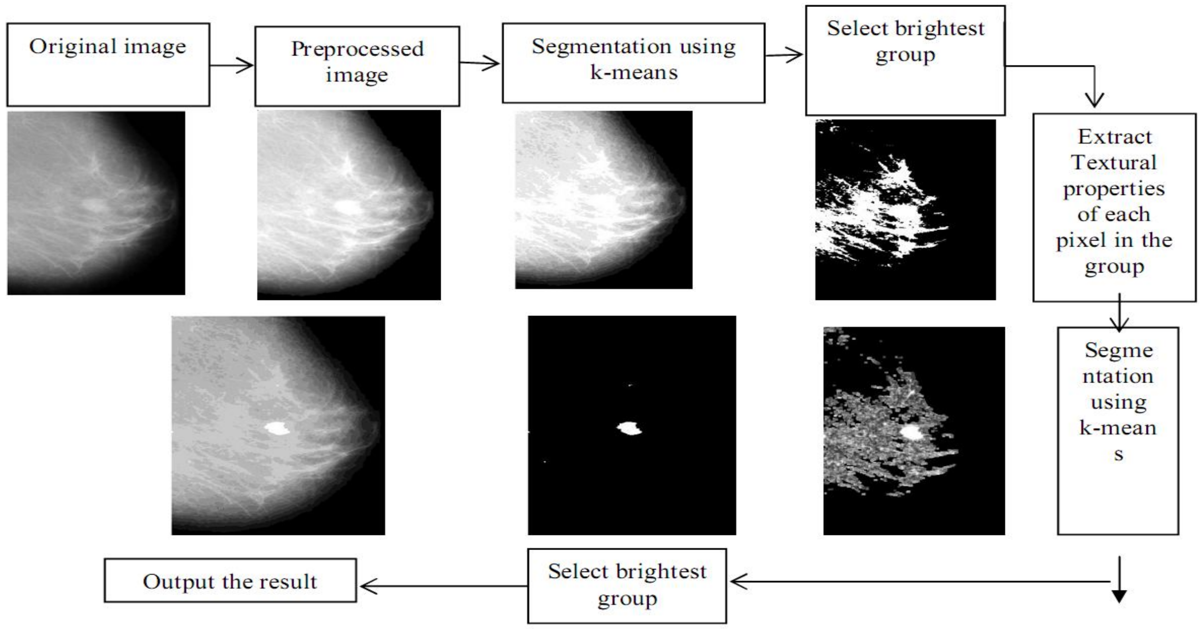

Fig. 1. The sequence k-means for mammogram segmentation

The methodology is explained in detail with a sample Image (matrix) in the following steps.

Step 1: In order to find accurate regions of masses, a pre-processing is necessary for each mammogram. In this work, each mammogram is pre-processed using mathematical morphology 28. Assume that the pre-processed image is as follows.

Image = [

24	25	25	24	25	26	27	28	29	27	26	26	27	27	29
24	25	25	25	25	26	27	28	29	27	27	27	27	28	29
24	25	25	25	25	26	27	29	29	29	28	28	29	29	32
25	26	25	25	26	26	28	29	31	31	31	29	31	32	33
25	26	26	26	26	27	28	29	31	31	31	31	31	32	33
26	26	26	26	26	27	28	31	31	31	31	31	31	32	33
26	27	26	26	27	27	29	31	32	29	29	29	29	31	32
26	27	27	27	27	28	29	31	32	29	28	28	29	31	32
26	27	27	27	28	28	29	29	29	31	31	32	33	33	34
26	27	27	28	28	28	29	29	29	31	31	32	33	34	34

]

Step 2: Apply the k-means clustering considering pixel intensities. Resultant segmentation of the "Image", Clust1, is as follows assume k=2.

Clust1=[

2	2	2	2	2	2	2	2	1	2	2	2	2	2	1
2	2	2	2	2	2	2	2	1	2	2	2	2	2	1
2	2	2	2	2	2	2	1	1	1	2	2	1	1	1
2	2	2	2	2	2	2	1	1	1	1	1	1	1	1
2	2	2	2	2	2	2	1	1	1	1	1	1	1	1
2	2	2	2	2	2	2	1	1	1	1	1	1	1	1
2	2	2	2	2	2	1	1	1	1	1	1	1	1	1
2	2	2	2	2	2	1	1	1	1	2	2	1	1	1
2	2	2	2	2	2	1	1	1	1	1	1	1	1	1
2	2	2	2	2	2	1	1	1	1	1	1	1	1	1

]

Step 3: Select the brightest group. Make all pixels in brightest group to "1" and remaining as "0". The resultant matrix for the labelled matrix in step 2 is as follows.
Bright=[

0	0	0	0	0	0	0	0	1	0	0	0	0	0	1
0	0	0	0	0	0	0	0	1	0	0	0	0	0	1
0	0	0	0	0	0	0	1	1	1	0	0	1	1	1
0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
0	0	0	0	0	0	1	1	1	1	1	1	1	1	1
0	0	0	0	0	0	1	1	1	1	0	0	1	1	1
0	0	0	0	0	0	1	1	1	1	1	1	1	1	1
0	0	0	0	0	0	1	1	1	1	1	1	1	1	1

]

Step 4. Construct GLCM for each pixel by selecting a neighbourhood window of size 5x5 for each pixel. Extract textural feature, homogeneity, and place in the pixel position. The resultant matrix is as follows.
Texture_HOM=[

0	0	0	0	0	0	0	0	0.5	0	0	0	0	0	0.66
0	0	0	0	0	0	0	0	0.5	0	0	0	0	0	0.66
0	0	0	0	0	0	0	0.5	0.5	0.5	0	0	0.66	0.66	0.66
0	0	0	0	0	0	0	0.5	0.5	0.5	0.66	0.66	0.66	0.66	0.66
0	0	0	0	0	0	0	0.5	0.5	0.5	0.66	0.66	0.66	0.66	0.66
0	0	0	0	0	0	0	0.5	0.5	0.5	0.65	0.65	0.65	0.65	0.65
0	0	0	0	0	0	0.5	0.5	0.5	0.5	0.65	0.65	0.65	0.65	0.65
0	0	0	0	0	0	0.5	0.5	0.5	0.5	0	0	0.65	0.65	0.65
0	0	0	0	0	0	0.5	0.5	0.5	0.5	0.65	0.65	0.65	0.65	0.65
0	0	0	0	0	0	0.5	0.5	0.5	0.5	0.65	0.65	0.65	0.65	0.65

]

Step 5: Segmentation using k-means,assume k=2. The result for the example is as follows.
Clust2=[

0	0	0	0	0	0	0	0	1	0	0	0	0	0	2
0	0	0	0	0	0	0	0	1	0	0	0	0	0	2
0	0	0	0	0	0	0	1	1	1	0	0	2	2	2
0	0	0	0	0	0	0	1	1	1	2	2	2	2	2
0	0	0	0	0	0	0	1	1	1	2	2	2	2	2
0	0	0	0	0	0	0	1	1	1	2	2	2	2	2
0	0	0	0	0	0	1	1	1	1	2	2	2	2	2
0	0	0	0	0	0	1	1	1	1	0	0	2	2	2
0	0	0	0	0	0	1	1	1	1	2	2	2	2	2
0	0	0	0	0	0	1	1	1	1	2	2	2	2	2

]

Step6: Output the brightest group. The final output for the selected matrix is as follows.

0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
0	0	0	0	0	0	0	0	0	0	0	0	1	1	1
0	0	0	0	0	0	0	0	0	0	1	1	1	1	1
0	0	0	0	0	0	0	0	0	0	1	1	1	1	1
0	0	0	0	0	0	0	0	0	0	1	1	1	1	1
0	0	0	0	0	0	0	0	0	0	1	1	1	1	1
0	0	0	0	0	0	0	0	0	0	0	0	1	1	1
0	0	0	0	0	0	0	0	0	0	1	1	1	1	1
0	0	0	0	0	0	0	0	0	0	1	1	1	1	1

1. Experimental Results

The experiments are conducted on each image of MIAS database. The results in each phase of the proposed sequential k-means are reported in the following table Table1.

Original	Preprocess	Intensity based Clustering	Texture based clustering(Brightest group)	Brightest group	Final output

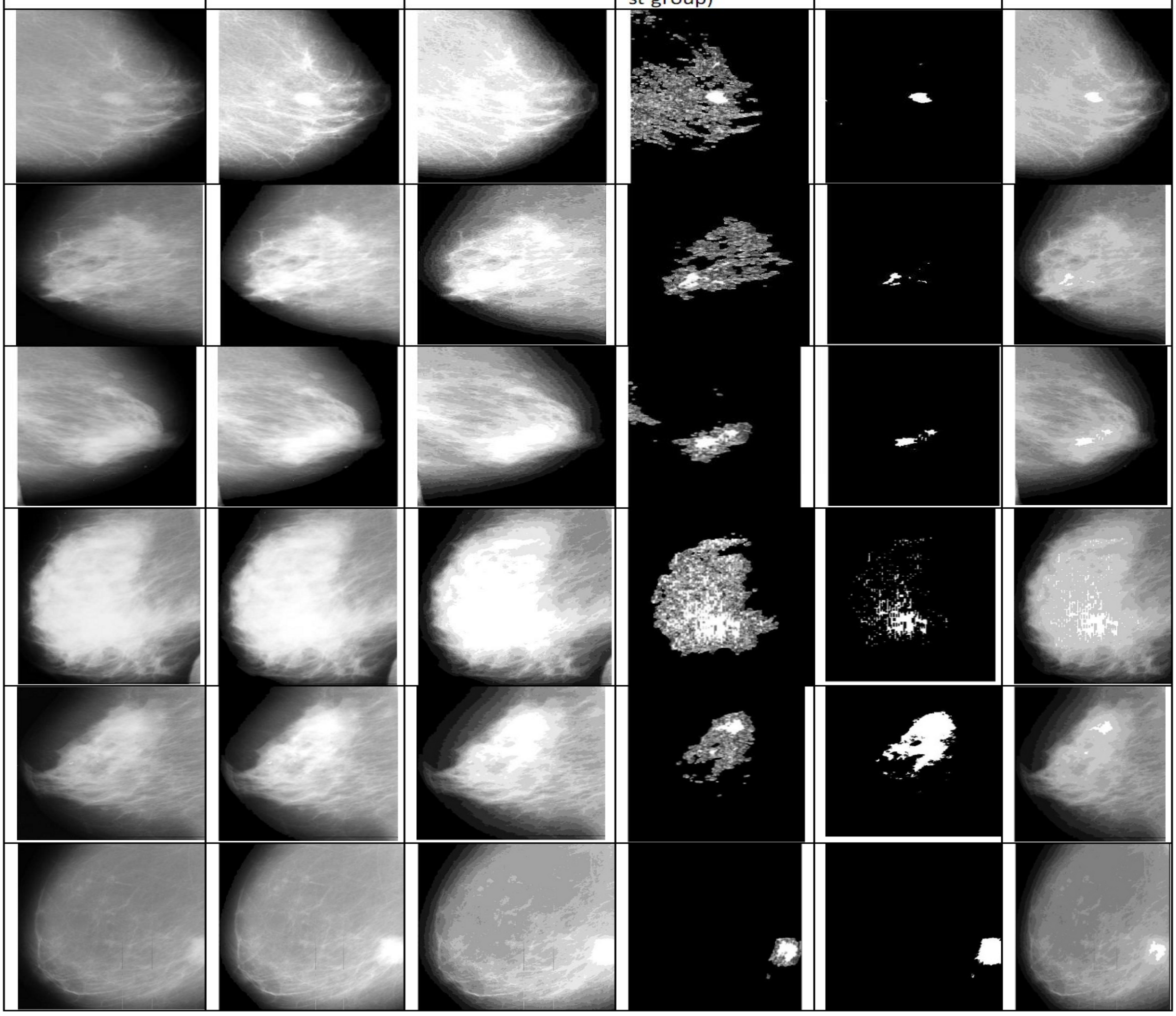

Table1 demonstrates the efficiency and accuracy in finding Regions of interest.

CONCLUSIONS

A simple sequential k-means is proposed to find suspicious masses in the mammograms. Comparative to the other methods, combining more number of features in segmentation, the sequential k-means is very simple. The experimental results demonstrate the efficiency of the proposed algorithm in determining masses accurately from mammograms. Incorporating intensity, texture and shape in clustering process is our future endeavour.

ACKNOWLEDGEMENTS

This work is funded by DST, New Delhi.

REFERENCES

1. Pisano , ED., Cole , EB., Hemminger , BM., Yaffe , MJ., Aylward , SR., Maidment , ADA., Eugene Johnston , R., Williams , MB., Niklason , LT., Conant , EF., Fajardo , LL., Kopans , DB., Brown , ME.,&Pizer , SM. (2000)."Image processing algorithms for digital mammography", a pictorial essay, J Radiogr vol 20,issue 5,400–420 .
2. Claudio Marrocco, Mario Molinara, Ciro D'Elia,& Francesco Tortorella. (2010). "A computer-aided detection system for clustered microcalcifications", Artificial Intelligence in Medicine, vol 50, 23–32.
3. Jiang , J., Yao , B.,& Wason, A.M. (2007). "A genetic algorithm design for microcalcification detection and classification in digital mammograms", Computerized Medical Imaging and Graphics ,vol 31, 49–61.
4. Brijesh Verma, Peter McLeod, & Alan Klevansky. (2009). "A novel soft cluster neural network for the classification of suspicious areas in digital mammograms", Pattern Recognition ,vol 42 , 1845 – 1852.
5. Haralick , RM., Shanmugam, K., &Dinstein, I. (1973)."Textural features for image classification", IEEE Trans Syst Man Cybern ,vol 3,issue:6, 610–621.
6. Chang , HD., Shi , XJ., Min , R., Hu , LM., Cai , XP.,& Du , HN. (2006)." Approaches for automated detection and classification of masses in mammograms", Pattern Recognit ,vol 39, 646–668.
7. Ke, L., Mu, N.,& KangY, Y. (2010). "Mass computer-aided diagnosis method in mammogram based on texture features" Biomedical engineering and informatics (BMEI), 3rd international conference, IEEE explore, 146–149)
8. Jain, A.K., Murty, M.N., Flynn, P.J., 1999. Data clustering: a review. ACM: Computing Surveys 31 (3), 264–323.
9. MacQueen, J.B., 1967. Some methods of classification and analysis of multivariate observations. In: Berkeley Symposium on Mathematical Statistics and Probability, vol. 1, pp. 281–297.
10. Wu, X., V. Kumar, J.R. Quinlan, J. Ghosh, D.J. Hand and D. Steinberg et al., 2008. Top10 algorithms in data mining. Knowl. Inform. Syst. J., 14: 1-37.
11. A.K. Jain "Data Clustering: 50 Years Beyond K-Means" , Pattern Recognition letters, 31, 2010, pp 651-666
12. Sahiner, B., Chan, H.P., Wei, D., Petrick, N., Helvie, M.A., Adler, D.D., Goodsit, M.M., 1996. Image feature selection by a genetic algorithm: application to classification of mass and normal breast tissue. Med. Phys. 23, 1671–1684.
13. Sahiner, B., Chan, H.P., Petrick, N., Helvie, M.A., Goodsit, M.M., 1998. Computerized characterization of masses on mammograms: the rubber band straightening transform and texture analysis. Med. Phys. 25 (4), 516–526.
14. Li, L., Qian, W., Clarke, L.P., 1997. Digital mammography: computer assisted diagnosis method for mass detection with multi-orientation and multiresolution wavelet transform. Acad. Radiol. 4 (11), 724–731.
15. Li, L., Qian, W., Clarke, L.P., Clark, R.A., Thomas, J.A., 1999. Improving mass detection by adaptive and multiscale processing in digitized mammograms. In: Proceedings of SPIE, vol. 3661, pp. 490–498.
16. Bezdek, J.C., 1981. Pattern Recognition with Fuzzy Objective Function Algorithms. Plenum Press, New York.
17. Velthuizen, R.P., 2000. Computer diagnosis of mammographic masses. In: Workshop on Applied Imagery Pattern Recognition, pp. 166–172.
18. Chen, C.H., Lee, G.G., 1997. On digital mammogram segmentation and microcalcification detection using multiresolution wavelet analysis. Graph. Models Image Process. 59 (5), 349–364.
19. S.C. Kim, T.J. Kang, Texture classification and segmentation using wavelet packet frame and Gaussian mixture model, Pattern Recogn. 40 (2007) 1207– 1221.
20. B.S. Manjunath, W.-Y. Ma, Texture features for browsing and image retrieval, IEEE Trans. Pattern Anal. 18 (1996) 837–842.
21. D. Puig, M.A. Garcia, Automatic texture feature selection for image pixel classification, Pattern Recogn. 39 (2006) 1996–2009.
22. Liao, S., & Chung, A. C. S. (2007). Texture classification by using advanced local binary patterns and spatial distribution of dominant patterns. International Conference on Acoustics, Speech and Signal Processing, 1, 1221–1224.
23. Zohra Haliche, Kamal Hammouche, "The gray level aura matrices for textured image segmentation", Analog Integr Circ Sig Process (2011) 69:29–38
24. D.A. Clausi K-means Iterative Fisher (KIF) unsupervised clustering algorithm applied to image texture segmentation, Pattern Recognition 35 (2002) 1959–1972
25. Jitendra malik, serge belongie, thomas leung?and jianbo shi, Contour and Texture Analysis for Image Segmentation, International Journal of Computer Vision 43(1), 7–27, 2001
26. G. Wiselin Jiji, L. Ganesan, A new approach for unsupervised segmentation, Applied Soft Computing 10 (2010) 689–693
27. Padraig Corcoran, Adam Winstanley, Peter Mooney Complementary texture and intensity gradient estimation and fusion for watershed segmentation, , Machine Vision and Applications (2011) 22:1027–1045
28. Jawad Nagi, Sameem Abdul Kareem, Farrukh Nagi, Syed Khaleel Ahmed. (2010) Automated Breast Profile Segmentation for ROI Detection Using Digital Mammograms, IEEE EMBS Conference on Biomedical Engineering & Sciences (IECBES 2010), Kuala Lumpur, Malaysia, 87-92.

★★★