JOURNAL OF INFORMATION, KNOWLEDGE AND RESEARCH IN ELECTRONICS AND COMMUNICATION ENGINEERING

ARCHITECTURE OF HIERARCHICAL K-MEANS CLUSTERING FOR BREAST TUMORS.

[1] MR. VIPUL R. CHAUDHARI, [2] ASST.PROF. N.B.GOHIL , [3] ASST.PROF. VANDANA SHAH

[1]M.E.[EC] Student, Electronics And Communication, S.S.Engineering College,Bhavnagar,Gujarat
[2] Asst.Professor Department Of EC Engineering, S.S.Engineering College,Bhavnagar,Gujarat
[3] Asst.Professor Department Of EC Engineering, SCET, Surat,Gujarat

mevipul_ec@yahoo.com,icc_narendra@yahoo.co.in,vandshah@gmail.com

***ABSTRACT*:** *Mammography is a special case of CT scan who adopts X-ray method & uses the high resolution film so that it can detect the tumors in the breast. Mammography is the widely used for the detection of the breast tumor. The sensitivity of mammography is the percentage of breast cancers detected in a given population, when breast cancer is present. Sensitivity depends on tumor size, conspicuity, and hormone sensitivity as well as breast tissue density. For detection of breast tumor from the image we proposed the Hierarchical K-means clustering algorithm. Clustering for breast tumor image to detection of the which part infected of the breast. Developing algorithms and software to analyse these images may also assist physicians in there daily work.*

Keywords—Mammography, CT scan, X-ray.

I: INTRODUCTION

Breast cancer is the second most common cause of cancer death particularly for women in all over the world. According in india "Breast cancer in India rising rapidly", it is rapidly becoming the number one cancer in females and pushing the cervical cancer to second place. Cancer is one of the biggest threats to human life. It is expected to become the leading cause of death in future. Masses and micro calcification are two important signs that appear in mammogram. Mass detection is more difficult than micro calcification, because masses may have similar density as normal breast tissue. Micro calcification is just the collection of calcium cells. Mass will have different shapes and ill defined boundaries than micro calcification[@]. Mammography is the best available technique to detect cancer cell in its earlier stages. Many other secondary methods are available such as MRT, CT, Ultrasonic. The accordance rate between these instruments and histopathological feature is low, but between mammography and histopathologic diagnosis the rate is quite high. Ultrasonic produces good contrast images but does not contain detailed information.

1.1. Different techniques Used For Breast cancer.

Breast image analysis can be performed using X-rays, magnetic resonance, nuclear medicine or ultrasound [1].

1.1.1 X-ray Mammography

X-ray mammography is mostly used for the clinical practice diagnostic and screening purposes [2]. Mammography uses special X-ray images to detect abnormal growths or changes in the breast tissue.

1.1.2 MRI Of The Breast

Magnetic Resonance Imaging is the most attractive alternative to Mammography for detecting some cancers which could be missed by mammography. In addition, MRI can help radiologists and other specialists determine how to treat breast cancer patients by identifying the stage of the disease [1, 2]. Breast MRI is increasingly used as an adjunct to conventional imaging modalities, particularly in diagnostic problem cases, but also for pre-operative staging. It is an extremely sensitive technique, with relatively limited specificity. To avoid both, false-positive, but particularly false-negative diagnoses, it is imperative to be familiar with the varying MRI appearance of benign and malignant breast tumors. MRI is more sensitive and it can lead to false diagnosis.

II: RESARCH GOAL

The overall in this paper to detect the breast tumor by using the hierarchical k-means algorithm. Since mammography is the currently used main test for detection of the breast tumor, a huge number of mammograms need to be examined by a limited number of radiologists, resulting misdiagnoses due to human errors by visual fatigue. There are several image processing methods proposed for the detection of tumors in mammograms.[3]In this paper new Hk-means clustering algorithm proposed for efficient detection the area of the breast tumor. Figure 1 show different image processing techniques can be used to detect the cancer mass [8].

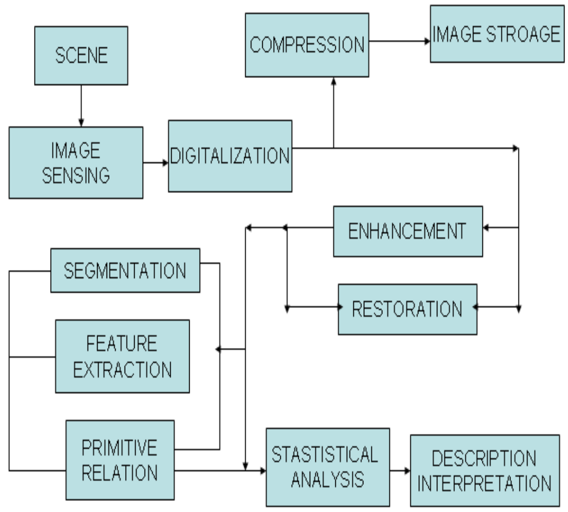


Fig.1 Schematic diagram of different stages of image processing and analysis technique.

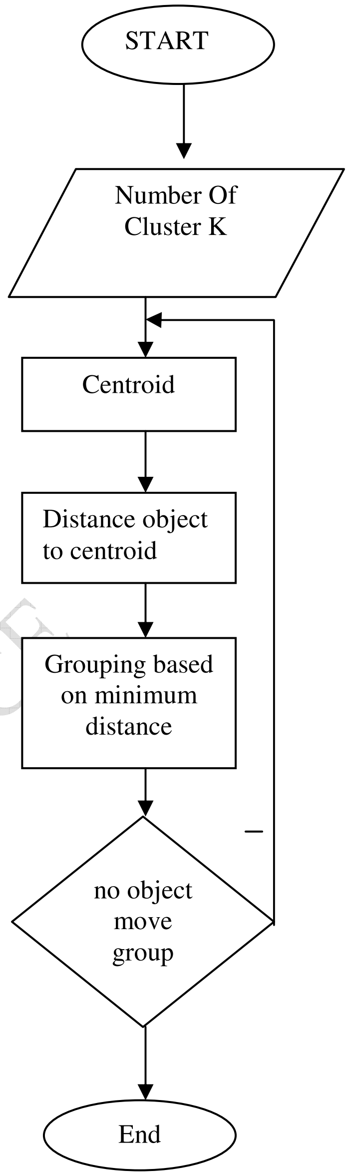
Various technologies such as thresholding, intensity level slicing, contrast-stretching, image-negative, power-transform, logarithm-transform and segmentations have been designed for analysis of tumor and tumor like structures [2].

III PROPOSED HK-MEANS ALGORITHM FOR BREAST CANCER

Hierarchical clustering is a traditional clustering method that generates a tree structure, or a dendrogram, in the clustering process, and there are top-down and bottom-up clustering approaches. Since there are no clustering methods that are suitable for all the problems, many complementary clustering algorithms that combine K-Means with hierarchical clustering are proposed. Wang *et al.*

K-means support vector representation machine (SVRM) clustering. HK-Means clustering algorithm, which is divided into two stages. In the first stage, the agglomerative hierarchical clustering algorithm is employed, and then second stage k-means algorithm used for the improve the quality of result of clustering for the produced first stage and given the final result. The concept of the HK-means algorithm applied to reduce the computational time in the software algorithm when the cluster is large. First the cluster are split in to two small cluster using the Euclidean space[5]. The added benefit of spatial information could further improve algorithm efficiency, allowing potentially simpler and more efficient motion tracking and object occlusion detection. A number of segmentation algorithms for building a hierarchical image representations have been proposed. Although these hierarchical segmentation algorithms perform well in some specific areas, all of them suffer from various drawbacks. However, their methods can lead to excessive iterations, generating a large number of hierarchy levels.

IV. FLOW CHART



Clustering refers to the process of grouping pixels of an image such that pixels which are in the same group (cluster) are similar among them and are dissimilar to the pixels which belong to the other groups (clusters). Let the feature vectors derived from I clustered data. The generalized algorithm initiates k cluster centroids by randomly selecting k feature vectors from X. Later, the feature vectors are grouped into k clusters using a selected distance measure such as Euclidean . The next step is to recompute the cluster centroids based on their group members and then regroup the feature vectors according to the new cluster centroids. The clustering procedure stops only when all cluster centroids tend to converge. K means algorithm starts clustering by determining k initial central points, either at random or using some heuristic data. It then groups each image pixel under the central point it is closest to. Next, it calculates new central points by averaging the pixels grouped under each central point. The two former algorithmic steps are repeated alternately until convergence (central points no longer change by averaging). The limitations of K-means clustering are many iterative rounds may be required. This work strives to reduce that limitation.

Step 1) In the root level, perform K-Means on all the input vectors with cluster number k=2 , and the input vectors are all separated into two child clusters with sequence number 0 and 1 in the first level. Since it is the first step of the algorithm, at least two clusters will be generated in the HK-Means clustering process. Set the current level n=1 and the current sequence number t=0 for the root level. Go to **Step 2**.

Step 2) The stopping sequence number is calculated by using the following equation:

$T_s=K-2^{(n-1)}-1$ 3.1

Where n represents the current level.

The process of HK-Means clustering terminates when the following two conditions are satisfied:

$T= T_{s.}$ and $n= \log_2 k$ 3.2

where t denotes the current sequence number.

When the clustering process terminates, set the final level $n_{final} = n$ and the final stopping sequence number $T_{S,fina\ l} = T_S$ for the label representation of clustering results. Otherwise, go to **Step 3**.

Step 3) If the current sequence number t in the (n-1)th level is equal to $2^{(n-1)}-1$, it means that 2^n clusters are generated in the current level. Jump to the next level by increasing n by 1, set t=0 and go to **Step 4**. Otherwise, increase t by 1 without altering the value of n and go to **Step 4**.

Step 4) Perform K-Means on the input vectors that belong to the t^{th} cluster in the (n-1)level. The cluster number of K-Means is set to . After the K-Means clustering based on the t^{th} cluster in the (n-1)th level is finished, two new clusters with sequence number 2t and 2t+1 are generated in the current level (the nth level). Go to **Step 2**[3].

V. HK means Clustering process

The cluster number in K-Means is becoming larger and larger for newly developed algorithms. Although the related works and previous works offer various options for K-Means clustering with different applications, the large cluster number is still a design challenge for K-Means hardware architectures. In the traditional K-Means, the algorithm complexity is proportional to the cluster number[3] .

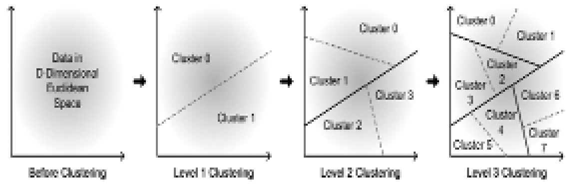


Fig 5.1 clustering processes

VI SIMULATION RESULT

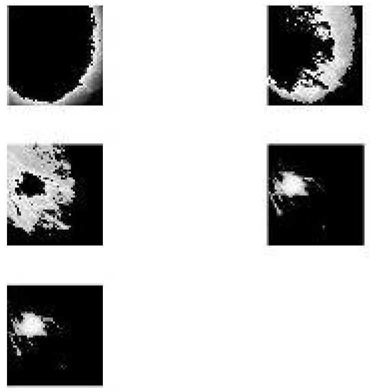


Fig 6.1 image segmentation of breast tumor with 5 cluster.

REFRENCES

[1]. “Detection of breast cancer tumor using mathematical morphology and wavelet analysis”, Mohiy Hadhoud, ohamed Amin, Walid Dabbour in GVIP 05 conference, 19-21 December 2005, CICC, Cairo, Egypt.

[2]. Characterization of micro-calcifications and mass on female breast using processing in full field digital mammography (FFDM)”, Kanaga, K.C.1, Anandan, S.2, Chin, M.Y.1 & Laila, S.E.1, symposium sains kesihatan kebangsaan ke 7, hotel legend, Kuala Lumpur, 18-20 June 2008: 183-187

[3]. Flexible Hardware Architecture of Hierarchical K-Means Clustering for Large Cluster Number. Tse-Wei Chen and Shao-Yi Chien, Member, IEEE, IEEE Transactionn on very large sacle integration(vlsi) systems.

[4]. Bill Andreopoulos, Aijun An, Xiaogang Wang, and Michael Schroeder. “A roadmap of clustering algorithms: finding a match for a biomedical application”, *Brief Bioinform*, February 2009

[5]. Y.C. Hu and M.-G. Lee, “K-means-based color palette design scheme with the use of stable flags,” *J. Electron. Imag.*, vol. 16, no. 3, pp. 033003 1–11, 2007.

[6]. A. G. d. S. Filho, A. C. Frery, C. C. de Araújo, H. Alice, J. Cerqueira, J. A. Loureiro, M. E. de Lima, M. d. G. S. Oliveira, and M. M. Horta,“Hyperspectral images clustering on reconfigurable hardware using

the K-means algorithm," in *Proc. Symp. Integr. Circuits Syst. Des.*, Sep.2003, pp. 99–104.

[7] Zhang, Q.,Goldman,S. A., Yu,W.,and Fritts, J.E..”Content based image retrieval using multiple instance learning”.In Proc.ICML,2002

[8]. T.-W. Chen, C.-H. Sun, J.-Y. Bai, H.-R. Chen, and S.-Y. Chien, “Architectural analyses of K-means silicon intellectual property for image segmentation,” in *Proc. IEEE Int. Symp. Circuits Syst.*, May 2008,pp. 2578–2581.

[9]. Aijuan Dong and Baoying Wang “Feature selection and analysis on mammogram classification“‟, Communications, Computers and Signal Processing, 2009. Pac Rim 2009. IEEE Pacific Rim Conference on 23-26 Aug. 2009

[10].Thangavel, K.Mohideen, A.K. Dept. of Comp.Sci., Periyar Univ., Salem, India, “Semi-supervised k-means clustering for outlier detection in mammogram classification” appears in Trandz in information sciences & computing, (2010).