ISSN: 0976-3031

International Journal of Recent Scientific Research

Impact factor: 5.114

APPLICATION OF DWT FOR MEDICAL IMAGE PROCESSING AND DETECTION ABNORMALITY IN MAMMOGRAPHIC IMAGE

Snehali D. Sable and Alpana Deshmukh

Volume: 6 Issue: 10

THE PUBLICATION OF
INTERNATIONAL JOURNAL OF RECENT SCIENTIFIC RESEARCH
http://www.recentscientific.com
E-mail: recentscientific@gmail.com

Available Online at http://www.recentscientific.com

International Journal of Recent Scientific Research
Vol. 6, Issue, 10, pp. 7052-7056, October, 2015

ISSN: 0976-3031

International Journal of Recent Scientific Research

RESEARCH ARTICLE

APPLICATION OF DWT FOR MEDICAL IMAGE PROCESSING AND DETECTION ABNORMALITY IN MAMMOGRAPHIC IMAGE

Snehali D. Sable[1] and Alpana Deshmukh[2]

[1]VLSI and Embedded GHRIET, Pune India
[2]Department of E&TC, GHRIET, Pune India

ARTICLE INFO

Article History:

Received 05th July, 2015
Received in revised form 08th August, 2015
Accepted 10th September, 2015
Published online 28st October, 2015

Key words:

Breast Cancer, Mammograms, Clustering Technique, Fuzzy C means (FCM), Fuzzy K Means (FKM).

ABSTRACT

Mammographic mass detection is an important task for the early diagnosis of breast cancer. However, it is difficult to distinguish masses from normal regions because of their abundant morphological characteristics and ambiguous margins. To improve the mass detection performance, it is essential to effectively pre-process mammogram to preserve both the intensity distribution and morphological characteristics of regions.

Breast cancer is the second most common cause of cancer death in women. Early detection is the only way to reduce the mortality. Mammography is the best available technique used for earlier detection. Mammography is a special case of CT scan who adopts X-ray method & uses the high resolution film so that it can detect well the tumors in the breast. Low radiation is the strength of this method. Mammogram breast cancer images have the ability to assist physicians in detecting disease caused by cells abnormal growth. Developing algorithms and software to analyze these images may also assist physicians in their daily work. The real-time implementation of this paper can be implemented using data acquisition hardware and software interface with the mammography systems.

In the proposed work breast tumor detection by using fuzzy K-means & fuzzy C-means clustering technique is proposed. K Means algorithm is Centroid Based and Fuzzy C Means is Representative Object Based. These two algorithms are to be implemented and their performance is to be analyzed based on their clustering result quality.

INTRODUCTION

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

X-Ray Mammography

X-Ray Mammography is commonly used in clinical practice for diagnostic and screening purposes [2]. Mammography provides high sensitivity on fatty breast and excellent demonstration of micro calcifications; it is highly indicative of an early malignancy.

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]. To improve the mass detection performance, it is essential to effectively pre-process mammogram to preserve both the intensity distribution and morphological characteristics of regions.

Breast Ultrasound

Ultrasound, also known as *sonography*, uses sound waves to look inside a part of the body. A gel is put on the skin of the breast and a handheld instrument called a transducer is rubbed in the gel and pressed against the skin. The transducer transmits the sound waves through the breast. Echoes from the sound waves are picked up and converted by a computer into a black and white picture that is shown on a computer screen. This test is painless and does not expose you to radiation.

Ultrasound has become a valuable tool to use along with mammograms because it is widely available, non-invasive, and costs less than other options. But the value of an ultrasound test depends on the operator's level of skill and experience. Although ultrasound is less sensitive than MRI (that is, it detects fewer tumours), it has the advantage of costing less and being more widely available.

Corresponding author:* **Snehali D. Sable
VLSI and Embedded GHRIET, Pune India

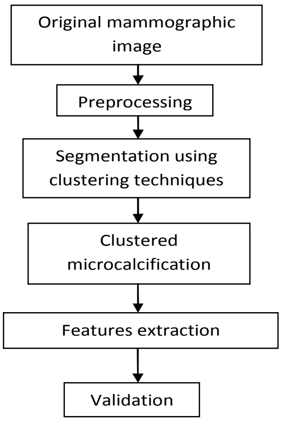
METHODOLOGY

Pre-processing

Mammograms are medical images that are difficult to interpret. Hence pre-processing is essential to improve the quality. It will prepare the mammogram for the next two process segmentation and feature extraction. Digitization noise and high frequency components in the mammography images are removed by using median filter. The selective median filters have proved to be good because they have some very interesting properties:

1. They can smooth the transient changes in signal intensity (e.g., noise);
2. They are very effective for removing the impulsive noises from the signals;
3. They can preserve the edge information in the filtered signal;
4. They can be implemented by using very simple digital nonlinear operations. Edges are the more important factor in the segmentation of mammogram.

Block Diagram



Segmentation

The goal of segmentation is to find out the suspicious or abnormal mass region from mammogram. A mass is space occupying lesion and usually appears as a bright region on a mammogram. So contrast enhancement is implemented in order to extract the brighter region.

Discrete Wavelet Transform (DWT)

The wavelet transform (WT) has gained widespread acceptance in signal processing and image compression. Because of their inherent multi-resolution nature, wavelet-coding schemes are especially suitable for applications where scalability and tolerable degradation are important. Wavelet transform decomposes a signal into a set of basis functions. These basis functions are called wavelets. Wavelets are obtained from a single prototype wavelet y (t) called mother wavelet by dilations and shifting: where a is the scaling parameter and b is the shifting parameter

$$\cdot_{a,b}(t) = \frac{1}{\sqrt{a}} \cdot \left(\frac{t-b}{a}\right)$$

D Wavelet Analysis

- Images are treated as two dimensional signals, they change horizontally and vertically, thus 2-D wavelet analysis must be used for images. 2-D wavelet analysis use the same 'mother wavelets' but requires an extra step at every level of decomposition
- The 1-D analysis filtered out the high frequency information from the low frequency information at every level of decomposition, so only two sub signals were produced at each level.
- In 2-D, the images are considered as matrices with N rows and M columns. At every level of decomposition the horizontal data is filtered, and then the approximation and details produced from this are filtered on columns.
- First apply a 1-D filter bank to the rows of the image. Then apply same transform to the columns of each channel of result. Therefore, obtained 3 high pass channels corresponding to vertical, horizontal, and diagonal, and one approximation image. At every level, four sub images are obtained; the approximation, the vertical details, the horizontal details and the diagonal details.
- The first upper half block (LL) shows the approximation, while second upper half (HL) is shows horizontal detail. First lower level block (LH) shows vertical detail and second lower level block (HH) shows diagonal detail of original image.'

In algorithm there is shown one level discrete wavelet transform. You can also increase the level of DWT by applying this process more than one time. Second and third level DWT gives the better compression ratio. But it will come with loss of some information. First level DWT is quite reasonable for both achieving high compression ratio and also got quality (less MSE).

Inverse Discrete Wavelet Transform

Inverse Discrete Wavelet Transform (IDWT) is applied to the image which is obtained at the output of DWT. But before applying IDWT to the image, four frequency bands of the image which are obtained after performing DWT operation are De- noised i.e. any kind of synthetic or natural noise that may appear in an image is removed, using image de noising functions which are present in the MATLAB 7.9. After performing IDWT operation on an image, we get the recomposed image which more clearly reveals the mass containing lesion in the original medical image of a breast.

K Means Clustering Technique

The main idea behind the k-means algorithm is the minimization of an objective function usually taken up as a function of the deviations between all patterns from their respective cluster centres. The most popular and widely used algorithm among partitioning algorithms is K-means algorithm. It is simple, easy to implement and very fast with linear time complexity.

The K-means algorithm partitions a dataset into k predefined number of clusters that will try to minimize the intra-cluster distance based on Euclidean distance (Jain *et al.*, 1999). K-mean algorithm is very fast and simple algorithm [15]. In statistics and machine learning, k-means clustering is a method of cluster analysis which aims to partition ‘n’ observations in to ‘k’ clusters in which each observation belongs to the cluster with the nearest mean [7]. For a given set of observation (x_1, x_2,...x_n) , where each observation is a d-dimensional real vector, then k-means clustering aims to partition the ‘n’ observations in to ‘k’ sets (k<n), {S= S_1, S_2,...S_n} so as to minimize the within cluster sum of squares (WCSS) in eqn (1).

$$\text{Arg min}_{\,s} \sum_{i=1}^{k} \sum_{x_i \in S_i} \| x_j - \mu_i \|^2 \qquad ...(1)$$

Where, μ_i is the mean of S_i. The number of cluster k is assumed to be fixed in k-means clustering.

Standard algorithm

Given an initial set of k-means which may be specified randomly or by some heuristic, the algorithm produces by alternating between two steps.

Assignment Step

Assign each observation to the cluster with the closest mean (i.e. partition the observation according to the voronoi diagram generated by the means) in equation (2).

$$S_i = \left\{ xj : \| xj - mi\hat{\ } t \| \leq \| xj - mi'\hat{\ } t \| \ \forall\, i' = 1 \ldots k \right\} \qquad ...(2)$$

Update Step

Calculate the new means to be the Centroid of the observations in the cluster in equation (3)

$$m_i^{t+1} = \frac{1}{|Si\hat{\ }t|} \sum_{xj \in Si\hat{\ }t} xj \qquad ...(3)$$

The algorithm is usually very fast, it is common to run it multiple times with different starting conditions. Theoretically it has seen that there exist certain point sets on which k-means takes super-polynomial time, but practically it is not so far.

1. *K* initial "means" (in this case *k*=3) are randomly selected from the data set (shown in colour)
2. *K* clusters are created by associating every observation with the nearest mean. The partitions here represent the Voronoi diagram generated by the means.
3. The Centroid becomes the new means
4. Steps (b) and (c) are repeated until convergence has been reached.

Fuzzy C Means Clustering

Fuzzy C-means (FCM) is a method of clustering technique which allows one piece of data to belong to two or more clusters. This method was developed by Dunn in 1973 and improved by Bezdek in 1981 and it is frequently used in pattern recognition. FCM clustering is one of well known unsupervised clustering Techniques, which can be used for unsupervised image segmentation. The measurement data considered from an unsupervised fuzzy clustering technique is only used to reveal the underlying structure of the data and segment the image in regions with similar spectral properties, so this method has not relationship between pixels in spatial domain; but it just depends on the spectral domain. [16]

The Fuzzy C-means algorithm, also known as fuzzy ISODATA, is one of the most frequently used methods in pattern recognition. Fuzzy C-means (FCM) is a method of clustering which allows one piece of data to belong to two or more clusters [7]. It is based on the minimization of objective function to achieve a good classification. ‘j_m’ is a squared error clustering criterion, and solutions of minimization are least squared error stationary point of J in equation (4).

$$jm = \sum_{i=1}^{k} \quad \sum_{j=1}^{c} uij \ \| xi - cj \| \hat{\ } 2 \qquad ...(4)$$

Where, 1≤m≤ ∞, m is any real number greater than 1, u_{ij} is the degree of membership of x_i in the cluster j, x_i is the d-dimensional measured data, c_j is the dimension centre of the cluster and ‖*‖ is any norm expressing the similarity between any measured data and the centre. Fuzzy partitioning is carried out through an iterative optimization of the objective function shown above, with the update of member ship u_{ij} in equation (5) and the cluster centres’ c_j by equation (5) (6)

$$uij = \frac{1}{\sum_{k=1}^{c} \left[\frac{\| xi - cj \|}{\| xi - ck \|} \right]^{\frac{2}{m-1}}} \qquad ...\ (5)$$

$$cj = \frac{\sum_{i=1}^{k} uij\, xi}{\sum_{i=1}^{k} uij} \qquad ...\ (6)$$

The iteration will stop when

$$\max_{ij} = \{ | u_{ij}^{k+1} - u_{ij}^{k} | \} < \in \qquad ...(7)$$

Where ∈ is the termination criterion between 0 & 1, whereas k is the iteration steps. This procedure converges to a local minimum or a saddle point of j_m.

The fuzzy c means algorithm composed of following steps.

1. Initialize U= [u ij] matrix, $U^{(0)}$
2. At k-step calculate the centre vectors $C^{(k)} = [C_j]$ with $U^{(k)}$.
3. Update, $U^{(k)}$, $U^{(k+1)}$ $uij = \frac{1}{\sum_{k=1}^{c} \left[\frac{\| xi - cj \|}{\| xi - ck \|} \right]^{\frac{2}{m-1}}}$

   If $\| u_{ij}^{k+1} - u_{ij}^{k} \| < \in$ then STOP, otherwise return to step 2

Features Extraction

Texture feature is useful in differentiating normal and abnormal pattern. Texture is an alteration and variation of surface of the image. Texture is characterized as the space distribution of gray levels in neighbourhood. There are two types of texture measures first order and second order. In the first order texture measure are statistics calculated from individual pixel. In second order relationship between Neighbour pixels are considered. In the proposed method Spatial Gray Level Dependence (SGLD) matrix is used for feature extraction which comes under second order texture measure.

Second order statics can be used to model the relationship between pixels within the breast region by constructing SGLD matrix. A SGLD matrix is the joint probability of occurrences of gray levels i and j for the two pixels with a defined spatial relationship in an image. Spatial relationship is defined in terms of distance d and angle θ. SGLD matrix is constructed at a distance d=1, 2, 3, 4 and for angles $\theta=0^0$, 45^0, 90^0 and 135^0. If the texture is course and distance d is small then pair of points at distance d should have similar gray levels. If the texture is fine and distance d is comparable to the texture size then gray level of the two points would be different. Hence texture coarseness should be analyzed with various values of distance d. From SGLD matrices a variety of features may be extracted. Texture descriptors derived from SGLD are contrast, Energy, Homogeneity and Correlation.

Contrast= $\sum_{i,j=0}^{n-1} \text{Pij}\,(\text{i}-\text{j})\hat{}2$

Energy = $\sum_{i,j=0}^{n-1} (\text{Pij})\hat{}2$

Homogeneity = $\sum_{i,j=0}^{n-1} \frac{\text{Pij}}{1+(i-j)\hat{}2}$

Correlation = $\sum_{i,j=0}^{n-1} \text{Pij}\, \frac{(i-\mu)(j-\mu)}{\sigma\hat{}2}$

P ij= Element i, j of the normalized symmetrical GLCM
N is number of gray levels in the image
The GLCM mean, calculated as:

μ = $\sum_{i,j=0}^{N-1} i\, P\, ij$

σˆ2 The variance of the intensities

σˆ2 = $\sum_{i,j=0}^{N-1} Pij\,(i-j)$

Where,

Contrast is the contrast between a pixel and its neighbour. Energy is the sum of squared elements in SGLD or uniformity. Homogeneity is closeness of the distribution of elements in SGLD.

Correlation shows how correlated a pixel is to its neighbour over the whole image.

Approximate Reasoning

In the approximate reasoning step the tumour area is calculated using the binarization method. That is the image having only two values either black or white (0 or1). Here 256x256 jpeg image is a maximum image size. The binary image can be represented as a summation of total number of white and black pixels.

Image, I = $\sum_{W=0}^{255} \sum_{H=0}^{255}[\,f(0)+F(1)]$
Pixels = Width X Height = 256 X 256
F (0) = white pixel (digit 0)
F (l) = black pixel (digit 1)
No of white pixel; P = $\sum_{W=0}^{255} \sum_{H=0}^{255}[f(0)]$
Where,

P = number of white pixels (width*height)
1 Pixel = 0.264 mm
The area calculation formula is
Size of tumour, S= $[(\sqrt{P})0.264]\ mm\hat{}2$
P=no of white pixels; W=width; H=height.

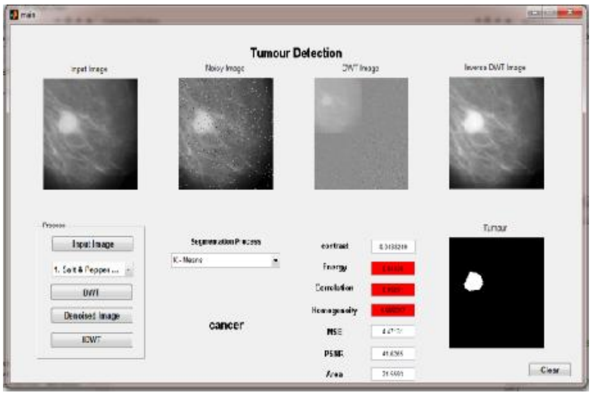


Fig. 1 GUI for Tumour Detection

RESULTS AND DISCUSSION

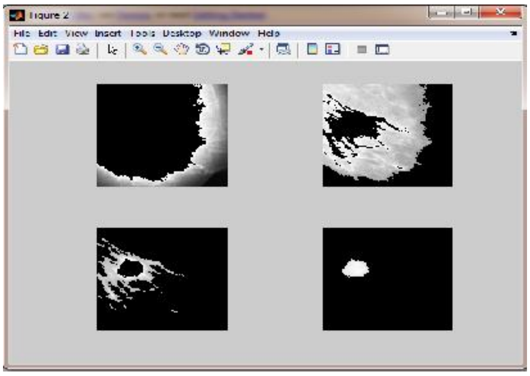


Fig. 2 Result of K Means algorithm

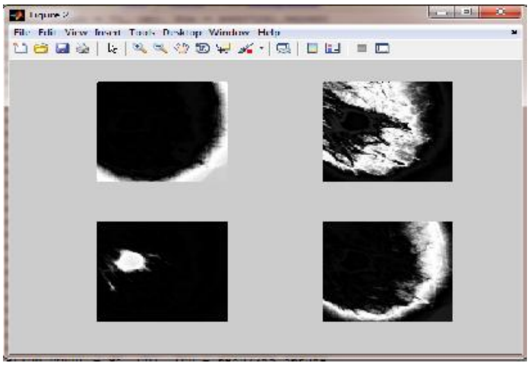


Fig. 3 Result of Fuzzy C means Algorithm

In this paper I have presented a novel approach to identify the presence of breast cancer mass and calcification in mammograms using K-means and Fuzzy C-Means clustering for clear identification of clusters. Combining these we have successfully detected the breast cancer area in raw mammograms images. The results indicate that this system can facilitate the doctor to detect the breast cancer in the early stage of diagnosis as well as classify the total cancer affected area. This will help doctor to take or analyze in which stage of cancer the patient have and according to which he/she can take necessary and appropriate treatment steps.

Table I Result of feature extraction process

Img Id	Img Class	Method	Contrast	Energy	Correlati	Homogen	MSE	PSNR	Area
MAM 3	Cancer	K Means	0.055	0.921	0.985	0.999	8.524	38.824	54.34
	Cancer	FCM	0.029	0.815	0.994	0.994	8.524	38.824	72.63
MAM 4	Cancer	K Means	1.653	0.848	0.716	0.97	147.599	26.44	66.23
	Normal	FCM	0.321	0.812	0.901	0.956	147.59	26.44	56.91
MAM 5	Cancer	K Means	0.058	0.92	0.984	0.998	12.32	37.224	54.6
	Cancer	FCM	0.032	0.828	0.992	0.993	12.32	37.22	62.41
MAM 6	Cancer	K Means	0.046	0.949	0.98	0.999	4.575	41.526	21.47
	Cancer	FCM	0.016	0.864	0.994	0.992	4.575	41.526	26.28
MAM7	Normal	K Means	0.204	0.534	0.99	0.996	9.117	38.534	162.347
	Normal	FCM	0.062	0.466	0.997	0.994	8.99	38.593	197.886
MAM8	Cancer	K Means	0.353	0.74	0.971	0.993	4.685	41.423	53.05
	Cancer	FCM	0.038	0.757	0.991	0.982	3.82	42.309	34.31
MAM9	Normal	K Means	0.228	0.652	0.986	0.995	8.063	39.065	47.18
	Normal	FCM	0.052	0.445	0.997	0.994	8.063	39.065	198.43
MAM10	Cancer	K Means	0.133	0.885	0.975	0.997	7.211	39.55	65.76
	Cancer	FCM	0.03	0.762	0.993	0.986	7.211	39.55	96.71
MAM11	Cancer	K Means	0.066	0.923	0.981	0.998	10.507	37.915	53.3
	Cancer	FCM	0.029	0.784	0.994	0.99	10.507	37.915	72.79
MRI D2	Cancer	K Means	0.107	0.984	0.833	0.833	0.998	29.885	9.93
	Cancer	FCM	0.091	0.942	0.939	0.939	0.989	29.885	17.91
MAM SIDE M2	Cancer	K Means	0.057	0.967	0.961	0.961	0.998	9.032	16.05
	Cancer	FCM	0.069	0.833	0.979	0.979	0.979	9.032	27.98
MAM TOP D2	Cancer	K Means	0.367	0.94	0.854	0.854	0.993	40.263	20.56
	Cancer	FCM	0.178	0.874	0.935	0.935	0.981	40.263	24.96

Performance Evaluation of K-Means & Fuzzy C-Means

The main difference is that, in FCM, each point has a weighted associated with a particular cluster. So, a point doesn't in a cluster as much as has a weak or strong association to cluster, which is determined by the inverse distance to the centre of the cluster. FCM will tend to run slower than K-means, since it is actually doing more work [7]. Each point is evaluated with each cluster, and more operations are involved in each evaluation. K-means just needs to do a distance calculation, whereas FCM needs to do a full inverse-distance weighting. In this proposed work we make these differences or weakness our strong point for full detection of breast cancer. From this we were able to find out the masses as well as the cancerous area i.e. how far the cancer has affected the breast [5]. A real-time system can be implemented using suitable data acquisition software and hardware interface with digital mammography systems.

Acknowledgment

'Also to The medical images used in the experiment of the proposed work were taken from P. H. Medical Center, Pune and Sudarshan technologies.'

References

1. Rafael C.Gonzalez and Richards E.Woods, 2007. "Digital Image Processing", Prentice Hall.
2. Nalini Singh Asst. Professor (AE&I Dept.),Silicon Institute of Technology, Bhubaneswar, India., "Breast Cancer Mass Detection in Mammograms using K-means and Fuzzy C-means Clustering", *International Journal of Computer Applications* (0975 – 8887) Volume 22– No.2, May 2011
3. J. selvakumar A. Lakshmi T. Arivoli Department of ECE, Kalasalingam University, India., "Brain Tumor Segmentation and Its Area Calculation in Brain MR Images using K-Mean Clustering and Fuzzy C-Mean Algorithm", IEEE-International Conference On Advances In Engineering, Science And Management (ICAESM -2012) March 30, 31, 2012
4. S. Meenalosini "Computer Aided Diagnosis of Malignancy in mammograms, European Journal of Scientific Research, ISSN 1450-216X Vol.72 No.3 (2012), pp. 360-368.
5. M. P. Sampat, M. K. Markey, and A. C. Bovik, "Computer-aided detection and diagnosis in mammography," in Handbook of Image and Video Processing, 2nd ed. New York: Academic, 2005, pp. 1195–1217
6. R. M. Rangayyan, F. J. Ayres, and J. E. Leo Desautels, "A review of computer-aided diagnosis of breast cancer: Toward the detection of subtle signs," *J. Franklin Inst., vol. 344, no. 3–4, pp. 312–348, May/Jul. 200*
7. American Cancer Society. (2008). "Cancer facts and figures 2008," in *Statistics for 2008. Available: http://www.cancer.org/downloads/*
8. Performance Evaluation of K-Means and Fuzzy C-Means Clustering Algorithms for Statistical Distributions of Input Data Points, European Journal of Scientific Research ISSN 1450-216X Vol.46 No.3 (2010), pp.320-330
9. 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)
10. Cahoon, T.C .Sutton, M .A. Bezdek "Brest cancer detection using image processing techniques", J.C. Dept. of Comp. Sci. Univ. of West Florida, Pensacola, FL Fuzzy IEEE 2000.The 9th IEEE conference
11. Sameti,M.; Ward,R.K.; Morgan-Parkes,J.; Palcic,B.; "Image feature extraction in the last screening mammograms prior to detection of breast cancer" in IEEE journal 2009.
12. S. Zulaikha Beevi M, Mohamed Sathik "An Effective Approach for Segmentation of MRI Images: Combining Spatial Information with Fuzzy C-Mean Clustering," *European Journal of Scientific Research*, ISSN I450-2I6X Vol. 41 No.3 pp.437-45I
13. K.S. Ravichandran and 2B. Ananthi(2009), "Color Skin Segmentation Using K-Means Cluster," *International Journal of Computational and Applied Mathematics ISSN 1819-4966 Volume 4 Number* 2,pp. 153 -157
14. R.Nithya and B.Santhi, 2011. "Comparative Study on Feature Extraction Method For Breast Cancer Classification", *Journal of Theoretical and Applied Information Technology* 33, pp. 220-226.

****

International Journal of Recent Scientific Research

ISSN 0976-3031

9 770976 303009