Automated Segmentation of Breast Cancer Lesions in Ultrasound Images Using Modified Fuzzy Possibilistic C-Means with Replusions Clustering and Generalized Gradient Vector Flow Snake Algorithm

Alamelumangai. N[1], Dr. Devishree. J[2]

1 Assistant Professor, Department of MCA, Karpagam College of Engineering, Coimbatore.
Email: alamelumangai.n@gmail.com
2 Assistant Professor (SG), Department of Electrical and Electronics Engineering, Coimbatore Institute of Technology, Coimbatore.

Abstract: This paper presents a methodology to incorporate Modified Fuzzy Possibilistic C-Means (MFPCM) with Repulsions clustering and Generalized Gradient Vector Flow (GGVF) snake algorithm for lesion contour segmentation on breast ultrasound (US) images. Preprocessing US images is mandatory and involves two stages- Memetic Adaptive Neuro Fuzzy Inference System (ANFIS) to remove speckle noise and Hough Transformation for contrast enhancement. Segmentation is performed on preprocessed US images based on region-of-interest (ROI) and the lesion is separated from background using MFPCM with Repulsions. To further improve the shape, boundary of the lesion is identified using active contour method, GGVF snake algorithm. The proposed methodology is compared with the Fuzzy C-Means clustering and traditional contour detection methods for a database of 50 lesions (18 benign and 32 malignant cases) and yield better results in detecting the boundary of the lesion. The shape and size of the lesion classify the lesion as, benign, pre-malignant and malignant. The proposed method assist radiologists to determine and improve the efficiency of ultrasound breast cancer image elucidation.
[Alamelumangai. N[1], Dr. Devishree. J. **Automated Segmentation of Breast Cancer Lesions in Ultrasound Images Using Modified Fuzzy Possibilistic C-Means with Replusions Clustering and Generalized Gradient Vector Flow Snake Algorithm.** *Life Sci J* 2013;10(6s):360-367] (ISSN:1097-8135). http://www.lifesciencesite.com. 55

Keywords: Modified Fuzzy Possibilistic C-Means (MFPCM) with Repulsions, GGVF snake algorithm, Ultrasound Images, Breast Cancer Detection

1. INTRODUCTION

Cancer is a most severe disease [1] and one in eight deaths across the world is because of this disease. Cancer is the second leading cause of death in developed countries and the third leading cause of death in developing countries. The most frequently diagnosed cancer in women aged between 40 and 60 is Breast cancer [2]. Based on the report released by World Health Organization (WHO) there are about 7.6 million deaths occurred across the world because of cancer every year, among them around 502,000 deaths are resulted by breast cancer alone . By investigating these data, breast cancer is considered to be the most deadly cancer. In recent years, researchers have been concerned to find out the best technique to diagnosis breast cancer. Triumphant healing is a means to reduce the elevated death rate. To effectively treat a patient with breast cancer it is required to diagnose it as soon as possible. Cancers in their starting stages are vulnerable to treatment while cancers in their extremely developed stages are characteristically almost impossible to cure [3].

Even though breast cancer has very high occurrence and death rate, the reason for breast cancer is still not determined. There is no efficient technique to prevent the occurrence of breast cancer. As a result, early detection is the primary vital process towards diagnosing breast cancer. It acts as a main function in breast cancer diagnosis and treatment[4].

There are some factors which are utilized by the physicians to identify whether a breast nodule is benign or malignant. The usage of images instead of mammography [5, 6] Ultrasound (US) images possesses the following advantages:

- Breast ultrasound examinations can produce any section image of breast and examine the breast tissues in real-time and dynamically.
- Ultrasound imaging can depict small, early-stage malignancies of dense breasts, which is complex for mammography to attain.
- Sonographic equipment is portable and relatively low cost and has no ionizing radiation and side effects.

The proposed CAD system for classification and detection of breast cancer from the ultrasound image comprises the following phases:

- Speckle Noise Reduction
- Image Enhancement
- Clustering
- Contour Detection

In the first phase, the US image [7] is processed to remove the speckle noise with the use of Memetic

ANFIS Algorithm. During the second phase, the image is contrast enhanced [8] using Hough transformation for better visibility. Then the segmentation is performed using the clustering technique called MFPCM with Repulsion factor. The segmented image from proposed segmentation technique is given to the GGVF snake to determine the shape of the lesion.

2. Materials and Methods

Figure 1 shows the diagrammatic representation of the flow of the system. The proposed system performs an automated segmentation of Ultrasound images used as input. In the computerized segmentation section, MFCM with repulsion clustering based method is used to produce an initial segmentation of the input image, while the generalized gradient vector flow (GGVF) snake model is applied to the initial segmented image to obtain the final segmented image.

2.1 Preprocessing Using Memetic ANFIS and Hough Transformation

The capabilities of the Evolutionary Algorithm are added to the ANFIS-Memetic Algorithm. Memetic algorithm is used to optimize and set the neuro-fuzzy parameters. This acts as a filter to despeckle [9] the ultrasound image. The system uses the neural network ability to learn and knowledge is in fuzzy form. The noise is detected based on local statistical features, the memetic algorithm is used to decide and optimize the parameters to the network. After the speckle noise is removed, the enhancement process is carried out. It is complicated to simplify the processes for the ultrasound images which huge variations in the gray hence the image is normalized [10,11]. The fuzzy set of S function is utilized to map the image to the fuzzy domain, and also to use the membership values of its gray levels for further operations. A point is indulgenced as meeting points of infinite number of lines in the Hough Plane [12,13]. Identify the edges in the images with lines drawn in the plane. A triangular filter is created to filter the needless signal. Finally, the intensities are adjusted with inverse S function.

2.2. Clustering Using MFCM with Repulsions

The fuzzified version of the k-means algorithm is Fuzzy C-Means (FCM) [14]. It is a clustering approach which allows one piece of data to correspond to two or more clusters [15].

The algorithm is an iterative clustering approach [16] that brings out an optimal c partition by minimizing the weighted within group sum of squared error objective function J_{FCM}:

$$J_{FCM}(V,U,X)=\sum_{i=1}^{c}\sum_{j=1}^{n}u_{ij}^{m}d^2(X_j,v_i),1<m<+\infty \qquad (1)$$

$X = \{x_1, x_2,...,x_n\} \varepsilon R^p$ is the data set in the p-dimensional vector space, ***p*** represents the number of data items, **c** represents the number of clusters with $2 \le c \le n-1$. $V = \{v_1, v_2, \ldots, v_c\}$ is the vector space, c - centers or prototypes of the clusters, v_i - represents the p-dimension center of the cluster i, $d^2(x_j, v_i)$ represents a distance measure between object x_j and cluster centre v_i. $U = \{\mu_{ij}\}$ represents a fuzzy partition matrix with $u_{ij} = u_i(x_j)$ is the degree of membership of x_j in the i^{th} cluster, x_j is the j^{th} of p-dimensional measured data.

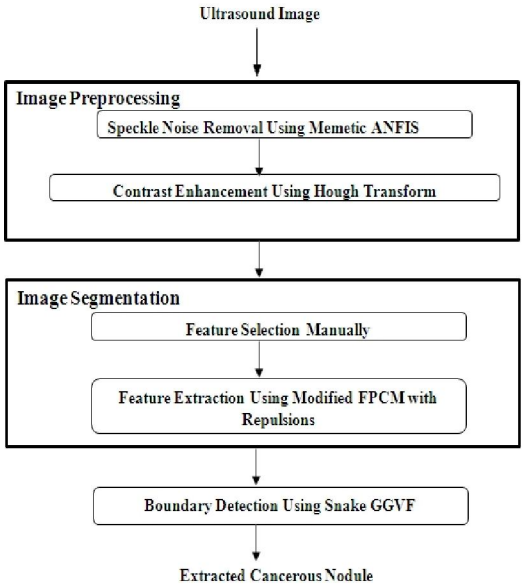

Figure 1: System Methodology

The fuzzy partition matrix satisfies the constrained conditions:

$$0 < \sum_{j=1}^{n} \mu_{ij} < n, \qquad \forall_i \in \{1,\ldots,c\}$$

$$\sum_{i=1}^{c} \mu_{ij} = 1, \qquad \forall_j \in \{1,\ldots,n\}$$

where, m is a weighting exponent parameter on each fuzzy membership function. It establishes the amount of fuzziness of the resulting classification. It is a fixed number greater than one. Under the constraint of U the objective function J_{FCM} is minimized. When J_{FCM} with respect to u_{ij} and v_i and zeroing, is necessary. The sufficient condition for J_{FCM} to be at its local extreme is as the following:

$$\mu_{ij}=\left[\sum_{k=1}^{c}\left(\frac{d(X_j v_i)}{d(X_j v_k)}\right)^{2/(m-1)}\right]^{-1},1\le i\le c,\quad 1\le j\le n$$

$$v_i = \frac{\sum_{k=1}^{n} \mu_{ik}^m x_k}{\sum_{k=1}^{n} \mu_{ik}^m}, \quad 1 \le i \le c$$

In reality, in a noisy environment, membership functions of FCM are not correspondent with the belonging data. It is inaccurate as real data involves some inbuilt noise.

To recover built-in noise of FCM, the constrained condition is not considered. It is replaced by PCM for unsupervised clustering[17]. The component generated by the PCM belongs to a dense region in the data set. Each cluster is independent of the other clusters in the PCM strategy. The objective function of the PCM is

$$J_{PCM}(V,U,X) = \sum_{i=1}^{c}\sum_{j=1}^{n} \mu_{ik}^m d^2(X_j, v_i) + \sum_{i=1}^{c} \eta_i \sum_{j=1}^{n} (1-u_{ij})^m \qquad (2)$$

Where $\eta_i = \frac{\sum_{j=1}^{n} \mu_{ik}^m \| x_j - v_i \|^2}{\sum_{j=1}^{n} \mu_{ij}^m}$ η_i is the scale parameter at the i[th] cluster represents, the possibilistic typicality value of training

$$u_{ij} = \frac{1}{1 + \left[\frac{d^2(x_j, v_i)}{\eta_i} \right]^{\frac{1}{m-1}}}$$

x_j belong to the cluster i, m belongs to [1,∞] is a weighting factor said to be the possibilistic parameter.

PCM is based on initialization typical of other cluster approaches. Clusters do not have mobility as each data point is classified as only one cluster at a time rather than all the clusters simultaneously. Initialization is of algorithm is done to converge to global minimum. Fuzzy and Possibilistic C-Means approaches are incorporated to do this. Memberships and typicalities are important factors for feature of data substructure in clustering problems.

The objective function in the FPCM depending on both memberships and typicalities is represented with two constraints.

$$J_{FPCM}(U,T,V) = \sum_{i=1}^{c}\sum_{j=1}^{n} \left(\mu_{ij}^m + t^n\right) d^2(X_j, v_i) \qquad (3)$$

with the following constraints :

$$\sum_{i=1}^{c} \mu_{ij} = 1, \qquad \forall_j \in \{1, \ldots, n\}$$

$$\sum_{j=1}^{n} t_{ij} = 1, \qquad \forall_i \in \{1, \ldots, c\}$$

A solution of the objective function can be obtained through an iterative process where the degrees of membership, typicality and the cluster centers are update with the equations as follows

$$\mu_{ij} = \left[\sum_{k=1}^{c} \left(\frac{d(X_j, v_i)}{d(X_j, v_k)} \right)^{2/(m-1)} \right]^{-1}, \quad 1 \le i \le c, 1 \le j \le n$$

$$t_{ij} = \left[\sum_{k=1}^{n} \left(\frac{d(X_j, v_i)}{d(X_j, v_k)} \right)^{2/(\eta-1)} \right]^{-1}, \quad 1 \le i \le c, 1 \le j \le n$$

$$v_i = \frac{\sum_{k=1}^{n} \left(\mu_{ik}^m + t_{ik}^\eta\right) X_K}{\sum_{k=1}^{n} \left(\mu_{ik}^m + t_{ik}^\eta\right) X_K}, 1 \le i \le c$$

Hybridization of possibilistic c-means (PCM) and fuzzy c-means (FCM) is the PFCM avoids various problems of PCM, FCM and FPCM. The noise sensitivity defect of FCM is solved by PFCM, and FCM overcomes the coincident clusters problem of PCM. But the estimation of centroids is influenced by the noise data.

Modified Fuzzy Possibilistic C-Means Technique (FPCM)

Objective function is very much necessary to enhance the quality of the clustering results. Modified Suppressed Fuzzy C- Means (MS-FCM) improves the performance of FCM due to parameter α a prototype-driven learning.

$$\alpha = \exp\left[- \min_{i=k} \frac{\| v_i - v_k \|^2}{\beta} \right]$$

$$\beta = \frac{\sum_{j=1}^{n} \| x_j - \bar{x} \|^2}{n} \quad where \quad \bar{x} = \frac{\sum_{j=n}^{n} x_j}{n}$$

β is a normalized term so it is chosen as a sample variance. The common value used by this parameter by all the data at each of the iteration, induces error. Therefore, a new parameter is added to suppress the common value of α and replaces it by a new parameter like a weight to each vector. Every point of the data set possesses a weight in relation to every cluster. Consequently the weight permits to have a better classification especially in the case of noise data. Weights are calculated as:

$$w_{ij} = \exp\left[- \frac{\| x_j - v_i \|^2}{\left[\sum_{j=1}^{n} \| x_j - \bar{v} \|^2 \right] * c / n} \right]$$

w_{ji} represents weight of the point j in relation to the class i. To alter the fuzzy and typical partition, this weight is used. The objective function is composed of two expressions - the first is the fuzzy function and uses a fuzziness weighting exponent, the second is possibilistic function and uses a typical weighting exponent. The two coefficients in the objective

function are only used as exhibitor of membership and typicality. A new relation decreases the function and increases the membership, and then typicality will tend towards 1 and decrease this degree when they tend toward 0.

$$J_{MFPCM}=\sum_{i=1}^{c}\sum_{j=1}^{n}\left(\mu_{ij}^{m}w_{ij}^{m}d^{2m}(x_j,v)+t_{ij}^{\eta}w_{ij}^{\eta}d^{2\eta}(x_j,v_i)\right) \quad (4)$$

U = {μij} represents a fuzzy partition matrix, is defined as:

$$\mu_{ij}=\left[\sum_{k=1}^{c}\left(\frac{d(X_j,v_i)}{d(X_j,v_k)}\right)^{2m/(m-1)}\right]^{-1}$$

T = {tij} represents a typical partition matrix, is defined as:

$$t_{ij}=\left[\sum_{k=1}^{n}\left(\frac{d(X_j,v_i)}{d(X_j,v_k)}\right)^{2\eta/(\eta-1)}\right]^{-1}$$

V = {vi} represents c centers of the clusters, is defined as:

$$v_i=\frac{\sum_{j=1}^{n}(\mu_{ij}^{m}w_{ji}^{m}+t_{ij}^{\eta}w_{ji}^{\eta})*X_j}{\sum_{j=1}^{n}(\mu_{ik}^{m}w_{ji}^{m}+t_{ik}^{\eta}w_{ij}^{\eta})}$$

Penalized and Compensated Constraints based Modified Fuzzy Possibilistic C-Means(PCMFPCM)

The Penalized and compensated constraints are embedded with the Modified Fuzzy Possibilistic C-Means algorithm to construct the objective function of PCMFPCM. The penalized constraint is represented as

$$\frac{1}{2}v\sum_{x=1}^{n}\sum_{i-1}^{c}(\mu_{x,i}^{m}\ell n\alpha_i+t_{x,i}^{n}\ell n\beta_x)$$

where

$$\alpha_i=\frac{\sum_{x=1}^{n}\mu_{xi}^{m}}{\sum_{x=1}^{n}\sum_{i=1}^{c}\mu_{xi}^{m}},\qquad i=1,2,\ldots,c$$

$$\beta_x=\frac{\sum_{i=1}^{c}t_{x,i}^{\eta}}{\sum_{x=1}^{n}\sum_{i=1}^{c}t_{x,i}^{\eta}},\quad x=1,2,\ldots,n$$

where α_i is a proportional constant of class i; β_x is a proportional constant of training vector z_x, and v (v≥0); τ (τ≥0) are also constants. Membership and typicality for the penalize is given as:

$$(\mu_{x,i})_P=\left(\sum_{i=1}^{c}\frac{(\|z_x-\varpi_i\|^2-v\ell n\alpha_i)^{1/(m-1)}}{(\|z_x-\varpi_i\|^2-v\ell n\alpha_i)^{1/(m-1)}}\right)^{-1}$$

$$x=1,2,\ldots,n\qquad i=1,2,\ldots,c$$

$$(t_{x,i})_P=\left(\sum_{y=1}^{n}\frac{(\|z_x-\varpi_i\|^2-v\ell n\beta_x)^{1/(\eta-1)}}{(\|z_x-\varpi_i\|^2-v\ell n\beta_y)^{1/(\eta-1)}}\right)^{-1}$$

$$\varpi_i=v_i=\frac{\sum_{k=1}^{n}(\mu_{ik}^{m}+t_{ik}^{\eta})X_K}{\sum_{k=1}^{\eta}(\mu_{ik}^{m}+t_{ik}^{\eta})},1\le i\le c$$

ϖ_i in the expression is the centroid. The compensated constraints can represented as follows

$$\frac{1}{2}\tau\sum_{x=1}^{n}\sum_{i=1}^{c}(\mu_{x,i}^{m}\tanh\alpha_i+t_{x,i}^{\eta}\tanh\beta_x)$$

where Membership and typicality for the compensation is presented below

$$(\mu_{x,i})_c=\left(\sum_{i=1}^{c}\frac{(\|z_x-\varpi_i\|^2-\tau\tanh(\alpha_i))^{1/(m-1)}}{(\|z_x-\varpi_i\|^2-\tau\tanh(\alpha_i))^{1/(m-1)}}\right)^{-1},$$

$$x=1,2,\ldots,n\qquad i=1,2,\ldots,c$$

To obtain an efficient clustering the penalization term must be removed and the compensation term must be added to the basic objective function of the existing FPCM.

$$(t_{x,i})_c=\left(\sum_{y=1}^{n}\frac{(\|z_x-\varpi_i\|^2-\tau\tanh(\beta_x))^{1/(\eta-1)}}{(\|z_x-\varpi_i\|^2-\tau\tanh(\beta_y))^{1/(\eta-1)}}\right)^{-1},$$

$$x=1,2,\ldots,n\qquad i=1,2,\ldots,c$$

The Final Objective Function is same as objective function of PCFPCM as given in the equation 5. The centroid of i^{th} cluster is calculated in the similar way as the defined earlier.

Clustering Enhancement using Repulsion

In the above described clustering technique, objective function is truly minimized only if all the centroids are identical (coincident), since the typicality of a point to a cluster, depends only on the distance between the point to that cluster.

$$\begin{aligned}J_{PCFPCM}=&\sum_{i=1}^{c}\sum_{j=1}^{n}\left(\mu_{ij}^{m}w_{ij}^{m}d^{2m}(x_j,v)+t_{ij}^{\eta}w_{ij}^{\eta}d^{2\eta}(x_j,v_i)\right)\\&-\frac{1}{2}v\sum_{x=1}^{n}\sum_{i-1}^{c}(\mu_{x,i}^{m}\ell n\alpha_i+t_{x,i}^{n}\ell n\beta_x)\\&+\frac{1}{2}v\sum_{x=1}^{n}\sum_{i-1}^{c}(\mu_{x,i}^{m}\tanh\alpha_i+t_{x,i}^{n}\tanh\beta_x)\end{aligned} \quad (5)$$

The usage of repulsion aims to minimize the intra-cluster distances, while maximizing the inter-cluster distances, without using implicitly the restriction, but by adding a cluster repulsion term to the objective function.

$$J_{MFPCM}=\sum_{i=1}^{c}\sum_{j=1}^{n}\left(\mu_{ij}^{m}w_{ij}^{m}d^{2m}(x_j,v)+t_{ij}^{\eta}w_{ij}^{\eta}d^{2\eta}(x_j,v_i)\right)$$
$$-\frac{1}{2}v\sum_{x=1}^{n}\sum_{i-1}^{c}(\mu_{x,i}^{m}\ell n\alpha_i+t_{x,i}^{n}\ell n\beta_x)$$
$$+\frac{1}{2}v\sum_{x=1}^{n}\sum_{i-1}^{c}(\mu_{x,i}^{m}\tanh\alpha_i+t_{x,i}^{n}\tanh\beta_x) \quad (6)$$
$$+\sum_{i=1}^{c}\eta_i\sum_{k=1}^{n}(1-u_{ik})^m+\gamma\sum_{i=1}^{c}\sum_{k=1,k=i}^{c}\frac{1}{d^2(v_i,v_k)}$$

where γ is a weighting factor, and u_{ik} satisfies $u_{ik}\in[0,1],\forall_i$. When clusters are close the repulsion term is relevant. When distance grows it becomes smaller until it is compensated by the attraction of the clusters. When the clusters are sufficiently spread out Inter-cluster distance decreases, the attraction of the cluster is compensated only by the repulsion term. Minimization of objective function with respect to cluster prototypes leads to

$$v_i=\frac{\sum_{j=1}^{n}u_{ij}x_j-\gamma\sum_{k=1,k=i}v_k\frac{1}{d^2(v_k,v_i)}}{\sum_{j=1}^{n}u_{ij}x_j-\gamma\sum_{k=1,k=i}\frac{1}{d^2(v_k,v_i)}}$$

Singularity occurs when one or more of the distances $d^2(v_k,v_i)=0$ at any iteration. In such a case, v_i cannot be calculated. When this happens, assign zeros to each nonsingular class (all the classes except i) and assign 1 to class i, in the membership matrix U. An alternative repulsion term to minimize the objective function is given by

$$J_{MFPCM}=\sum_{i=1}^{c}\sum_{j=1}^{n}\left(\mu_{ij}^{m}w_{ij}^{m}d^{2m}(x_j,v)+t_{ij}^{\eta}w_{ij}^{\eta}d^{2\eta}(x_j,v_i)\right)$$
$$-\frac{1}{2}v\sum_{x=1}^{n}\sum_{i-1}^{c}(\mu_{x,i}^{m}\ell n\alpha_i+t_{x,i}^{n}\ell n\beta_x)$$
$$+\frac{1}{2}v\sum_{x=1}^{n}\sum_{i-1}^{c}(\mu_{x,i}^{m}\tanh\alpha_i+t_{x,i}^{n}\tanh\beta_x) \quad (7)$$
$$+\sum_{i=1}^{c}\eta_i\sum_{k=1}^{n}(1-u_{ik})^m+\gamma\sum_{i=1}^{c}\sum_{k=1,k=i}^{c}e^{-d^2(v_k,v_i)}$$

The weighting factor γ is used to balance the attraction and repulsion forces, i.e., minimizing the intra-distances inside clusters and maximizing the inter-distances between clusters.

The proposed segmentation technique is applied to mammary gland image segmentation. The pixel values are the inputs of the clustering algorithm, and the pixels are clustered based on the optimum centers of clustering. The values of the pixels contained in the lesion are very low, the cluster of pixels with the lesser intensities can be considered as the lesion-like pixels. The mammary gland region is determined by the following formula:

$$bw(i,j)=\begin{cases}0, & g(i,j)\in C_1\\ 255, & otherwise\end{cases} \quad (8)$$

where, *g(i, j)* - pixel in mammary gland region at the location (i, j), C_1 - cluster with the lesser intensities. B_w - binary mammary gland image after segmentation. After the mammary gland is segmented, the round-like regions are kept as the lesion-like regions and the others are rejected for further processing to identify contours.

2.3. Contour Detection Using GGVF Snakes

A *traditional snake* is a curve $x(s)=[x(s),y(s)]$, $s\in[0,1]$, that moves through the spatial domain of an image to minimize the energy functional [18] as in equation (9)

$$E=\int_0^1\frac{1}{2}\left(|x'(s)|^2+\beta|x''(s)|^2+E_{ext}(x(s))\right)ds \quad (9)$$

where ∝ and β are the weighing parameters to control tension and rigidity of the snake. $x'(s)$ and $x''(s)$ indicate the first and second order derivates of the $x(s)$ with respect to s. The external energy function E_{ext} is derived from the image so that it takes on the smaller values at the features of interest, such as boundaries[19]. A typical external energy functions for gray–level image, $I(x,y)$ is $\pm G_\sigma(x,y)*I(x,y)$ for lines and $-|\nabla(G_\sigma(x,y)*I(x,y))|^2$ for step edges [1,2] where $G_\sigma(x,y)$ is the 2-D Gaussian function with σ as standard deviation and ∇ as gradient operator.

A snake that minimizes E must satisfy Euler equation (10).

$$\propto x''(s)-\beta x(s)-\nabla E_{ext}=0 \quad (10)$$

To minimize equation 9, the snake is made dynamic by treating x as a function of time t and s, as $x(s,t)$. Then, the partial derivative of x with respect to t is then set equal to left-hand side of equation 2 as follows:

$$x_t(s,t)=\propto x''(s,t)-\beta x(s,t)-\nabla E_{ext} \quad (11)$$

When the solution $x(s,t)$ stabilizes, the term $x_t(s,t)$ vanishes and we achieve a solution of equation 9.

Equation 11 is the starting point to define a new snake, called *GVF Snake*. The external force term - ∇E_{ext} in equation 10 with GVF $\tilde{v}(x,y)$ field defined as equilibrium solution for the following differential equation 4:

$$v_t=\mu\nabla^2v-(v-\nabla f)|\nabla f|^2 \quad (12)$$

where μ denotes the partial derivative of $\tilde{v}(x,y,t)$ with respect to *t*, and $\nabla^2 = \frac{\delta^2}{\delta x^2} + \frac{\delta^2}{\delta y^2}$ is the Laplacian operator where f is the edge map derived from the gray-scale image $I(x,y)$, having the property that it is larger near edges. This edge map can be gray-level or binary-valued. It can be computed using $\pm G_\sigma(x,y) * I(x,y)$ or $-\left|\nabla\left(G_\sigma(x,y) * I(x,y)\right)\right|^2$, or any conventional image edge detectors.

It is difficult for GVF to force a snake into long, thin boundary indentations due to excessive smooth effect caused by the co-efficient μ in equation 12. On introducing a weighting function, instead of constant and decreasing the smooth effect near the gradient solves the problem. This is *generalized gradient vector flow (GGVF)*, μ and $|\nabla f|^2$ is replaced in equation 12 by weighing functions. The GGVF equilibrium solution of the vector partial differential equation:

$$v_t = g(|\nabla f|)\nabla^2 v - h(|\nabla f|)(v - \nabla f) \quad (13)$$

The first term on the right is referred to as the smoothing term since this term alone will produce a smoothly varying vector field. The second term is referred as the data term since it encourages the vector field v to be close to ∇f computed from the data. The weighting functions *g(.)* and *h(.)* apply to the smoothing and data terms, respectively. Since these weighting functions are dependent on the gradient of the edge map which is spatially varying, the weights themselves are spatially varying, in general. Since we want the vector field v to be slowly varying (or smooth) at locations far from the edges, but to conform to ∇f near the edges, *g(.)* and *h(.)* should be monotonically non-increasing and non-decreasing functions of $|\nabla f|$ respectively.

The equation 5 reduces to that of GVF when

$$g(|\nabla f|) = \mu \quad (14)$$

$$h(|\nabla f|) = |\nabla f|^2 \quad (15)$$

Since *g(.)* is constant here, smoothing occurs everywhere; however, *h(.)* grows larger near strong edges, and should dominate at the boundaries. Thus, GVF should provide good edge localization. The effect of smoothing becomes apparent, however, when there are two edges in close proximity, such as when there is a long, thin indentation along the boundary. In this situation, GVF tends to smooth between opposite edges, losing the forces necessary to drive an active contour into this region.

To address this problem, weighting functions can be selected such that *g(.)* gets smaller as *h(.)* becomes larger. Then, in the proximity of large gradients, there will be very little smoothing, and the effective vector field will be nearly equal to the gradient of the edge map. There are many ways to specify such pairs of weighting functions. In this paper, we use the following weighting functions for GGVF:

$$g(|\nabla f|) = e^{-(|\nabla f|/K)} \quad (16)$$

$$h(|\nabla f|) = 1 - g(|\nabla f|) \quad (17)$$

The GGVF field computed using this pair of weighting functions will conform to the edge map gradient at strong edges, but will vary smoothly away from the boundaries. The specification of *K* determines to some extent the degree of tradeoff between field smoothness and gradient conformity.

As in GVF [16], the partial differential equation 5 specifying GGVF, can be implemented using an explicit finite difference scheme, which is stable if the time step Δt and the spatial sample intervals Δx and Δy satisfy

$$\Delta t = \frac{\Delta x \Delta y}{4 g_{max}} \quad (18)$$

where g_{max} is the maximum value of *g(.)* over the range of gradients encountered in the edge map image. While an implicit scheme for the numerical implementations of Equation 13 would be unconditionally stable and therefore not need this condition, the explicit scheme is faster.

3. Results and Discussion

The experiments are conducted on the proposed computer-aided diagnosis systems [20] with the help of real time breast ultrasound images. Figure 2 -10 show the results of various process involved while detecting a cancer in its benign state. Table 1 summarizes the standard deviations, Accuracy and areas from the lesion contours which were segmented by the FCM, MFPCM with Repulsion and GGVF-MFPCM with Repulsion. Among the computerized segmentation methods, the GGVF-MFPCM with repulsion method achieves the better segmentation performance. Thus proposed method is adopted.

Table 1: Comparison of Various Methods

Segmentation Method	Std. Deviation (px)	Accuracy (%)	Area (px)
FCM	1.2	83.3	4.263
MFPCM	2.1	87.1	4.721
GGVF-MFPCM with Repulsion	2.3	89.6	4.711

4. Conclusion

The proposed MFPCM with repulsion and GGVF snake shows better results as applied to detection breast tumors in ultrasound images. The method provides accuracy with the reference to the conventional techniques. The model works better with regard to the distance between the initial contour and the object boundary. The method can be applied with a variety of segmentation methods.

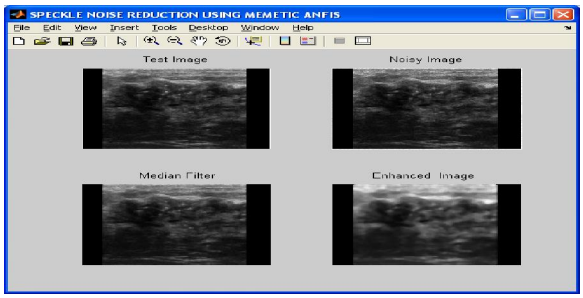

Figure 2: Speckle Noise Reduction

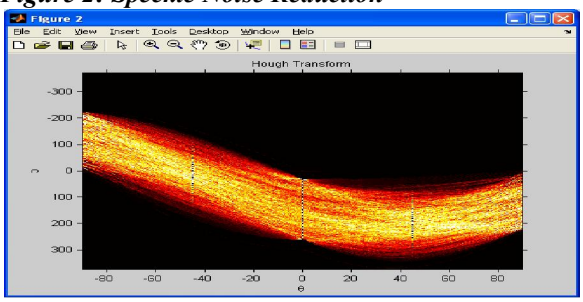

Figure 3: Hough Transform Curve

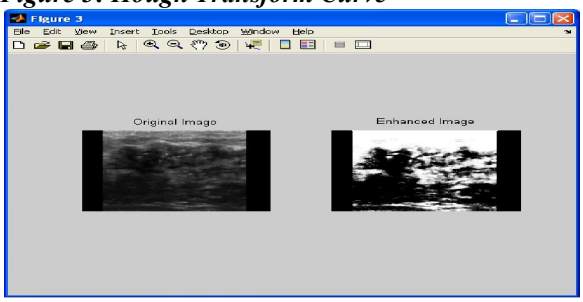

Figure 4: Hough Enhanced Image

Figure 5: Region of Interest Selection

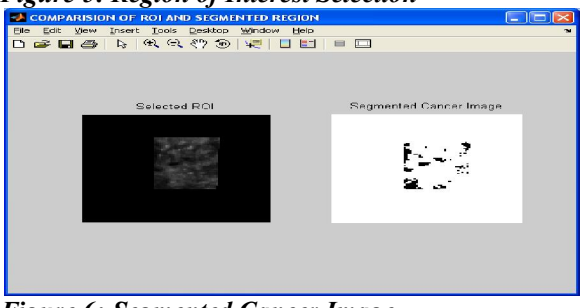

Figure 6: Segmented Cancer Image

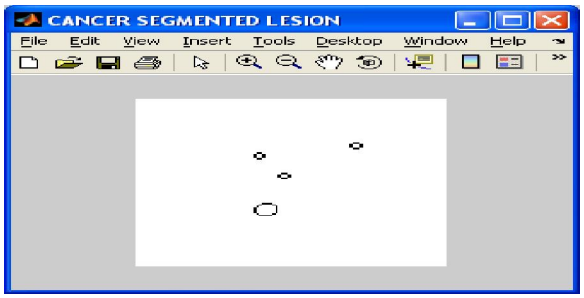

Figure 6: Segmented Cancer Lesions

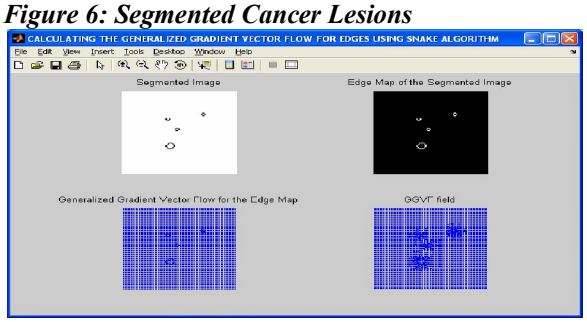

Figure 7: Boundary Detection Using Snake GGVF

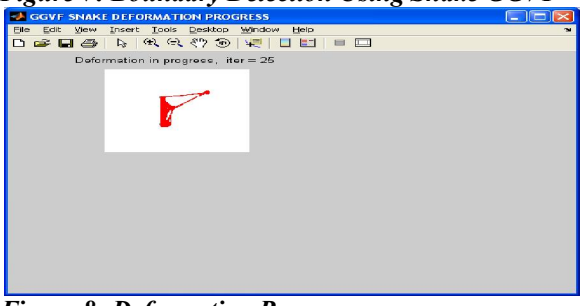

Figure 8: Deformation Process

Figure 9: Calculating Cancer Size

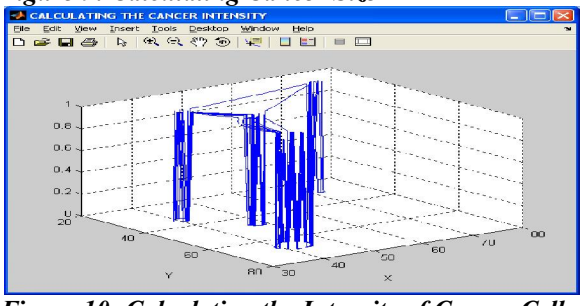

Figure 10: Calculating the Intensity of Cancer Cells

References

[1] Oelze, M.L.; O'Brien, W.D.; Zachary, J.F, "11B-4 Quantitative Ultrasound Assessment of Breast Cancer Using a Multiparameter Approach," IEEE Ultrasonics Symposium, Pp. 981 – 984, 2007.

[2] Gefen, S.; Tretiak, O.J.; Piccoli, C.W.; Donohue, K.D.; Petropulu, A.P.; Shankar, P.M.; Dumane, V.A.; Lexun Huang; Kutay, M.A.; Genis, V.; Forsberg, F.; Reid, J.M.; Goldberg, B.B., "ROC analysis of ultrasound tissue characterization classifiers for breast cancer diagnosis," IEEE Transactions on Medical Imaging, Vol. 22, No. 2, Pp. 170 – 177, 2003.

[3] Winder, A.A.; Jadidian, B.; Muratore, R., "Synthetic Structural Imaging (SSI): A new ultrasound method for tracking breast cancer morphology," 39th Annual Ultrasonic Industry Association Symposium (UIA), Pp. 1 – 4, 2010.

[4] American Cancer Society, Breast Cancer Facts & Figures 2007-2008, American Cancer Society, 2008.

[5] Cheng, H.D., Cai, X., Chen, X., Hu, L. and Lou, X. Computer-aided detection and classification of microcalcifications in mammograms: a survey. Pattern Recognition 36, 12 (2003), 2967-2991.

[6] Breastcancer.org. Mammograms. 2009. http://www.breastcancer.org/symptoms/testing/types/mammograms/. June 2009.

[7] Drukker, K., Giger, M.L., Vyborny, C.J. and Mendelson, E. B. Computerized detection and classification of cancer on breast ultrasound. Academic Radiology 11, 5 (2004), 526-535.

[8] Li, X. and Liu, D.C. Ultrasound image enhancement using dynamic filtering. In 4th International Conference on Image and Graphics, 2007, 106-109.

[9] Deshmukh, K.S. and Shinde, G.N. An adaptive color image segmentation. Electronic Letters on Computer Vision and Image Analysis 5, 4 (2005), 12-23, 2005.

[10] Noble, J.A. and D. Boukerroui, D. Ultrasound image segmentation: A survey. IEEE Transactions on Medical Imaging 25, 8 (2006), 987-1010.

[11] Adam, D., Beilin-Nissan, S., Friedman, Z. and Behar, V. The combined effect of spatial compounding and nonlinear filtering on the speckle reduction in ultrasound images. Ultrasonics 44, 2 (2006), 166.

[12] Juan Wachs, Oren Shapira and Helman Stern, "A Method to Enhance the Possibilistic C-Means with Repulsion Algorithm based on Cluster Validity Index", Advances in Intelligent and Soft Computing, Springerlink, Vol. 34, Pp. 77-87, 2006

[13] Zhang, L.C., Wong, E.M.C., Zhang, F. and Zhou, J. Adaptive pyramid filtering for medical ultrasound image enhancement. In 3rd IEEE International Symposium on Biomedical Imaging: Nano to Macro, 2006, 916-919.

[14] Shi, X., Cheng, H.D., Hu, L., Ju, W. and Tian, J. Detection and classification of masses in breast ultrasound images. Digital Signal Processing 20, 3 (2010), 824- 836.

[15] Jihene Malek, et al, Automated Breast Cancer Diagnosis Based on GGVF Snake Segmentation, Wavelet Feature Extraction and Fuzzy Classification. Journal of Signal Processing System, 55(1-3):49-66 (2009).

[16] Noble, J.A. and Boukerroui, D. Ultrasound image segmentation: A survey. IEEE Transactions on Medical Imaging 25, 8 (2006), 987.

[17] Chen, C.-Y. and Ye, F. Particle swarm optimization algorithm and its application to clustering analysis. In IEEE International Conference on Networking, Sensing and Control, 2004, 789-794.

[18] Chenyang Xu, Jerry L. Prince, "Gradient Vector Flow: A New External Force for Snake", IEEE. Proc, conf. on comp, Vis. Patt. Recog., 1997.

[19] S. Haykin, Neural Networks, Prentice-Hall, Englewood Cliffs, NJ, (1998).

[20] Elif Derya Ubeyli, Inan Guler, Teaching Automated Diagnostic systems for Doppler ultrasound blood flow signals to biomedical engineering students using Matlab, International Journal of Engineering Education, Vol 21 (4) (2005),pp.649-667.

3/17/2013