International Journal of Innovative Computing, Information and Control
Volume **8**, Number **7(B)**, July **2012**

ICIC International ©2012 ISSN 1349-4198
pp. 5179–5197

HYBRID PARTICLE SWARM OPTIMIZATION WITH WAVELET MUTATION BASED SEGMENTATION AND PROGRESSIVE TRANSMISSION TECHNIQUE FOR MRI IMAGES

Arunava De[1], Anup Kumar Bhattacharjee[1], Chandan Kumar Chanda[2] and Bansibadan Maji[1]

[1]Department of Electronics and Communication
National Institute of Technology
Durgapur, India
arunavade@yahoo.com

[2]Department of Electrical Engineering
Bengal Engineering and Science University
Shibpur, Howrah, India

Received May 2011; revised September 2011

Abstract. *A Hybrid particle swarm algorithm that incorporates a wavelet theory based mutation operation is used for segmentation and progressive transmission for Magnetic Resonance Images. The concept of irrelevancy reduction is used to reduce the load on the transmission network. We use Entropy maximization using Hybrid particle swarm algorithm with Wavelet based mutation operation to get the region of interest of the Magnetic Resonance Image. It applies the Multi-resolution Wavelet theory to enhance the Particle Swarm Optimization in exploring the solution space more effectively for a better solution. A variable rectangular mask is used to reduce the noise in the segmented image. Precision and Recall is used to evaluate segmentation accuracy. We use Discrete Cosine Transform to make progressive transmission system very efficient. Twin Distribution Entropy Coding is employed to compress the DCT coefficients. We use varying percentages of DCT coefficients of segmented MRI image for progressive image transmission. Clustering of the segmented MRI image using k-means algorithm results in predominantly two clusters namely that of lesions and background. At the receiver's end the doctor or a radiologist identifies a particular class of lesions and may ask for the entire un-segmented Magnetic Resonance Image dataset of a particular patient. Irrevalency reduction thus helps to reduce the load on the system by choosing not to send the images without lesions.*

Keywords: Medical resonance imaging, Particle swarm optimization, Within-class variance, Intensity contrast, Entropy, Discrete cosine transform, K-means algorithm, Progressive image transmission

1. **Introduction.** This paper proposes a scheme for segmentation of Magnetic Resonance Images (MRI) and thereby progressively transmits it for proper diagnosis at the receiver's end.

Thresholding is one of the very important techniques for image segmentation. Here in this paper we deal with multi-modal MRI images. The histogram of multi-modal images has multiple peaks as opposed to two peaks for a bi-modal image as depicted in Figure 1. We separate the lesions from the healthy part of an MRI image and then we propose a way to perform progressive transmission of the segmented MRI image.

In our technique we segregate the image based on levels of intensity, because lesion portion of the MRI image will have a different intensity value with that of a multimodal MRI image without lesions. Particle Swarm optimization (PSO) [1] is a population based stochastic optimization algorithm. [2] found that using Entropy maximization using PSO

algorithm gives better results in comparison with other methods [3]. However, PSO converges sharply in the earlier stages of the search process but it saturates or even terminates in the later stages. We perform entropy maximization using Hybrid Particle Swarm Optimization with Wavelet Mutation Operation (HPSOWM) algorithm to get a threshold value which segregates the lesions from the healthy portion of the MRI image. This gets us the region of interest (ROI). The threshold value is further modified using the concept of variable mask in which the mask is incrementally applied on the ROI. Depending on the similarity of the neighborhood pixels the mask is incremented. Only the MRI data of a patient with diseased lesions are selected for transmission. We perform Discrete Cosine Transform (DCT) [4] of the segmented MRI image. The DCT coefficients are compressed using Twin Distribution Entropy Coding (TDEC) and then varying percentages of the DCT coefficients together with k-means clustering [5, 6] is used to perform progressive transmission and reconstruction of the MRI image.

If the segmented MRI does not have lesions, then those MRI data of the patients are transmitted only on demand. The proposed technique results in transmission of only those MRI data of the patient which have diseased lesions in them. The purpose of irrelevancy reduction is to reduce the effective load of the transmitter. Since the images can be viewed progressively a high degree of compression and efficiency can be achieved by not sending the irrelevant details of the image. The devised system described here intends to provide a fully systematic progressive transmission solution from the primary input MRI images.

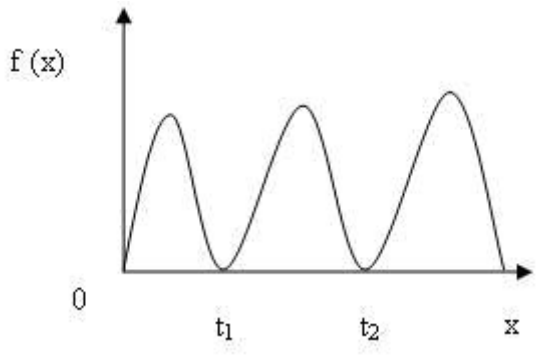


FIGURE 1. Histogram of multimodal image

2. **Related Work.** [7] uses Tchebycheff orthogonal matrix polynomials for reconstructing 3-D images when they are progressively transmitted. Interpolating the 2-D slices will result in progressively-improving 3-D image as slices are received. [8] presents a hierarchical coding algorithm for 3-D medical images. It uses the Kronecker product and sparse matrix techniques. The salient feature of the work was that arbitrary cross-section images can be progressively displayed without reconstruction of the whole 3D image. A new Shape Adaptive Integer Wavelet Transform (SAIWT) based progressive transmission coding scheme for 2-D and 3-D MRI was proposed by [9]. [10] implemented a prototype system consisting of Java based image viewer and a web server extension component for transmitting MRIs to an image viewer. Set Partitioning in Hierarchical Trees (SPHIT) image compression algorithm was used by [10] for compression and decompression and analysis of the performance of the SPHIT image compression algorithm in a tele-radiology application were done. [11] proposed a way of progressive image transmission using Haar Wavelets. [12] proposed a method of progressive image transmission of two-dimensional Gel Electrophoresis images and was designed such that it uses more than one threshold to detect the more important parts of an image.

Valley detection and criteria optimization [13, 14] are one of the many thresholding approaches that are already defined. Minimum error thresholding method (MinError) assumes the normal distribution of object and background [15, 16]. [17] estimated the parameters of normal distribution from a truncated normal distribution corresponding to object and background. [18] was the first to propose a thresholding method based on maximum entropy principle whereas [19] employed cross entropy for threshold selection. [20] applied fuzzy c-partition to gray level images and maximized fuzzy entropy to select the threshold. [21] introduced an energy criterion formulated by intensity-based class uncertainty and region homogeneity. The threshold is selected by minimizing the energy. [22] applied the idea of maximizing the between class variance into histogram-based thresholding. The method shows satisfactory results in various applications. However, it tends to split the larger part when the sizes of object and background are unequal [23]. [24] took advantage of the knowledge about the range of background proportion to the ROI to confine the range of threshold selection and achieved reliable results in segmenting magnetic resonance (MR) and computed tomography (CT) images of the human brain. [25] demonstrated that threshold can be obtained by optimizing the weighted sum of the within-class variance and the intensity contrast.

3. **Proposed Algorithm.**

3.1. **Irrelevancy reduction.**

Transform coding denotes a procedure in which the image is subjected to an invertible transform before transmission with an aim to convert the statistically dependent image elements to independent coefficients. Thus the data are compressed by redundancy reduction. However, in many cases to achieve further compression, the transformed data are also subjected to irrelevancy reduction [26]. Data that are considered irrelevant and do not convey a lot of information about the image are discarded. Thus the redundancy reduction is reversible, whereas the irrelevancy reduction is not.

Progressive image transmission involves an approximate reconstruction of the image whose fidelity is built up gradually until either the viewer decides to abort the transmission or perfect reconstruction is achieved.

We segment an MRI image and get the lesions. The lesions in the MRI can be anything from blood clot to tumor formation. The lesions are treated as relevant information whereas the rest of the image is irrelevant. The progressive transmission of the MRI occurs if lesions are detected. To reduce load on the transmission line the viewer may choose to abort transmission if lesions are not detected.

3.1.1. *Entropy.* First, we compute the normalized histogram $h(n)$ for a gray image $f(x, y)$.

$$P_n = h(n) = f_n/N, \quad n = 0, 1, 2, \ldots, 255 \tag{1}$$

where f_n is the observed frequency of gray level n (or f_n is the number of pixel having gray level n) and N is the total number of pixels in the picture. For multimodal image we want to divide the total image into $(k+1)$ number of homogeneous zones and for that we consider the threshold gray levels at $t_1, t_2, t_3, \ldots, t_k$. Shannon Entropy is defined as

$$H = -\sum_{n=0}^{t_1} P_{1n} \ln P_{1n} - \sum_{n=t_1+1}^{t_2} P_{2n} \ln P_{2n} - \ldots \sum_{n=t_n+1}^{255} P_{kn} \ln P_{kn} \tag{2}$$

where

$$P_{1n} = P_n \Big/ \sum_{n=0}^{t_1} P_n \quad \text{for } 0 \le n \le t_1,$$

$$P_{2n} = P_n \Big/ \sum_{n=t_1+1}^{t_2} P_n \quad \text{for } t_1 < n \le t_2,$$

$$P_{kn} = P_n \Big/ \sum_{n=t_k+1}^{255} P_n \quad \text{for } t_k < n \le 255$$

We have obtained the function H for the threshold gray level t_1 to t_k using HPSOWM algorithm [27]. Now we apply this expert knowledge that the gray level of diseased lesion zone of an MRI image vary from range say Tx to Ty. Using this expert knowledge, we apply our concept of variable low pass filter mask after optimizing the threshold range using PSO and Wavelet Mutation which is explained in subsequent section.

3.1.2. *Optimization using HPSOWM algorithm.* We optimize the basis functions obtained using Entropy maximization technique using the concept of HPSOWM algorithm.

PSO algorithm as applied to our technique of HPSOWM [27]:

To maximize function $f(X)$ with $X^l \le X \le X^u$ where X^l and X^u denote the lower and upper bounds on X. PSO works as follows:

1. Assume the size of the swarm (number of particles) is N.
2. Generate the initial position of X in the range X^l and X^u randomly as $X_1, X_2, X_3, \ldots, X_N$.
3. Velocity is assigned for each particle as $V_1, V_2, V_3, \ldots, V_N$.
4. Evaluate objective function values corresponding to the particles. Initially all the particle velocities are assumed to be zero and set the iteration number as $i = 1$.
5. In ith iteration find the following two parameters. Calculate the historical best value of $X_j(i)$ called $P_{best,j}$. Also find G_{best}, the highest value of objective function encountered in all the previous iterations by any of the N particles.
6. Update the velocity as follows:

$$V_j(i) = \theta V_j(i-1) + c_1 r_1 (P_{best,j} - X_j(i-1)) + c_2 r_2 (G_{best} - X_j(i-1)), \quad j = 1, 2, \ldots, N \tag{3}$$

7. Update the position of the particle as

$$X_j(i) = X_j(i-1) + V_j(i), \quad j = 1, 2, \ldots, N \tag{4}$$

8. Check the convergence of the current solution. If the convergence criteria is not satisfied then repeat steps 5-7.

The Hybrid PSO algorithm with Wavelet Mutation is used for optimizing the initial value of threshold to be used for segmenting the MRI image using variable mask.

3.1.2.1. Hybrid particle swarm optimization with wavelet mutation operation (HPSOWM). We observe that PSO algorithm works well in the early stages but usually presents problems on reaching the near optimal stage. If a particle's current position coincides with the global best position, the particle will move away from this point if its inertia weight and velocity are different from zero. If their velocities are very close to zero all the particles will stop moving once they catch up with the global best particle which may lead to a premature convergence and no further improvement can be obtained. This phenomenon is stagnation.

[28] proposed a hybrid PSO with the integration of Genetic Algorithm's mutation with a constant mutating space. However, in that approach, the mutating space is kept unchanged throughout the search, so the space of permutation of particles in PSO is also unchanged.

To overcome the deficiencies of PSO algorithm HPSOWM is used. In Hybrid Particle Swarm Optimization with Wavelet mutation operation (HPSOWM) [27] larger mutating space is set in the early stages of the search to find a solution from the solution space whereas it fine tunes to a better solution in the later stage of the search by setting a smaller mutating space based on the properties of Multi-resolution Wavelet Analysis [29]. The Pseudo code is given in Figure 2, in which the mutation on particles is performed after updating their velocities and positions.

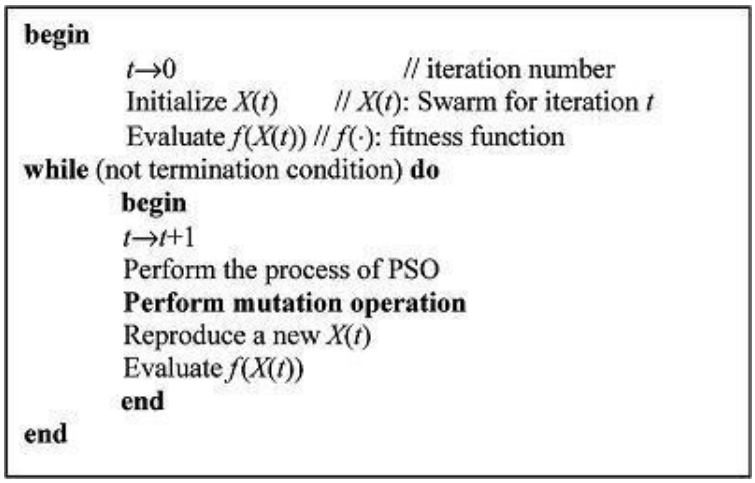


FIGURE 2. Pseudo code for HPSO with wavelet mutation operation (HPSOWM)

3.1.2.1.1. Basic wavelet theory. Certain seismic signals can be modeled by combining translations and dilations of an oscillatory function with a finite duration called "wavelet". Wavelet transform can be divided in two categories: continuous wavelet transform and discrete Wavelet transforms. The continuous wavelet transform $W(a,b)$ of function $f(x)$ with respect to a mother wavelet $\Psi(x) \in L^2(\Re)$ is given by the following equation [30,31].

$$W_{a,b}(x) = \frac{1}{\sqrt{C_\Psi}} \int_{-\infty}^{+\infty} f(x)\Psi^*_{a,b}(x)dx \tag{5}$$

where

$$\Psi_{a,b}(x) = \frac{1}{\sqrt{a}}\Psi\left(\frac{x-b}{a}\right), \quad x \in \Re,\ a,b \in \Re\, a > 0 \tag{6}$$

In Equation (5), $(*)$ denotes the complex conjugate, a is the dilation (scale) parameter, and b is the translation (shift) parameter. It is to be noted that a controls the spread of the wavelet and b determines its control position. A set of basis function $\Psi_{a,b}(x)$ is derived from scaling and shifting the mother wavelet. The basis function of the transform is called the daughter wavelet. The mother wavelet has to satisfy the following admissibility condition

$$C_\Psi = 2\pi \int_{-\alpha}^{+\alpha} \frac{\left|\bar{\Psi}(\omega)\right|^2}{\omega} d\omega < \alpha \tag{7}$$

where $\bar{\Psi}(\omega)$ is the Fourier Transform of $\Psi(\omega)$ and given by the following equation:

$$\bar{\Psi}(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\alpha}^{+\alpha} \Psi(x) \times e^{-j\omega x} dx \tag{8}$$

A continuous function $\Psi(x)$ is a "mother wavelet" or "Wavelet", if it satisfies the following properties.
Property I:

$$\int_{-\alpha}^{+\alpha} \Psi(x) dx = 0 \tag{9}$$

Equation (9) demonstrates that the total positive momentum of $\Psi(x)$ is equal to the total negative momentum of $\Psi(x)$.
Property II:

$$\int_{-\alpha}^{+\alpha} |\Psi(x)|^2 \, dx < \alpha \tag{10}$$

From Equation (10), it may be inferred that most of the energy $\Psi(x)$ is confined to a finite domain and is bounded.

3.1.2.1.2. Association of wavelet based mutation with PSO. It is proposed that every element of the particle of the population will have a chance to mutate, governed by a user defined probability $p_m \in [0, 1]$. For each element, a random number between 0 and 1 will be generated such that if it is less than or equal to p_m, the mutation will take place on that element. Among the population, a randomly selected ith particle and its jth element within the limits $[S_{j,\min}, S_{j,\max}]$ at kth iteration $\left(S_{i,j}^{(k)}\right)$ will undergo mutation as given in the following equation:

$$\left(S_{i,j}^{(k)}\right) = \begin{cases} S_{i,j}^{(k)} + \sigma \times \left(S_{j,\max} - S_{i,j}^{(k)}\right), \text{ if } \sigma > 0 \\ S_{i,j}^{(k)} + \sigma \times \left(S_{i,j}^{(k)} - S_{j,\min}\right), \text{ if } \sigma \leq 0 \end{cases} \tag{11}$$

where

$$\sigma = \Psi_{a,0}(x) = \frac{1}{\sqrt{a}} \Psi\left(\frac{x}{a}\right).$$

A Morlet wavelet (mother wavelet) defined in the following equation may be shown as in Figure 3.

$$\Psi(x) = e^{-\frac{x^2}{2}} \cos(5x) \tag{12}$$

Thus,

$$\sigma = \frac{1}{\sqrt{a}} e^{-\frac{\left(\frac{x}{a}\right)^2}{2}} \cos\left(5\left(\frac{x}{a}\right)\right) \tag{13}$$

Different dilated Morlet wavelets are shown in Figure 4. From this figure, it is clear that as the dilation parameter a increases, the amplitude of $\Psi_{a,0}(x)$ will be scaled down. To enhance the searching performance, this property will be utilized in mutation operation.

As more than 99% of the total energy of the mother wavelet function is contained in the interval $[-2.5, 2.5]$ (Property II), x can be randomly generated from $[-2.5 \times a, 2.5 \times a]$. The value of the dilation parameter a is set to vary with the value of k/K to meet the fine-tuning purpose, where k is the current iteration number and K is the maximum number of iterations. To perform a local search when k is large, the value of a should increase as k/K increases to reduce the significance of the mutation. Hence, a monotonic increasing function governing a and k/K may be written as given in the following equation [27]:

$$a = e^{-\ln(g) \times \left(1 - \frac{k}{K}\right)^{\xi_{\omega m}} + \ln(g)} \tag{14}$$

where $\xi_{\omega m}$ is the shape parameter of the monotonic increasing function, and g is the upper limit of the parameter a. The value of a is between 1 and 10,000.

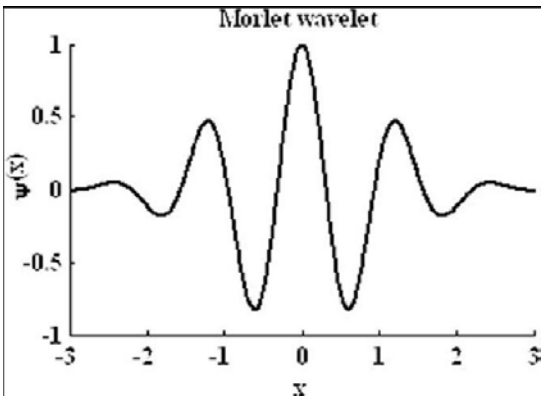


FIGURE 3. Morlet wavelet

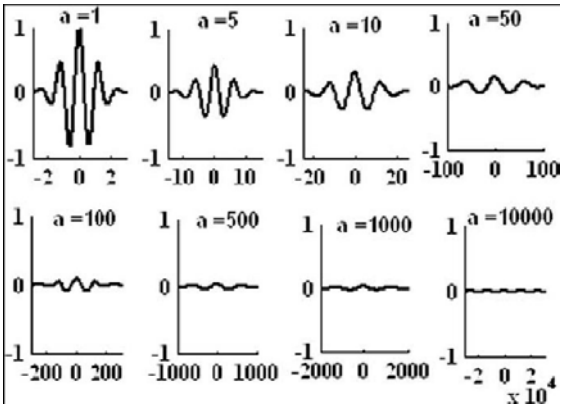


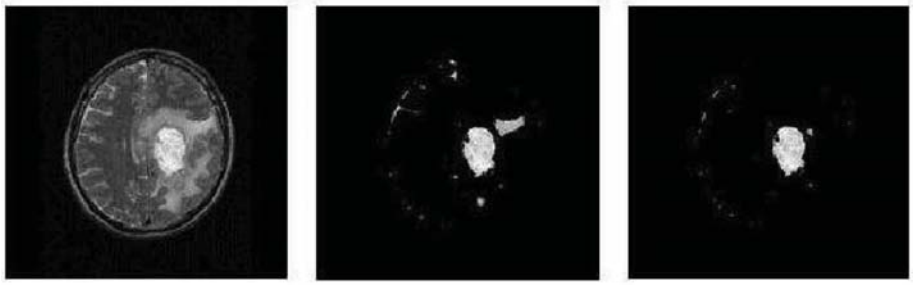
FIGURE 4. Various dilated Morlet wavelets [x-axis, y-axis: $\Psi_{a,0}(x)$]

The HPSOWM algorithm is used for optimizing the initial value of threshold. Using the threshold value we get the ROI which contains the lesions and some noise, the results of the process is shown in Figure 5(b) and Figure 6(b). The original un-segmented brain MRI images of the two patients is displayed in Figure 5(a) and Figure 6(a).

3.1.3. *Concept of variable mask.* We start with a 3×3 mask. All the pixel positions having a value more than the threshold value obtained using Entropy maximization forms the ROI. The mask is moved over the ROI, if the neighbourhood pixels display a similar value, we increase the size of the mask. The mask is grown in all the four directions depending on the similarities of the pixel values. If the neighbourhood pixels have different value, then we again start afresh in the new region of the ROI. The variable mask is applied in a region growing manner on the ROI to reduce the noise and get the final segmented MRI image as shown in Figure 5(c) and Figure 6(c).

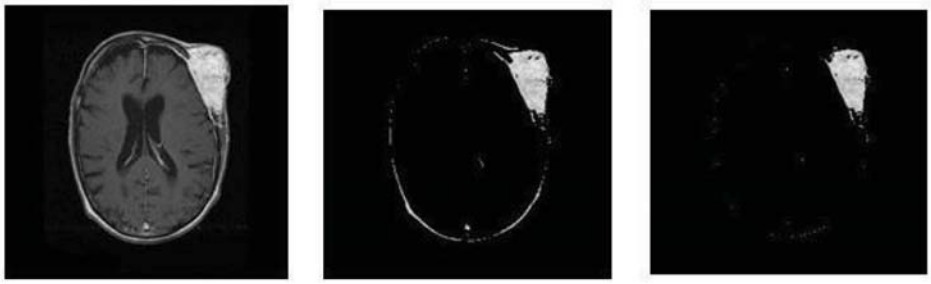
The goal of irrelevancy reduction using HPSOWM technique is to segment the MRI images and transmit only those MRI images which have lesions in them.

3.2. **Progressive transmission of MRI data.** In the field of medical imaging and other allied areas of imaging science the size of MRI images poses a problem in analysis and transmission. For large numbers of data and in our case 2-D MRI data poses problems in network transmissions. During progressive viewing of the digital data in the receiver end a way has to be developed to stop the transmission if the object seems undesirable.



(a) (b) (c)

FIGURE 5. Patient dataset-1: (a) original un-segmented brain MRI, (b) optimized MRI image using HPSOWM, (c) segmentation using the proposed method



(a) (b) (c)

FIGURE 6. Patient dataset-2: (a) original un-segmented brain MRI, (b) optimized MRI image using HPSOWM, (c) segmentation using the proposed method

We define a general process of progressive transmission of MRI image having lesions. We take a set of 19 MRI images in the Axial T2 view which mainly gives the pathology of the disease. The slice thickness is 5.0 mm and the gap between two slices was 1.5 mm. Each slice is having a resolution of 256×256. We denote this digital object as $\{(y_s, N_s)\}_{s=0}^{18}$ where $y_0 < y_1 < y_2 < \ldots < y_{18}$ provide ordering of the slices, from top to bottom of the head and N_s are 256×256 matrices.

A very simple scheme for progressive transmission at the transmitter end can be taking each slice from top to bottom from $y = 0$ to $y = 18$, one at a time and get the segmented MRI image using Entropy maximization using HPSOWM algorithm and variable mask as discussed in Subsections 3.1.1, 3.1.2 and 3.1.3. After getting the segmented image of each slice of MRI image, we perform Discrete Cosine Transform of the segmented MRI slice. DCT does a better job of concentrating energy into lower order coefficients and are much more efficient (fewer are needed to approximate a typical signal). DCT is purely real and produces coefficients that are similar to the frequency domain coefficients produced by a Discrete Fourier transform operation. For most images, after transformation the majority of signal energy is carried by just a few of the low order DCT coefficients. These coefficients can be more finely quantized than the higher order coefficients. Since DCT uses only real values and because they can be finely quantized we use DCT for the purpose of progressive transmission of the segmented MRI image.

3.2.1. *Implementation of twin distribution entropy coding (TDEC).* In this subsection a new technique for efficient lossless data compression using arithmetic coding for the purpose of compressing the DCT coefficients is presented. Arithmetic coding is an entropy

source coding method. In principle when a message is coded using this method, the number of bits in the coded string is the same as the entropy of that message with respect to the model (probability distribution of data) used for coding. This method is superior to Huffman Coding because it dispenses with the restriction that each symbol must translate to an integer number of bits, thus coding more efficiently.

The images are assumed to have $n.L \times m.L$ pixels, where L is the length of the square blocks and n, m are the number of blocks in horizontal and vertical directions, respectively. An image can then be represented by 4-tuple $Y(k,l,i,j)$.

$$Y(k,l,i,j) = \left\{ Y_{i,j}^{k,l},\ k,l = 0,1,2,\ldots,L-1 \right\} \\ i = 0,1,\ldots,n-1, \quad j = 0,1,\ldots,m-1, \quad (15)$$

where k, l are the indexes of the pixels within the block and i, j are the indexes of the block shown in Figure 7.

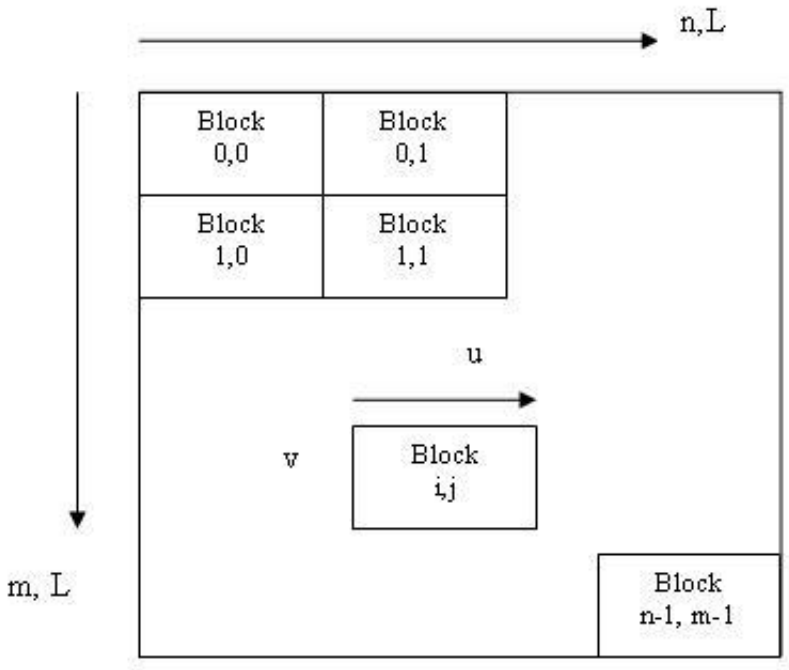


FIGURE 7. Image blocks in the DCT domain

The blocks, $Y_{i,j}$ then undergo an orthogonal and unitary block DCT transform $A(A^t = A^{-1})$.

$$Q_{i,j} = AY_{i,j}A^t \quad (16)$$

where $Y_{i,j}$ and $Q_{i,j}$ are $L \times L$ blocks of pixel data and DCT coefficients, respectively, and A is the $L \times L$ DCT coefficient matrix [4]. The resulting coefficients, $Q_{i,j}^{u,v}$ are rounded to C bit precision before arithmetic coding and transmission. The quantized coefficients are denoted by $X_{i,j}^{u,v}$. After the DCT transform, the original image $Y(k,l,i,j)$ is converted into the transformed image $X(u,v,i,j)$ with coefficient elements $X_{i,j}^{u,v}$.

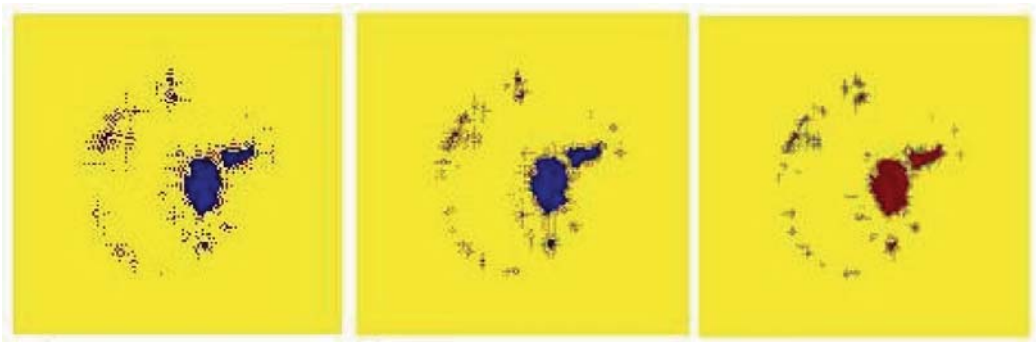
TDEC involves using the lesion information to split the $L \times L$ sources with the DCT coefficients into two subsources, one with lesion and other the background. The multiple probability distribution of subsources are then utilized to arithmetically code the data. If multiple distributions are used, the available *apriori* knowledge about the original source is higher than that before partition. Thus the average entropy of the data is now smaller,

(see Appendix A for a proof) and this results in a lower coding cost. From Appendix A it is clear that no coding benefit is produced if the subsources have identical distributions.

The idea of splitting sources into subsources is not new. [32] used ac energy classification as the criteria to split the DCT coefficients into subsources and then designed quantizers for each subsource. [33] used a maximum-likelihood approach for the splitting. From the findings in [32, 34], it is clear that if ac energy classification is used, the DCT coefficients of different classes have different variances and thus different distributions. Thus based on the proof in Appendix A, TDEC will always yield a coding gain.

In practice the splitting based on the lesion information has the disadvantage that it requires the transmission of the classification map to the receiver. Let the DCT coefficients in each class (lesion and background) be represented as $[X_{i,j}^{u,v}]_z$ for $z = 1, 2$ with z the number of subsources. In order to decode the transmitted data the receiver must know the subsource from which each piece of data came.

Progressive transmission of coded information results in the progressively reconstructed images. Percentages of coded data (DCT coefficients) are transmitted and at the receiver end they are decoded and progressively reconstructed by performing Inverse DCT. K-means clustering is then used to get the clustered view of the each of the progressively reconstructed images. This method of progressive transmission helps us to transmit the lesions of the MRI images for further diagnosis by the doctor. Reconstructed MRI slice with 30%, 40% and 50% of the transmitted coefficients are displayed in Figures 8(a)-8(c)



(a) (b) (c)

FIGURE 8. Reconstruced slices of Figure 5(c) with (a) 30% coefficients, (b) 40% coefficients, (c) 50% coefficients

(a) (b) (c)

FIGURE 9. Reconstructed slices of Figure 6(c) with (a) 30% coefficients, (b) 40% coefficients, (c) 50% coefficients

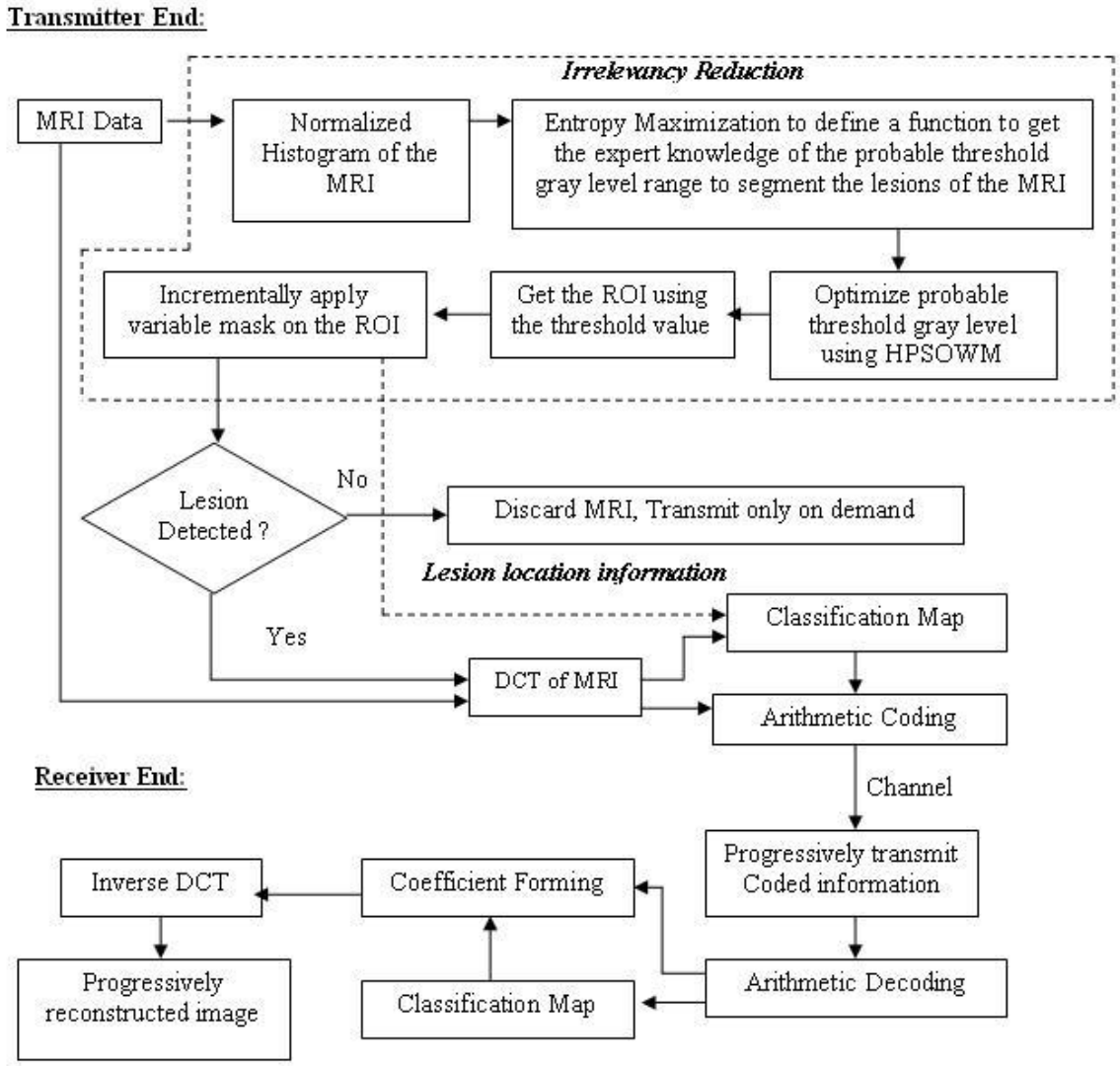


FIGURE 10. Flow diagram of the proposed method

and Figures 9(a)-9(c) respectively. The flow diagram and algorithm are depicted in Figure 10 and Subsection 3.2.2 respectively.

Presence of lesions may result in the progressive transmission of the segmented MRI image by removing the irrelevant details whereas absence of lesions may lead to aborting the transmission. At the receiver's end the doctor identifies the progressively transmitted and segmented MRI image and then may ask for the un-segmented MRI dataset of a particular patient with lesions. Thus effectively the system reduces the load on the transmission network by not sending MRI images without lesions.

3.2.2. *Algorithm.* The algorithm of the proposed method is detailed below.

For a single patient data do the following:

Step 1: Select an MRI slice in a random manner from the data set.

Step 2: Get the ROI by performing Entropy Maximization using HPSOWM algorithm.

Step 3: Expert knowledge is applied to select one or more ROI's from all the computed ROI's.

Step 4: Apply variable mask in a region growing manner on the ROI to get a segregated image containing the lesions of the MRI image.

Step 5: If the lesions are detected then perform Discrete Cosine Transform on the segregated image obtained in Step 4. If lesions are absent then the user may choose to abort transmission and go to Step 1 to choose another MRI image.

Step 6: Use arithmetic coding to compress the DCT coefficients and thereby implement Twin Distribution Entropy Coding.

Step 7: Perform progressive transmission of coded information.

Step 8: Perform arithmetic decoding of the coefficients and perform Inverse Discrete Cosine Transform to progressively re-construct the images.

Step 9: Perform k-means clustering on the images that are obtained in Step 8 for the purpose of analysis. k-means clustering results in predominantly two clusters namely that of lesions and the background.

Step 10: The Progressively reconstructed segmented images are viewed by the receiver and the receiver may identify a particular class of lesions and ask for the transmission of the unsegmented MRI dataset of the patient. Thus the irrelevancy reduction reduces the load on the transmission network by not sending MRIs without lesions in them.

4. Experiments.

4.1. Irrelevancy reduction.

In HPSOWM larger mutating space is set in the early stages of the search whereas it fine tunes to a better solution in the later stage of the search by setting a smaller mutating space based on the properties of Multi-resolution Wavelet Analysis.

A perfect balance between the exploration of new regions and the exploitation of the already sampled regions in the search space is expected in HPSOWM. This balance, which critically affects the performance of the HPSOWM, is governed by the right choices of the control parameters, e.g., swarm size (n_p), the probability of mutation (p_m), and the shape parameter of wavelet mutation (WM) ($\xi_{\omega m}$) as already defined in Subsection 3.1.2.1.2. Changing the parameter $\xi_{\omega m}$ will change the characteristics of the monotonic increasing function of WM. The dilation parameter a will take a value to perform fine tuning faster as $\xi_{\omega m}$ increases. In general, if the optimization problem is smooth and symmetric, it is easier to find the solution, and the fine tuning can be done in early iteration. Thus, a larger value of $\xi_{\omega m}$ can be used to increase the step size (σ) for the early mutation.

The MRI images are multimodal images as discussed earlier in Section 1. In the case of Figure 5(a), using HPSOWM algorithm we take ten particles for each of three dimensional spaces taken, where each dimension of space represents a threshold value. We get three threshold values 25, 115, 160. Using expert knowledge we observe that diseased lesions of the MRI image lie above the threshold value of 160. For Figure 6(a) we obtain the threshold values of 20, 52, 180. We choose the threshold value of 180 for the purpose of detection of the diseased lesions.

Because of space constraints we have displayed the segmentation result of two test images (Figure 5(a) and Figure 6(a)) each from two different patient datasets (Figure 11 and Figure 12) having lesions in the brain MRIs. Each patient dataset has 19 slices of MRI data. Evaluation of segmentation accuracy is done for the two patient datasets using Precision and Recall at the end of this section.

Thus in irrelevancy reduction, Entropy Maximization using HPSOWM is performed to get the expert knowledge of the probable threshold gray level range which correctly segments the lesions in an MRI image. After the threshold value is found, it is used to get the ROI of the MRI image. The variable mask is applied in a region growing manner on the ROI to remove the noise and get the segmented MRI image.

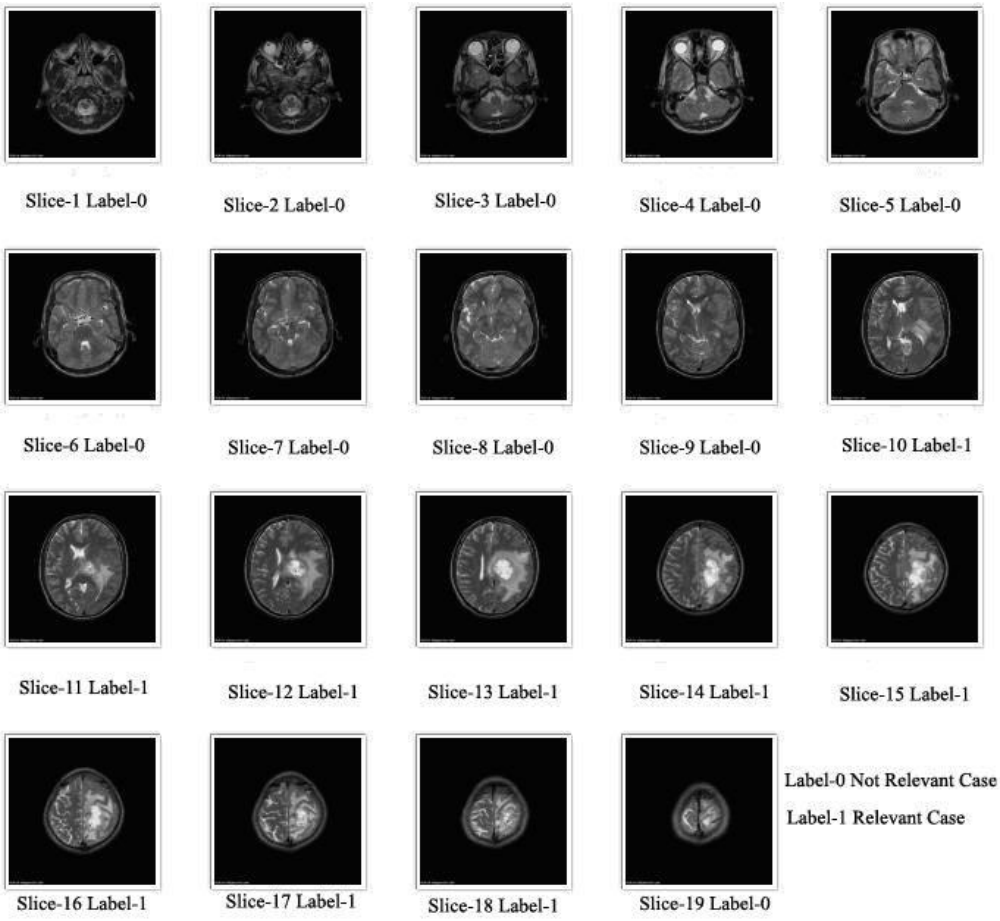


FIGURE 11. Patient dataset-1

4.2. **Progressive transmission.** For analysis we use 2 sets of data (Figure 11 and Figure 12). Each set of data refers to MRI images of brain of a separate patient. Each dataset is denoted by $\{(y_s, N_s)\}_{s=0}^{18}$, where $y_0 < y_1 < y_2 < \ldots < y_{18}$ provide ordering of the slices, from top to bottom of the head and N_s are 256×256 matrices. Thus a total of 38 MRI images were tested, 19 each of two separate patients to arrive at the values of segmentation accuracy. The two patients have presence of lesions in some or all of their MRI images of brain.

The original image is 256×256 gray scale image with horizontal and vertical resolution of 96 dpi. The bit depth is 8. Each DCT coefficient was rounded to 8-b precision, that is, the ac coefficients are integers in the range $[-128, 127]$. The dc coefficients are integers in the range $[0, 255]$. The images were divided into 16 by 16 blocks and the blocks split into 2 classes namely the lesions and the background. The concept of TDEC is used where the blocks are split into two classes which results in lower coding costs of DCT coefficients as explained in Subsection 3.2.1 and Appendix A.

Transmission of a sequence of samples is commonly performed in the natural order defined by the underlying time or spatial index. We use the concept where the percentage of coded DCT coefficients of the MRI data forms the basis of progressive transmission. Here we use percentage of coded DCT coefficients of the MRI data (namely 30% to 100% of coded DCT coefficients) in steps of 10 for progressive transmission. We omit the 10% and 20% of DCT coefficients data for analysis and progressive transmission purpose because these do not result in meaningful reconstruction of the lesions of the MRI image. The image is decoded and reconstructed using inverse DCT. K-means clustering is performed

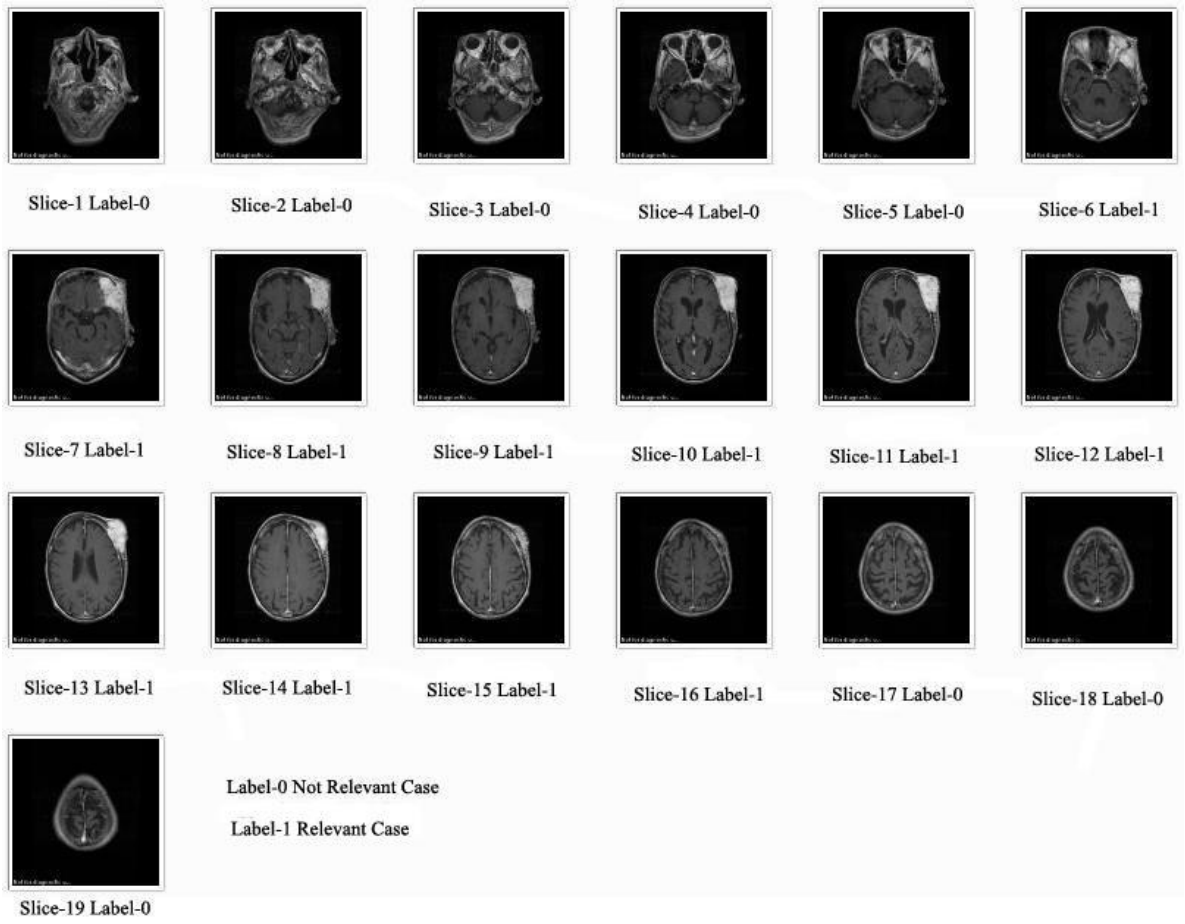


FIGURE 12. Patient dataset-2

to get the clustered view of each of the progressively reconstructed image with lesions. For progressively transmitted images, Figure 5(c) and Figure 6(c), reconstructed images using *k-means* clustering algorithm with 30%, 40% and 50% of the DCT coefficients are displayed in Figure 8 and Figure 9. The reconstructed images become progressively clearer as the percentage of DCT coefficients used for reconstruction are increased.

4.3. **Precision and recall.** We use Precision and Recall to evaluate segmentation accuracy. The precision for a class is the number of true positives (that is the number of items correctly labeled as belonging to the positive class) divided by the total number of elements labeled as belonging to the positive class (i.e., the sum of true positives and false positives, which are items incorrectly labeled as belonging to the class). Recall is defined as the number of true positives divided by the total number of elements that actually belong to the positive class (that is the sum of true positives and false negatives, which are items which are not labeled as belonging to the positive class but should have been). The slices as depicted in Figure 11 and Figure 12 are labeled as belonging to relevant class or that of Not relevant class. The relevant case is the presence of lesions and the Not relevant case is the absence of lesions. The values needed for evaluating segmentation accuracy for the two patients are listed in tabular form in Table 1.

TN/True Negative: Case was negative and predicted negative.

TP/True Positive: Case was positive and predicted positive.

FN/False Negative: Case was positive but predicted negative.

FP/False Positive : Case was negative but predicted positive.

$$Precision = \frac{TP}{TP + FP} \tag{17}$$

$$Recall = \frac{TP}{TP + FN} \tag{18}$$

The Precision and Recall values for Patient Dataset-1 (Figure 11) are 87.5% and 77.78 % respectively whereas for Patient Dataset-2 (Figure 12) the values are 88.89% and 72.72% respectively.

TABLE 1. Evaluation of segmentation

Patient	Expert Knowledge	Total Cases		Predicted Negative	Predicted Positive
Patient Dataset-1	Presence of Lesions in some or all of the MRI's	19	Negative Cases	9	1
			Positive Cases	2	7
Patient Dataset-2	Presence of Lesions in some or all of the MRI's	19	Negative Cases	7	1
			Positive Cases	3	8

Note: Total Cases: Total 19 MRI slices were tested for each patient.
Positive Cases: MRI's with Lesions.
Negative Cases: MRI's without Lesions.

5. Discussions, Conclusions and Future Research.

5.1. **Motivation of the practical use of the theoretic results.** [2] found that using Entropy maximization using PSO algorithm gives better results in comparison with other methods [3]. However, PSO converges sharply in the earlier stages of the search process but it saturates or even terminates in the later stages. The behaviour of PSO presents problems with the velocity update. If a particle's current position coincides with the global best position, the particle will only move away from this point if its inertia weight and velocity are different from zero. If their velocities are very close to zero, then all the particles will stop moving once they catch up with the global best particle, which may lead to premature convergence, and no further improvement can be obtained. This phenomenon is known as stagnation [35].

The mutation operation is used to mutate the elements of particles. [28] proposed to integrate the Genetic Algorithm's mutation operation into the PSO, which aids to break through stagnation. This variation of PSO is called HPSOM (Hybrid Particle Swarm Optimization with Mutation). The mutation operation starts with a randomly chosen particle in the swarm, which moves to different positions inside the search area through the mutation. The following mutation operation is used in the HPSOM:

$$mut(x_j) = \begin{cases} x_j - \omega, & r < 0 \\ x_j + \omega, & r \geq 0 \end{cases} \tag{19}$$

where x_j is randomly chosen element of the particle from the swarm, and ω is randomly generated within the range $[0, 0.1 \times (S_{j,\max} - S_{j,\min})]$ representing one tenth of the length of the search space. r is a random number between $+1$ and -1 and $S_{j,\max}$ and $S_{j,\min}$ are the upper and lower boundaries of each particle element, respectively.

It is noticed from (19) that the mutating space in the HPSOM is limited by ω. It is not the best approach in fixing the size of the mutating space all the time along the search. It can be further improved by a dynamic mutation operation in which the mutating space dynamically contracts along the search. Thus Wavelet Mutation that varies the mutating space based on the wavelet theory is used by us along with PSO (HPSOWM) for MRI segmentation purpose.

In this paper a method for segmentation and progressive transmission of MRI images using Hybrid Particle Swarm Optimization with Wavelet Mutation Operation (HPSOWM) is proposed. Everything other than the lesion in the MRI is treated as irrelevant for segmentation purpose. The MRI images which do not contain lesions are treated as irrelevant. The novelty of the method is that it discards the irrelevant information of the MRI image and thus reduces the effective size of the data to be transmitted over a network. Progressively transmitted MRI are viewed by a doctor or a radiologist at the receiver's end. After the doctor identifies a particular MRI image from a set of patient data, the un-segmented MRI data may be sent over the network for further diagnosis.

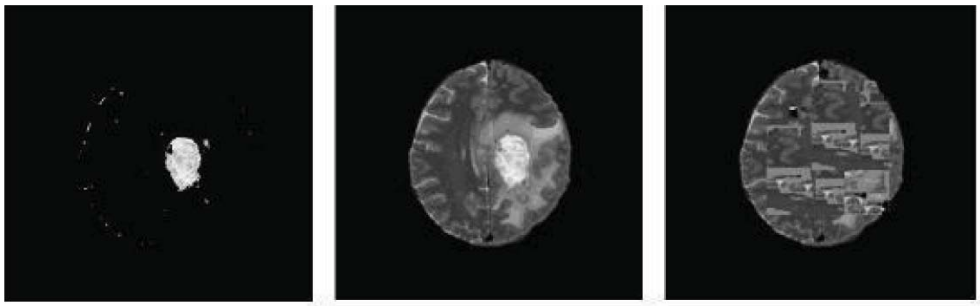
5.2. **Unique features of this work.** A hybrid particle swarm optimization (PSO) that incorporates a wavelet-theory-based mutation operation (HPSOWM) is proposed to be used in the segmentation of MRI image. It applies the wavelet theory to enhance the PSO in exploring the solution space more effectively for a better solution. HPSOWM employs a dynamic mutation operation which overcomes the stagnation phenomenon occurring in PSO.

A Twin Distribution Entropy Coding (TDEC) technique for efficient lossless data compression using arithmetic coding for the purpose of compressing the DCT coefficients is presented. The multiple probability distribution of sub-sources are utilized to arithmetically code the data. If multiple distributions are used, the available a priori knowledge about the original source is higher than that before partition. Thus the average entropy of the data is now smaller and this results in a lower coding cost (proof in Appendix A). Also no coding benefit is produced if the sub-sources have identical distributions (from Appendix A). Although more sub-sources implies larger coding gains, for the ac energy splitting, in practice, this coding gain must be weighted against the increase in overhead required to transmit the block classification map.

Our method of progressive transmission does not produce visually meaningful reconstruction of the lesions when the percentage of DCT coefficients used for transmission is less (we omit the 10% and 20% of DCT coefficients data for analysis and progressive transmission purpose). Also the Precision and Recall value for calculating segmentation accuracy is to be improved upon. Work is underway to improve the segmentation accuracy and to use the method of irrelevancy reduction in abnormalities of Lymph nodes.

5.3. **Comparative study.** [2] used Particle Swarm Algorithm together with entropy maximization for segmenting lesions in an MRI image. The result of the segmentation using [2] is depicted in Figure 13(a). The segmentation result using the proposed method, Figure 5(c) is visually comparable with Figure 13(a).

Before performing the segmentation using [3], the skull is removed so that only the brain tissues remain. Here Axial T2 sequence Figure 13(b) (the same as Figure 5(a) but without skull) was used to bring out the usefulness of the multimodal approach for single sequences only. We consider $K = 4$ corresponding to four classes white matter, gray matter, Cerebro-Spinal fluid and the lesion. The result depicted in Figure 13(c) is not comparable to Figure 5(c) and Figure 13(a).



(a) (b) (c)

FIGURE 13. (a) Segmentation using [2], (b) axial T2 with brain cells only, (c) segmentation using [3]

REFERENCES

[1] S. S. Rao, *Engineering Optimization: Theory and Practice*, 4th Edition, John Wiley and Sons, 2009.

[2] A. De, R. L. Das, A. K. Bhattacharjee and D. Sharma, Masking based segmentation of diseased MRI images, *Proc. of International Conference on Information Science and Applications*, Seoul, Korea, pp.230-236, 2010.

[3] Y. Kabir, M. Dojat, B. Scherrer, F. Forbes and C. Garbay, Multimodal MRI segmentation of ischemic stroke lesions, *Proc. of the 29th Annual International Conference of the IEEE EMBS, Cite Internationale*, Lyon, France, 2007.

[4] K. R. Rao and P. Yip, *Discrete Cosine Transform: Algorithms, Advantages, Applications*, Academic, New York, 1990.

[5] J. B. MacQueen, Some methods for classification and analysis of multivariate observations, *Proc. of the 5th Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley, CA, USA, pp.281-297, 1967.

[6] R. O. Duda, P. E. Hart and D. G. Stork, *Pattern Classification*, 2nd Edition, John Wiley and Sons, 2004.

[7] E. Defez, A. Hervs, A. Law, J. Villanueva-Oller and R. J. Villanueva, Progressive transmission of images: PC-based computations, using orthogonal matrix polynomials, *Mathematical and Computer Modelling*, vol.32, no.10, pp.1125-1140, 2000.

[8] T. Sigitani, Y. Iiguni and H. Maeda, Progressive cross-section display of 3D medical images, *Medical and Biological Engineering and Computing*, vol.38, pp.140-149, 2000.

[9] A. Mehrotra, R. Srikanth and A. G. Ramakrishnan, Shape adaptive integer wavelet transform based coding scheme for 2-D/3-D MR images, *Proc. of the Conference on Data Compression*, pp.552, 2004.

[10] M. S. Atkins, R. Hwang and S. Tang, Performance analysis of algorithms for retrieval of magnetic resonance images for interactive teleradiology, *Proc. of SPIE – The International Society for Optical Engineering*, vol.4319, pp.671-680, 2001.

[11] T.-C. Lu and C.-C. Chang, A progressive image transmission technique using Haar wavelet transformation, *International Journal of Innovative Computing, Information and Control*, vol.3, no.6(A), pp.1449-1461, 2007.

[12] M.-H. Tsai, S.-F. Chiou and M.-S. Hwang, A progressive image transmission method for 2D-GE image based on context feature with different thresholds, *International Journal of Innovative Computing, Information and Control*, vol.5, no.2, pp.379-386, 2009.

[13] J. M. S. Prewitt and M. L. Mendelsohn, The analysis of cell images, *Ann. N. Y. Acad. Sci.*, vol.128, no.3, pp.1035-1053, 1966.

[14] J. S. Weszka and A. Rosenfeld, Histogram modification for threshold selection, *IEEE Transactions on Systems, Man and Cybernetics*, vol.9, no.1, pp.38-52, 1979.

[15] J. Kittler and J. Illingworth, Minimum error thresholding, *Pattern Recognition*, vol.19, no.1, pp.41-47, 1986.

[16] S. Cho, R. Haralick and S. Yi, Improvement of Kittler and Illingworth's minimum error thresholding, *Pattern Recognition*, vol.22, no.5, pp.609-617, 1989.

[17] J. S. Lee and M. C. K. Yang, Threshold selection using estimates from truncated normal distribution, *IEEE Trans. Syst. Man Cybern.*, vol.19, no.2, pp.422-429, 1989.

[18] T. Pun, Entropic thresholding: A new approach, *CVGIP: Graphical Models Image Process.*, vol.16, pp.210-239, 1981.
[19] C. H. Li and C. K. Lee, Minimum cross entropy thresholding, *Pattern Recognition*, vol.26, no.4, pp.617-625, 1993.
[20] H. D. Cheng, J. R. Chen and J. G. Li, Threshold selection based on fuzzy c-partition entropy approach, *Pattern Recognition*, vol.31, no.7, pp.857-870, 1998.
[21] P. K. Saha and J. K. Udupa, Optimum image thresholding via class uncertainty and region homogeneity, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol.23, no.7, pp.689-706, 2001.
[22] N. Otsu, A thresholding selection method from gray-level histograms, *IEEE Trans. Syst. Man Cybern*, vol.9, no.1, pp.62-66, 1979.
[23] J. Kittler and J. Illingworth, On threshold selection using clustering criteria, *IEEE Trans. Syst. Man Cybern*, vol.15, no.5, pp.652-655, 1985.
[24] Q. Hu, Z. Hou and W. L. Nowinski, Supervised range-constrained thresholding, *IEEE Trans. Image Process.*, vol.15, no.1, pp.228-240, 2006.
[25] Y. Qiao, Q. Hu, G. Qian, S. Luo and W. L. Nowinski, Thresholding based on variance and intensity contrast, *Pattern Recognition*, vol.40, pp.596-608, 2007.
[26] H. Musmann, P. Pirsch and H. Gallert, Advances in picture coding, *Proc. of IEEE*, pp.523-542, 1982.
[27] S. H. Ling, H. H. C. Iu, F. H. F. Leung and K. Y. Chan, Improved hybrid particle swarm optimized wavelet neural network for modeling the development of fluid dispensing for electronic packaging, *IEEE Transactions on Industrial Electronics*, vol.55, no.9, pp.3447-3460, 2008.
[28] A. A. E. Ahmed, L. T. Germano and Z. C. Antonio, A hybrid particle swarm optimization applied to loss power minimization, *IEEE Transactions on Power Systems*, vol.20, no.2, pp.859-866, 2005.
[29] I. Duabechies, Ten lectures on wavelets, *Society for Industrial and Applied Mathematics*, PA, 1992.
[30] S. H. Ling, C. W. Yeung, K. Y. Chan, H. H. C. Iu and F. H. F. Leung, A new hybrid particle swarm optimization with wavelet theory based mutation operation, *Proc. of IEEE Congress on Evolutionary Computation*, Singapore, pp.1977-1984, 2007.
[31] C. W. Yeung, S. H. Ling, Y. H. Chan, S. H. Ling and F. H. F. Leung, Restoration of half-toned color-quantized images using particle swarm optimization with wavelet mutation, *Proc. of IEEE TENCON*, pp.19-21, 2008.
[32] W. H. Chen and C. H. Smith, Adaptive coding of monochrome and color image, *IEEE Trans. Commun.*, vol.25, pp.1285-1292, 1977.
[33] G. W. Wornell and D. H. Staelin, Transform image coding with a new family of models, *Proc. of IEEE ICASSP*, New York, pp.777-780, 1988.
[34] R. C. Reininger and J. D. Gibson, Distribution of the two-dimensional DCT coefficients for images, *IEEE Trans. Commun.*, vol.31, pp.835-839, 1983.
[35] R. C. Eberhart and Y. Shi, Comparison between genetic algorithms and particle swarm optimization, *Evolutionary Programming VII*, vol.1447, pp.611-616, 1998.

Appendix A. In this appendix, it is shown that the entropy of a source is reduced if it is partitioned into subsources with different probability distributions. Let X_i be a source sequence of $N > 1$ independent samples $X_1, X_2, \ldots, X_N$ with time or spatial indexing, with a discrete range of integer values $1 \leq X_i \leq R$, and with known $P\{X_i = r\}$ or P_r. As usual, the source entropy $H(X_i) = \sum_{r=1}^{R} -P_r \log_2 P_r$.

Multiple distribution entropy coding divides the source X_i to M subsource sequences X_i^k where $1 \leq k \leq M$. X_i^k is a sequence of L_k independent samples $X_1^k, X_2^k, \ldots, X_{L_k}^k$ with time or spatial indexing, and with known $P\{X_i^k = r\}$ or P_r^k. That is, the X_i^k are disjoint, nonempty, and $U_k\{X_i^k\} = \{X_i\}$. Because X_i^k are subsources of X_i, it is clear that

$$P_r = \sum_{k=1}^{M} \frac{L_k}{N} . P_r^k, \quad 1 \leq r \leq R \tag{A-1}$$

After dividing the source, the average entropy of the sub-sources based on their multiple probability distributions is

$$\bar{H}(X_i^k) = \sum_{k=1}^{M} H(X_i^k) = \sum_{k=1}^{M} \frac{L_k}{N} \left\{ \sum_{r=1}^{R} -P_r^k \log_2 P_r^k \right\} \tag{A-2}$$

the mean of the entropy of the sub-sources.

Lemma 1. *$\bar{H}(X_i^k) \leq H(X_i)$ for known P_r^k.*

Proof: Because $\bar{H}(X_i^k)$ is a function of P_r^k, the maximum value of $\bar{H}(X_i^k)$ can be found from

$$\frac{d\bar{H}(X_i^k)}{dP_r^k} = 0 \text{ for } k = 1, 2, \ldots, M \tag{A-3}$$

with the constraint that

$$P_r^M = \frac{N}{L_M} P_r - \sum_{k=1}^{M-1} \frac{L_k}{L_M} P_r^k, \quad 1 \leq r \leq R \tag{A-4}$$

Equations (A-3) and (A-4) yield

$$\begin{aligned} \frac{dH(X_i^k)}{dP_r^k} &= \frac{d\left[-P_r^k \log_2 P_r^k - P_r^M \log_2 P_r^M\right]}{dP_r^k} \\ &= -P_r^k \left(\frac{1}{P_r^k}\right) \log_2 e - \log P_r^k + P_r^M \left(\frac{1}{P_r^M}\right) . \log_2 e + \log P_r^M \\ &= \log_2 \frac{P_r^k}{P_r^M} \end{aligned} \tag{A-5}$$

From (A-5), it is clear that the derivative is zero if $P_r^k = P_r^M$ for $k = 1, 2, \ldots, M$. Then (A-1) yields

$$P_r = P_r^k, \quad 1 \leq k \leq M \tag{A-6}$$

combining (A-2) and (A-6) yields

$$[\bar{H}(X_i^k)]_{\max} = H(X_i) \tag{A-7}$$

Lemma 1 verifies that the entropy of source is reduced or remains constant if we divide it to $M > 1$ sub-sources. Also, from the proof it is clear that no coding advantage is gained if the sub-sources have the same probability distributions.