TABLE OF CONTENTS

LIST OF FIGURES

Figure Number **Page**

LIST OF TABLES

Table Number **Page**

LIST OF ACRONYMS

AWGN	Wavelet Transform
2D	Two Dimensional
DWT	Discrete Wavelet Transform
IDWT	Inverse Discrete Wavelet Transform
MSE	Mean Square Error
PSNR	Peak Signal to Noise Ratio
GMM	Gaussian Mixture Model
EM	Expectation Maximization
HMM	Hidden Markov Model
HMT	Hidden Markov Tree
IMM	Independent Mixture Model
PMF	Probability Mass Function
PDF	Probability Density Function
FFT	Fast Fourier Transform
MAP	Maximum a Posteriori
ML	Maximum Likelihood
MRF	Markov Random Field
MRA	Multi Resolution Analysis
PGM	Probabilistic Graphical Models

CHAPTER 1

INTRODUCTION

This chapter concentrates on recent trends in statistical image processing arena mainly image modelling and signal processing. Both these trends have huge applications in fast transmission of accurate and reliable information (image, audio and video). The real world signals are random in nature and they cannot be modelled by traditional Fourier transform techniques due to their time-varying limitations. To counter these spatial domain problems, wavelet transform recognized itself as better tool for statistical signal processing [1]. Wavelet transform performs a variety of tasks on real time signals such as signal estimation, detection, synthesis and most importantly compression [2].

In the field of medical image processing, two main challenges that occur are; image enhancement and image reconstruction. Image enhancement deals with getting a better quality image than original image. Image reconstruction handles refinement of corrupt image and obtains noise free image by using previous observations. In both scenarios, a clean image for medical diagnostic purposes is obtained by using image denoising. Image denoising is the mechanism of removing noise artefacts from desired image while keeping key attributes of image.

Wavelet based image denoising algorithms are emerged as powerful tools to obtain noise suppressed images. These wavelet based techniques perform scalar transforms on individual wavelet coefficients by exploiting inter-scale dependencies among wavelet coefficients. The objective of this thesis is to develop a wavelet based denoising scheme using Hidden Markov Model (HMM) that will capture the non-Gaussian statistics of wavelet coefficients by using intrinsic characteristics of wavelet transform.

1.1. Existing Research Techniques

Medical images are produced by techniques such as *Magnetic Resonance Imaging (MRI), X-ray, Computed Tomography (CT) and Ultrasounds.* These images can be corrupted by noise during acquisition or transmission. Denoising of medical images especially ultrasounds and CT scans that are corrupted by non-Gaussian noise is a very difficult task since finer details in a medical image enclosing diagnostic information should not be destroyed during noise suppression. Current wavelet based denoising models exploits primary properties of wavelet transform that are; locality, compression and multiresolution. These models treats wavelet coefficients as statistically independent or jointly Gaussian [3]. For signal estimation, research techniques mostly revolves around capturing the non-Gaussian characteristics of wavelet coefficients and ignores their inter-scale dependencies [4].

Image processing experts usually lacks the medical expertise to distinguish the diagnostically relevant information from denoising results. For instance, in case of CT scans, speckle noise may contain useful information for medical professionals [5]. Speckle suppression is achieved by enhancing edges of image by using curvelet denoising and Wavelet based image fusion [6]. Thus, robust and versatile denoising techniques are needed instead of methods that are optimal under very specific conditions. This notion of robustness for medical images denoising algorithms is proposed in [7]. A wavelet based technique employing maximum a posteriori estimation (MAP) for speckle noise reduction in medical images while retaining image boundaries is proposed in [6] and [7]. In spatial domain, commonly used medical image denoising techniques are Weiner filter, bilateral filter and hybrid median filter [8]. In medical diagnosis, ultrasonography is regarded to be one of the most powerful medical tools for imaging superficial muscles and ligament tissues. It is preferred over other image processing techniques because of its portability, versatility and non-invasive nature. But ultrasound images are usually corrupted with multiplicative noise known as 'speckle' caused by scattering phenomenon that degrades major image quality. X-Ray is good for visualizing bone fractures and joint spaces. X-Ray images are usually corrupted by Poisson distributed noise [9]. Many wavelet based denoising techniques have been developed to address noise removal especially speckle suppression in case of CT scans. In [7], Pizurica et al. proposed a robust multi-scale wavelet technique based on generalized likelihood ratio for speckle removal from medical image. For image restoration, Discrete Wavelet Transform (DWT) is used because of its sparsity and multiresolution properties [10]. To avoid shift-sensitivity of DWT, a Dual Tree Complex Daubechies Wavelet Transform based denoising technique is proposed in [11]. A wavelet based Hidden Markov Model for effective image denoising is developed in [12]. A hierarchical Hidden Markov Tree Model (HHMT) for medical image denoising that uses contextual Hidden Markov Model to avoid time consuming HMT parameters training process is proposed in [13], [14], [15]. In [16], a thresholding technique using neighbouring wavelet statistics for medical image denoising is proposed. MRI is used for visualizing soft tissues, ligaments and intervertebral discs. For MRI, the complex MRI components are usually modeled by additive white-Gaussian noise [17], [18].

1.2. Synopsis/Thesis Statement

The goal of this research is to develop an efficient wavelet based denoising technique using Hidden Markov Model for noise corrupted medical images. This denoising algorithm uses Probabilistic Graphical Model and exploits quad tree structure of multidimensional Gaussian Mixture Models known as Hidden Markov Tree (HMT) to capture non-Gaussian statistics and inter-scale dependencies of medical images in wavelet domain.

1.3. Objectives of Research

The primary objective of this thesis is to develop wavelet based denoising algorithm using Hidden Markov Model for 2D medical images. The proposed model captures non-Gaussian statistics of wavelet coefficients and their inter-scale dependencies. The noise distributions of images are represented by Probabilistic Graphical Models in wavelet domain. The secondary objective is to apply Expectation Maximization (EM) algorithm on Hidden Markov Tree model to achieve denoising.

1.4. Applied Research Methodology

In this thesis, research is carried out in two stages:

- In first stage, primary properties of wavelet transform are achieved. Then secondary properties of wavelet transform are captured using Probabilistic Graphical Models.
- In second stage, an iterative optimization algorithm (EM) combined with Hidden Markov Model is used for medical image denoising.

1.5. Scope of Research

Medical image denoising is critical when an image is transmitted to medical specialist at remote location for medical diagnosis. An effective denoising model is needed so that diagnostically important information embedded in medical image don't get destroyed during denoising process. The proposed framework based on Hidden Markov Model reduces computational complexity as compared to traditional wavelet based denoising techniques. The proposed model can also be used for image compression, synthesis and estimation.

1.6. Use in Other Areas of Application

The proposed framework has found its usage in variety of applictions such as;

- Image compression
- Video denoising
- For synthesis of sound in audio signal processing
- For synthesis of speech in speech signal processing
- In image processing for features extraction

1.7. Organization of Thesis

The thesis consists of following chapters;

- Chapter 1 gives introduction of thesis topic, research background, problem statement, used research methodology, scope and objectives of proposed work.
- In Chapter 2 provides literature review of Discrete Wavelet Transform (DWT) and image decomposition carried out by DWT.
- Chapter 3 explains Hidden Markov Models, their properties, characteristics and implementation.

- Chapter 4 provides in depth analysis of proposed denoising framework for medical images.
- Chapter 5 gives experimental results based on the proposed denoising framework and evaluation of its performance by using real world medical images.
- Chapter 6 points out future work and give thesis conclusion.

CHAPTER 2

DISCRETE WAVELET TRANSFORM

2.1 Introduction of Discrete Wavelet Transform

Fourier transform represents random signals that are stochastic in nature either in time domain or in frequency domain. Fourier transform is localized in frequency but does not give any space or time information of the statistical signal. The widowing method of Fourier transform known as short time Fourier transforms (STFT) is a trade-off between localization of frequency and time. Failure of Fourier transform in spatial domain has given rise to the wavelet transform. Wavelets works perfectly for time varying signals. Wavelets are specific group of functions that are used to analyse given signal in both time and frequency domains. Mallat in [19] proposed that a family of wavelets can be constructed using mother wavelet on specific criterion that makes wavelet transform very favourable for handling non-stationary signals. The main advantage of wavelet transform is that it performs analysis of signal at various scales of resolution attributing to phenomenon multiresolution that makes wavelet transform very useful tool for image processing applications.

Discrete Wavelet Transform (DWT) is based on the concept of subband coding. DWT is a fast variant of wavelet transform in terms of computations. DWT based decomposition techniques were proposed in 1976. DWT uses different digital filtering schemes to represent the signal in wavelet domain that have different scales for signal analysis. DWT has vast applications such as image segmentation, compression, denoising, edge detection and pattern recognition in image processing field [20]. The compact nature of DWT makes it computationally efficient.

2.2 DWT Multiresolution Analysis

Discrete Wavelet Transform (DWT) is preferred over other wavelet transforms because it keeps temporal details of stochastic signal alongside providing amplitude and frequency of said signal. DWT is usually expressed as;

$$\psi_{m,n}(t) = l_0^{-\frac{m}{2}} \psi(l_0^{-m} t - s f_0), \quad m, n \in \mathbb{Z} \tag{2.1}$$

Where *l* and *s* are two random signals discretised in time domain as;

$$l = l_0^m, m \in \mathbb{Z} \tag{2.2}$$

$$s = n l_0 s_0^m, m \in \mathbb{Z} \tag{2.3}$$

Where m and n denote scale and translation statistics of DWT. The DWT of function $f(t)$ becomes:

$$f(t) = \sum_m \sum_n \omega_f\,(m, n) \Psi_{m,n}(t) \tag{2.4}$$

Multi resolution theory presents a stochastic technique for generation of wavelets. MRA works by approximating function $f(t)$ at different levels of resolution. Consider two basic functions; the scaling function $\varphi(t)$ and a mother wavelet $\alpha(t)$. The corresponding scaled and translated versions of scaling function are given by (2.5). For definite values of m, the relevant set of scaling function $\varphi(t)$ and $\alpha(t)$ are orthonormal.

$$\varphi_{m,n}(t) = 2^{-\frac{m}{2}}\varphi(2^{-m}t - n) \tag{2.5}$$

MRA finds wavelet functions by exploiting hierarchical nature of increasing resolutions of scaling functions. By linear combining scaling function with its translations, MRA achieve a set of functions to represent any stochastic signal [21].

$$f(t) = \sum_{n} \alpha_n \, \varphi_{m,n}(t), f(t)\epsilon V_m \tag{2.6}$$

2.2.1 Signal Decimation using DWT

DWT is analysed by using a cascade of high and low pass filters. DWT carries out decomposition by taking advantage of the property that it connects continuous time multiresolution to discrete time filters [22]. In figure 2.1, a random signal $f(n)$ is passed through analysis filter bank consisting of $h_o[n]$ and $h_1[n]$, that is used to break the input signal $f(n)$ into two half-length subbands $h_{lp}[n]$ and $h_{hp}[n]$. Analysis filter $h_o[n]$ is a low pass filter having output subband $f_{lp}[n]$, that is known as approximation of $f(n)$. The output of high pass filter $h_1[n]$ denoted by $f_{hp}[n]$, is known as detail part of $f(n)$. Each subband is decimated by a factor of two so that the output data coincides with original signal. The decimation and filtering scheme continues till best signal levels are achieved.

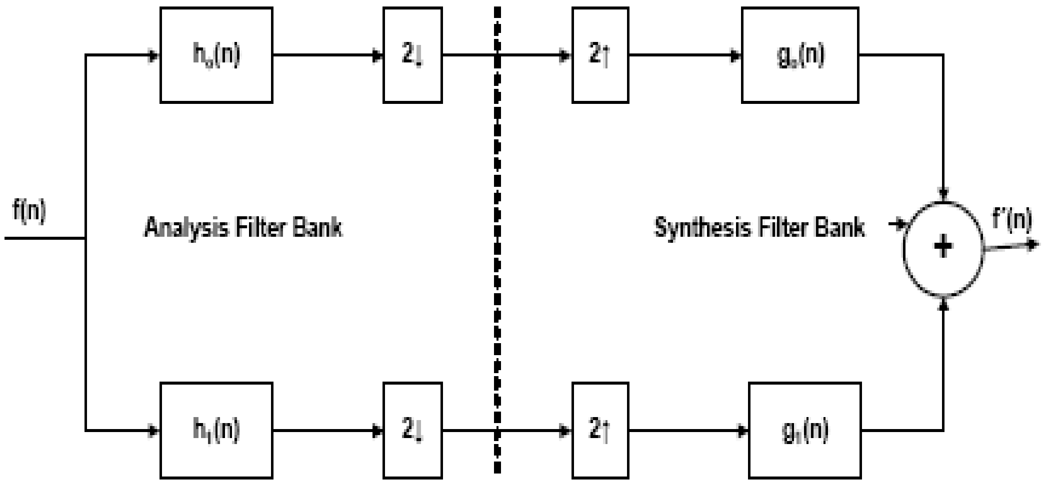


Figure 2-1 Wavelet decomposition using filter banks (1-level)

2.2.2 Signal Reconstruction using DWT

To perform synthesis on signal, it is up sampled by a factor of 2 by inserting zeros between consecutive samples. Reconstruction is opposite of decimation process. Approximation $f_{lp}[n]$ and detail $f_{hp}[n]$ coefficients are passed through synthesis filters $g_o[n]$ *and* $g_1[n]$ separately and then combined together to re-construct the best possible replica $f'(n)$ of original signal.

The objective of MRA based subband decomposition is to select $h_o[n]$, $h_1[n]$, $g_o[n]$ *and* $g_1[n]$ such that $f\acute{[n]} = f[n]$.That is; by using DWT, reconstructed image is best possible copy of original image.

In subband decimation theory, synthesis filters are regarded as modulated versions of analysis filters [22]. In order to obtain best restoration, the impulse responses of the synthesis and analysis filters have to satisfy conditions;

$$g_0[n] = (-1)^n h_1[n] \tag{2.7}$$

$$g_1[n] = (-1)^{n+1} h_0[n] \tag{2.8}$$

MRA based DWT decomposition can be extended to two dimensional 2D medical images by processing rows in one dimension followed by columns processing in other dimension as discussed in the next section.

2.3 DWT Image Decomposition

The compact property of DWT dictates that most of the diagnostic information embedded in medical image is concentrated onto few wavelet coefficients with large magnitudes. Using this compact property, diagnostic details of CT scan can be captured from those large wavelet coefficients which is useful for image restoration. Discrete Wavelet transform (DWT) of a signal is computed by using a cascade of low and high pass filters with a subsampling by a factor of 2. For images, 2D DWT is implemented by using 1D DWT along the rows of the image first and then on columns of the image. DWT decomposes the 2D image into four subbands as shown in figure 2.2. The LL band carries of approximation coefficients, LH band contains horizontal details, HL band contains vertical details whereas HH band carries diagonal details of image. Usually, most of the important information is concentrated in LL subband of highest level. DWT alleviates the noise artefacts such as speckles by using its overlapping nature.

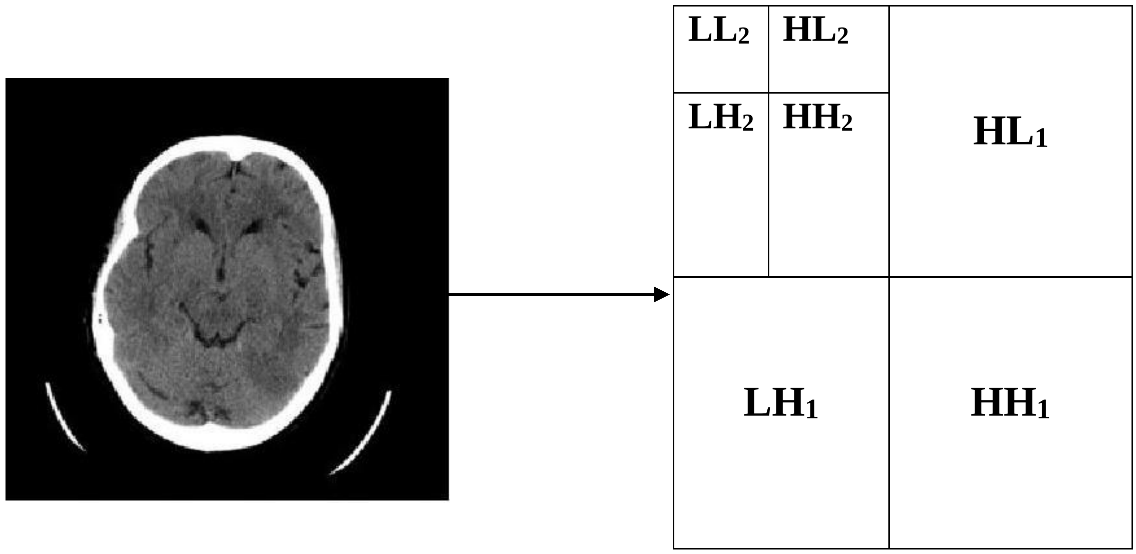


Figure 2-2 Image decomposition of 2D CT scan of brain using DWT (2 levels)

DWT decomposition of image $c(t)$ having dimensions (NXM) and $(scales = R)$ is given by;

$$c(t) = \sum_{i\epsilon Z^2} \mu_{R.i} \emptyset_{R,i}^{LL}(t) + \sum_{b\epsilon\beta} \sum_{r=1}^{R} \sum_{i\epsilon Z^2} \omega_i^b \, \psi_{r,i}^b(t) \quad (2.9)$$

Where $\emptyset^{LL}$ represents scaling function for image. Image decomposition can be achieved by extending 1D Discrete Wavelet Transform to 2D (two dimensions) conditioned on fact that image should be separable. To perform image decomposition, one 2D scaling function $\varphi(a, b)$ and three 2D mother wavelets $\psi^1(a, b)$, $\psi^2(a, b)$ and $\psi^3(a, b)$ are required. Every wavelet function is computed by assuming that is the product of two separable 1D scaling functions and mother wavelets.

$$\varphi(a,b) = \varphi(a)\varphi(b) \tag{2.10}$$

$$\psi^1(a,b) = \psi(a)\varphi(b) \tag{2.11}$$

$$\psi^2(a,b) = \varphi(a)\psi(b) \tag{2.12}$$

$$\psi^3(a,b) = \psi(a)\psi(b) \tag{2.13}$$

Equation (2.10) computes approximation coefficients whereas equation (2.11), (2.12) and (2.13) find directional details of wavelet coefficients along horizontal, vertical and diagonal directions of medical image respectively. Extension of 2D wavelet from 1D wavelet for image decomposition is determined by equations (2.14) and (2.15).

$$\varphi_{r,m,n}(a,b) = 2^{r/2}\psi(2^r a - m, 2^r b - n) \tag{2.14}$$

$$\psi_{r,m,n}^{i}(a,b) = 2^{r/2}\psi^r(2^r a - m, 2^r b - n) \tag{2.15}$$

Where parameter $i = 1,2,3$ identifies the directions (horizontal, vertical and diagonal) of wavelet coefficients. DWT decomposition of image $c(a,b)$ can be found by modifying equation (2.9).

$$W_{\varphi}(r_0,m,n) = \frac{1}{\sqrt{MN}} \sum_{A=0}^{M-1} \sum_{B=0}^{N-1} f(a,b)\, \varphi_{r_0,m,n}(a,b) \tag{2.16}$$

$$W_{\psi}^{i}(r,m,n) = \frac{1}{\sqrt{MN}} \sum_{A=0}^{M-1} \sum_{B=0}^{N-1} f(a,b)\, W_{i,m,n}^{r}(a,b) \qquad (2.17)$$

Where r_0 is an arbitrary starting wavelet scale and $W_{\varphi}(r_0, m, n)$ imposes approximation for image $c(a,b)$ at scale r_0. The coefficients of $W_{\psi}^{i}(r,m,n)$ computes horizontal, vertical and diagonal directional details using scales $r \geq r_0$. Standard approximations are $r_0 = 0$, $N = M = 2^r$ for $r = 1,2,3 \dots, r-1$ and $m = n = 0,1,2, \dots, 2^{r-1}$.

The inverse 2D Discrete Wavelet Transform is performed as;

$$f(a,b) = \frac{1}{\sqrt{MN}} \sum_{A=0}^{M-1} \sum_{B=0}^{N-1} W_{\varphi}(r_0,m,n) \varphi_{r_0,m,n}(a,b) + \frac{1}{\sqrt{MN}} \sum_{i=H,V,D} \sum_{R=R_0}^{\propto} \sum_{m} \sum_{n} W_{\psi}^{i}(r,m,n)\ \psi_{r,m,n}^{i}(a,b) \qquad (2.18)$$

The real image decomposition of CT scan of brain using matlab wavelet package is shown in figure 2.3. It decomposes real medical image using 3-levels.

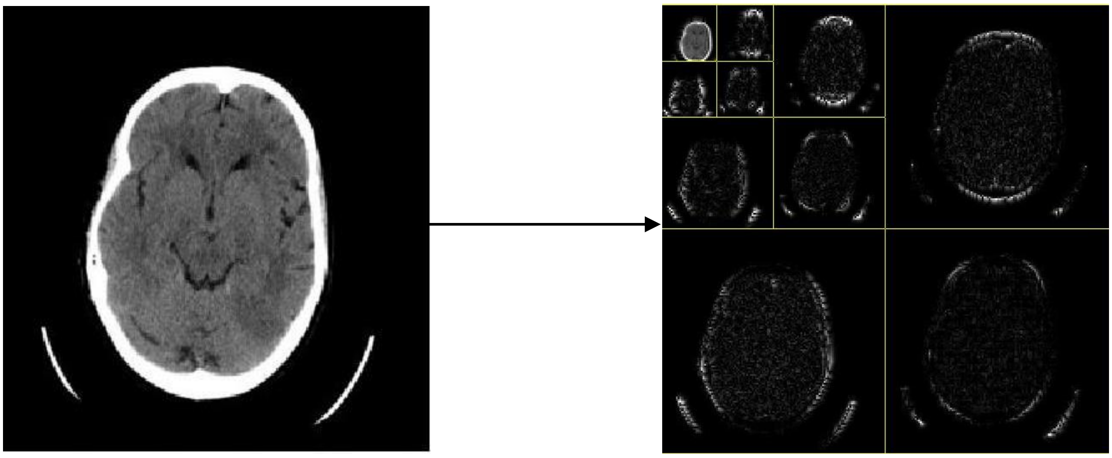
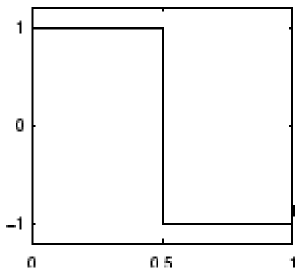


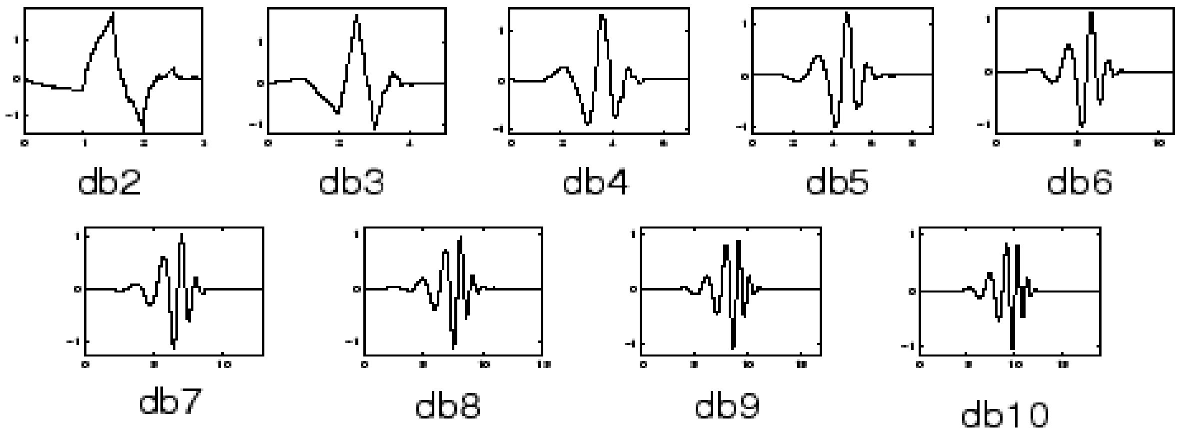
Figure 2-3 Three level image decomposition of 2D CT scan of brain using DWT

2.4 Wavelet Families

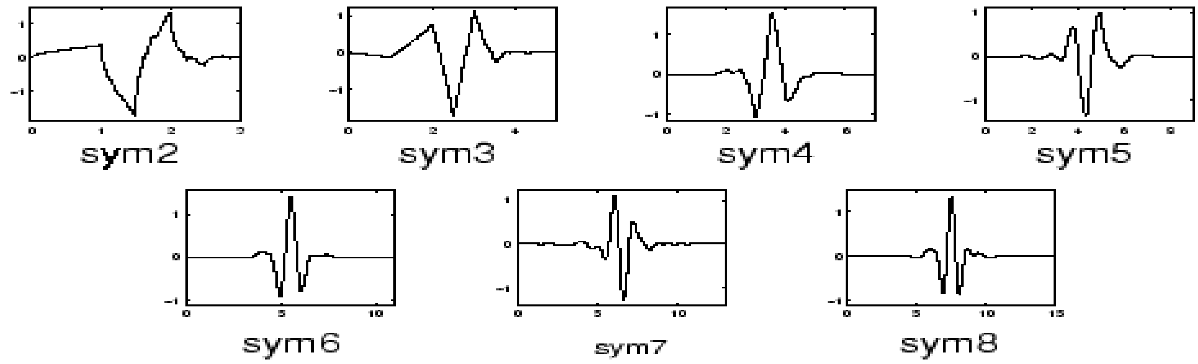
Wavelet transform becomes more efficient by using most suitable type of mother wavelet for particular application. In wavelet analysis, selection of suitable parent wavelet is very important because parent wavelet generates wavelet functions by using scaling and translations. Figure 2.4 depicts psi distribution of different wavelet families.



(a)



(b)



(c)

Figure 2-4 Different Wavelet Families (a) Haar (b) Daubechies (c) Symlets

Wavelet packages Haar, Daubechies and Symlets usually provide compact orthogonal wavelets. Best restoration of image is possible by using these techniques. Other wavelet families such as Demyer, Morlet and Mexican Hat follow symmetrical distributions. The choice of selecting a particular wavelet family is dependent on psi distribution of a wavelet scaling function and their capability to analyse image effectively in different image processing applications.

CHAPTER 3

HIDDEN MARKOV MODEL

3.1. Introduction to Markov Models

Hidden Markov Model (HMM) is powerful tool for statistical signal processing. HMM stochastically models a system that has unobserved states also known as hidden states. HMM can be graphically represented by Bayesian network or Probabilistic Graphical Models (PGM) as discussed in next sections. HMM has found vast applications in image modeling, speech recognition and data tagging [23].

Markovian models perform stochastic signal processing by using two mechanisms namely Markov Chain Models and Hidden Markov Models. In most simple Markov Chain Models, unobserved state is directly visible to user so model is only characterized by single parameter; state transition probabilities matrix. In Hidden Markov Model, desired state is not visible and only hidden state output based on state is available for stochastic modeling. Each state has Gaussian probability distribution over possible outputs. Computation of probability distribution for each and every state is impractical approach for data modeling. Hidden Markov Model based PGM are used to solve this problem.

Hidden Markov Model is regarded as generalization of output mixture models known as Gaussian Mixture Models (GMM) [24]. Latent variables are used to determine the type of mixture component to be selected for each and every observation. HMM links latent variables by using Markovian process instead of leaving them statistically independent. In Markov chain models, the particular state of given random process and its associated probability of existing in another state at next instant only depends on current state of process. But the state of random process is hidden in case of HMM. The latent state is linked with one of other probability distributions that is known to observer. Hidden Markov Model is characterized by parameters given in table 3.1. The complete Markov random process is defined as;

$$M = \{E, F, \pi\} \tag{3.1}$$

Parameters	Description
D	Length of sequence observed
S	Number of states in Markov Model
O	Number of Observations
$Q = (q_1, q_2, \ldots . q_N)$	Observed States
$V = (v_1, v_2, \ldots . v_M)$	Set of possible Observations
$E = \{e_{ij}\}$	State Transition Probability matrix
$F = \{f_i(k)\}$	Output Symbols Probability Matrix
π_i	Initial State Vector Distribution

Table 3-1 Characteristic Parameters of Hidden Markov Process

3.2. Basic Components of Markov Models

Hidden Markov Model uses particular algorithms and assumptions to perform effectively.

3.2.1 Bayes Rule combined with Markov Assumption

Bayes rule combined with Markov assumption (Markov Model Order =1) is an excellent computational tool for stochastic systems. It is used to determine state transition probability matrix and observation symbols probabilities. The state transition probabilities are represented as;

$$P\left(S_m \overset{x}{\rightarrow} O_k S_{m+1}\right) = P(O_m|S_m)P(S_{m+1}|S_m) \tag{3.2}$$

The probability of transitioning from state S_m to S_{m+1} conditioned on symbol O_m is given in equation (3.2). The probabilities of observed symbol are defined by;

$$P(O_m) = \sum_m P(O_m, S_m) \tag{3.3}$$

By using Bayes chain rule on equation (3.3), we get;

$$P(O_m) = P(O_1)P(O_2|O_1)P(O_3|O_2, O_1) \cdots\cdots\cdots\cdots P(O_m|O_{m-1}, O_{m-2}, \cdots O_1) \tag{3.4}$$

The probability distribution given by equation (3.4), is computationally inefficient in terms of observing them in probability distribution table. Equation (3.5) uses the concept of marginalization where a new variable R known as margin variable, is linked with observation symbols parameter O by acquiring all possible outcomes of Markov process and perform summation on them.

$$P(O_m) = \sum_r P(O_m, S_m) \quad (3.5)$$

$$P(O_m) = \sum_r P(S_m) P(O_m|S_m) \quad (3.6)$$

3.3. Fundamental Probability Rules for Markov Models

Bayes Chain Rule combined with marginalization property provides the computational ability of estimating Maximum Likelihood. The guiding probability rules for efficient modelling of Markov processes are discussed in next sections.

3.3.1 Bayes Chain Rule

By applying Bayes rule on long input chains, we can easily decompose them into computationally feasible scalar products.

$$P(T_1, T_2 \cdots \cdots T_m) = P(T_1)P(T_2|T_1)P(T_3|T_2, T_1) \cdots \cdots P(T_m|T_{m-1}, T_{m-2}, \cdots T_1) \quad (3.7)$$

3.3.2 Marginalization Property

The concept of marginalization provides the ability to represent any event 'T' as summation of all sub events of Markov random process as shown in equation (3.8).

$$P(T) = \sum_{m} P(T, X_1, X_2, \cdots \cdots X_m) \quad (3.8)$$

3.4 Fundamental Properties of Hidden Markov Models

Hidden Markov Model (HMM) is developed on basis of two characteristic properties of Markovian processes as discussed in next sections.

3.4.1 Limited Horizon Property of Hidden Markov Model

This property dictates that a current state ***n*** is statistically independent of transitioning from state 1 to $(n - m + 1)$, given past state ***z***.

$$P(T_z = n | T_{z-1}, T_{z-2} \cdots T_1) = P(T_z = n | T_{z-1}, T_{z-2} \cdots T_{z-m}) \quad (3.9)$$

Equation (3.9) is conditioned on fact that beyond *M* states which comes before z^{th} state, every other state can be ignored. This rule is known as Limited Horizon or window property of m order Markovian process.

3.4.2 Time Invariance Property of Hidden Markov Model

This property dictates that statistical dependence of current state on last state can be analysed over complete sequence for Markov processes. It is simplified as; conditional probability of current state is time invariant and it does not vary from position to position for given sequence T.

$$P(T_z = n|T_{z-1} = x) = P(T_2 = n|T_1 = x) = P(T_m = n|T_{m-1} = x) \qquad (3.10)$$

3.5 Implementation of Hidden Markov Model

Hidden Markov Model (HMM) faces three fundamental problems during its implementation. In this section, effective algorithms have been devised to counter these problems.

3.5.1 Likelihood Sequence Problem

Assume a Hidden Markov Model (HMM) is defined by equation (3.1), observation sequence is given by O, then it is required to find likelihood of observed sequence $P(O|M)$ of given model.

3.5.2 Determination of State Sequence Probability

For given HMM denoted by M, it is necessary to determine highest probability state sequence conditioned on observation sequence, that is to find unobserved states hiding in HMM.

3.5.3 Re-estimation Problem of Hidden Markov Model

It identifies the problem of computing state transition probabilities conditioned on output observation sequence probability for given HMM denoted by M. HMM wants to fit the observational data perfectly by re-estimating model parameters.

3.6 Solution to Implementation Problems of Hidden Markov Model

The implementation problems of Hidden Markov Model can be solved by using techniques mentioned in next sections.

3.6.1 Forward Probability Algorithm for Implementation Problem 1

Forward probability is defined as probability of being present in current state S_i after transitioning through states $O_1, O_2, O_3 \cdots O_m$ shown in equation (3.11).

$$F(m,i) = P(O_1, O_2, O_3 \cdots \cdots O_m \cdots, S_i) \tag{3.11}$$

$$P(O) = P(O_1, O_2, O_3 \cdots O_m) \tag{3.12}$$

By using concept of marginalization, observed sequence probability $P(O)$ for L number of states is computed in equation (3.12).

$$P(O) = \sum_{m=1}^{M} P\big(O_1, O_2, O_3 \cdots \cdots O_m \cdots, S_p\big) \tag{3.13}$$

Setting S_p to be final state that is achieved after transitioning through all the observation sequences, summation of final forward probability of observed sequence and corresponding state will give probability $P(O)$ by using rule of marginalization.

$$P(O) = \sum_{p=1}^{L} F(m,p) \tag{3.14}$$

In order to calculate forward probability $F(m,p)$ of complete observed sequence, a concise Markov model is developed in figure 3.1 using state transitions and respective observation probabilities. It starts from state S_1, transitions to next state S_2 that can be one of the unobserved states. It can transition to any state such that $S_p \rightarrow S_q$ observed on sequence O_m.

The state transition probability of being in state S_1 is defined by;

$$F(m,i) = P(O_1, O_2, O_3 \cdots\cdots O_m, S_i) \tag{3.15}$$

By using the concept of marginalization, probability of observed sequence $P(O)$ is found by;

$$P(O_1, O_2, O_3 \cdots\cdots O_m) = \sum_{i-1}^{L} F(m,i) \tag{3.16}$$

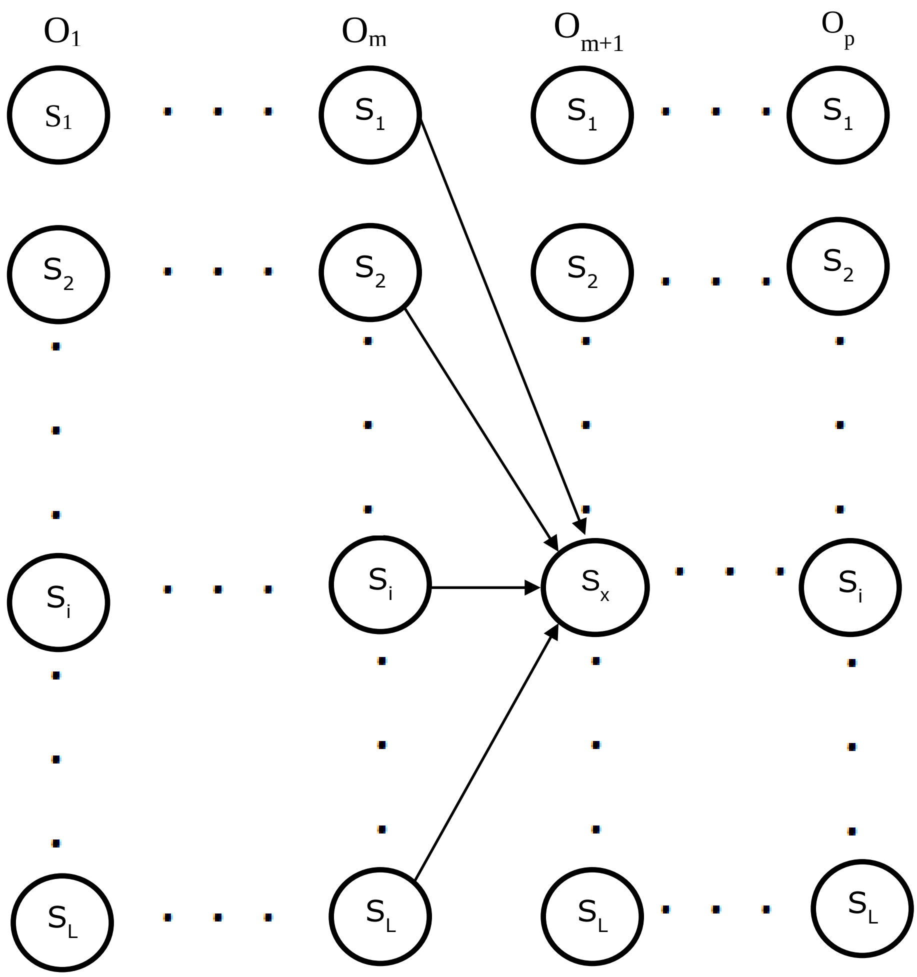


Figure 3-1 Forward Probability Algorithm

In order to perform transition from state S_p to state S_x, we use;

$$F(m,x) = P(O_1, O_2, O_3 \cdots\cdots O_m) \tag{3.17}$$

$$F(m,x) = P\left(O_1, O_2, O_3 \cdots\cdots O_m, S_q\right) \tag{3.18}$$

$$F(m,x) = P(O_1, O_2, O_3 \cdots\cdots O_{m-1}, O_m, S_x) \tag{3.19}$$

$$F(m,x) = \sum_{p=1}^{L} P\big(O_1, O_2, O_3 \cdots\cdots O_{m-1}, S_p, O_m, S_x\big) \qquad (3.20)$$

By using Bayes chain rule on equation (3.20), recursive form of forward probability function is finally computed.

$$F(m,x) = \sum_{p=1}^{L} P\big(O_1, O_{m-1}, S_p, O_m, S_x\big) \qquad (3.23)$$

$$F(m,x) = \sum_{p=1}^{L} P\big(O_1, O_{m-1}, S_p\big) P\big(O_m, S_x \big| O_{m-1}, S_p\big) \qquad (3.22)$$

$$F(m,x) = \sum_{p=1}^{L} F(m-1,x).\, P\big(O_m, S_x \big| S_p\big) \qquad (3.23)$$

$$F(m,x) = \sum_{p=1}^{L} F(m-1,x).\, P\left(S_p \overset{O_m}{\rightarrow} S_x\right) \qquad (3.24)$$

$$F(m,x) = \sum_{p=1}^{L} F(m-1,x) \qquad (3.25)$$

The generalized form for computing $F(m,x)$ is;

$$T_m = \sum_{m=1}^{L} T_{m-1} \qquad (3.26)$$

The boundary condition imposed on observed sequence for forward algorithm is given by;

$$F(0, x) = P(x) \tag{3.27}$$

Where $P(x)$ is forward probability of being in state S_x after transitioning from state S_p. Computation of forward probability is very easy because forward algorithm makes it linear time operation.

3.6.2 Backward Probability Algorithm for Implementation Problem 2

Backward probability algorithm is used to determine highest probability state sequence conditioned on observed sequence [25]. Backward probability is computed by observing sequences in backward order $O_m, O_{m+1}, O_{m+2} \cdots \cdots O_x$ based on initial state S_i, where t is length of observed symbol.

$$B(m, p) = P\left(O_m, O_{m+1}, O_{m+2} \cdots \cdots O_t \middle| S_p\right) \tag{3.28}$$

Again by using marginalization rule on observed sequence; that is introducing margin variable S_p, we will isolate observed symbol O_m to make this process a linear time computation as shown in following equations.

$$B(m, p) = P\left(O_{m+1}, O_{m+2} \cdots \cdots O_t, O_m, S_x \middle| S_p\right) \tag{3.29}$$

$$B(m,p) = \sum_{x=1}^{L} P\big(O_m, S_x \big| S_p\big). P\big(O_{m+1}, O_{m+2} \cdots \cdots O_t | O_m, S_x \big| S_p\big) \tag{3.30}$$

$$B(m,p) = \sum_{x=1}^{L} P\big(O_m \cdots, S_x \big| S_p\big). P(O_{m+1}, O_{m+2} \cdots \cdots, O_t | S_x) \tag{3.31}$$

$$B(m,p) = \sum_{x=1}^{L} B\ (m+1, x). P\big(S_p \to S_x\big) \tag{3.32}$$

For a given observed sequence and respective state, if an end location is set at m^{th} place, then forward probability can be determined up to any position in symbol sequence and backward probability from that location to the end of observation sequence [29]. The HMM is said to be in final state after it observed the last symbol in observation sequence stream. The boundary condition of backward algorithm is imposed on transition from S_i to S_{final} over observed sequence O_m as $\left(S_i \xrightarrow{O_m} S_{final}\right)$.

CHAPTER 4

IMAGE DENOISING USING HIDDEN MARKOV MODEL

4.1 Introduction of Proposed Denoising Framework

In this research work, a denoising algorithm for medical images is developed in wavelet domain using Hidden Markov Model. The proposed technique uses Discrete Wavelet Transform (DWT) for image decomposition and takes advantage of its hierarchical relationships between different subbands [22]. The non-Gaussian statistics of wavelet coefficients are modeled using Probabilistic Graphical Models. Multidimensional Gaussian Mixture Models (GMM) known as Hidden Markov Tree (HMT) model [26], are used to determine inter-scale dependencies among wavelet coefficients. Proposed framework models the wavelet coefficients using Probabilistic Graphical Models. HMT model combined with Expectation Algorithm (EM) is used for image denoising. EM is an iterative algorithm that converges HMT model parameters vector [27].

In this chapter concepts of Probabilistic Graphical Models, Gaussian Mixture Models and Expectation Maximization Algorithm for Hidden Markov Tree (HMT) model are discussed in detail. Block Diagram of proposed framework is shown in figure 4.1. Discrete Wavelet Transform (DWT) is used for decomposition of given medical image. DWT decomposition is explained extensively in chapter 2. Probabilistic Graphical Models are used for modelling of wavelet coefficients of decomposed medical images. EM algorithm is for adjusting observational data according to Hidden Markov

Model parameters. The proposed framework can also be used for compression and detection of medical images.

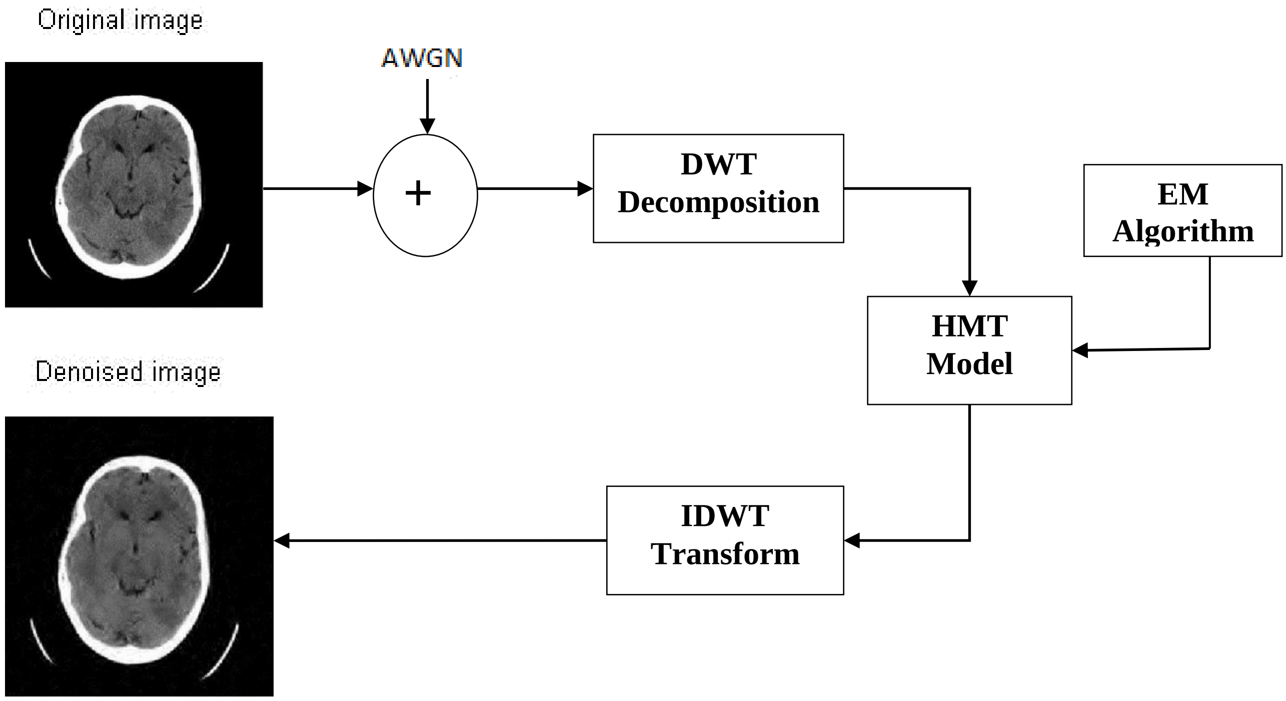


Figure 4-1 Proposed Wavelet Based Denoising Scheme for CT scan of brain using Hidden Markov Model

4.2 Image Modelling Using Wavelet Transform

Images are modelled by using Probabilistic Graphical Models based on assumption that images and their wavelet coefficients are stochastic in nature. Existing wavelet based denoising techniques considers wavelet coefficients as statistically independent or jointly Gaussian [28]. The proposed denoising framework assumes wavelet coefficients to be non-Gaussian in nature. The goal of framework is to construct wavelet based probabilistic model that captures inter-scale statistical dependencies and non-Gaussian statistics of wavelet coefficients. The idea behind this assumption of non-Gaussian nature lies in primary and secondary properties of wavelet transform that are discussed in next section.

4.2.1 Wavelet Transform Properties for Image Modelling

The primary properties of wavelet transform are used in many applications that are stochastic in nature including image estimation, detection and classification using different scales [29]. The primary and secondary properties of DWT are defined as;

Locality: Wavelet coefficients are localized in time and frequency at same time that allows them to adjust wide range of various components of given wavelet.

Multiresolution: This property generates a set of wavelet scaling functions for statistical signal processing by allowing wavelet transform to capture long and short duration signal components.

Compression: The sparsity of wavelet transform of medical images is very large due to formation of unconditional bases for wavelet scaling functions [22].

The secondary properties of wavelet transform are used to determine non-Gaussian statistics of wavelet coefficients.

Clustering: It dictates that inter-scale dependencies among wavelet coefficients can be computed by adjusting magnitudes of certain wavelet with respective to magnitudes of adjacent wavelets.

Persistence: It states that dependencies of wavelet coefficients can be propagated across scales vertically.

4.3 Gaussian Mixture Models for Image Modelling

The wavelet coefficients of real-world medical images are characterized by peaky marginal densities at zeroes and heavy-tail non-Gaussian densities [30]. Gaussian Mixture Model (GMM) is used to model distributions of these wavelet coefficients by using large number of mixture components. GMM increments complexity of model in order to improve flexibility of image modelling. For medical images, probability density function (pdf) of each wavelet coefficient w_i is approximated by Gaussian Mixture Model (GMM). To each wavelet coefficient w_i, we associate a discrete set of Hidden state S_i with it, that takes on values $m = S, L$ with probability mass function $P(S_i = m)$. Conditioned on $S_i = m$, w_i is Gaussian with mean $\mu_{i,m}$ and variance $\sigma_{i,m}^2$. The overall probability density function is given by;

$$f(w) = \sum_{m=1}^{M} P(S_i = m)\, f(w_i \,|S_i = m) \qquad (4.1)$$

The $f(w_i \,|S_i = m)$ is conditional pmf given by;

$$f(w_i|S_i = m) = \frac{1}{\sigma_{i,m}\sqrt{2\pi}} \exp\left(-\frac{\left(b-\mu_{i,m}\right)^2}{2\sigma_{i,m}^2}\right) \qquad (4.2)$$

$$P(S_i = S) + P(S_i = L) = 1 \qquad (4.3)$$

Due to stochastic nature of hidden state variable S_i, wavelet coefficient becomes Gaussian $S_i = m$. The dependencies of wavelet coefficients are usually captured by three models namely inter-scale, intra-scale and hybrid of inter and intra-scale statistical models [31]. These models work on the assumption of independence of wavelet coefficients. The simplest GMM that is used for modelling of sub-events from an overall event without needing to know that given observed sequence belongs to which sub-event, is known as Independent Mixture Model (IM) [32].

4.4 Hidden Markov Tree Model

Hidden Markov Tree (HMT) Model is the name given to a multidimensional Gaussian Mixture Model (GMM). HMT satisfies clustering property of wavelet transform by determining the non-Gaussian statistics of wavelet coefficients. HMT links magnitude of each wavelet coefficient $|w_i|$ with an unobserved (latent) state variable S_i. The compression property of wavelet transform dictates that a few number of wavelet coefficients with large magnitudes hold maximum information about medical image. Whereas smaller magnitude wavelet coefficients hold little amount of diagnostically relevant details about medical image but they are present in large numbers. Above notion guides us to the development of simple statistical model with only two states. One state is 'high' which represents large magnitudes and other state is 'low' which denotes smaller magnitudes of wavelet coefficients. This model is very useful for estimation purposes due to its computational simplicity.

The persistence property of wavelet coefficients dictates that inter-scale dependencies among wavelet coefficients can be computed when Gaussian Mixture Model (GMM) applies Markov Chain mechanism across scales of wavelet coefficients. The wavelet coefficients scales are considered tree structured because magnitude of wavelet coefficients is only dependent on respective magnitude of their parents only. This notion summarizes that only magnitude of parent coefficient is sufficient for determination of probability of child wavelet coefficients to be considered as ‘high’ or ‘low’. Wavelet-based HMT is developed in [32] by linking vertical connections of hidden states.

To summarize Hidden Markov Tree (HMT) model, it captures the non-Gaussian statistics of wavelet coefficients by establishing links between the hidden state of each wavelet coefficient and its four children known as inter-scale dependencies. Using links to identify dependencies, HMT model takes the form of quad-tree structure as shown in figure 4.2 where each white node represents a hidden state whereas adjacent black node represents complex wavelet coefficient associated with hidden state. Each hidden state variable is considered as parent that make links to its four child wavelet coefficients. The overall HMT model consists of an HMT for each of the three Discrete Wavelet Transform (DWT) subbands.

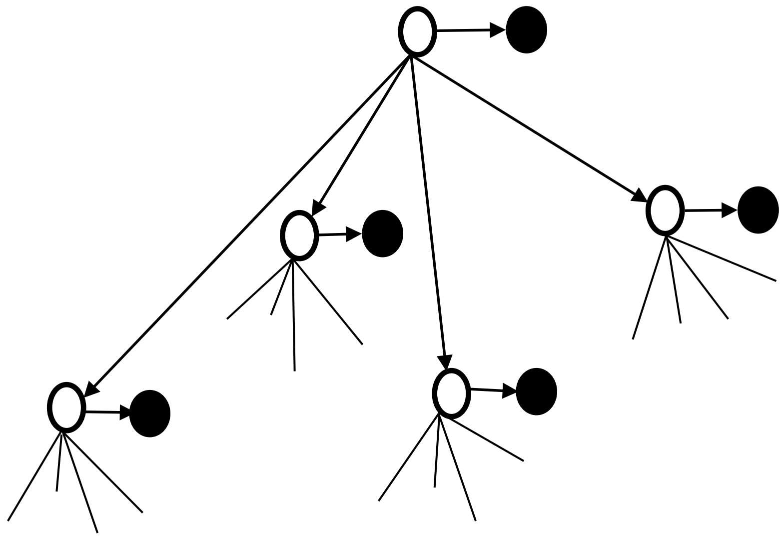


Figure 4-2 Wavelet based HMT model

Gaussian Mixture Model (GMM) is used to capture the non-Gaussian densities of wavelet coefficients. HMT is a multidimensional Gaussian Mixture Model (GMM). HMT models the wavelet coefficients as random variables having probability density function (pdf) as a mixture of zero mean Gaussian distributions by using a hidden state to denote large and small wavelet coefficient magnitudes [26]. HMT uses Probabilistic Graphical Model that finds Markovian dependencies between hidden states of two neighbouring scales [27]. Thus, due to persistence across scales, state transition probability matrix X_t denotes parent-child $state \rightarrow to \rightarrow state$ links between hidden states as;

$$X_t = \begin{bmatrix} p_t^{a\rightarrow a} & p_t^{a\rightarrow b} \\ p_t^{b\rightarrow a} & p_t^{b\rightarrow b} \end{bmatrix} \tag{4.4}$$

Where $p_t^{a \to a'}$ indicates given a' being the hidden state of parent coefficient, the child coefficient is in hidden state a. HMT model is characterized by parameters such as $P(S_i = m)$ pmf of root node S_1, X_t state transition probability matrix, mean $\mu_{i,m}$ and variance $\sigma_{i,m}^2$ of wavelet coefficient w_i conditioned on $S_i = m$. All these parameters are grouped together in the form of vector θ. However, each wavelet coefficient has different variances and state transition probabilities which leads to greater complexity in HMT model. This computational complexity can be reduced by a method known as tying within scale [15]. HMT is used for image analysis in image processing applications in [28].

The proposed Hidden Markov Tree (HMT) model combined with EM algorithm performs medical image denoising by satisfying secondary properties of wavelet transform. Expectation Maximization algorithm is explained in detail in next section.

4.6 Expectation Maximization Algorithm for Hidden Markov Tree Model

Expectation Maximization algorithm is an iterative scheme defined for finding maximum likelihood estimation of parameters associated with certain distribution that has missing data (unobserved states). Hidden Markov Model (HMM) is an ideal candidate for utilization of EM algorithm because it is characterized by missing variables known as latent or hidden variables. EM algorithm finds its usage in exploiting graphical structures associated with different type of distributions [33]. Usually, all the missing data is defined in context of probability distribution and it follows exponential progression, so EM is best candidate for handling such scenarios.

4.5 Implementation of Expectation Maximization Algorithm

Expectation Maximization (EM) algorithm has two functions. One is to perform estimation on data distribution that has some incomplete values due to observational restraints. Second function is to perform optimization of maximum likelihood estimation for given data distribution that have unobserved states (hidden variables) as evident in table 4-1.

EM Algorithm
1) *Initialize HMT parameters vector* θ_l . *Set iteration counter* $l = 0$ 2) *E-step: Calculate joint probability mass function* $P(S\|w,\theta_l)$ *for estimation* $E_S[\ln f(w,S\|\theta)\ \|w,\theta_l]$ 3) *M-step: Find maximum value by setting* $\theta_{l+1} = argmax_\theta . E_S[\ln f(w,S\|\theta)\ \|w,\theta_l]$ 4) *Set* $l = l + 1$ *and perform above steps until* θ_l *converges.*

Table 4-1 Expectation Maximization Algorithm Implementation

EM technique switches between E and M steps. Expectation (E) step is used for estimating log-likelihood function by generating expectation function. Maximization (M) step re-estimates HMT parameters by maximizing computed log-likelihood expectation function in E-step. This iterative process continues until parameters of Hidden Markov Tree (HMT) model are converged as shown in figure 4-3. The properties of EM algorithm for Hidden Markov Model are discussed in next section.

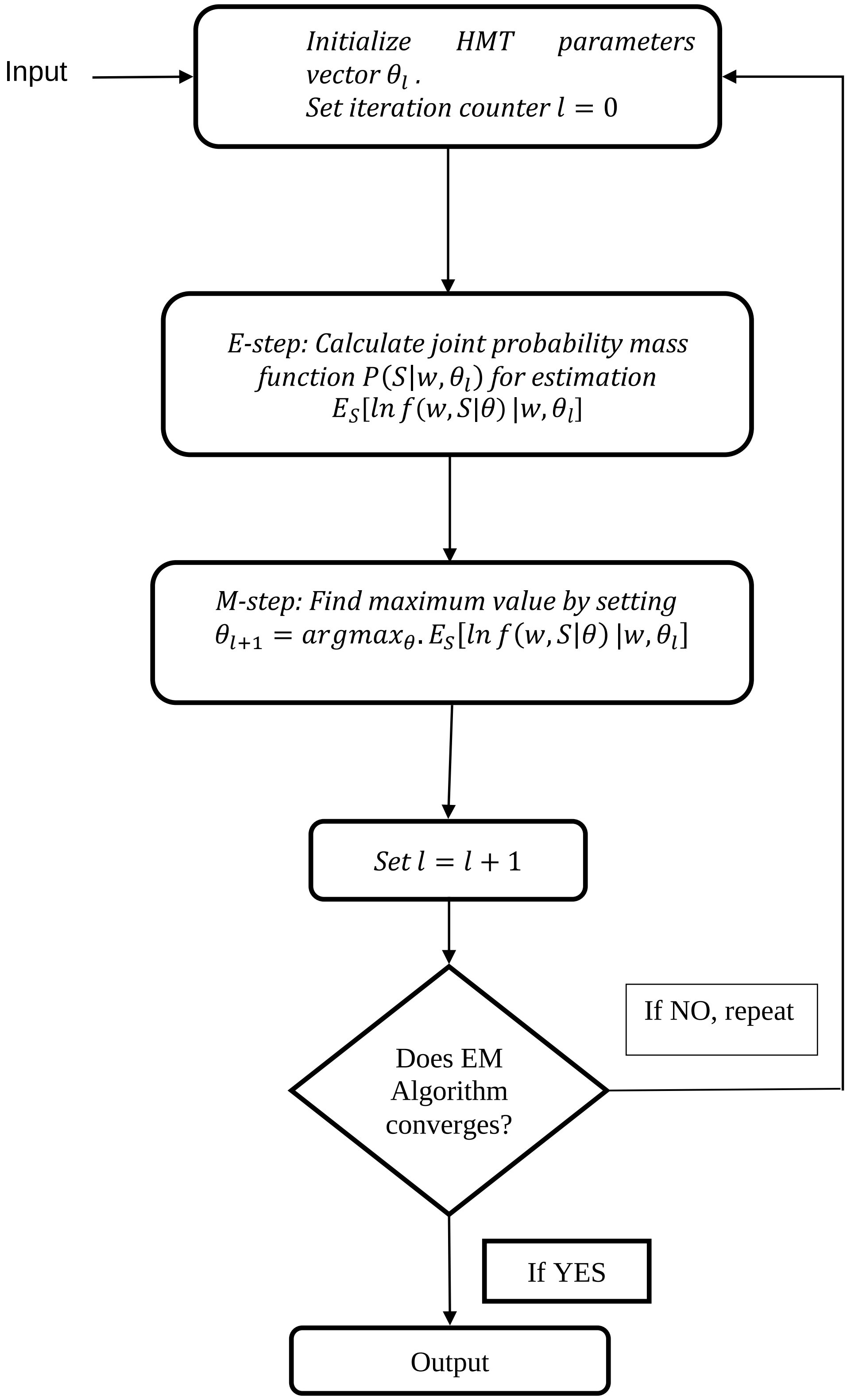


Figure 4-3 Flowchart of EM Algorithm for Hidden Markov Tree Model

4.6 Properties of Expectation Maximization Algorithm

Expectation Maximization (EM) algorithm has many statistically important properties that makes it attractive for Gaussian Mixture Models. EM algorithm has low computational complexity because it does not require to set a certain algorithm learning rate. This property makes EM algorithm useful to handle probabilistic graphical models and solve problems involving convergence. Sometimes, this algorithm is considered slightly slow relative to other optimization techniques but for large distributions involving mixture models, this drawback can be ignored. Newton's Gradient method [17] for convergence purposes perform extremely poor for statistically dependent mixture components. Thus, EM algorithm works very well under worst case scenarios and provide certain safety net.

Expectation Maximization algorithm is at its best when maximum likelihood function belongs to exponential family. E-step computes summation of all expected values of wavelet coefficient's distribution statistics, whereas M-step maximizes expected log-likelihood function. EM scheme usually computes maximum a posteriori (MAP) estimates to perform iterative optimization.

4.7 Expectation Maximization utilizing Gaussian Mixture Models

Expectation Maximization algorithm for Gaussian Mixture Models (GMM) can be formulated by assigning Hidden Markov Tree (HMT) model parameters vector θ, the statistics of wavelet transform such as vector α, means μ_m and co-variances v_m. Assume $P(a,b)$ is an exponential family and its relationship with wavelet statistics given by equation (4.6).

$$\theta = (\alpha, \{\mu_m\}, v_m) \tag{4.5}$$

$$P(a,b) = \prod_{m=1}^{k} \alpha_m N(a|\mu_m, v_m)^{z^m} \tag{4.6}$$

4.9.1 Expectation Step of Expectation Maximization Algorithm

Function $T(\theta, \theta_m)$ is used for estimation of log-likelihood function. To achieve simplicity, estimation is performed by finding expectation of θ_{m-1} to solve for θ_m with respect to conditional probability distribution of state variable A given B.

$$T(\theta, \theta_m) = E_{A|B,\theta_m}[\ln f(\theta; A, B)] \tag{4.7}$$

For a sequence of random variables $A = (a_1 a_2, \cdots \cdots, a_n)$ and hidden state variables $B = (b_1 b_2, \cdots \cdots, b_n)$, then relevant statistics of wavelet coefficients are found by;

$$P(a_1 a_2, \cdots \cdots, a_n, b_1 b_2, \cdots \cdots, b_n) = \prod_{i=1}^{n} P_\theta(a_i b_i) \tag{4.8}$$

4.9.2 Maximization Step of Expectation Maximization Algorithm

In M-step, optimization is achieved by finding θ_l that maximizes quantity given by;

$$\theta_{l+1} = arg_{\theta} max\, T(\theta, \theta_l) \tag{4.9}$$

The EM algorithm will go back to E-step, if θ_l does not converge.

4.9.3 Mathematical Analysis of Expectation Maximization Algorithm

Mathematically, the goal of EM algorithm is to maximize the incomplete log-likelihood function $\ln f(w|\theta)$ where w is incomplete training data. The statistics S_t of HMT model that are sufficient to determine the expected value for random variable U and hidden state variable H are defined by following relation;

$$E_{\theta_o}(S_t(U,H)|U = u)) = E_{\theta} S_t(U,H) \tag{4.10}$$

In M-step, the maximization of conditional probability mass function of hidden state S_i is given by;

$$p(S = m|u, \theta') = \frac{P(S_o = m)\mathrm{g}\big(\mathrm{u}; 0, \sigma_{i,m}^2\big)}{\sum_{l=0}^{1} p(S_o = l)\, \mathrm{g}\big(\mathrm{u}; 0, \sigma_{i,l}^2\big)} \tag{4.11}$$

The noise free medical image u_i is obtained by equation;

$$u_i = E[u|u',\theta] = \sum_{n} P[(S_0 = m)|u',\theta] \frac{\sigma_{i,m}^2}{\sigma_n^2 + \sigma_{i,m}^2} u' \qquad (4.12)$$

Thus, E-step and M-steps are iterated till convergence of Hidden Markov Tree (HMT) model parameters vector θ_l is achieved.

CHAPTER 5

EXPERIMENTAL RESULTS

In this research work, wavelet based denoising of medical images using Hidden Markov Model is carried out on real medical images taken from [34]. The performance of denoising framework is evaluated by using different type of wavelet families in terms of peak signal to noise ratio (PSNR) and visual quality of image.

This chapter consists of simulation results of proposed denoising framework tested on various medical images using Matlab version R2012b. Steps involving in implementation of developed denoising framework are discussed in detail in next section.

5.1 Implementation Steps of Proposed Framework

The denoising of medical images is achieved by following steps;

1. Read the medical image (256x256) by Matlab.
2. Corrupt the medical images by adding Gaussian noise (AWGN) of known variances $\sigma = 10, 20, 30$ to them.
3. Perform decomposition of medical images using different types of wavelet families (set decomposition up to 3 levels).
4. Initialize HMT model parameters vector θ for every subband.
5. Estimate HMT model parameters vector θ by using Expectation Maximization (EM) algorithm. E-step will compute expectation of log-

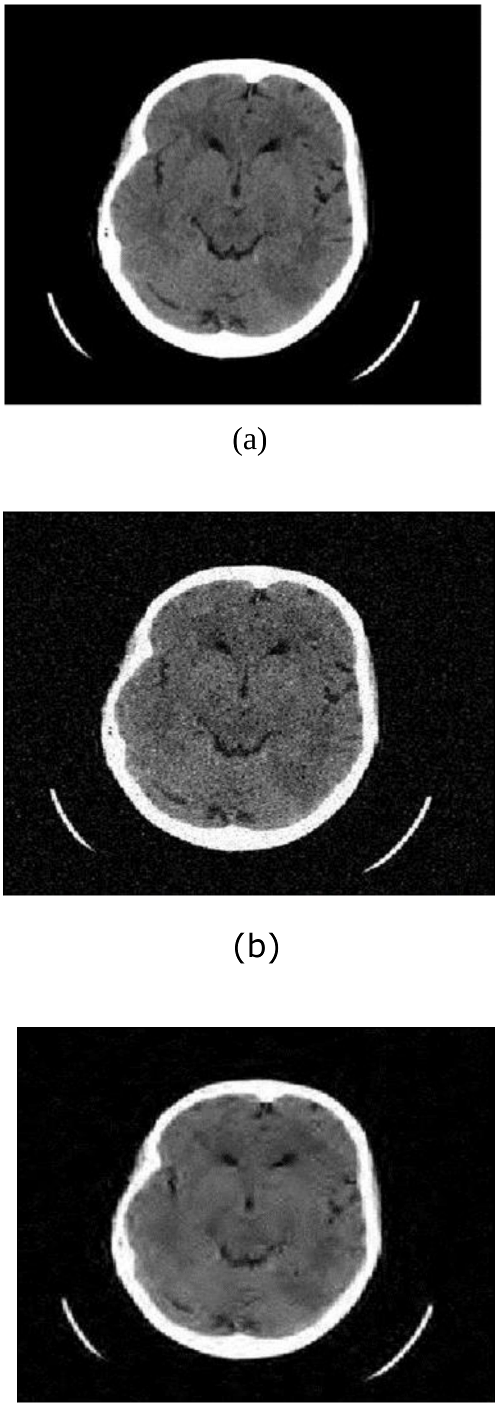
likelihood function using initial HMT model parameters vector. M-step maximizes estimated log-likelihood function.

6. Check for convergence of HMT model parameters vector θ.
7. Apply IDWT to get denoised medical image.

5.2 Objective Analysis of Proposed Denoising Framework

The proposed novel denoising scheme is applied on various medical images including MRI (coronal, sagittal views) and CT scan of human brain. Each medical image is 2D grayscale having dimensions (256x256). We have corrupted each medical image with Gaussian noise of known variance ($\sigma_n = 10,20,30$ db). The noisy medical image is decomposed into 3 levels by using variety of wavelet families (Haar, Daubechies, Symlets, Bior, Rbio, Coiflet etc). Twenty wavelet packages are applied with proposed denoising algorithm to perform comparison of our technique at different wavelet types and find out which wavelet family works better with our novel denoising framework in terms of PSNR (in db) and visual quality of image as presented in table 5-1.

The compression of medical images is also performed by using the mechanism developed in [14].

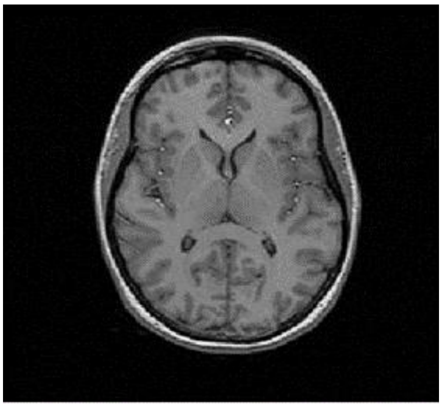


(a)

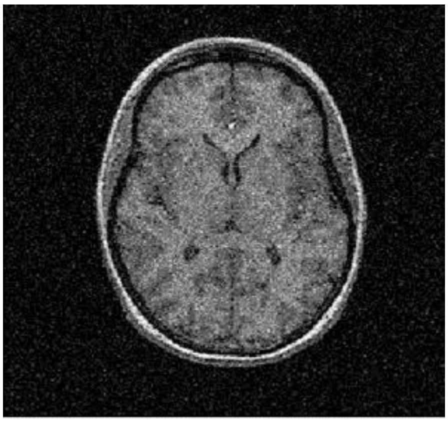
(b)

(c)

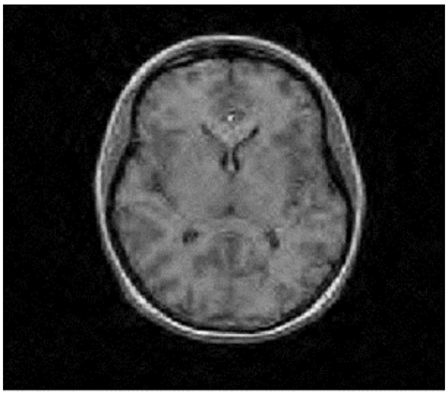
Figure 5-1 Denoising of CT scan of brain by proposed model (a) Original CT scan (b) Noise corrupted CT scan (c) Denoised CT scan



(a)



(b)



(c)

Figure 5-2 Denoising of MRI scan of brain by proposed model (a) Original MRI scan (b) Noise corrupted MRI scan (c) Denoised MRI scan

Image	Wavelet Type (3 Levels)	Variance σ_n					
		10		20		30	
		PSNR	Compression Ratio	PSNR	Compression Ratio	PSNR	Compressio n Ratio
MRI of Brain	Haar	32.07	65.52%	31.02	66.24%	30.22	67.23%
	Sym2	33.52	63.26%	32.76	65.03%	31.42	66.02%
	Sym3	32.98	61.78%	32.42	63.21%	31.04	64.43%
	Sym4	34.14	61.92%	33.23	62.79%	32.08	64.12%
	Sym6	32.13	62.02%	31.82	62.89%	30.89	63.34%
	Sym8	31.97	61.87%	31.49	62.76%	30.56	63.25%
	Db1	32.15	65.50%	31.06	66.26%	30.19	67.22%
	Db2	33.12	63.19%	32.45	65.03%	31.12	66.04%
	Db3	32.58	61.68%	31.95	63.13%	30.91	64.52%
	Db4	32.35	60.08%	31.26	61.23%	30.95	62.42%
	Db6	31.63	61.91%	31.28	62.73%	30.14	63.28%
	Db8	32.01	62.82%	31.49	63.56%	30.25	64.47%
	Coif1	32.19	62.44%	31.78	63.28%	30.87	64.32%
	Coif2	32.62	61.76%	31.79	62.64%	30.47	63.18%
	Coif3	32.13	60.15%	31.01	61.38%	30.16	62.54%
	Coif4	32.69	60.37%	31.78	61.59%	31.32	62.76%
	Bior1	32.07	65.52%	31.02	66.24%	30.22	67.23%
	Bior2	32.65	66.68%	31.77	67.41%	31.46	68.34%
	Bior3	33.95	66.46%	32.92	67.27%	31.96	68.08%
	Bior4	37.05	59.95%	35.91	60.17%	34.44	61.34%
	Rbio1	31.01	72.31%	30.68	73.18%	29.34	74.25%
	Rbio2	30.01	76.22%	28.93	77.13%	27.63	78.06%
	Dmey	32.12	61.87%	31.04	62.75%	30.78	64.08%

Table 5-1 Performance Measures of Proposed Denoising Scheme for MRI scan using different Wavelet Families

Image	Wavelet Type (3 Levels)	Variance σ_n					
		10		20		30	
		PSNR	Compression Ratio	PSNR	Compression Ratio	PSNR	Compression Ratio
CT Scan of Brain	Haar	30.08	67.52%	29.02	68.24%	28.22	69.13%
	Sym2	31.82	66.36%	30.76	67.03%	29.42	68.12%
	Sym3	32.03	63.98%	30.42	65.21%	28.04	66.43%
	Sym4	32.54	63.92%	31.23	64.89%	28.08	66.22%
	Sym6	30.37	64.01%	29.82	64.96%	28.79	65.34%
	Sym8	30.87	62.67%	29.51	64.76%	28.56	65.25%
	Db1	30.25	67.54%	29.06	68.26%	28.19	69.24%
	Db2	31.12	65.69%	30.45	67.13%	29.12	68.04%
	Db3	30.54	63.48%	29.85	65.13%	28.91	66.52%
	Db4	30.35	62.04%	29.26	63.65%	28.82	64.42%
	Db6	30.53	63.93%	29.38	64.78%	28.24	65.36%
	Db8	30.03	64.80%	29.37	65.51%	28.02	66.47%
	Coif1	30.29	64.34%	29.88	65.28%	28.76	66.35%
	Coif2	30.52	63.82%	29.79	64.54%	28.57	65.18%
	Coif3	30.43	61.15%	29.01	63.38%	38.26	64.44%
	Coif4	31.01	62.49%	30.08	63.57%	29.32	64.68%
	Bior1	30.17	67.53%	29.04	68.14%	28.23	69.21%
	Bior2	30.68	68.62%	29.79	69.44%	29.16	70.14%
	Bior3	31.92	68.56%	30.93	69.17%	29.96	70.01%
	Bior4	35.15	61.91%	33.83	62.15%	32.44	63.34%
	Rbio1	29.03	74.32%	28.78	75.28%	27.14	76.25%
	Rbio2	28.02	78.15%	26.85	79.27%	25.68	80.12%
	Dmey	30.12	63.77%	29.14	64.72%	28.78	66.16%

Table 5-2 Performance Measures of Proposed Denoising Scheme for MRI scan using different Wavelet Families

The calculated PSNR values of MRI scan of brain using different wavelet families as shown in table 5.1, indicates that 'Bior4' gives best result both in terms of PSNR and visual quality. On the other hand, 'Rbio2' gives lowest PSNR value and image is more distorted. These results conclude that when we are required best possible denoised image through our proposed wavelet based denoising algorithm, then it is best to decompose medical image using 'Bior4' wavelet family.

The preservation of diagnostic details in medical image after denoising is very critical. Thus, performance of our novel denoising framework is evaluated by using different wavelet packages to achieve best denoising of CT scan and MRI of brain as shown in table 5.2.

The graphical results of proposed denoising scheme are shown in figure 5.1 and 5.2.

The compression of CT scan and MRI image of brain is performed by employing wavelet based compression technique using Hidden Markov Model proposed in [14]. Results of compression is shown in figure 5-3.

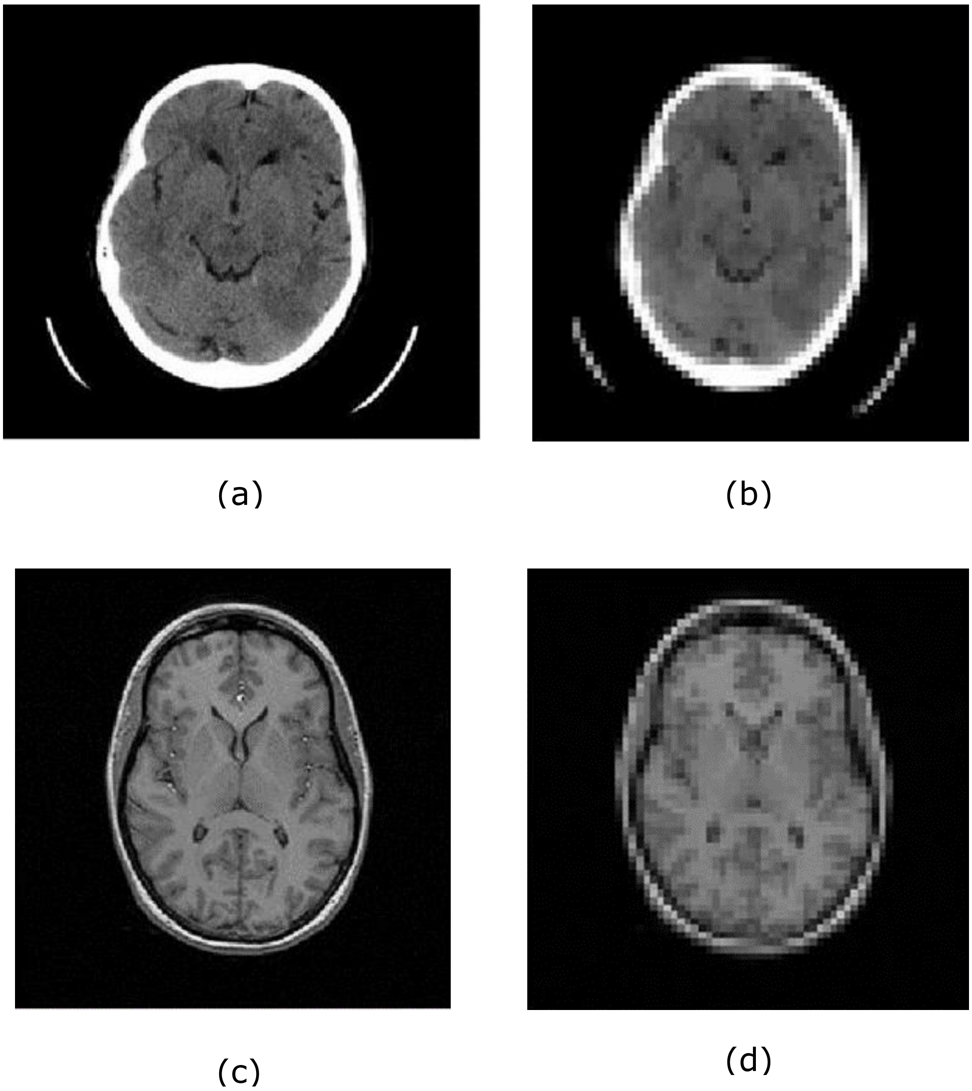


Figure 5-3 Results of compression performed on CT scan and MRI scan of brain by proposed model (a) Original CT scan (b) Compressed CT scan (c) Original MRI scan (d) Compressed MRI scan

5.3 Subjective Analysis of Proposed Denoising Framework

The method of mean opinion score (MOS) is used for subjective analysis to determine image quality. This technique refers to the averaged value of the opinions taken from the users (doctors). This method scores the quality of the image from 1 (worse) to 5 (excellent). Table 5.3 shows different classes of MOS ranging from 1 to 5. We have performed MOS analysis on CT scan of brain that is tested through our denoising framework using different wavelet families in order to determine whether diagnostic details embedded in medical image are preserved after denoising. Figure 5.4, shows the results taken from a group of doctors and then averaged over the total number of observations. The graph shows that visual quality of CT scan using our proposed denoising method is diagnostically comparable to that of original CT scan for wavelet family ‘Bior4’.

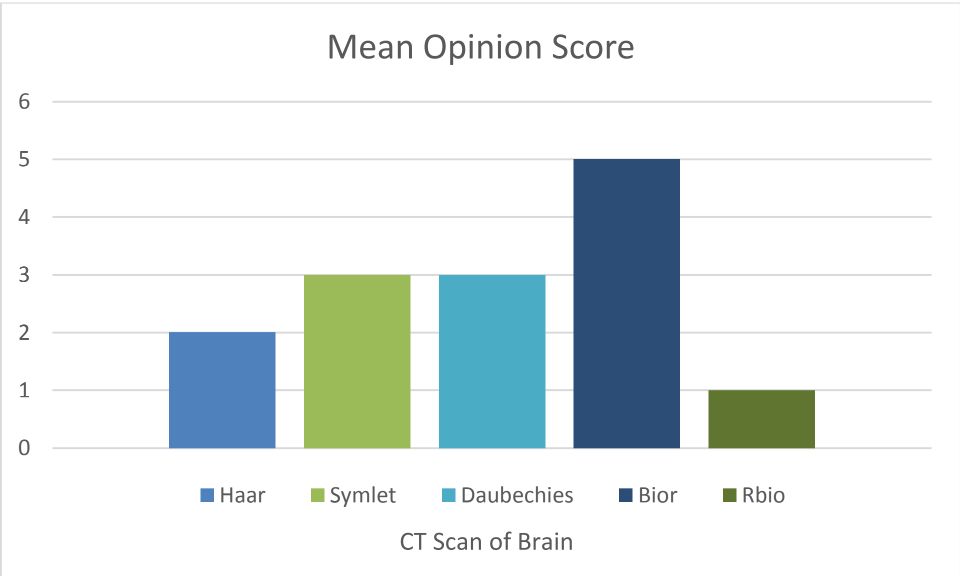


Figure 5-4 Mean Opinion Score for denoised CT scan using different wavelet families

CHAPTER 6

CONCLUSION AND FUTURE WORK

6.1 Conclusion

In this work, wavelet based HMM is developed that empolys Gaussian Mixture Models (GMM) and 2D DWT. The proposed technique uses DWT for image decomposition and takes advantage of its hierarchical relationships between different subbands. The non-Gaussian statistics of wavelet coefficients are modeled using Probabilistic Graphical Models. Multidimensional GMMs known as Hidden Markov Tree (HMT) model, are used to determine inter-scale dependencies among wavelet coefficients. Proposed framework models the wavelet coefficients using Probabilistic Graphical Models. HMT model combined with EM is used for image denoising. EM is an iterative algorithm that converges HMT model parameters vector. This denoising scheme is applied on MRI and CT scans and their performance is compared by using different wavelet families. Results of proposed technique are shown in terms of PSNR and image quality.

6.2 Future Work

The proposed denoising scheme can be extended to multidimensional Hidden Markov Tree (HMT) models due to their quad-tree structures among wavelet coefficients. HMT can be used for modelling of 3D medical images.

Hidden Markov Model developed in this framework can also be used for analysis of transforms other than wavelet using higher dimensions.

The proposed wavelet based framework using HMM can also be used for video denoising. This framework can also handle noise distributions other than Gaussian such as Poisson noise occurred in X-rays. Thus, this model can be modified for denoising of Poisson corrupted X-rays and MRI scans.

REFERENCES

[1] M. Basseville et al., “Modeling and estimation of multiresolution stochastic processes,” IEEE Trans. Inform. Theory, vol. 38, pp. 766–784, Mar. 1992.

[2] A.K. Jain, R. P. W. Duin, and J. Mao, “Statistical pattern recognition: a review,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, pp. 4–37, 2000.

[3] G. Y. Chen and T. D. Bui, “Multi-wavelet Denoising using Neighboring Coefficients” IEEE Signal Process. Lett., vol. 10, pp. 211-214, 2003.

[4] D. Cho, T. D. Bui and G. Y. Chen, “Image denoising based on wavelet shrinkage using neighbor and level dependency” International Journal of Wavelets, Multiresolution and Information Processing, vol. 7, no. 3, pp. 299-311, 2009.

[5] T. Jagadesh and R. J. Rani, "A novel speckle noise reduction in biomedical images using PCA and wavelet transform," *IEEE International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)*, pp. 1335-1340, 2016.

[6] D. Sale and S. Sawant, "Wavelet Based selection for fusion of Medical images," *2016 International Conference on Computing Communication Control and automation (ICCUBEA)*, pp. 1-6, 2016.

[7] Pizurica, W. Philips, I. Lemahieu and M. Acheroy, "A versatile wavelet domain noise filtration technique for medical imaging," in *IEEE Transactions on Medical Imaging*, vol. 22, no. 3, pp. 323-331, 2003.

[8] Y. A. Youzbaki and S. Paşca, "Study of mixed filters schemes for denoising of the medical images," *IEEE 9th International Symposium on Advanced Topics in Electrical Engineering (ATEE)*, pp. 289-294, 2015.

[9] V. V. Kumar Raju and M. P. Kumar, "Denoising of MRI and X-Ray images using dual tree complex wavelet and Curvelet transforms,"

IEEE International Conference on Communication and Signal Processing, pp. 1844-1848, 2014.

[10] J.Zhang, D. Wang and Q. N. Tran, “A wavelet-based multiresolution statistical model for texture” *IEEE Trans. Image Processing*, vol. 7, pp. 1621–1627, Nov. 1998.

[11] L. Mitiche, A. B. Houda Adamou-Mitiche and H. Naimi, "Medical image denoising using dual tree complex thresholding wavelet transform," *IEEE Conference on Applied Electrical Engineering and Computing Technologies (AEECT)*, pp. 1-5, 2013.

[12] Crouse, M.S., Nowak, R.D., and Baraniuk, R.G., "Wavelet-based statistical signal processing using Hidden Markov Models", IEEE Trans. on Signal Processing, Volume 46, No.4 , pp. 886-902, 1998.

[13] J. Zhang, X. Zhang and Z. Pei, "Medical Image Denoising Using Hierarchical Hidden Markov Model in the Wavelet Domain," *IEEE First International Workshop on Education Technology and Computer Science*, pp. 857-860, 2009.

[14] Riaz, M.U., Touqir, I., and Haider, M.,“Wavelet-based image modelling for compression using Hidden Markov model”, International Journal of Advanced Computer Science and Applications, Volume. 7, No. 8, pp. 1-7, August 2016.

[15] Haider, M., Touqir, I., Riaz, M.Usman, and Siddiqui, A.M., “Denoising in Wavelet Domain Using Probabilistic Graphical Models” International Journal of Advanced Computer Science and Applications, Volume. 7, No. 11, 2016.

[16] Romberg, J., Choi, H., and Baraniuk, R.G., "Bayesian tree-structured image modeling using wavelet-domain Hidden Markov Models", IEEE Trans. On Image Processing, Volume 10, No. 7, pp.1056-1068, 2001.

[17] B. Deepa and M. G. Sumithra, "Comparative analysis of noise removal techniques in MRI brain images," *IEEE International Conference on Computational Intelligence and Computing Research (ICCIC)*, pp. 1-4, 2015.

[18] M. Diwakar, and M. Kumar, “CT image noise reduction based on adaptive wiener filtering with Wavelet packet thresholding” In IEEE

International Conference on Parallel, Distributed and Grid Computing, pp. 94-98, 2014.

[19] S. Mallat, *Geometrical Grouplets, Applications, Computations and Harmonics Analaysis*, 2009.

[20] M. K. Mihcak, I. Kozintsev, and K. Ramchandran, “Low-complexity image denoising based on statistical modeling of wavelet coefficients” *IEEE Signal Processing Lett.*, vol. 6, pp. 300–303, Dec. 1999.

[21] H. Choi and R. Baraniuk, “Multiscale image segmentation using wavelet-domain hidden Markov models” *IEEE Trans. Image Processing*, vol. 10, pp. 1309–1321, 2001.

[22] Gilbert Strang and Truong Nguyen, *Wavelets and filter Banks*, 4th ed., Wellesley Cambridge Press, 1996.

[23] R. D. Nowak, “Multiscale hidden Markov models for Bayesian image analysis” in *Bayesian Inference in Wavelet Based Models*, P. Müller and B. Vidakovic, Eds. New York: Springer Verlag, 1999, pp. 243–266.

[24] Guoliang Fan, and Xiang-Gen Xia,“Improved Hidden Markov Models in the Wavelet-Domain”, IEEE Transactions on Signal Processing, Vol. 49, No, 1, January, 2001.

[25] H. Chipman, E. Kolaczyk, and R. McCulloch, “Adaptive Bayesian wavelet shrinkage” *J.Amer. Stat. Assoc.*, vol. 440, no. 92, pp. 1413–1421, Dec. 1997.

[26] H. Choi, J. K. Romberg, R. G. Baraniuk, and N. G. Kingsbury, “Hidden Markov tree modeling of complex wavelet transforms” in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, 2000.

[27] J. K. Romberg, H. Choi, and R. G. Baraniuk, “Bayesian tree-structured image modeling using wavelet-domain hidden Markov models” *IEEE Trans. Image Processing*, vol. 10, pp. 1056–1068, 2001.

[28] Wenge, Z., Suang,W., Fang,L., Xinbo, G., and Licheng, J.,“Image Denoising Using Bandelets and Hidden Markov Tree Models”, Chinese Journal of Electronics, Volume 19, No.4, 2010.

[29] Donoho, D. L., and Johnstone, 1. M., "Ideal Spatial Adaptation by wavelet shrinkage", Biomitrika, Volume 81, pp. 425-455, 1994.

[30] V. N. P. Raj and T. Venkateswarlu, "Denoising of medical images using undecimated wavelet transform," *IEEE Recent Advances in Intelligent Computational Systems*, pp. 483-488, 2011.

[31] L. Sendur and I. W. Selesnick, “Bivariate shrinkage functions for wavelet-based denoising exploiting interscale dependency” IEEE Transaction on Signal Processing, vol. 50, pp. 2744–2756, 2002.

[32] M. S. Crouse and R. G. Baraniuk, “Contextual hidden Markov models for wavelet domain signal processing” in IEEE *Proc. Conf. Signals, Systems, and Computers*, 1997.

[33] H. Lucke, “Which stochastic models allow Baum–Welch training,” *IEEE Trans. Signal Processing*, vol. 44, pp. 2746–2756, 1994.

[34] Biomedical Image and Signal analysis research group (www.biomisa.org).