STOCHASTIC AND BIOLOGICAL METAPHOR PARAMETER ESTIMATION ON THE GAUSSIAN MIXTURE MODEL AND IMAGE SEGMENTATION BY MARKOV RANDOM FIELD

A Thesis

Submitted to the Faculty

of

Purdue University

by

Nariman Majdi Nasab

In Partial Fulfillment of the

Requirements for the Degree

of

Doctor of Philosophy

May 2004

To my parents whom have taught me the most beautiful lessons in life.

ACKNOWLEDGMENTS

I wish to express my sincere thanks to my major professors Mostafa Analoui and Edward J. Delp for their guidance and support. I greatly appreciate their cares and dedications in constructively commenting on my work.

I would also like to thank the other members of my committee, professors Peter Doerschuk, Thomas Talavage, and Tie-Qiang Li for their support.

I wish to thank my parents (Maman and Baba), my sisters (Nastaran and Niloufar), my brothers (Nader and Navid), and brother-in-law (Hossein) for their constant support and encouragement.

I would like to thank the personnel at Indiana University School of Dentistry, Dr. Masatoshi Ando, Dr. Hafsteinn Eggertsson, and Dr. Ted Park for providing images and spending time to validate my data. I am thankful to Dr. Susan Zunt for her permission to use the dental school facilities, Julie LeHunt for her continuous help and guidance on any issues related to the dental school through my Ph.D. studies.

Finally, I wish to thank Estela Agredano and Cristian Constantinescu for their help and encouragement through all the years of my Ph.D. transition.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

Figure Page

Figure Page

Figure Page

Figure Page

Figure Page

ABSTRACT

Majdi Nasab, Nariman Ph.D., Purdue University, May, 2004. Stochastic and Biological Metaphor Parameter Estimation on the Gaussian Mixture Model and Image Segmentation by Markov Random Field. Supervisors: Ed. J. Delp and M. Analoui.

The model parameters of image in real life applications are usually unknown and are necessary for any image processing such as image segmentation. Parameter estimation, labels, can be done from observed image. We proposed use of probabilistic transition rules based on biological metaphor, Genetic Algorithm (GA), standard Expectation Maximization (EM), Simulated Annealing (SA) and mix of these methods for learning Gaussian mixture components to achieve accurate parameter estimation on images.

We also introduced modified implementations of SA for image segmentation. The segmentation procedure is based on Markov random field (MRF) model for describing regions within an image. We proposed a random cost function for computing a posterior energy function in SA. The proposed modified Simulated Annealing (SA-RCF) method depicts more robust performance for image segmentation than standard SA at the same computational cost. Alternatively, we proposed a multi-resolution (MR) approach based on MRF, which offers a robust segmentation for noisy images with significant reduction in the computational cost on phantom images.

This thesis proposes accurate and stable solution methods for both parameter estimation and image segmentation for dental images. All proposed methods were evaluated on CT phantom images and applied on μCT images.

1. INTRODUCTION

For more than a century, medical imaging has been developing and enhancing new techniques such as Magnetic Resonance Imaging (MRI), Computed Tomography (CT), echography, and nuclear medicine imaging to observe the inside of the human body. All these methods obtain a Two-Dimensional (2-D) slice of a specific part of the body. The goal of medical imaging is to allow a better understanding of the body's metabolism anatomy and pathology.

Image segmentation, which labels individual pixels within a data set according to pre-defined homogeneity criteria [1], is an essential tool for automatic image analysis and understanding. A number of studies have described various statistical approaches to segmenting an image into regions or classes of differing statistical properties [2-5].

Non-fuzzy segmentation of an image assigns each pixel of the image to only one of a finite set of classes (labels) depending on the statistical properties of the pixel and its neighborhood. The pixel of an observed image (2-D field) that contains pixel classification is referred to as a label field. The label field is unknown and is estimated from the observed image. Segmentation requires the following conditions:

- every pixel must be in a class
- pixels in a region must be connected and
- the classes must be disjointed.

Markov random field (MRF) provides a convenient and consistent way for modeling context-dependent entities such as image pixels and correlated features [6]. This modeling is achieved through characterizing the mutual influence among such entities by using MRF probabilities.

The practical use of MRF models is largely ascribed to the equivalence between MRFs and Gibbs distributions established by Hammersley and Clifford [7] and further

developed by Besag [8]. The MRF-Gibbs equivalence theorem shows that the joint distribution of an MRF is a Gibbs distribution. This enables us to model problems in the Bayesian framework [9-10]. From a computational perspective, the local property of MRF leads to algorithms that can be implemented in a local and massively parallel manner. MRF theory also provides a foundation for multiresolution computation [11]. For the above reasons, The MRF has been widely used to conduct image analysis in many different levels like image segmentation and restoration [12-19].

MRF theory is often used in conjunction with statistical decision and estimation theories, so as to formulate objective functions in terms of established optimality principles. In this approach, finding an approximate Maximum A Posteriori (MAP) estimate of the field segments the image. MRF and MAP criterion was advocated by Geman and Geman [10]. A particular MRF model is specified by its functional form and parameters. Two major parts of MAP-MRF modeling are to derive the form of the posterior distribution and to determine the parameters on it, so as to completely define the posterior probability. Another part is to design optimization algorithms for finding the maximum of the posterior distribution.

The recent success of stochastic models and methods in image processing, in particular MRF representations, has led to a wide variety of hierarchical MRF approaches and multi-grid stochastic algorithms for image analysis. Algorithms that incorporate stochastic relaxation to find solutions have achieved excellent results. However, these algorithms are either computationally expensive or lack robustness when used to analyze noisy images. Alternative approaches to MAP segmentation of images have been proposed by various investigators [20-22].

This thesis focuses on two issues: 1) parameter estimation of an image, basically its labels, using statistical and biological metaphor methods and 2) image segmentation by stochastic approach.

1) We present new methods to determine the Gaussian mixture model parameters; our methods use Genetic Algorithms (GAs), Simulated Annealing (SA), and Expectation Maximization (EM).

Obtaining the true Gaussian mixture model parameters with local algorithms like the standard Expectation Maximization (EM) method [23-26], which is frequently used for these kind of problems, can be a difficult task because there are no generally accepted methods for parameter initialization; hence, one is likely to get trapped in one of the many local maxima of the likelihood function. Improvements on EM algorithms have been offered and their performances evaluated by many investigators [27-29].

GAs are adaptive search techniques designed to find out near-optimal solutions of large-scale optimization problems with multiple local maxima. It has been shown that GAs are independent of initialization parameters and an efficient technique to optimize difficult functions in large search spaces [30, 31].

Like other investigators, we have shown that using the EM algorithm might not always be successful because the true number of mixing components is unknown and algorithms can get trapped in one of the many local maxima of the likelihood function and especially because there is no generally accepted method for parameter initialization. We present new methods to find the Gaussian mixture model parameters by GAs. We show that by using GAs we can find noisy and noiseless normal mixture model parameters fast and accurately [32]. The accuracy of EM results strongly depends on how close an EM's initial values are to the correct values; we introduced a solution for this problem, a two-step process, first estimating the parameters by GAs and getting *a priori* information and second using *a priori* information as an initial guess for an EM algorithm to find out the final parameters; thus the problem of parameters can be solved using both GA and EM (GA-EM) algorithms [33].

We have also aimed to enhance the accuracy of parameter estimation with GA by employing the Simulated Annealing (SA) concept. We proposed a Simulated Annealing version of Genetic Algorithm (GA-SA). GA-SA randomly replaces labels of each population with an individual from the previous generation that has a lower cost function at the certain temperature. We have also coupled the result of GA-SA as an initial guess for EM to estimate observed image parameters. This method was called GA-SA-EM.

Optimization of the mixture of GA and EM was the next method of estimation of the Gaussian mixture model parameters. In this work, we propose a combination of Genetic Algorithm with EM (Interlaced GA-EM) to improve estimation of the Gaussian mixture parameters. We proposed Interactive GA and EM to find the mixture of the Gaussian parameters. This method uses a population of a mixture model rather than a single mixture, iteratively in both GA and EM to determine the Gaussian mixture parameters. The Gaussian parameters are defined by GA at the beginning and refined by EM for a specific number of iterations; then finer parameters from EM are returned back to GA to make sure these values are not trapped in the local minimum. After evaluating by GA, they pass back to EM again. This procedure continues for a specific number of iterations or until a termination condition is reached.

2) To achieve better image segmentation, we have proposed modifications of the SA [10] method to achieve more robust performance than standard SA at the same computational cost. Previous approaches based on multiresolution models for the pixel labels have used fine resolution of an observed image or multiresolution representation of the observed image with the implicit assumption that the random variables at a given level of the observed image pyramid are independent of the random variables at the other levels [21, 34-36].

One of our modifications is a Multi-Resolution (MR) approach based on MRF [37]. In the MR approach, we obtained a fine multiresolution representation of the observed image; then we modeled this representation as a random stochastic process indexed by nodes of a fine resolution lattice. A second modification was to evaluate random effects of the pixel neighborhood cost function at fine resolutions as a random stochastic process indexed by nodes of a fine resolution lattice. We have called this technique SA-RCF [38]. We have also used the Minimum Gray Level Distance (MGLD) method based on Euclidean distance between pixel intensity and label regions of an image [37]. Image segmentation methods were compared among SA, SA-RCF, MGLD, and MR.

To assess the performance of the proposed methods, we have tested our methods, both parameter estimations and segmentation methods, on a series of the Gaussian CT phantoms, based on ‘the modified Shepp-Logan method’, and dental images. Performances of methods were discussed in the term of accuracy of results and computational cost.

We have also demonstrated the advantage of unsupervised image segmentation methods compared with manual caries detection of the dental images.

In chapter two, we introduce the EM, GA, GA-SA, GA-EM, GA-SA-EM, Interlaced GA-EM, and Interlaced GA-CGM (Conjugate Gradient Method) parameter estimation methods for the Gaussian mixture model. Our algorithms and simulation results on a series CT phantom are shown. In chapter three, we present MRF concepts and a brief review of previous MRF work on both single and multiresolution image segmentation. Our algorithms are introduced and segmentation results on a series CT phantom image are shown in this chapter. In the chapter four, mixture model parameter estimation and image segmentation methods are applied to a series of dental images. Chapter four also shows the interval error range of manual caries detection in two different times in statistical approach of estimation by a dentist. Chapter five concludes the performance and results of parameter estimation and image segmentation methods (unsupervised image segmentation methods) on dental and CT phantom images.

2. EXPECTATION MAXIMIZATION, SIMULATED ANNEALING, AND GENETIC ALGORITHMS ON PARAMETER ESTIMATION OF THE GAUSSIAN MIXTURE MODEL

This chapter describes Expectation Maximization (EM), Genetic Algorithm (GA), and GA with Simulated Annealing (SA) for parameter estimation on the Gaussian mixture model. Combinations of EM, GA, and SA are introduced. The concepts and algorithms of all these methods are presented in this chapter.

In section 2.1, the Gaussian mixture model is introduced. In section 2.2, we describe the standard EM algorithm. In section 2.3, GAs are described and use of GAs to estimate parameters of mixture the Gaussian model are introduced. In section 2.4 modifications to GAs by SA concepts are introduced (We have called it GA-SA). Sections 2.5 and 2.6 present sequential coupling of GA and GA-SA to EM respectively. In section 2.7, we introduce an interactive combination of GA with EM (Interlaced GA-EM). In section 2.8 introduces an interactive combination of GA with Conjugate Gradient Method (Interlaced GA-CGM). Section 2.9 demonstrates performance of above methods on set of the modified Shepp Logan phantoms. Conclusion and discussion are presented in section 2.10.

2.1 Mixture model

The Artificial Intelligence (AI) community has been interested in the problem of classification in unsupervised learning for several years. In the classification problem in unsupervised learning, it is desired to determine a classification into G known classes from unclassified data. Parameter estimation can be achieved easily when all of the variables of the model are directly observable in the data. When some variables are unknown or hidden, as is common in popular model classes such as Bayesian networks

with hidden variables and hidden Markov models, ML parameter estimation is much more involved. The one specific use of parameter estimation of unclassified data is for medial image analysis. In our overall research, image segmentation, we need to know an accurate estimation of image parameters to be able to classify abnormal and normal tissues in the image. In this section, we address some solutions to find accurate parameter estimations of unclassified data, which are used for image segmentation methods. We provide our image segmentation methods in the next section.

Decomposing a finite mixture of unclassified data distributions is a very challenging problem. Several authors have utilized the likelihood approaches of fitting parameters to mixture models. The common approaches for analyzing and estimating parameters for mixture models seek to optimize various aspects of the estimation process, such as accuracy and computation cost [39-42]. Finite mixture distributions provide a simple framework for modeling population heterogeneity. Multivariate observations on a set of *n* entities forming an array can be represented as $x_1,\ldots,x_n$, where each x_j is a vector of *p*-dimensions. Under the finite mixture models, each x_j can be viewed as arising from a super population *G*, which is a mixture of a finite number of populations $G_1,\ldots,G_g$ in some proportions $\pi_1,\ldots,\pi_g$, respectively, where each mixture probability density function (p.d.f.) of an observation x in G can be represented in the finite mixture form:

$$f(x)=\sum_{i=1}^{g} p(x,c_i) \tag{2.1}$$

where $p(x,c_i)$ is the p.d.f. corresponding to G_i. Equation 2.1 can be expressed as:

$$f(x;\phi)=\sum_{i=1}^{g} p(x,c_i)=\sum_{i=1}^{g} p(x\mid c_i)\,p(c_i)=\sum_{i=1}^{g} p(x\mid\theta_i)\pi_i \tag{2.2}$$

where $p(x\mid\theta_i)$ is component of *i*th model, θ_i, vector of all unknown parameters, and π_i, weight, denote all unknown parameters associated with the parametric forms that have all of these g component densities. Weights are distributed through the population as follow:

$$\sum_{i=1}^{g}\pi_i=1 \quad and\ \pi_i\geq 0 \quad (i=1,\ldots,g). \tag{2.3}$$

In the case multivariate of the Gaussian mixture model, component densities consist of the elements of the mean vectors μ_i and the distinct elements of the covariance matrices Σ_i for $i = 1,\ldots,g$. The vector, ϕ,

$$\phi = (\pi, \theta)^T \tag{2.4}$$

of all unknown parameters belongs to parameters' space Ω; "^{T}" denotes vector transpose. For example for g = 4 and *1*-dimension vector space, we have:

$$\phi = (\pi_1, \mu_1, \sigma_1, \pi_2, \mu_2, \sigma_2, \pi_3, \mu_3, \sigma_3, \pi_4, \mu_4, \sigma_4)^T \tag{2.5}$$

Figure 2.1. shows $p(x, \theta_i)$ for i=1,…,4 for a numeric value $\phi = (0.25, 50, 12, 0.15, 79, 12, 0.45, 130, 13, 0.15, 180, 22)^T$. Figure 2.2 shows the Gaussian mixture model of Eq. 2.5. All the Gaussian mixture model distributions for the above example were created noiseless.

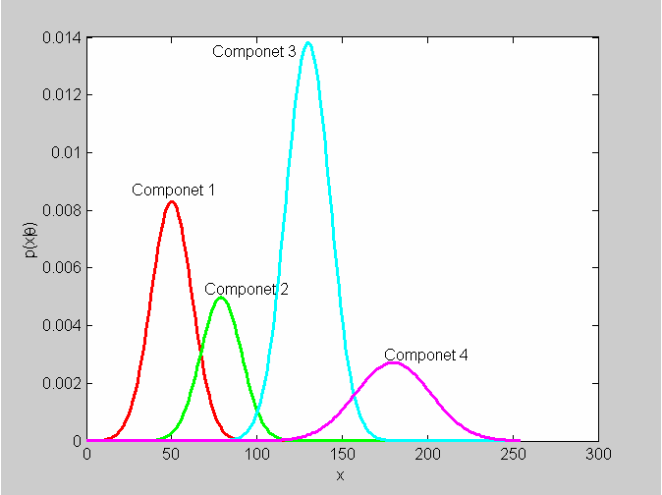


Figure 2.1. Four individual components of the Gaussian distribution function.

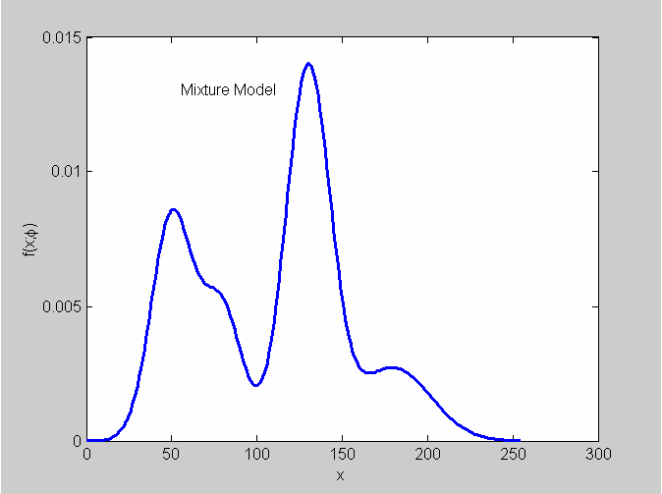


Figure 2.2. Mixture model for four components of the Gaussian distribution function.

Learning the mixture, namely, estimating the weights π_i and the parameters θ_i of each component is often carried out through likelihood maximization using the EM algorithm [40, 42]. However, known limitations of EM are the follow up:

(i) It assumes that the number of mixing components is known,

(ii) There is no widely accepted "good" method for initialization of the parameters, and

(iii) The algorithm is of local nature and thus can get trapped in local maxima of the likelihood function.

2.2 Expectation Maximization

In this section we describe the standard Expectation Maximization (EM) method. The EM algorithm [42] aims to determine a Maximum Likelihood (ML) an estimate of the parameter θ, by use of the estimation of the missing data.

The probability density function of the mixture the Gaussian model of an observation random variable, x, is defined in the finite mixture form as:

$$f(x;\phi) = \sum_{i=1}^{g} \pi_i . p_i(x;\theta) = \sum_{i=1}^{g} \pi_i . \frac{1}{(2\pi)^{1/2} \left|\Sigma_i\right|^{\frac{1}{2}}} \exp -\frac{1}{2}(x-\mu_i)^T \Sigma_i^{-1}(x-\mu_i) \qquad (2.6)$$

where π_i is the mixing parameter, $p_i(x;\theta)$ is the p.d.f. corresponding to G_i, and θ denotes the vector of all unknown parameters associated with parametric forms adopted for g

component densities. The vector $\phi = (\pi, \theta)^T$ of all unknown parameters is estimated using the EM algorithm.

We have obtained the following iterative equations of parametric mixture model by solving ML of Eq. 2.6 [40]:

$$\mu_i^{t+1} = \frac{\sum_{j=1}^{n} p^t(i \mid x_j) x_j}{\sum_{j=1}^{n} p^t(i \mid x_j)} \tag{2.7}$$

$$\Sigma_i^{t+1} = \frac{\sum_{j=1}^{n} p^t(i \mid x_j)(x_j - \mu_i^t)^T (x_j - \mu_i^t)}{\sum_{j=1}^{n} p^t(i \mid x_j)} \tag{2.8}$$

$$\pi_i^{t+1} = \frac{1}{n} \sum_{j=1}^{n} p^t(i \mid x_j) \tag{2.9}$$

where

$$p^t(i \mid x_j) = \frac{p^t(x_j \mid i)\pi^t(i)}{p^t(x_j)} \tag{2.10}$$

is the conditional probability of x_j random variable in the *i*th class at *t*th iteration. Other researchers have also introduced improvements to EM algorithms for image analysis [43-45].

Good abilities of GAs-- escaping from local minima, finding global minima, and dependency of GAs on initial values-- are important advantages of GAs in comparison with the EM algorithm.

GAs are adaptive search techniques designed to find near optimal solutions for large-scale optimization problems with multiple local maxima, and GAs are good candidates to find mixture model parameters. In the following subsections, we introduce use of GAs to estimate mixture model parameters.

2.3 Genetic Algorithms

The genetic Algorithm (GA) is a method that mimics the natural biological metaphor and tries to emulate some of the processes observed in natural evolution. GAs operate on a population of potential solutions, applying the principle of survival of the fittest to produce better and better approximations to a solution. At each generation, a new set of approximations is created by the process of selecting individuals according to the level of fitness in the problem domain and breeding them together using operators borrowed from natural genetics. This process leads to populations of individuals that are better suited to their environment than the individuals from which they were created, just as in natural adaptation.

GAs are an efficient technique to optimize difficult functions in large search spaces. By testing populations of solutions represented as strings of bits (called chromosomes) in an iterative process, a genetic algorithm is able to find a near-optimal solution in a robust manner, with the ability to produce a best ‘guess’ from incomplete or noisy data [30].

GAs are intended to mimic some of the processes observed in natural evolution in the following way [31, 46]:

- Evolution is a process that operates on encoding of biological entities, rather than on the living beings themselves.
- Natural selection is the link between a chromosome and its performance (measured on its decoded version). Nature obeys the principle of Darwinian “survival of the fittest”; the chromosomes with high fitness values will, on the average, reproduce more often than those with low fitness values. GAs try to mimic this procedure; highly fitted chromosomes will reproduce more often at the cost of lower fitted ones.
- Biological evolution has no memory; nature acts as the environment and biological entities are modified to adapt to their environment. Whatever nature knows about the evolution of good chromosomes is contained in the set of chromosomes possessed by current individuals and in the chromosome decoding procedure. Likewise, GAs also operate on chromosomes blindly. They use only objective function information, which acts as environment. Based on this information, each chromosome is evaluated;

during the selection process more importance is given to choosing chromosomes having high fitness values.

- Like natural evolution, variation of the entities in GAs is introduced when reproduction occurs. Crossover and mutation are the basic operators for reproduction.

GAs tackle the problem of minimizing a cost function, the objective function, of the vector $x_i^j = (x_i^1,\ldots,x_i^l)$. In a general formulation each x_i^j, *j=1, …,l,* is binary, taking the value 0 or 1. The basic GA starts by randomly generating an even number n of binary strings of length *l* to form an initial population (parents). Rules are then applied to create successive generations in turn.

These parents are considered in pairs and, for each pair, a crossover operation is performed with a pre-selected probability p_c. If the crossover occurs, an integer *k* is generated from the uniform distribution on $\{1,\ldots,l-1\}$, and the last *l-k* elements of each parent are exchanged to create two new strings; thus, one string takes values $x_i^1,\ldots,x_i^k$ from the first parent and $x_i^{k+1},\ldots,x_i^l$ from the second while the other takes values $x_i^1,\ldots,x_i^k$ from second parent and $x_i^{k+1},\ldots,x_i^l$ from the first. If crossover does not occur, the parents are copied unaltered into new strings. Finally, the mutation operator switches the value at each string position from 0 to 1 or vice versa with probability p_m, mutation occurring independently at each element of each string. The algorithm is allowed to continue for a certain number of generations. On termination, the string in the final population with the highest value of fitness can be returned as the solution to the optimization problem. However, since a good solution may be lost in this algorithm, a more efficient strategy is to note the best solution seen at any stage and return this as the solution.

A positive fitness F is calculated as a monotone decreasing function of objective function Λ for each individual in the current generation, and *m* parents for the next generation are selected with replacement, the probability F_n of choosing *n*th individual in the current population being proportional to its fitness Λ_n, *i.e.*

$$F_n = \frac{\Lambda_n}{\sum_{i=1}^{m} \Lambda_n} \tag{2.11}$$

The population size *m*, probability of crossover, probability of mutation, and fitness function F must be specified before the algorithm is applied, although trial runs may be helpful in choosing appropriate values.

The sensivity of the algorithm to the choice of p_m is evident. If p_m is too small, the algorithm can become trapped in a local minimum at an early age, but if p_m is too high, random mutations disrupt what would otherwise be good solutions and the algorithm struggles to create the very best solutions.

A simple genetic algorithm is given below, where $p(t)$ is a population of candidate solutions to a given problem at generation *t*.

t=0;

Initialize $p(t)$ *;*

Evaluate $p(t)$ *;*

While not <termination condition>

Begin

t=t+1;

Reproduce $p(t)$ *from* $p(t-1)$ *;*

Recombine $p(t)$ *;*

Evaluate $p(t)$ *;*

End

Implementation of GAs requires a positive valued fitness function F , monotically decreasing in the objective function $\Lambda(x)$ where,

$$\Lambda_n(x) = \arg\min_{\hat{\phi}}(f(x;\phi);\ \hat{f}_n(x;\hat{\phi})) \tag{2.12}$$

The hat sign, ^, demonstrates the values which were generated by GAs for the *n*th individual, *n=1, …, m,* in the current population. This function must provide sufficient

selectivity to ensure that the algorithm performs superior solutions to the extent that it eventually produces an optimal or near-optimal answer.

Mutation is randomly applied with low probability, typically in the range 0.001 and 0.01, and modifies elements in the chromosomes. Usually considered as a background operator, the role of mutation is often seen as providing a guarantee that the probability of searching any given string will never be zero and will act as a safety net to recover good genetic material that may be lost through the action of selection and crossover. In this work, population size *m* = 500 was selected. The value for mutation, $p_m = \frac{0.7}{Lind}$, is selected, where *Lind* is the chromosome length. Chromosome length is the bit value for each individual, which is the number of bits for ϕ in equation 3.4. The basic operator for producing new chromosomes in the GA is that of crossover. Like its counterpart in the nature, crossover produces new individuals that have some parts of both parent's genetic material. Normally, probability of crossover is selected between 0.6 - 0.9 and in this work, the value of 0.08 was selected.

Objective function plays an important role to find out accurate ϕ parameters. The objective function is used to provide a measure of how individuals have performed in the problem domain. In the minimization problem considered in this thesis, the fittest individuals will have the lowest numerical value of the associated objective function. We have proposed and simulated the following different types of objective functions:

$$\Lambda_n(x) = \left| f(x;\phi) - \hat{f}_n(x;\hat{\phi}) \right| \tag{2.13}$$

$$\Lambda_n(x) = \log(\left| f(x;\phi) - \hat{f}_n(x;\hat{\phi}) \right|) \tag{2.14}$$

$$\Lambda_n(x) = \left\| f(x;\phi) - \hat{f}_n(x;\hat{\phi}) \right\|^2 \tag{2.15}$$

$$\Lambda_n(x) = \left| 1 - \sum \hat{f}_n(x;\hat{\phi}) \right| + a \times \left| 1 - \sum_{i=1}^{g} \hat{\pi}_i \right| + \left| f(x;\phi) - \hat{f}_n(x;\hat{\phi}) \right| \tag{2.16}$$

where *a* is a weight factor. The *a* value belongs to set of $A = \left\{0, \frac{\sqrt{2}}{2}, 1, \sqrt{2}, \sqrt{2}\pi\right\}$.

$$\Lambda_n(x) = \left| FT(f(x;\phi)) - FT(\hat{f}_n(x;\hat{\phi})) \right| \tag{2.17}$$

where *FT* is *1*-dimensional Fourier transform operator.

$$\Lambda_n(x) = \max\left(\left| f(x;\phi) - \hat{f}_n(x;\hat{\phi}) \right| \right) \tag{2.18}$$

$$\Lambda_n(x) = |1 - \sum_{i=1}^{g} \hat{\pi}_i| + \exp|f(x;\phi) - \hat{f}_n(x;\hat{\phi})|^3 \tag{2.19}$$

Eventually the best of them, equations 2.17 and 2.19, with almost the same performance, were determined; equation 2.19 was eventually selected for simulation of our data since it requires less computational process. We have done simulations for a set of six-component normal mixture model of CT phantoms. The maximum number of generations in our experiments was selected to be 300.

The following table is used as distribution of bits for ϕ. These values are selected based on 8 bits per pixel on 'the Modified Shepp-Logan' phantom [47].
More explanation about phantom images will be given in the simulation section.

Table 2.1. Bit value assignation in 'the Modified Shepp-Logan' phantom.

ϕ_i ;*i=1,…,6*	π_i	μ_i	Σ_i
Bits	6	8	4

2.4 Genetic Algorithm and Simulated Annealing (GA-SA)

In this section, we propose a Simulated Annealing version of a Genetic Algorithm (GA-SA). "Strategy of GA-SA is very similar to a combination of steepest decent and Newton's methods to improve each other for the constrained optimization problem of the continuous type" [48, 49]. Simulated annealing (SA) is a stochastic computational technique derived from statistical mechanics to determine near-globally minimum-cost solutions to large optimization problems. Both SA and GAs are nature stochastic

computational techniques; the SA is based on thermodynamics and the GA is based on natural evaluation. The major advantages of these nature-based algorithms are their applicability, flexibility, ease of implementation, and potential to find near-optimal solutions. In simulated annealing a control parameter called temperature is used to control the minimization search, which may occasionally move uphill. The mean and variance of the cost function decrease during the course of the search process [49].

The SA algorithm generates a finite-length Markov chain generated at the certain temperature T_t (temperature at the *t*th iteration) with a single starting point and a single ending point in the large search space Ω. The energy function ΔU lower or higher than of the starting search point, which should depend on how much computation time has been spent at the temperature T_t.

It is desire to minimize the cost function of mixture model at the temperature T_t for *t*th iteration:

$$\arg\min_{\hat{\phi}}(f(x;\phi);\ \hat{f}_t(x;\hat{\phi})) \tag{2.20}$$

where $f(x;\phi)$ is the Gaussian mixture model of an observed image and $\hat{f}(x;\hat{\phi})$ is the Gaussian mixture model of an unobserved image. $f(x;\phi)$ for *1*-dimensional, *x*, can be shown:

$$f(x;\phi)=\sum_{i=1}^{g}\pi_i.p_i(x;\theta)=\sum_{i=1}^{g}.\frac{\pi_i}{(2\pi)^{1/2}\sigma_i}\exp-(\frac{(x-\mu_i)^2}{2\sigma_i^2}) \tag{2.21}$$

GA-SA, like GAs, is independent of initial seed values. The first population of a candidate solution is generated completely randomly. We have used Matlab random generation function for this purpose. Cooling rate of the system at the *k*th generation was selected:

$$T_k=0.976\times T_{k-1} \tag{2.22}$$

The following portion was added to the reproduction section of the standard GA algorithm:

1. Perturb $p(t-1)$ into $p(t)$

   Let $\Delta U = U(\hat{f}_t \mid f) - U(\hat{f}_{t-1} \mid f)$.

   If ΔU <0 then replace $p(t-1)$ by $p(t)$; else if $\Delta U \geq 0$

   replace $p(t-1)$ by $p(t)$ with probability $e^{-\Delta U / T_t}$

2. Replace T_{t-1} by T_t

2.5 Genetic Algorithm and Expectation Maximization (GA-EM)

In our previous works [32, 33], we have shown that using EM algorithm might not always be successful because the true number of mixing component is unknown. Furthermore the algorithm can get trapped in one of the many local maxima of the likelihood function and there is not a generally accepted method for parameter initialization. The result of EM is strongly dependent on its initial values. In practice one does not have any knowledge of those parameters. An alternative approach to this problem is to have a two-step process, first estimating the parameters by GAs to obtain crude *a priori* information. Then the *a priori* information is used as initial guess for the EM algorithm to find the final parameters. So the problem of parameter estimations can be solved in coupling GA to EM (GA-EM) algorithms.

2.6 Genetic Algorithm, Simulated Annealing, and Expectation Maximization (GA-SA-EM)

The Genetic Algorithm, Simulated Annealing and Expectation Maximization (GA-SA-EM) method has the same procedure as GA-EM. It includes two steps to boost accuracy of mixture model parameter estimations. First estimating the parameters by GA-SA and getting *a priori* information and then using the *a priori* as starting values for the EM algorithm. Thus the problem of parameters can be solved using both GA-SA and EM algorithms.

The parameters, conditions, objective function are the same for both GA-SA-EM and GA-EM methods.

2.7 Interlaced Genetic Algorithm and Expectation Maximization (Interlaced GA-EM)

In the previous couple sections, we introduced coupling GA and GA-SA to EM to increase accuracy of parameter estimation but in this section, we achieve parameter estimation by interlacing GA and EM [50].

We propose a new Genetic version of EM (Interlaced GA-EM) to improve parameter estimation of Gaussian mixture model. We define a new method that uses a population of a mixture model (rather than a single mixture). Interactively merging GA with EM boosts accuracy of parameter estimation because we combine the global search algorithm, GA, with a local optimization method, EM. The Gaussian parameters are defined by GA at the beginning and refined by EM for a specific number of iterations; then finer parameters from EM are returned back to GA to make sure these values are not trapped in the local minimum. After a few iterations, they pass back to EM again. This procedure continues for a specific number of iterations or until a termination condition is reached.

2.8 Interlaced Genetic Algorithm and Conjugate Gradient Method (Interlaced GA-CGM)

In this subsection, we have used Conjugate Gradient Method [51, 52] instead of EM. The Interlaced Genetic algorithm and Conjugate Gradient method (Interlaced GA-CGM) is very similar to Interlaced GA-EM except that CGM was used instead EM. The idea is to use the change value the parameters of EM iteration to find better conjugate gradient. The method uses the information provided by the EM iteration to reshape the function and improve convergence speed when close to a solution. The first direction d^o is selected randomly at the beginning and it updates at each iteration by maximizing likelihood of:

$$\hat{\phi}^k = \arg\max_{\gamma} \text{Likelihood}(\hat{\phi}^{k-1} + \gamma d^k) \tag{2.23}$$

where γ is step size [51]. Golden section search method was used to find the best conjugate [53].

2.9 Simulations and discussion

In this section, we assess the performance of seven algorithms: 1) GA, 2) EM, 3) GA-SA, 4) GA-EM, 5) GA-SA-EM, 6) Interlaced GA-EM, and 7) Interlaced GA-CGM for four sets of phantom CT images. The images are called Image 1, Image 2, Image 3, and Image 4 and were created based on 'the Modified Shepp-Logan' method [47]. There was not any preference to select image parameters to build up a phantom except having different Gaussian model parameters for each image. Each phantom has six classes, and the class parameters are not the same among phantoms. Although the shapes look alike, no two have the same values.

Tables 2.2 to 2.5 summarize phantom parameters for Image 1 to Image 4 respectively. The table is divided in eight blocks. The first top block shows the parameters of image 1; rest of table shows results of EM, GA, GA-SA, GA-EM, GA-SA-EM, Interlaced GA-EM, and Interlaced GA-CGM. Four phantom images and their mixture models versus model parameters are shown in figures 2.3 (a and b) to figure 2.6 (a and b).

There is always an error between the p.d.f. of an image and the simulated Gaussian mixture model. We called this error "Model error". Model error is minimum error that can be obtained with a true mixture of model parameters. This error is produced because of finite generation of the Gaussian data. Each image's size is 256×256 pixels with 8 bits per pixel.

Empirically, we determined that if we used bigger size of images, such as an image of 1024×1024 pixels, model error would be zero. We have used the smaller size of image to reduce computational cost; furthermore, for this kind of mixture model, a smaller image is more similar to the true re-constructed image histogram.

Table 2.2. Parameters (ϕ) of image 1 and results of parameter estimation for seven different algorithms

	π	μ	σ	
Image 1	0.5818	30	5	Model error= 0.0315
	0.0014	60	15	
	0.3293	90	8	
	0.0434	120	6	
	0.0008	150	10	
	0.0434	210	12	
Result of EM on image 1	0.5817	30	5.0172	EM error= 0.0311
	0.3255	90	7.9312	
	0.0075	99	30.5937	
	0.0420	120	5.9463	
	0.0382	210	12.0632	
	0.0052	210	12.0681	
Result of GA on image 1	0.0632	30	5	GA error= 0.0337
	0.5205	30	5	
	0.1735	89	7	
	0.1596	91	9	
	0.0390	120	6	
	0.0446	210	12	
Result of GA-SA on image 1	0.3053	29	5	GA-SA error= 0.0431
	0.2812	31	5	
	0.1568	88	8	
	0.1652	91	7	
	0.0502	119	6	
	0.0418	210	12	
Result of GA-EM on image 1	0.5191	30	5.0273	GA-EM error= 0.0343
	0.0630	30	5.0273	
	0.2128	90	7.0856	
	0.1179	90	9.9262	
	0.0435	120	6.4735	
	0.0439	210	13.0658	
Result of GA-SA-EM on image 1	0.2400	30	4.8463	GA-SA-EM error= 0.0335
	0.3420	30	5.0149	
	0.2586	90	7.3343	
	0.0722	90	10.8886	
	0.0433	120	6.4906	
	0.0438	210	13.0620	
Result of Interlaced GA-EM on image 1	0.5820	29.9953	5.0245	Interlaced GA-EM error= 0.0308
	0.0010	61.9149	5.8352	
	0.3287	89.9085	7.9615	
	0.0439	119.8271	6.1029	
	0.0010	144.3430	15.6647	
	0.0434	209.8939	12.0539	
Result of Interlaced GA-CGM on image 1	0.5820	29.9960	5.0232	Interlaced GA-CGM error= 0.0314
	0.0039	73.7743	11.0701	
	0.1712	86.6216	6.9350	
	0.1536	93.6629	6.9843	
	0.0455	119.6979	6.6343	
	0.0438	209.3901	13.0473	

Table 2.3. Parameters (ϕ) of image 2 and results of parameter estimation for seven different algorithms

	π	μ	σ	
Image 2	0.5818	60	14	Model error= 0.0387
	0.0014	91	15	
	0.3293	122	8	
	0.0434	153	6	
	0.0008	184	10	
	0.0434	220	10	
Result of EM on image 2	0.3334	59	13.4930	EM error= 0.0383
	0.2494	62	14.6530	
	0.3300	122	8.0102	
	0.0430	153	6.0467	
	0.0013	190	13.3320	
	0.0430	220	10.0869	
Result of GA on image 2	0.1346	56	14	GA error= 0.0385
	0.4472	61	14	
	0.3294	122	8	
	0.0437	153	6	
	0.0010	201	15	
	0.0437	220	10	
Result of GA-SA on image 2	0.3043	58	13	GA-SA error= 0.0396
	0.2060	61	15	
	0.0724	66	14	
	0.3266	122	8	
	0.0446	153	6	
	0.0455	219	11	
Result of GA-EM on image 2	0.1350	57	13.1661	GA-EM error= 0.0381
	0.4477	61	14.1592	
	0.3300	122	8.0127	
	0.0431	153	6.0917	
	0.0006	184	5.9086	
	0.0435	220	10.2472	
Result of GA-SA-EM on image 2	0.3050	58	13.3976	GA-SA-EM error= 0.0387
	0.2059	61	14.5437	
	0.0718	65	13.8458	
	0.3297	122	7.9950	
	0.0435	153	6.2493	
	0.0441	220	10.8734	
Result of Interlaced GA-EM on image 2	0.5806	59.8816	13.9446	Interlaced GA-EM error= 0.0368
	0.0018	89.9659	2.9977	
	0.3281	121.9958	7.9967	
	0.0017	130.9313	1.9716	
	0.0436	152.9176	6.2744	
	0.0441	219.4674	10.8715	
Result of Interlaced GA-CGM on image 2	0.5760	59.7217	13.8557	Interlaced GA-CGM error= 0.0383
	0.0067	84.1689	9.2335	
	0.3300	122.0786	8.0064	
	0.0433	152.9802	6.1892	
	0.0330	218.8063	11.7471	
	0.0111	221.5140	7.6616	

Table 2.4. Parameters (ϕ) of image 3 and results of parameter estimation for seven different algorithms

	π	μ	σ	
Image 3	0.5818	120	13	Model error= 0.0338
	0.0014	147	15	
	0.3293	174	8	
	0.0434	201	6	
	0.0434	220	10	
	0.0008	228	5	
Result of EM on image 3	0.0001	66	1.4239	EM error= 0.0338
	0.1494	118	12.2838	
	0.4336	121	13.1791	
	0.3279	174	7.9048	
	0.0496	201	6.6086	
	0.0395	222	9.3253	
Result of GA on image 3	0.2329	118	14	GA error= 0.0453
	0.309	121	12	
	0.039	125	16	
	0.3285	174	8	
	0.0761	206	10	
	0.014	229	6	
Result of GA-SA on image 3	0.1270	118	12	GA-SA error= 0.05173
	0.4444	122	13	
	0.1746	174	7	
	0.1746	177	10	
	0.0317	207	7	
	0.0476	220	13	
Result of GA-EM on image 3	0.0816	119	11.8259	GA-EM error= 0.0335
	0.4233	119	13.0396	
	0.0780	126	12.6687	
	0.3287	174	7.9369	
	0.0463	201	6.2403	
	0.0421	221	9.7119	
Result of GA-SA-EM on image 3	0.1316	117	12.5511	GA-SA-EM Error= 0.0335
	0.4513	121	13.0494	
	0.1765	173	7.2815	
	0.1549	175	8.8051	
	0.0423	201	5.8208	
	0.0434	220	9.8736	
Result of Interlaced GA-EM on image 3	0.5820	120.0371	12.9970	Interlaced GA-EM error= 0.0334
	0.0010	156.0088	10.0427	
	0.3241	173.8624	7.8639	
	0.0229	201.2974	4.8747	
	0.0515	206.8387	13.7603	
	0.0184	224.9357	8.0625	
Result of Interlaced GA-CGM on image 3	0.0010	92.3897	1.7805	Interlaced GA- CGM error= 0.0333
	0.5819	120.1148	12.9788	
	0.3277	173.9287	7.9026	
	0.0532	201.7626	6.9325	
	0.0343	222.0469	8.1415	
	0.0019	236.1254	7.0341	

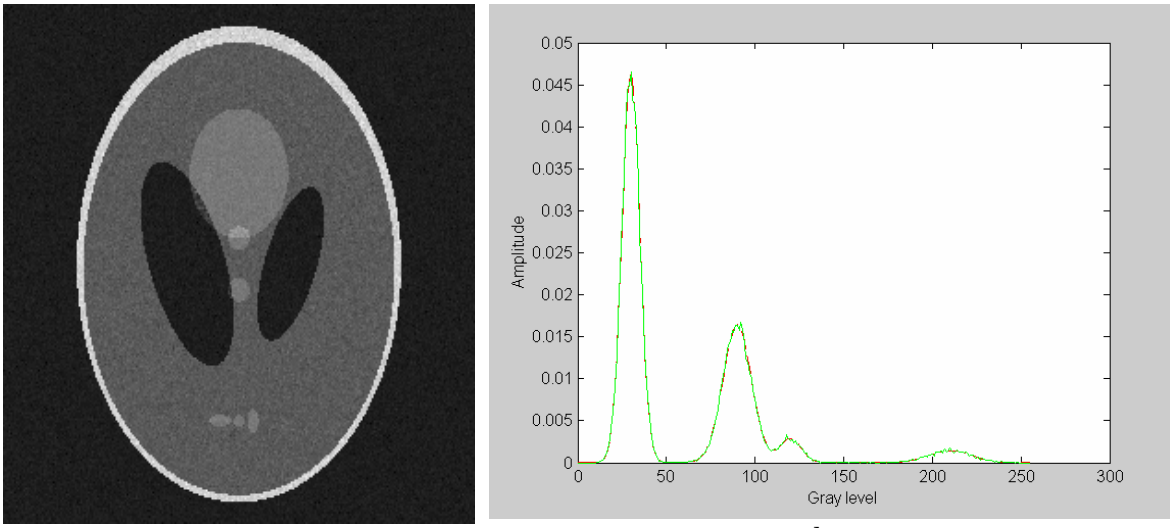
Table 2.5. Parameters (ϕ) of image 4 and results of parameter estimation for seven different algorithms

	π	μ	σ	
Image 4	0.5818	17	3	Model error= 0.0247
	0.0014	47	10	
	0.3293	77	8	
	0.0434	107	6	
	0.0008	137	2	
	0.0434	180	5	
Result of EM on image 4	NaN	NaN	NaN	EM error= Not defined
	NaN	NaN	NaN	
	NaN	NaN	NaN	
	NaN	NaN	NaN	
	NaN	NaN	NaN	
	NaN	NaN	NaN	
Result of GA on image 4	0.5789	17	3	GA error 0.0397
	0.232	76	7	
	0.0511	77	9	
	0.0799	92	15	
	0.0149	109	4	
	0.0427	180	5	
Result of GA-SA on image 4	0.4361	17	3	GA-SA error= 0.026
	0.1439	17	3	
	0.3266	77	8	
	0.0038	80	16	
	0.0437	107	6	
	0.0455	180	5	
Result of GA-EM on image 4	0.2911	16	2.5772	GA-EM error= 0.0514
	0.1118	17	3.5104	
	0.1789	19	2.3955	
	0.2688	76	7.1013	
	0.1059	92	16.4299	
	0.0434	180	5.0417	
Result of GA-SA-EM on image 4	0.5817	16.9881	3.0147	GA-SA-EM error= 0.0252
	0.0576	76.6340	7.8432	
	0.2652	77.0170	7.9572	
	0.0130	86.9548	24.8438	
	0.0391	106.9537	5.6985	
	0.0434	180.1640	5.0393	
Result of Interlaced GA-EM on image 4	0.5818	16.9883	3.0156	Interlaced GA-EM error= 0.0238
	0.0032	62.7381	15.3590	
	0.3275	77.0214	7.9838	
	0.0433	106.8804	6.0810	
	0.0010	136.6655	1.9943	
	0.0434	180.1616	5.0418	
Result of Interlaced GA-CGM on image 4	0.5818	16.9900	3.0166	Interlaced GA- CGM error= 0.0264
	0.3019	76.7496	7.6346	
	0.0419	86.8334	16.0000	
	0.0310	107.5523	5.1172	
	0.0422	179.7915	4.8450	
	0.0013	190.2144	2.9232	

Performance of each algorithm is calculated based on minimum squared errors between mixture models of each image with mixture model that is obtained from each proposed methods, Eq. 2.24.

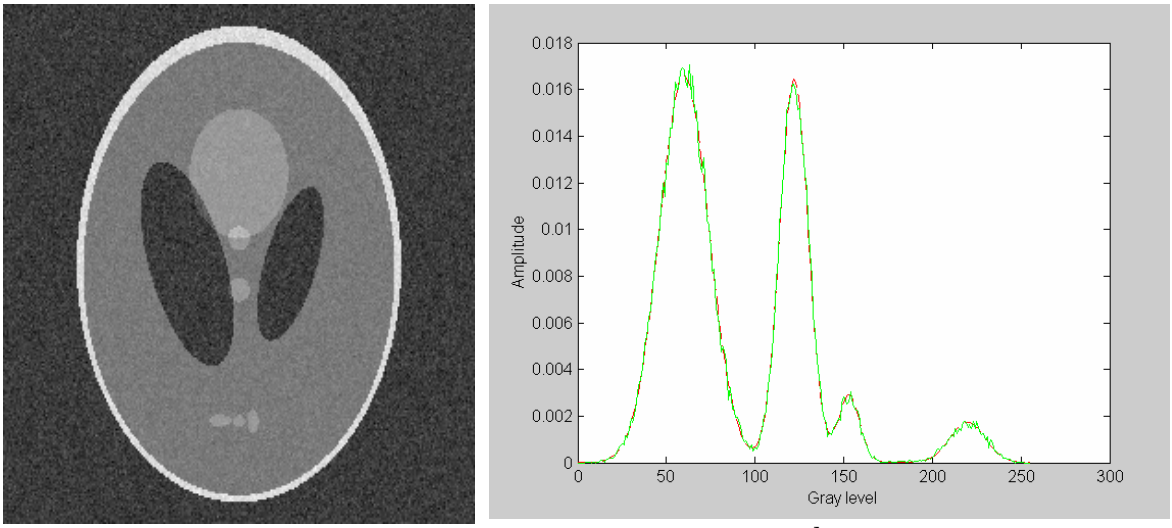
$$Error = \sum_{x} \left| f(x;\phi) - \hat{f}_{\text{algorithm}}(x;\hat{\phi}) \right| \tag{2.24}$$

where $\hat{f}_{\text{algorithm}}(x;\hat{\phi})$ is the mixture model resulting from EM, GA, GA-EM, GA-SA, GA-SA-EM, Interlaced GA-EM, and Interlaced GA-CGM. Equation 2.24 offers one plausible method for assessing accuracy of estimation process.



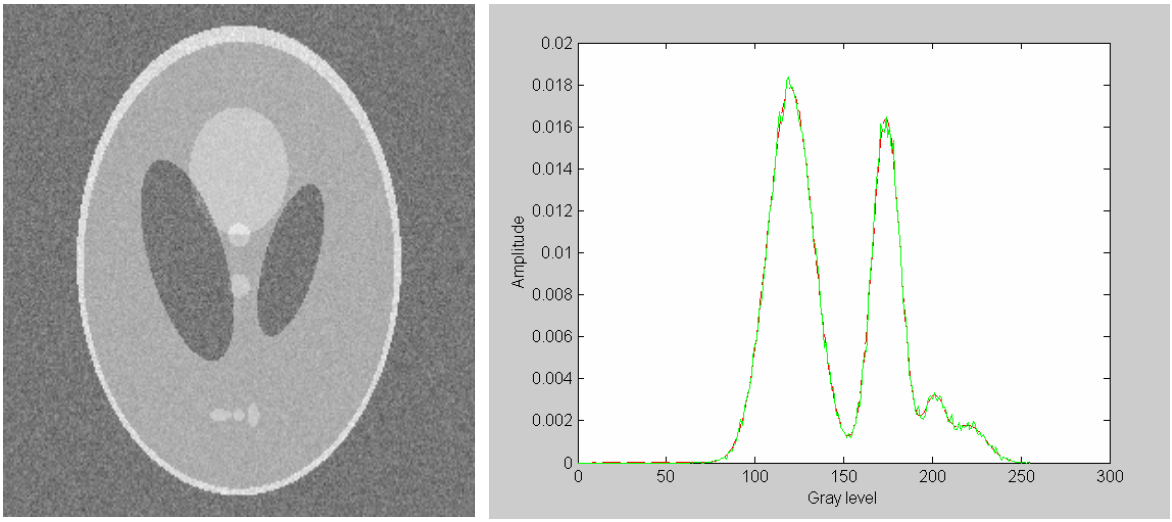
a b

Figure 2.3 a) Image 1 b) Comparison between mixture models of image 1 vs. its Gaussian mixture model, solid line is image mixture model and '-' line is the Gaussian mixture model.



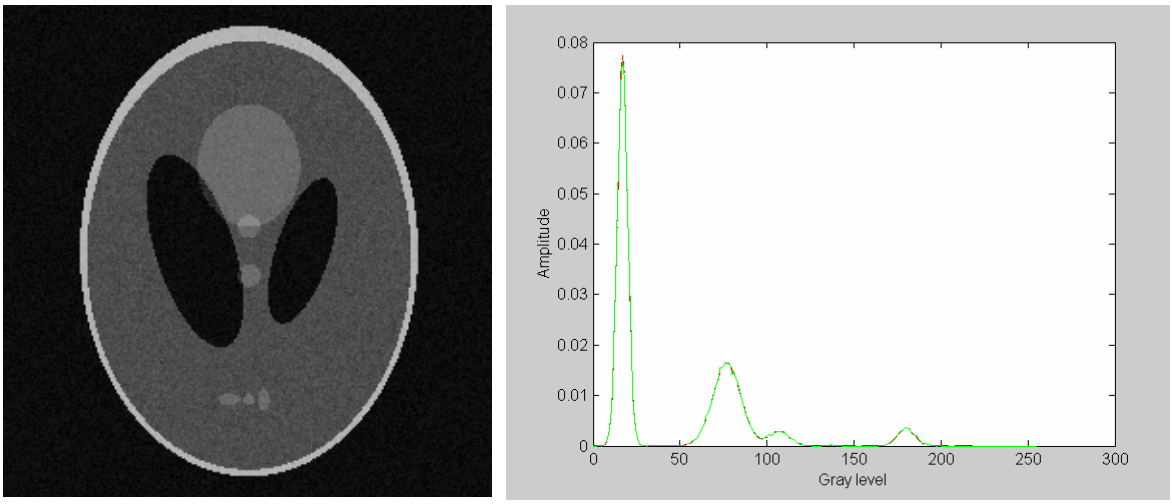
a b

Figure 2.4 a) Image 2 b) Comparison between mixture models of image 2 vs. its Gaussian mixture model, solid line is image mixture model and '-' line is the Gaussian mixture model.



a b

Figure 2.5 a) Image 3 b) Comparison between mixture models of image 3 vs. its Gaussian mixture model, solid line is image mixture model and '-' line is the Gaussian mixture model.



a b

Figure 2.6 a) Image 4 b) Comparison between mixture models of image 4 vs. its Gaussian mixture model, solid line is image mixture model and '-' line is the Gaussian mixture model.

2.10 Conclusion

In this chapter, the GA was used to estimate and enhance parameter estimation methods with EM for the Gaussian mixture model. One key advantage of GAs is that the algorithm does not depend on the initial condition, which is one of the limitations of EM algorithm in finding the mixture model parameters.

The performance of GAs is highly dependent on how the objective function defines the transition system. We tested several types of objective function and we presented the best of them in this thesis for the simulations.

In this work, we proposed combination of GA and EM algorithms (GA-EM) for determining the Gaussian mixture components to achieve accurate parameter estimation independent of initial values. GA-SA as simulated annealing version of GA was introduced and coupled to EM (as well as GA) to find the Gaussian mixture model parameters. The interactive GA and EM algorithm, which incorporates both GA and EM, was introduced.

To assess the performance of the proposed method, a series of the Gaussian phantoms, based on 'the modified Shepp-Logan method', were created.

All seven methods: EM, GA, GA-EM, GA-SA, GA-SA-EM, Interlaced GA-EM and Interlaced EM-CGM were employed to estimate region parameters in the four CT phantoms. The results indicate that the EM algorithm, as expected, is heavily affected by the initial values, and the EM failed to calculate mixture model parameters for Image 4.

Coupling GA and GA-SA with EM not only improves the overall accuracy, it also provides estimates that are independent of initial seed values.

The best result was obtained from the Interlaced GA-EM on both minimum error value and accuracy of obtained parameters.

EM had lowest while Interlaced GA-CGM had highest computational time. Almost all, the computational time for the rest of the methods were the same, higher than EM and much less than Interlaced GA-EM.

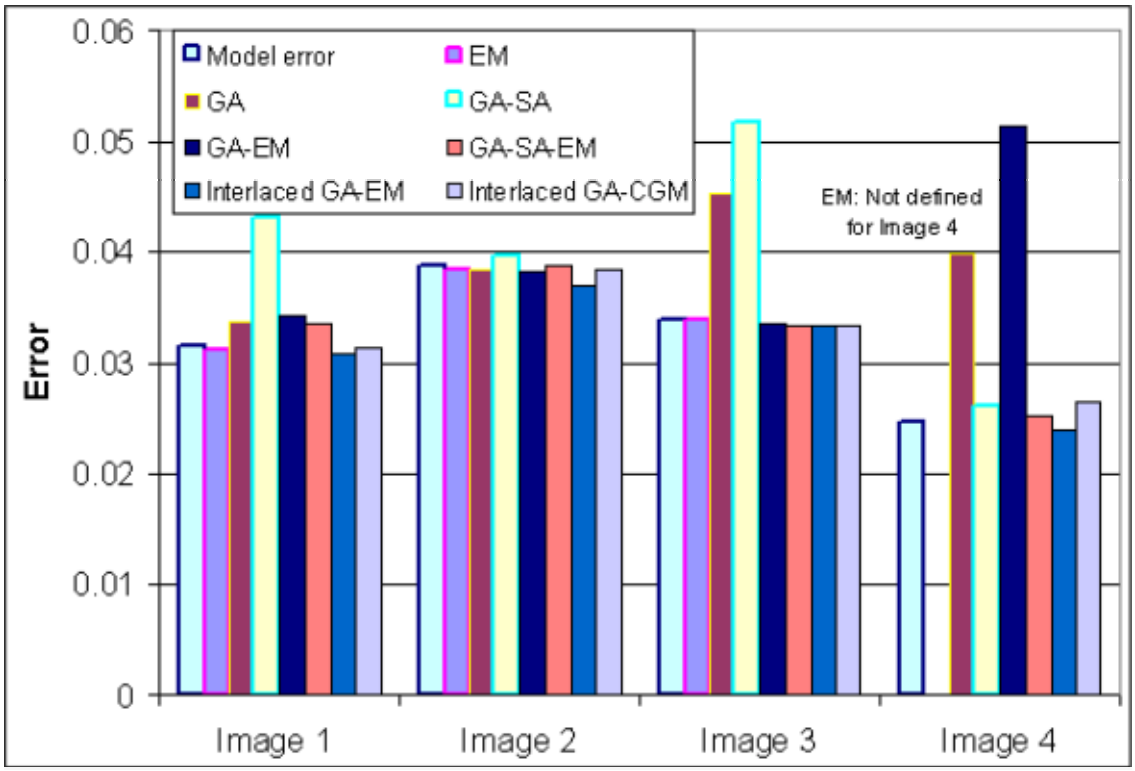


Figure 2.7. Errors of parameter estimation of EM, GA, GA-EM, GA-SA, GA-SA-EM, Interlaced GA-EM, and Interlaced GA-CGM methods for four sets of ‘ the Modified Shepp-Logan’ phantoms.

The proposed methods offer accurate and stable technique for estimating parameters of the Gaussian mixture models, with higher likelihood of achieving global optimal. The Interlaced GA-EM method offers an accurate parameter estimation for the Gaussian mixture models, with a higher chance of achieving global optimal minima. Obtaining such accurate parameter estimations is a key requirement for several image segmentation approaches, which rely on *a priori* knowledge of tissue distribution.

3. MARKOV RANDOM FIELD MODEL, ALGORITHMS ON IMAGE SEGMENTATION

This chapter describes Markov Random Field (MRF) model for image segmentation. We review some previous works on image segmentation by MRF and describe our algorithms and their simulation results. In section 3.1 Markov random field concepts and definition are introduced. In section 3.2 definitions of MRF methods on multiresolution in image segmentation are described. In section 3.3 our methods and their mathematical concepts are introduced. In section 3.4 contains simulation results on both accuracy of result and computational time on a series of the Gaussian phantoms, based on a modified Shepp-Logan method of CT images. Section 3.5 provides a discussion and conclusion on proposed algorithms.

3.1 Image model

In this chapter the label field is denoted **X** and the observed image is denoted **Y**. Location of each element in the spatial space is defined in the rectangular pixels of *S* on **X** and **Y** in which $(i, j) \in S$. The random variable is denoted by $X(i, j)$ or $X_{i,j}$. This notation is also true for **Y**. To simplify the notation, the random variables in the **X** and **Y** may also be single letters, as in a lexicographical ordering, in which case the *s*th random variable in **X** is X_s. Through out this thesis, sample realizations of both **X** and **Y** are shown with small letters and italics so that $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ and $\mathbf{y} = (y_1, y_2, \ldots, y_n)$ representing sample realization of $\mathrm{X} = (X_1, X_2, \ldots, X_n)$ and $\mathrm{Y} = (Y_1, Y_2, \ldots, Y_n)$, where *n* is the total number of pixels in *S*. The space of all possible realizations of **X** and **Y** will be denoted Ω_x and respectively Ω_y.

3.1.1 Markov random field model

In this section we review Markov random field model concepts. The specific model is defined and will be used throughout this thesis. More details about MRF models may be found in [8, 54] for interested readers.

3.1.1.1 Graphs and pyramidal structures

A graph structure on a set of image sites allows representation of local properties in images. More specifically, finite simple non-directed graphs provide the natural tool to support the definition of the local Markov property.

1) Neighborhood system and graph: A simple finite non-directed graph $G = (S, \eta)$ is defined on a finite site set *S* and a collection $\eta = \{\eta_s, s \in S\}$ of nonempty subsets of *S* such that:

$$\begin{cases} i) \ \forall s \in S, s \notin \eta_s, \\ ii) \ \forall \{s_1, s_2\} \subset S, s_1 \in \eta_{s_2} \Leftrightarrow s_2 \in \eta_{s_1}). \end{cases} \tag{3.1}$$

This collection η is referred to as a neighborhood system on s [55].

In image analysis, the most common structure is the Two-Dimensional (2-D) lattice (*i.e.*, a finite subset of Z^2) defined with 4-connexity or 8-connexity neighborhoods.

2) Trees and pyramidal graphs: A chain of *G* is a sequence $(s_1, ..., s_n)$ of successively neighboring sites $(s_{i+1} \in \eta_{s_i}$ for $i = 1, ..., n-1)$. In image analysis, quad-tree, in which each site has four children, is commonly used. Figure 3.1 shows an example of a quad-tree. The extra connections may concern only pairs of sites of the same level [56].

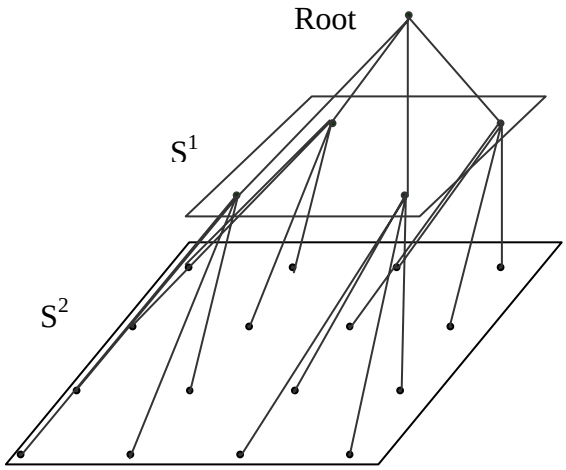


Figure 3.1. Quad-tree.

3.1.1.2 Markov random fields on graphs

Definition: Let Γ be a finite set, with probability of occurrence P, representing the possible values for the image feature at each site (gray levels for example). A *random field* on S with single site state space Γ is a vector $X = (X_s)_{s \in S}$ indexed by sites of S, and defined by a probability space (Ω, Γ, P), where Ω is the set of all possible configurations.

A realization of X (usually called a configuration) will be denoted by $x = (x_s)_{s \in S}$. A random field X is a Markov random field with respect to η, or on graph $G = (S, \eta)$ if and only if, $\forall s \in S, \forall x \in \Omega$:

$$\begin{aligned} &i)\ P\{X = x\} > 0, \\ &ii)\ P\{X_s = x_s \,|\, X_r = x_r, r \in S - \{s\}\} = P\{X_s = x_s \,|\, X_r = x_r, r \in N_s\}. \end{aligned} \tag{3.2}$$

The Markov property, Equation 3.2-*i*, is always assumed to be true for a random variable on a 2-D image and the Markov property in Equation 3.2-*ii* states certain neighborhood knowledge of random variables x_s and is sufficient to capture the local conditional distribution of x_s given realizations of all the other components of X.

A fundamental characterization of MRF is given by Hammersley and Clifford's theorem [7], in term of global distribution $P\{X = x\}$. This theorem states that *X* is a MRF with respect to η if and only if it follows a Gibbs distribution [54]:

$$P\{X = x\} = \frac{1}{Z}\exp\{-U(x)/T\}, \forall x \in \Omega \tag{3.3}$$

where *Z* is a normalization constant (also known as patrician function), *U* is the energy function, and *T* stands for temperature, a variable used to control the degree of change in the estimated probability function. The energy function is commonly expressed by:

$$U(x) = \sum_{c \in C} V_c(x) \tag{3.4}$$

where V_c is a clique potential (function) associated with clique order and depends only on variables $\{x_s, s \in C\}$. A clique is a non-empty subset *c* of *S*. It consists of either a single site $c = \{i\}$, or a pair of neighboring sites $c = \{i, i'\}$, or a triple of neighboring sites, and so on $c = \{i, i', i''\}$ [57].

The information of interest for image segmentation is modeled by a random vector (unobserved image) $X = (X)_{s \in S}$, the probability distribution function of which depends on an observed image, Y, as well as on *a priori* information (parameter vector) θ. The observed image is either directly measured (*e.g.* one or several images) or obtained as the output of some low-level image pre-processing. Given y and θ, X is assumed to be Markovian with respect to a relevant graph structure on S. Its posterior distribution is, therefore, of the following form:

$$P\{X = x \mid y, \theta\} = \frac{1}{Z(y,\theta)}\exp\{\underbrace{-\frac{1}{T}\sum_{c \in C} V_c(x; y, \theta)}_{U(x;y,\theta)}\} \tag{3.5}$$

Equation 3.5 leads to an optimization problem, Equation 3.6, by relating it to the standard Maximum A Posteriori (MAP) estimator for which the "best" estimate is the most likely, given y and θ on the (finite) set of all possible configurations of x, Λ.

$$\begin{cases} P(x \mid y, \theta) \propto \exp\{-U(x; y, \theta)\} \\ \textit{Find} \qquad \arg\min_{x \in \Lambda} U(x; y, \theta) \end{cases} \tag{3.6}$$

The problem can be formulated as optimizing criteria, explicitly or implicitly. The extensive use of optimization principles is due to various uncertainties in imaging modalities. Noise and other degradation factors are sources of uncertainties. We can generalize three basic issues in the optimization-based approach for image analysis [6]:

- *Representation*
- *Objective function (Cost function)*
- *Optimization algorithm.*

Representation concerns how to present image features and object shapes. In this thesis, for image segmentation, we use a chain of boundary locations to represent the solution.

Objective function is how to formulate an objective function for the optimization. The objective function maps a solution to a real number measuring the quality of the solution in terms of some goodness or cost.

Optimization algorithm is how to search for optimal solution in the admissible space. Two major concerns are (1) the problem of local minima existing in the non-convex functions and (2) the efficiency of algorithms in space and time. Our proposed algorithms (both MR and SA-RCF) for image segmentation are intended to overcome *objective function* and *optimization algorithm* issues. In section 3.3, MR and SA-RCF methods will be explained.

Sampling and global minimization of the posterior distribution are intimately interleaved, although to some extent distinct [58]. For instance, they can be approached using iterative methods sharing the same philosophy. The resolution of both these issues can be considered challenging since Λ is in general extremely large (or of large dimension), and the energy function may be non-convex. An extremely large solution space makes the computation of the partition function $Z(y,\theta)$ intractable in practice. One possible approach, the maximization of the *posterior* distribution, can be carried out with iterative relaxation algorithms, which take advantage of the local dependency structure of the energy to operate low-cost local updating. The methods discussed in the following sections fall into the general category of iterative relaxation approaches.

3.2 Multiresolution on MRFs

There are now a wide variety of image segmentation techniques, some considered general purpose and some designed for specific classes of images [57-59]. All of the images processing algorithms try to increase the accuracy of the image processing task while reducing computational processing. Multiresolution techniques first aimed at increasing the computational speed of relaxation algorithms in numerical analysis [60]. There is a strong need to reduce the computational complexity of most image processing algorithms in numerical analysis [61]. In the following sub-section, we are going to review the two categories that fall into hierarchical Markov random field models. Those are Induced and Explicit models.

3.2.1 Induced hierarchical model

The renormalization method widely used in statistical mechanics [11] is based on a hierarchical conditional random process. Site sets for such a hierarchy are S^0, $S^1,...,S^n$ and $S^n \equiv S$ (with $|S^0| < ... < |S^n|$ and S^n is finest resolution); new label fields (X^{n-1} *on* $S^{n-1},...,X^0$ *on* S^0) are defined from $X^n \equiv X$ through fine-to-coarse transition probability laws $P\left\{X^{k-1} = x^{k-1} \middle| X^k = x^k\right\}$ between two consecutive random processes.

Decimation and block scaling are particular cases of this general procedure, although the most important ones. Decimation successively restricts the random process to a decreasing sequence of subsets of the original lattice S ($S^{k-1} \subset S^k$):

$$P\left\{X^{k-1} = x^{k-1} \middle| X^k = x^k\right\} = \prod_{s \in S^{k-1}} \delta(x_s^{k-1}, x_s^k) \qquad (3.7)$$

where $\delta(a,b) = 1$ if a = b, and 0 otherwise. In block scaling, each set S^k is partitioned into "blocks", each of them corresponding to a unique site of S^{k-1}. Then for each site *s* of S^{k-1}, the attached random variable X_s^{k-1} is the only function of components of X^k lying within the associated block $B^k(s) \subset S^k$. For instance, block averaging is defined as the following process:

$$P\left\{X^{k-1}=x^{k-1}\middle|X^{k}=x^{k}\right\}=\prod_{s\in S^{k-1}}\delta(x_s^{k-1},\frac{1}{\left|B^k(s)\right|}\sum_{r\in B^k(s)}x_s^k) \quad (3.8)$$

In all cases, the induced posterior probability law at level *k*-1 is given by [20, 55, and 62]:

$$P\left\{X^{k-1}=x^{k-1}\middle|Y=y\right\}=\sum_{x^k}\left(P\left\{X^{k-1}=x^{k-1}\middle|X^{k}=x^{k}\right\}\right)\times P\left\{X^{k}=x^{k}\middle|Y=y\right\} \quad (3.9)$$

The hierarchical MAP estimation method based on these transformations generally transforms the energy function of fine resolution at all scales; then optimization algorithms are applied to the coarse resolution, and the result of each resolution passes to the next fine resolution in the order of coarse-to-fine optimizations:

$$\hat{x}^k=\underset{x^k\in\Lambda^k(\hat{x}^{k-1})}{\arg\min}\ U^k(x^k;y),\quad \text{with}\ \Lambda^k(x^{k-1})\equiv\underset{x^k}{\arg\max}\ \Pr\left\{X^{k-1}=x^{k-1}\middle|X^{k}=x^{k-1},y\right\} \quad (3.10)$$

Stochastic optimization methods such as Simulated Annealing (SA) are generally used to this end although deterministic schemes have also been proposed [11, 63]. Significant gains on CPU time have been observed, with respect to standard (non-hierarchical) optimization techniques.

Inspired from multigrid techniques in numerical analysis, the use of nested constrained subspaces has been proposed in [34, 64]. This method allows defining consistent hierarchical MRFs starting from a single full-resolution MRF. The principle of such subspaces consists in considering a sequence of embedded configuration subspaces $\Lambda^0\subset\Lambda^1\subset\ldots\Lambda^n\equiv\Lambda$. The MAP estimation, optimization problem, is successively considered within these nested configuration sets:

$$search\ \hat{x}^{(k)}=\underset{x\in\Lambda^k}{\arg\min}U(x;y)\qquad \text{from}\ k=0\ \ to\ \ k=n. \quad (3.11)$$

The algorithm is associated with these nested configuration spaces by searching Λ^0 for a configuration $\hat{x}^{(0)}$ that minimizes *U(x;y)*. The estimated field $\hat{x}^{(0)}$ is next used as an initial guess for the optimization in Λ^1. Because $\Lambda^0\subset\Lambda^1$, Λ^1 is then searched for a minimum $\hat{x}^{(1)}$ until the initial configuration space $\Lambda^n\equiv\Lambda$ is reached. The final estimate

$\hat{x}^{(k)}$ at level *k* corresponds to a local minimum of the initial energy function $U(x; y)$ (generally of low energy).

This optimization procedure on nested constrained configuration spaces is easily transformed into a coarse-to-fine optimization procedure on a hierarchy of MRF models defined by multigrid energy functions $U^k(.;y)$, *k=0,…,n*. Any $x^{(k)} \in \Lambda^k$ can, indeed, be completely described through a reduced set of variables $x^k = \{x_s^k, s \in S^k\}$, with $|S^k| < |S|$, interacting with original data *y* according to a "coarse" energy function derived from *U(x;y)* by:

$$U^k(x^k; y) \equiv U(x^{(k)}; y) \tag{3.12}$$

So the sequence of a coarse-to-fine optimization problem can be expressed as:

$$\hat{x}^k = \arg\min_{x^k} U^k(x^k; y) \text{ from } k = 0 \text{ to } k = n. \tag{3.13}$$

An ordered family of constrained configuration spaces of the above model was first considered by Bouman *et al*. [34] for texture segmentation. Bouman *et al.* [15, 34] used causal Auto Regressive (AR) models to describe texture information within regions. Bouman proposed the following recursion for horizontal, vertical or diagonal neighboring pixels (in practice, different parameters $\beta_{hv} > 0$, $\beta_d > 0$ are specified for horizontal or vertical cliques and diagonal cliques respectively) for specifying the parameters of energy function U^k associated to the MRF at level *k*:

$$\begin{aligned} \beta_{hv}^{(k)} &= 2(\beta_{hv}^{(k+1)} + \beta_d^{(k+1)}) \\ \beta_d^{(k)} &= \beta_d^{(k+1)} \end{aligned} \tag{3.14}$$

However, the authors [34] eventually maintained constant interaction parameters through scale in their implementations (i.e., $\beta^{(k)} = \beta$ for every *k*).

Experiments in different application fields show that the constrained configuration-based hierarchical modeling approach yields fast optimization schemes, significantly faster than other standard optimization procedures such as ICM [65] or stochastic relaxation [10]. Typically, when deterministic minimization procedures are used to

perform the search of $\hat{x}^k$, $k=0,\ldots, n$ (Equation 3.13), configurations close to those obtained by stochastic methods are reached with the hierarchical approach, while the convergence is much faster.

3.2.2 Explicit hierarchical model

For explicit hierarchical-based models, the hierarchy is embedded in the model definition itself. Single-model approaches essentially rely on a heuristic association of MRF models with multiresolution decomposition of image data. Multiresolution image transforms based on the Gaussian filters or wavelet decompositions are used to decompose the image data y on a pyramidal structure, providing a data set $(y^0, y^1, ..., y^n \equiv y)$. Given the original MRF model of energy $U(x; y)$, a similar reduced MRF model of energy $U^k(x^k; y^k)$ is then defined at each resolution level k (same neighborhood structure and same potential functions as the original model, *i.e.*, U^k "similar" to U). Standard stochastic or deterministic relaxation algorithms provide Bayesian estimates of the label field at each resolution level. These algorithms are generally associated with a coarse-to-fine exploration of the hierarchical structure, starting from the lowest resolution and propagating the estimates from coarse-to-fine scales. In most of the cases, the final estimate obtained at a given level is interpolated and used as an initialization for the relaxation process at the next finer level [66, 67].

$$\hat{x}^k = \arg\min_{x^k} U^k(x^k; y^k) \text{ and } \text{Interpol}_{k\to k+1}(\hat{x}^k) = \text{initial guess at level } k+1. \quad (3.15)$$

This algorithm is very time consuming but has provided results of high quality in supervised and unsupervised image classifications [68].

3.3 Random cost function as a robust solution on MRF for image segmentation

Through the past sections we defined MRF concepts and MRF's two-principle category for image segmentation. In the following sub sections, 3.3.1 and 3.3.2, our algorithms for image segmentation are introduced. We might embed these algorithms in the Induced rather than Explicit category, however, they are completely new and no work

has been done in this field. We call those algorithms Multi Resolution and Simulated Annealing-Random Cost Function (MR and SA-RCF respectively). We introduce these algorithms and we make a comparison between these algorithms with two other types of algorithms, Simulated Annealing (SA) [10] and Minimum Gray Level Distance (MGLD) [37], which are introduced in subsections 3.3.3 and 3.3.4. The segmentation accuracy of each algorithm is evaluated by set of CT phantoms.

3.3.1 Multi Resolution (MR), mathematical concept and algorithm

Throughout the following sections we use the same definition and notation of Markov model that were presented in last sections. If there is any change in notations, we mention it when we use it.

Let $\eta_u \subset S$ be the neighborhood of *u*th site such that for all $u \notin \eta_u$ and $z \in \eta_u \Leftrightarrow u \in \eta_z$. An example for η is illustrated in Figure 3.2. A first-order MRF is one for which the neighbor set consists of four pixels constituting the north, south, east, and west neighbors of a given pixel. A second order MRF is that for which the neighbor set is the first layer of eight pixels surrounding the center pixel, *i.e.* all pixels marked 1 or 2. A third-order neighbor set consists of all pixels marked 1, 2, or 3, etc. We have used a second order clique for all of our simulations in this thesis.

5	4	3	4	5
4	2	1	2	4
3	1	x_u	1	3
4	2	1	2	4
5	4	3	4	5

Figure 3.2. Clique orders around *u*th pixel (site).

It is assumed that the unobserved image is corrupted by additive noise which can be expressed by $Y_{ij} = F(X_{ij}) + W_{ij}, (i, j) \in S$, where $F(\cdot)$ is a simple mapping of the region type to the corresponding gray-level and W_{ij} is i.i.d (independent identical distribution)

Gaussian noise with zero-mean and variance σ^2, *i.e.*, $W_{ij} \approx N(0,\sigma)$. It is also assumed that θ, *a priori* information is known for image segmentation in this section.

MR performs segmentation on the coarse resolution. The fine (intensity) image specifies the gray levels for all pixels defined as $(2^m+1)\times(2^{m'}+1)$ in the rectangular lattice ($M \times N$). The lattice S is $S = \{s=(p,q),\ p=0,...2^m, 0...2^{m'}\}$[11, 69, and 70]. We define coarse grids S^i by:

$$S^i = \{s=(2^i k, 2^i k'),\ k=0,...2^{m-i},\ k'=0,...2^{m'-i}\} \tag{3.16}$$

Γ^i denotes the space of configurations defined on S^i. The neighborhood of a single site $s=(p,q)\in S$ at S^i is defined as $s_1=(p',q')$, $s_2=(p',q'+2^i)$, $s_3=(p'+2^i,q')$, $s_4=(p'+2^i,q'+2^i)$ with $p'=2^i\,Int(p/2^i)$ and $q'=2^i\,Int(q/2^i)$ where $Int(\psi)$ designates the integer part of ψ. See Figure 3.3.

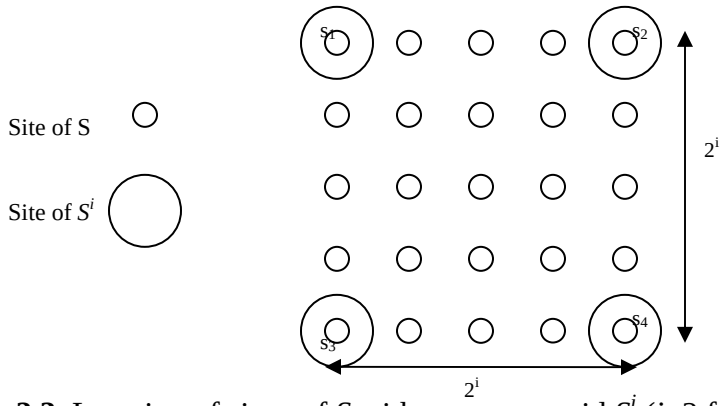


Figure 3.3. Location of site *s* of *S* with respect to grid S^i (*i*=2 for MR method).

Its label can define random variable x on S^i:

$$\begin{aligned} \Lambda^i &\xrightarrow{\Phi^i} \Omega^i \\ x &= \Phi^i(x^i) \end{aligned} \tag{3.17}$$

If *y* has a homogenous block $2^i \times 2^i$ distribution, the *y* can be calculated from $y=\Phi^i(y^i)$ directly. The pixel relation on an observed image at each grid is shown in Figure 3.4.

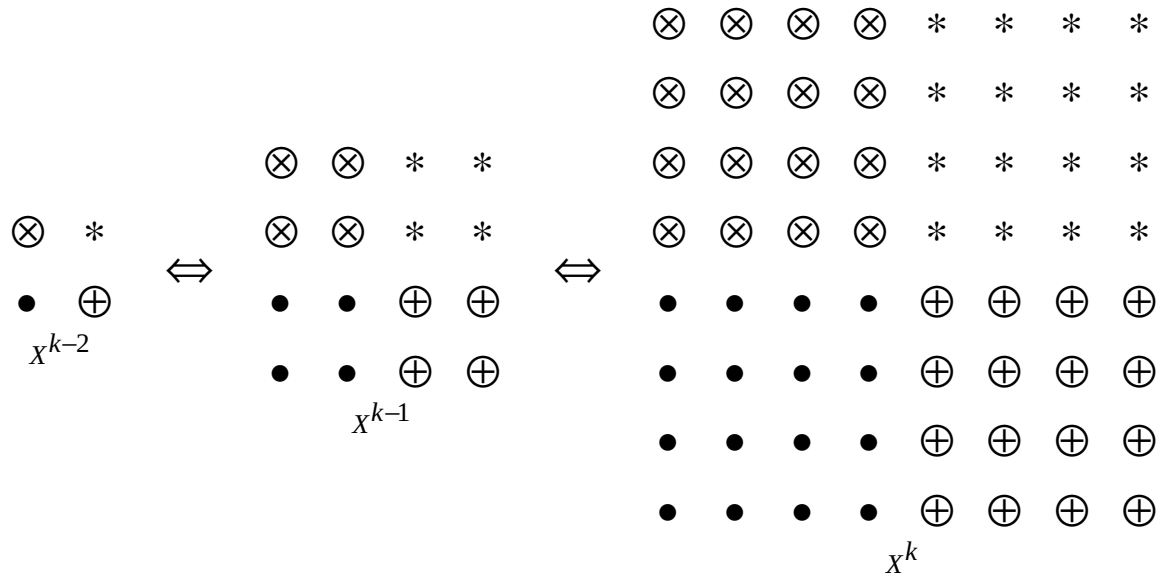


Figure 3.4. Pixel relations on different resolutions.

We are looking to find out $\hat{x}^{(k-2)}$ on MR. From equation 2-9 we learn:

$$P(X^{k-2} \mid y) = [\sum_{x^k} P(X^{k-2} \mid X^k)]P(X^k \mid y) \tag{3.18}$$

and we obtain:

$$\begin{aligned} \hat{x}^{(k-2)} &= \arg\max_{x^2} P(X^{k-2} \mid y) \\ &= \arg\max_{X^2}([\sum_{x^k} P(X^{k-2} \mid X^k)]P(X^k \mid y)) \\ &= \arg\max([\sum_{x^k} \prod_{s\in S^{k-2}} \delta(X_s^{k-2}, \frac{1}{|B^k(s)|} \sum_{r\in B^k(s)} x_r^k)]P(X^k \mid y)) \end{aligned} \tag{3.19}$$

where $B^k(s)$ is a block of pixels at *k* resolution that is related to site *s* to *k-2*th resolution. Since we have assumed that the 2×2 blocks are homogenous we can expect to obtain $\hat{x}^k$ from $\hat{x}^{k-2}$ close to its true values but this hypothesis is not always true because we always have some noises or in homogeneity in the observed image. As a matter of fact, the true demonstration of Figure 3.4 is most likely similar to Figure 3.5.

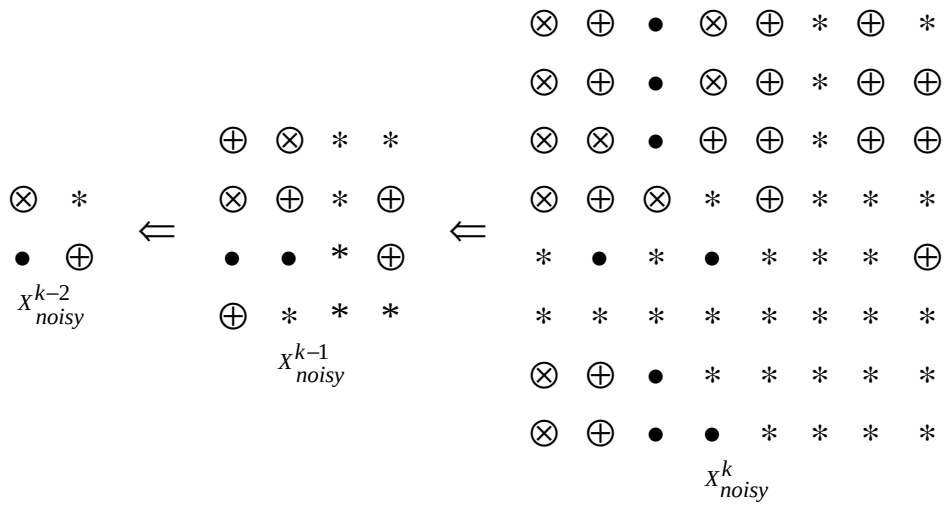


Figure 3.5. Pixel relations on different resolutions, in the presence of noise.

As we observe through Figure 3.4 and figure 3.5, we find that X^{k-2}_{noisy} and X^{k-2} are similar; however, both X^{k}_{noisy} and X^{k} are completely different. To overcome this problem we define conditional probability of randomness selection:

$$P(X^{k}_{s\ rand} \mid y) \tag{3.20}$$

where $X^{k}_{s\ rand}$ is a uniform random selection, equally distributed, $\frac{1}{16}$, of neighborhood of single site of *s* at the *k*th resolution for each $B^{k}(s)$.

We claim that for a degraded image:

$$P(X^{k} \mid y) \le P(X^{k}_{s\ rand} \mid y) \tag{3.21}$$

is always true.

Proof: Both $P(X^{k}_{s\ rand} \mid y)$ and $P(X^{k} \mid y)$ belong to the same resolution and each of them presents the same rule as another, but the probability of $P(X^{k}_{s\ rand} \mid y)$ is higher than $P(X^{k} \mid y)$ at iterative mode. As we saw in Figure 3.3, for the probability $P(X^{k} \mid y)$, s_1 represents one of $2^i \times 2^i$ (for $i = 2 \rightarrow 16$) pixels. That is, the probability that s_1 presents

the correct pixel in the related $2^i \times 2^i$ block at *k* resolution is $\frac{1}{16}$ and through whole of the observed image is $\frac{1}{16}^{\frac{m}{2^i}\times\frac{m'}{2^i}}$. For the probability $P(X_{s\ rand}^{k} \mid y)$, this value is $\frac{u}{16}^{\frac{m}{2^i}\times\frac{m'}{2^i}}$ where *u* is the number of pixels that were not affected by noise and $u \geq 1$ is always true. So we can obtain $\hat{x}^{(k-2)}$ from equation 3.22 with higher accuracy.

$$\hat{x}^{k-2} = \arg\max([\sum_{x^k} \prod_{s\in S^{k-2}} \delta(X_s^{k-2}, \frac{1}{|B^k(s)|} \sum_{r\in B^k(s)} x_r^k)].P(X_{s\ rand}^{k} \mid y)) \tag{3.22}$$

The maximization of the argument of equation of 3.22 is the minimization of $U^{k-2}(x)$. $U^{k-2}(x)$ is the maximum likelihood of the Bayesian relation of observed and unobserved image [71]:

$$P(X \mid Y) = \frac{P(Y \mid X)P(X)}{P(Y)} \tag{3.23}$$

Thus we can express the equivalent minimization of posterior energy function of $U(x \mid y)$ for equation 3.22:

$$\begin{aligned} \hat{x}^{k-2} = \arg\min_x\{U(x_{s\ rand}^{k-2} \mid y)\} &= \arg\min_x\{U(y \mid x_{s\ rand}^{k-2}) + U(x^{k-2})\} \\ &= \sum_{t\in\Omega}\sum_{s\in S^k} f(y_t \mid x_{s\ rand}^{k-2}) + \beta\sum_{c\in C} V_c(x^{k-2}) \end{aligned} \tag{3.24}$$

where f, β, T are the cost function, a fixed control parameter, and temperature, respectively. The first term in Equation 3.24 reflects a penalty for misclassification, while the second term represents smoothness constraints. Temperature and parameter, β, offer mechanisms for adjusting the relative importance of smoothness of the final solution versus labeling error. In this work, we used a piecewise linear potential function for each label (Figure 3.6). Vector θ contains 3*l*+1 parameters: *l* labels, *s*, and *d* (deviation parameters for each label), and a fixed β for entire of image. The vector of θ has three elements for each class, a label and two constant deviation parameters, *s* and *d* [72], with *s* and *d* defined experimentally. If the distance between a given data pixel and the label of the *l*th class is lower than s_l, then there is a high probability that this pixel belongs to the

class *l*. In contrast, if the distance between the data pixel and the label of the *l*th class is greater than d_l , then there is a low probability that this pixel belongs to the class *l*.

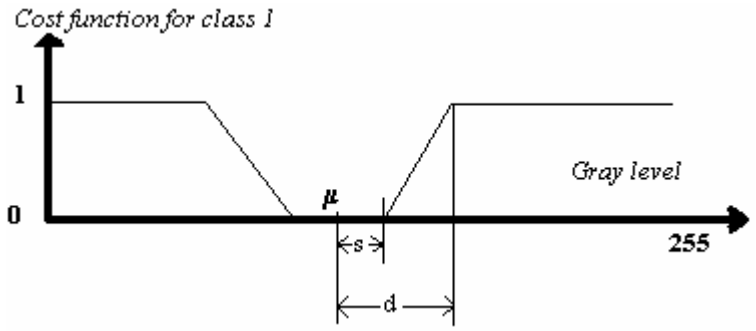


Figure 3.6. Cost function defining the data-driven term for one typical label.

The steps involved in MR algorithm can be summarized as the following:

1. Choose an initial temperature T_o .
2. Initialize $\hat{x}^{k-2}$ randomly.
3. Perturb $\hat{x}^{k-2}$ into $\hat{z}^{k-2}$ and set a pointer, h, randomly between 1 through 16 with equal probability for each pixel per iteration.

   Let: $\Delta = U(\hat{z}^{k-2}_{h\ rand} \mid y) - U(\hat{x}^{k-2}_{h\ rand} \mid y)$. If $\Delta < 0$ then replace $\hat{x}^{k-2}$ by $\hat{z}^{k-2}$; else replace $\hat{x}^{k-2}$ to $\hat{z}^{k-2}$ with probability $e^{-\Delta / T_k}$.
4. Replace T_k by T_{k+1} .
5. Repeat 3-4 until no changes happen, but stop after $Itr_{\max}$ iteration.
6. Interpolate $\hat{x}^{k-2}$ to size of $\hat{x}^k$ at fine resolution.

$Itr_{\max}$ was set to 700. The choice of cooling function defines the cooling schedule. Experimentally, the initial temperature was selected T_o =4 and $T_{k+1} = 0.95T_k$ [57].

3.3.2 Simulated Annealing – Random Cost Function (SA-RCF) algorithm

While MR applies to coarse (*k-2*th) resolution, SA-RCF applies only to fine (*k*th) resolution. In SA-RCF each iteration for all pixels of the observed image, a pixel is selected randomly from a block of 3×3 pixels in its neighborhood. It is hypothesized that

such a uniform random selection can increase robustness of the algorithm when the image is corrupted by noise. It is believed that in the iterative mode, the chance of finding a pixel in its 3×3 matrix neighborhood that was not corrupted or was less affected by noise is higher. Then, equation 3.24 can be modified as:

$$\hat{x}^k = \arg\min_x\{U(x^k_{s\ rand\,(3\times3)} \mid y)\} = \arg\min_x\{U(y \mid x^k_{s\ rand\,(3\times3)}) + U(x^k)\} \qquad (3.25)$$
$$= \sum_{t\in\Omega}\sum_{s\in S^k} f(y_t \mid x^k_{s\ rand\,(3\times3)}) + \beta\sum_{c\in C} V_c(x^k)$$

where $x^k_{s\ rand\,(3\times3)}$ refers to random selection of a grid from a block 3×3 surrounded around s. We have named this method Simulated Annealing with Random Cost Function (SA-RCF).

The steps involved in SA-RCF algorithm can be summarized as the following:

1. Choose an initial temperature T_o.
2. Initialize $\hat{x}^k$ randomly.
3. Perturb $\hat{x}^k$ into $\hat{z}^k$ and set a pointer, h, randomly between 1 through 9 with equal probability for each pixel per iteration.

   Let: $\Delta = U(\hat{z}^k_{h\ rand\,(3\times3)} \mid y) - U(\hat{x}^k_{h\ rand\,(3\times3)} \mid y)$. If $\Delta < 0$ then replace $\hat{x}^k$ by $\hat{z}^k$; else replace $\hat{x}^k$ to $\hat{z}^k$ with probability $e^{-\Delta/T_k}$.
4. Replace T_k by T_{k+1}.
5. Repeat 3-4 until no changes happen, but stop after $Itr_{\max}$ iteration.

3.3.3 Simulated Annealing (SA) algorithm

We have used this method, which was introduced by Geman [10], to compare with our methods. SA applies to the *k*th resolution, and *a priori* information of the introduced algorithms is the same for SA-RCF. The cost function of SA is similar to equation 3.25 except that there is no random search at each site.

$$\hat{x}^k = \arg\min_x\{U(x^k \mid y)\} = \arg\min_x\{U(y \mid x_s{}^k) + U(x^k)\} \qquad (3.26)$$

$$= \sum_{t \in \Omega} \sum_{s \in S^k} f(y_t \mid x_s^k) + \beta \sum_{c \in C} V_c(x^k)$$

Computational steps and algorithm are identical to SA-RCF method as outlined before.

3.3.4 Minimum Gray Level Distance (MGLD) algorithm

MGLD tends to segment images by assigning a pixel value by a label, which has minimum difference with all labels, with the smoothness constraint imposed by the clique potential function. Intuitively, this approach is similar to K-means clustering [73] with an added smoothness constraint.

$$U(x^k \mid y) = \sum_{t \in \Omega} \sum_{s \in S} (y_t - x_s^k)^2 + \beta \sum_{c \in C} V_c(x^k) \tag{3.27}$$

Equation 3.27 calculates its maximum value l times (l distinct labels) for each pixel in the image per iteration. This procedure continues for all pixels per iteration.

3.4 Simulation results and discussion

To assess accuracy and computational efficiency of the proposed algorithms, we generated a series of CT head phantoms, using a Modified Shepp-Logan [47]. This is a variant of the Shepp-Logan phantom in which the contrast is improved for better visual perception. Throughout this thesis, we use these phantom images to evaluate our simulation results.

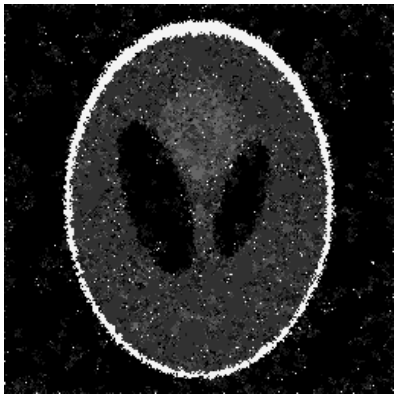
Additive white Gaussian noise was added to each image. The maximum number of iterations for MR, SA-RCF, and SA was set to 700. Sometimes algorithms converged at fewer than 700 iterations especially when the signal-to-noise rate of the image was high. MGLD ran until it converged, which happened at an iteration much lower than 700. The initial temperature was selected as 4.0 for all of the algorithms and *a priori* information was the same for all.

Table 3.1 summarizes the segmentation error of the complete set as a function of noise and β. β values are 0.15, 0.18, 0.2, 0.22, 0.25, and 0.4. The segmentation error is computed as the percentage of pixels misclassified per the total pixels in the image. For the original noise-free image, SA and MGLD achieved the lowest segmentation error among the four methods. As the noise increases, the segmentation errors for SA and

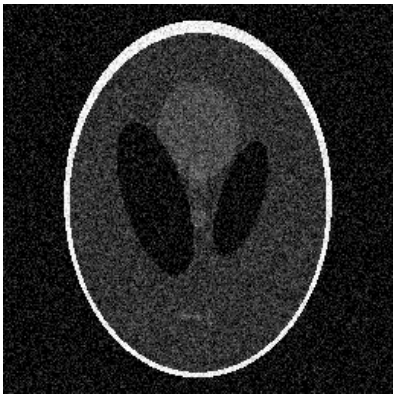
MGLD consistently worsen compared to those for MR and SA-RCF. It is also interesting to note that segmentation errors for noisy images decrease inversely with β. As was noted, β was introduced in equation 3.24 to control the relative contribution of smoothness constraint in the total posterior energy function. We expected to observe less dependency of SA-RCF and MR to changes in β for smaller values since these algorithms try to find the minimum energy by selecting the best pixel value inside of its neighborhood. We have done our simulations on noiseless images and for five CT phantoms that have degraded by different noise levels which are N(0,0.057), N(0,0.081), N(0,0.1), N(0,0.141), N(0,0.173). Since the result of each simulation for each image occupies lot of space in this thesis, we only posted the result of image segmentation for one noisy phantom image, which was corrupted with N(0, 0.081).

Segmentation performance was evaluated for all six phantoms with respect to noise and β. The dynamic range of original phantom image is among 0 and 1. The results of simulation are shown for MR, SA-FCF, SA, and MR for sets of β values. Segmentation results on noisy image N(0,0.0812) is shown from figure 3.7 to 3.12.

a. Original (Phantom) image.



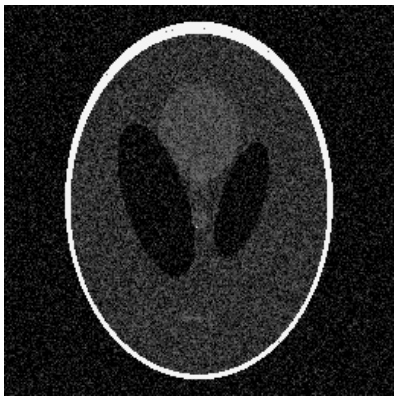
d. SA-RCF



b. Noisy image, N(0,0.081)

e. MR

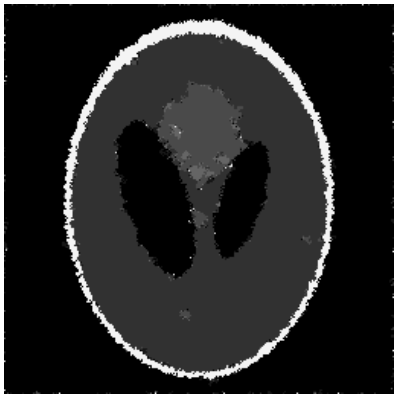
c. SA



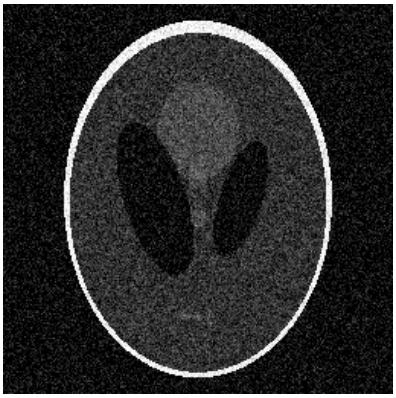
f. MGLD

Figure 3.7. a) Phantom image. b) Noisy image, phantom image corrupted by Gaussian white noise N (0,0.0812). c) Result of SA. d) Result of SA-RCF. e) Result of MR. f) Result of MGLD. β =0.15 was selected for all these images.

a. Original (Phantom) image.



d. SA-RCF



b. Noisy image, N(0,0.081)

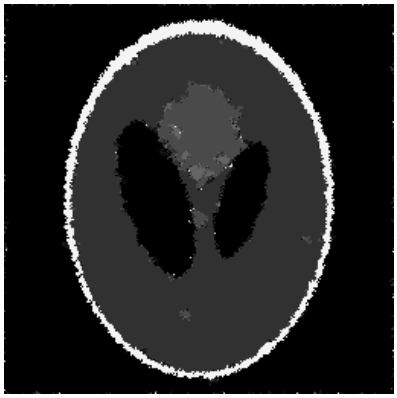
e. MR

c. SA

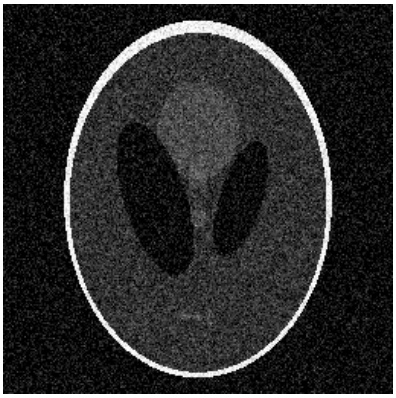
f. MGLD

Figure 3.8. a) Phantom image. b) Noisy image, phantom image corrupted by Gaussian white noise N (0,0.0812). c) Result of SA d) Result of SA-RCF. e) Result of MR. f) Result of MGLD. β =0.18 was selected for all these images.

a. Original (Phantom) image.



d. SA-RCF



b. Noisy image, N(0,0.081)

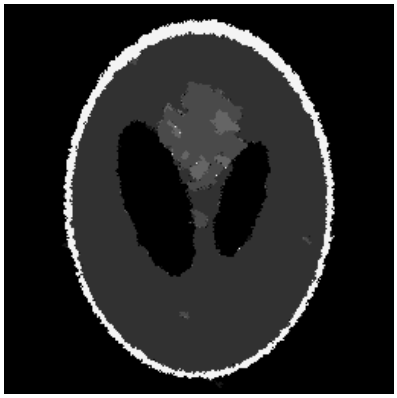
f. MR

c. SA

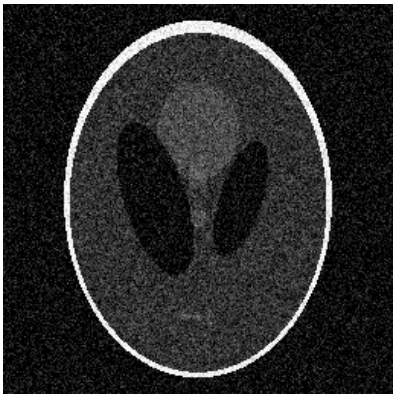
f. MGLD

Figure 3.9. a) Phantom image. b) Noisy image, phantom image corrupted by Gaussian white noise N (0,0.0812). c) Result of SA. d) Result of SA-RCF. e) Result of MR. f) Result of MGLD. β =0.20 was selected for all these images.

a. Original (Phantom) image.



d. SA-RCF



b. Noisy image, N(0,0.081)

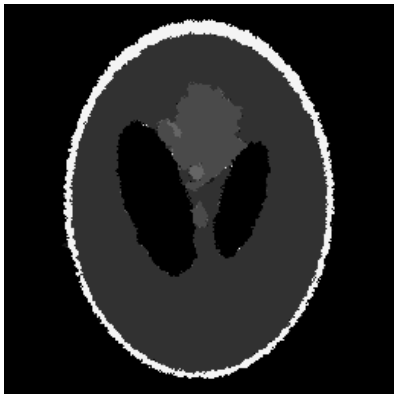
e. MR

c. SA

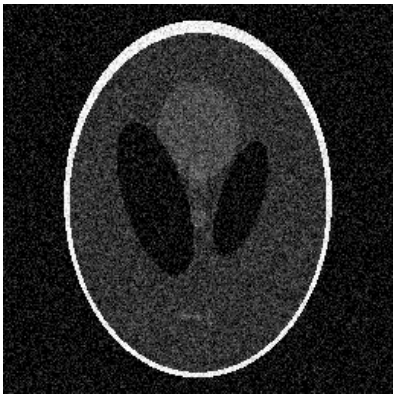
f. MGLD

Figure 3.10. a) Phantom image. b) Noisy image, phantom image corrupted by Gaussian white noise N (0,0.0812). c) Result of SA. d) Result of SA-RCF. e) Result of MR. f) Result of MGLD. β =0.22 was selected for all these images.

a. Original (Phantom) image.



d. SA-RCF



b. Noisy image, N(0,0.081)

e. MR

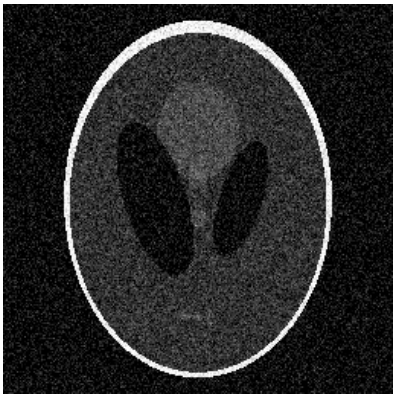
c. SA

f. MGLD

Figure 3.11. a) Phantom image. b) Noisy image, phantom image corrupted by Gaussian white noise N (0,0.0812). c) Result of SA. d) Result of SA-RCF. e) Result of MR. f) Result of MGLD. β =0.25 was selected for all these images.

a. Original (Phantom) image.

d. SA-RCF



b. Noisy image, N(0,0.081)

e. MR

c. SA

f. MGLD

Figure 3.12. a) Phantom image. b) Noisy image, phantom image corrupted by Gaussian white noise N (0,0.0812). c) Result of SA. d) Result of SA-RCF. e) Result of MR. f) Result of MGLD. β =0.4 was selected for all these images.

Through figure 3.13 and figure 3.18, we evaluate performance of four algorithms (MR, SA, SA-RCF, and MGLD) for a set of β and standard deviation (σ) of noise.

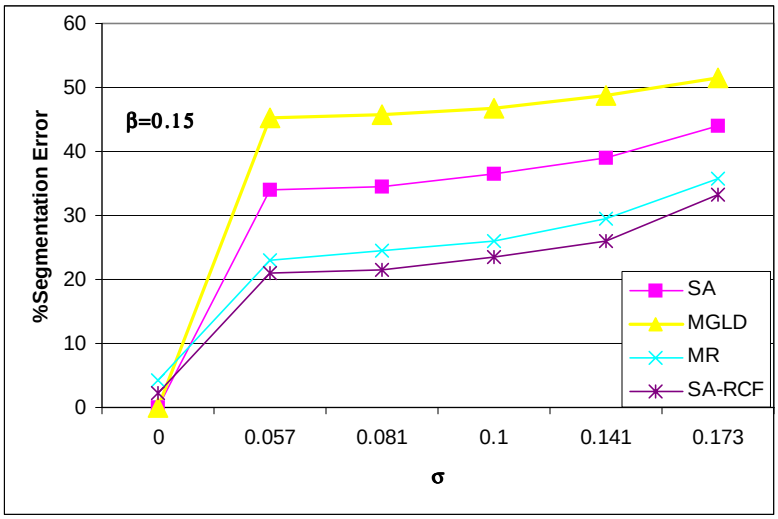


Figure 3.13. Segmentation accuracy of SA, MGLD vs. MR, and SA-RCF as a function of noise standard deviation for β=0.15.

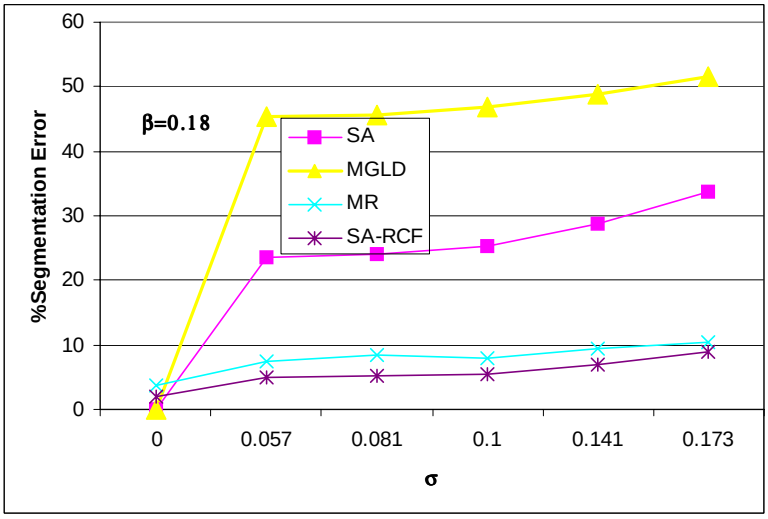


Figure 3.14. Segmentation accuracy of SA, MGLD vs. MR, and SA-RCF as a function of noise standard deviation for β=0.18.

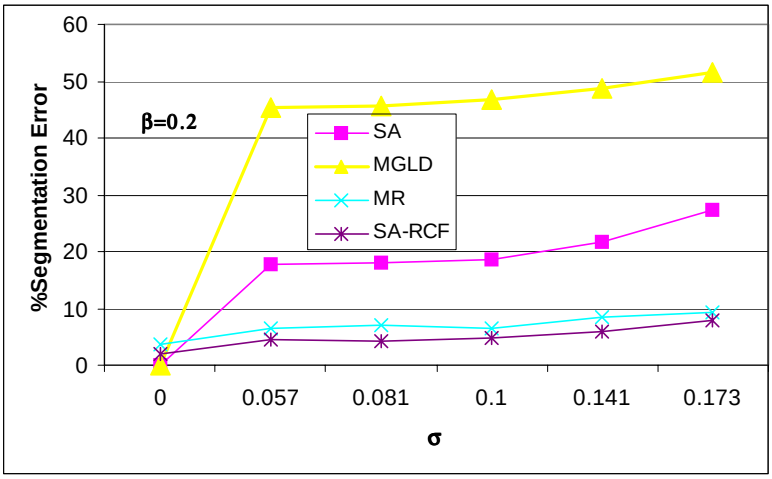


Figure 3.15. Segmentation accuracy of SA, MGLD vs. MR, and SA-RCF as a function of noise standard deviation for β =0.2.

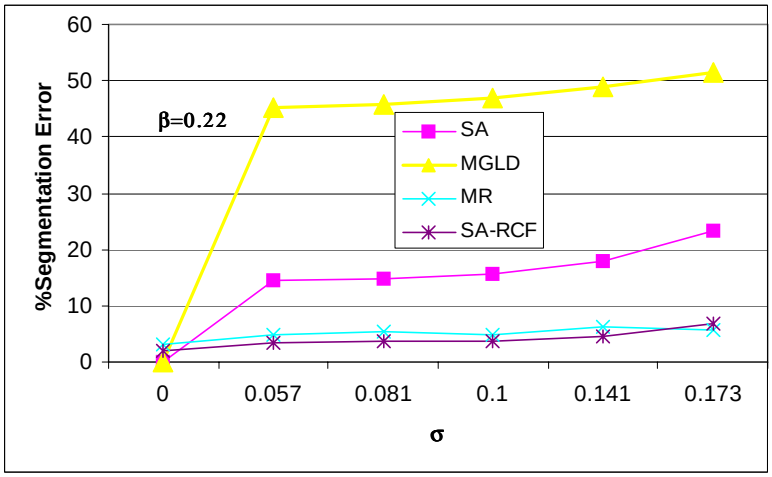


Figure 3.16. Segmentation accuracy of SA, MGLD vs. MR, and SA-RCF as a function of noise standard deviation for β =0.22.

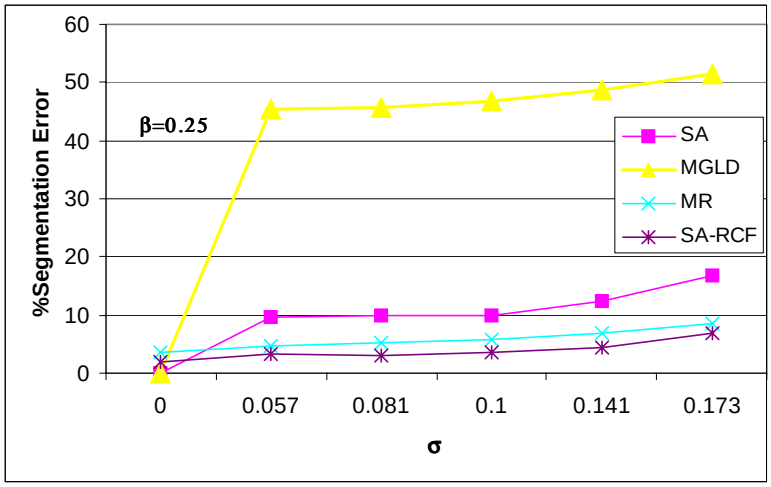


Figure 3.17. Segmentation accuracy of SA, MGLD vs. MR, and SA-RCF as a function of noise standard deviation for β=0.25.

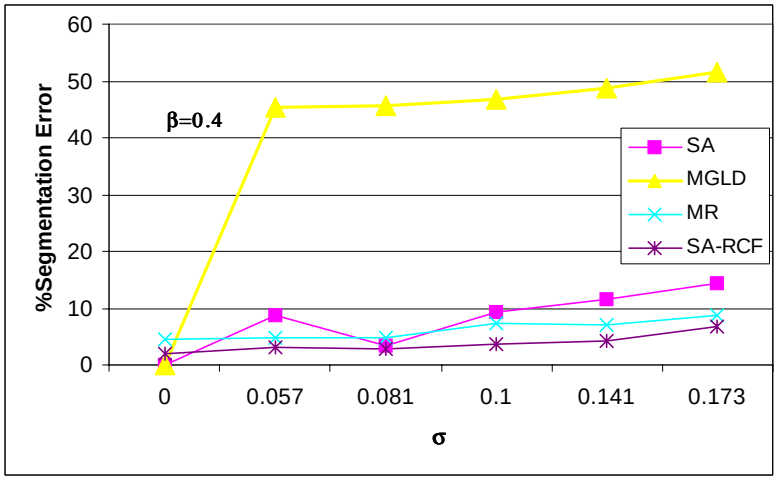


Figure 3.18. Segmentation accuracy of SA, MGLD vs. MR, and SA-RCF as a function of noise standard deviation for β=0.4.

Figure 3.19 demonstrates the percent of segmentation error based on different values of β for a noisy image which was corrupted by N(0,0.141). However, because figure 3.19, can be inferred from Figures 3.13 to 3.18, we posted only the result for σ = 0.141 to prevent showing redundant information.

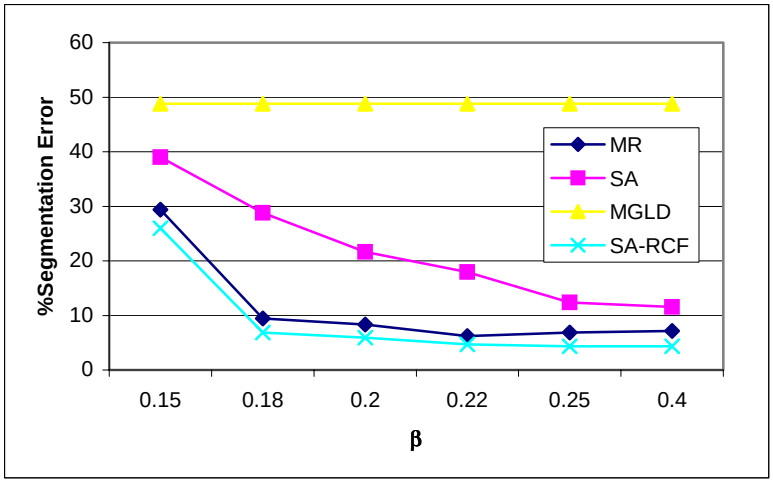


Figure 3.19. Result of segmentation error on a phantom image which was corrupted by N(0,0.141) vs. different β.

Since computational cost is almost completely independent of β, we set a stop value for iteration of our algorithm; we showed the computational cost of four algorithms for a set of β. Figure 3.20 shows the computational cost for each method in terms of the number of floating point operations (FLOP). The computational cost for three methods (SA, SA-RCF and MGLD) increases for noisy images. The computational cost of the proposed methods depicts a trade-off between SA and MGLD. Note that due to the wide range of difference in the computational cost among the four techniques, it is expressed in semi-log scale.

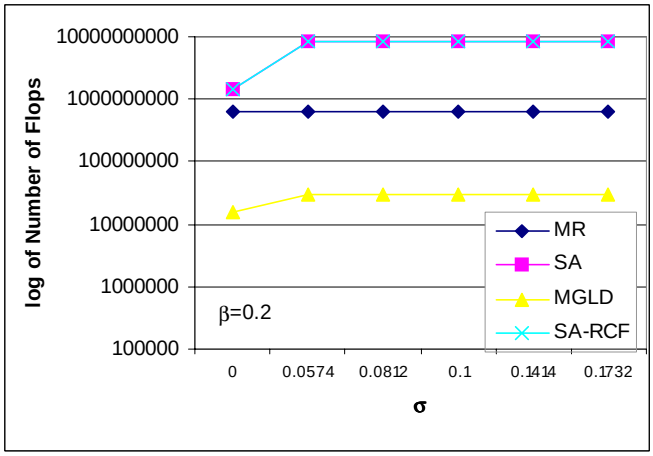


Figure 3.20. Computational cost of four methods studied in this work. Note that vertical axis is in log-scale.

Effects of different noise and β values for all four algorithms are shown separately in Figures 3.21 to 3.24. These figures show high dependency of SA on both β and σ and less dependency of MR and SA-RCF on higher β and σ. Figures 3.21 to 3.24 prove that our proposed algorithms (both MR and SA-RCF) are more successful than SA and MGLD in searching for minimum global energy.

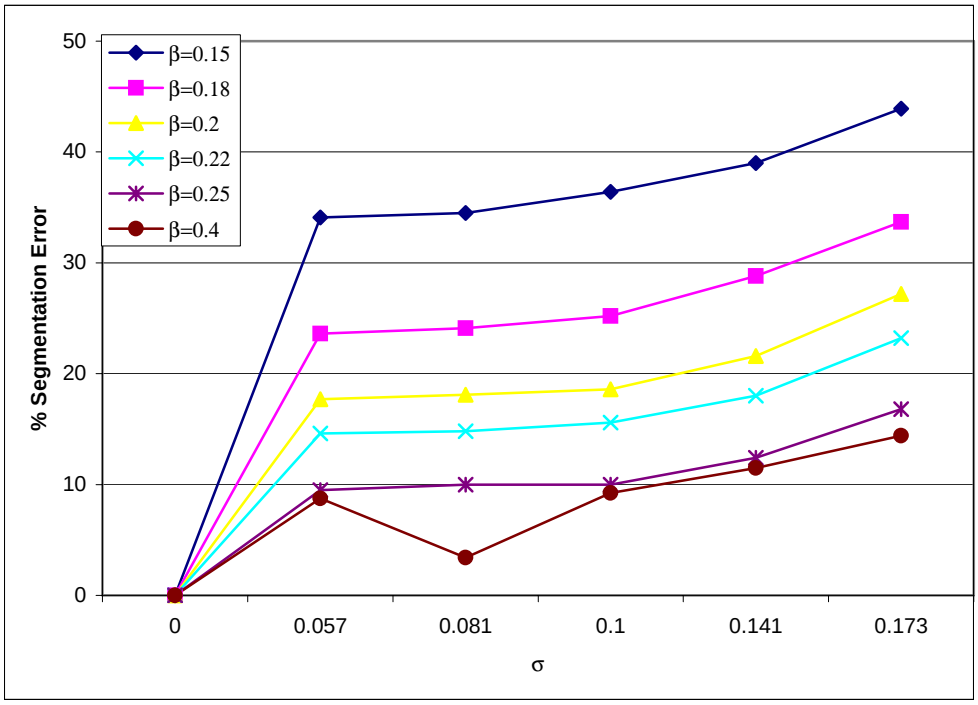


Figure 3.21. Result of segmentation error for SA based on different β and σ values.

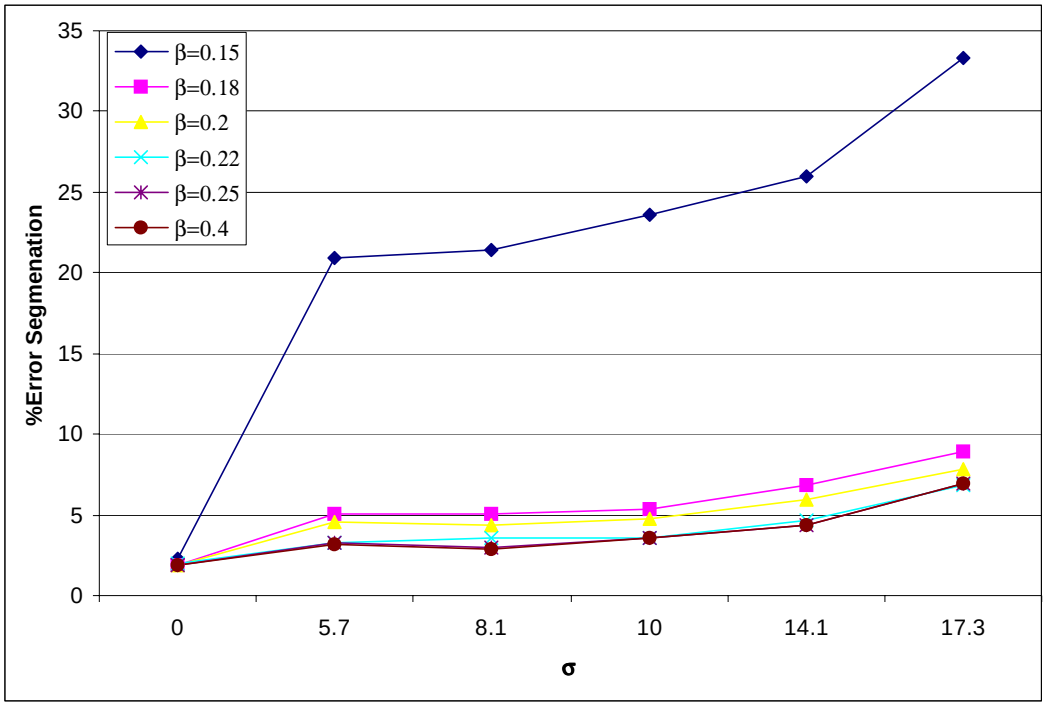


Figure 3.22. Result of segmentation error for SA-RCF based on different β and σ values.

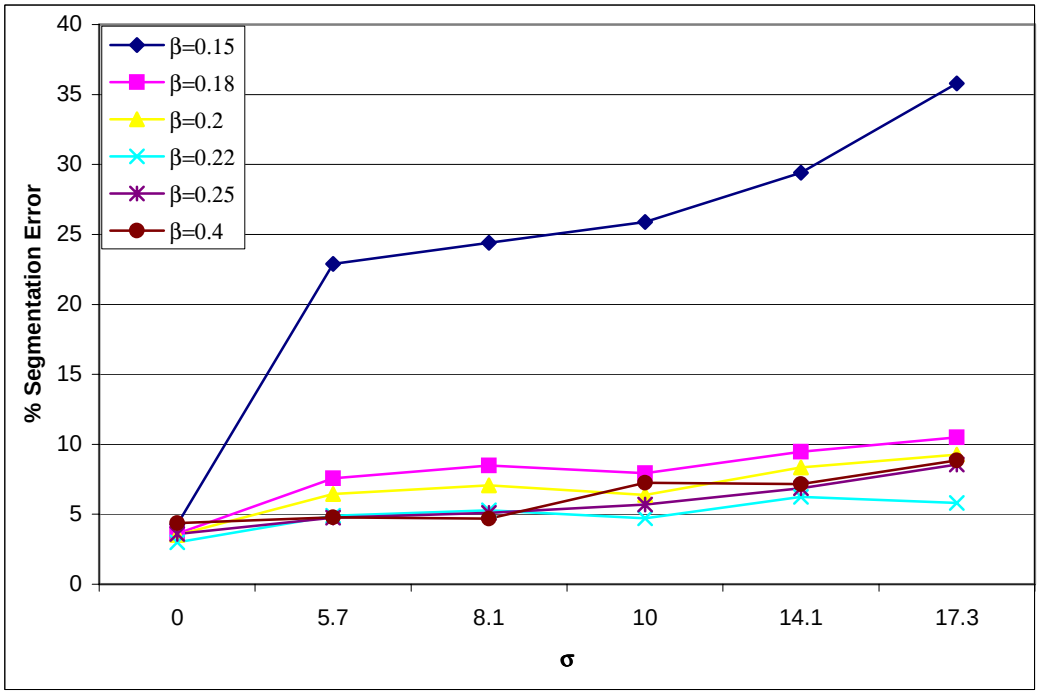


Figure 3.23. Result of segmentation error for MR based on different β and σ values.

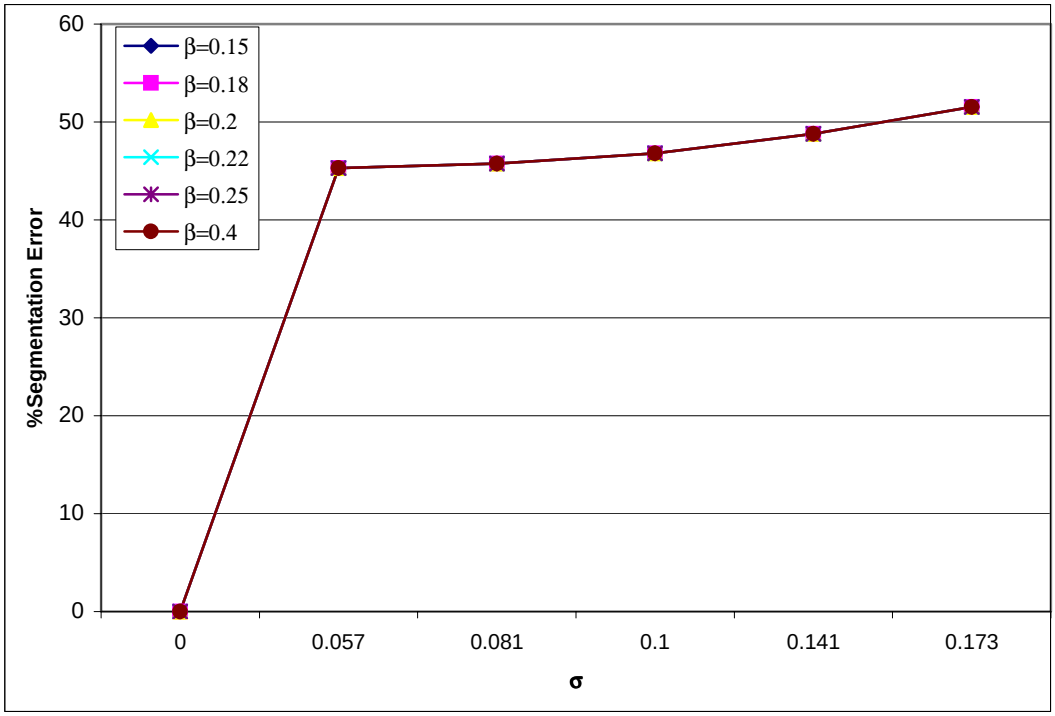


Figure 3.24. Result of segmentation error for MGLD based on different β and σ values.

Table 3.1 summarizes segmentation error of the complete set as a function of noise and β. β values are 0.15, 0.18, 0.2, 0.22, 0.25, and 0.4. σ has values:0, 0.057, 0.081, 0.1, 0.141, and 0.173. For each block of β, the intersection of the algorithm's name (row) with the noisy phantom (column) shows its segmentation error.

Table 3.1. The complete set of noise levels and its segmentation error with different values of β and σ demonstrated.

β	Segmentation Error						
		Noise free	N(0,0.057)	N(0,0.081)	N(0,0.1)	N(0,0.141)	N(0,0.173)
0.15	MGLD	0.00	45.30	45.74	46.79	48.78	51.53
	SA	0.00	34.10	34.50	36.40	39.00	43.90
	MR	4.27	22.90	24.40	25.90	29.40	35.80
	SA-RCF	2.24	20.90	21.40	23.60	26.00	33.30
0.18	MGLD	0.00	45.30	45.74	46.79	48.78	51.53
	SA	0.00	23.60	24.10	25.20	28.80	33.70
	MR	3.62	7.56	8.49	7.93	9.47	10.50
	SA-RCF	1.92	5.04	5.09	5.34	6.84	8.97
0.20	MGLD	0.00	45.30	45.74	46.79	48.78	51.53
	SA	0.00	17.70	18.10	18.60	21.60	27.20
	MR	3.53	6.45	7.06	6.39	8.34	9.26
	SA-RCF	1.93	4.60	4.35	4.79	5.91	7.87
0.22	MGLD	0.00	45.30	45.74	46.79	48.78	51.53
	SA	0.00	14.60	14.80	15.60	18.00	23.20
	MR	3.00	4.90	5.29	4.70	6.25	5.80
	SA-RCF	2.03	3.32	3.59	3.60	4.69	6.80
0.25	MGLD	0.00	45.30	45.74	46.79	48.78	51.53
	SA	0.00	9.50	9.98	10.00	12.40	16.80
	MR	3.59	4.76	5.10	5.69	6.85	8.55
	SA-RCF	1.86	3.26	3.00	3.54	4.34	6.90
0.4	MGLD	0.00	45.30	45.74	46.79	48.78	51.53
	SA	0.01	8.74	3.41	9.24	11.50	14.40
	MR	4.37	4.79	4.68	7.24	7.16	8.84
	SA-RCF	1.86	3.22	2.84	3.54	4.34	6.90

3.5 Conclusion

This chapter proposed two improvements to simulated annealing to perform an image segmentation using Markov random fields. These are the random neighbor selection of a site at fine resolution (SA-RCF) to improve performance of the algorithm in the presence of noise on an image at fine resolution and the use of stochastic sampling within a multiresolution scheme (MR) to reduce the computational complexity.

We compared MR, SA-RCF, SA, and MGLD methods for segmentation of images into a discrete set of classes. The SA-RCF and MR algorithms achieve higher segmentation accuracy in the present of noise. MGLD provides excellent segmentation accuracy for noise-free images, with comparable computational time. The MR method achieves higher segmentation accuracy at significantly lower computational cost compared to SA and MGLD. However, due to its multiresolution decomposition of image, segmentation error is not distributed uniformly across regions of various sizes, *i.e.* error in small regions would be bigger than in larger region. SA and SA-RCF have the same computational complexity, while SA-RCF offers a more robust segmentation for noisy images. Because of random pixel selection, we always obtain uncertainty around the borders of a labeled site. Since MR and SA-RCF look for the best label in observed images to find global minimum energy, they could be less dependent on small changes of β.

4. EXPERIMENTAL RESULTS: PARAMETER ESTIMATION AND IMAGE SEGMENTATION ON DENTAL IMAGES

In this chapter we present the result of our parameter estimation and image segmentation algorithms on set of nine dental images obtained with a μCT (Micro Computed Tomography) scanner. In section 4.1, we present a background in different imaging modalities for dental images and the reason for using μCT imaging. In section 4.2, we show some tooth images and caries as Regions Of Interest (ROI), as well as regions on the dental images that were classified by the faculty members at Indiana University School of Dentistry at Indianapolis. In section 4.3, we evaluate the performance on caries detection by a dentist at two different time spots and show the correlation results. In section 4.4, seven proposed parameter estimation algorithms are applied to these tooth images and the Gaussian mixture model parameters of the images are obtained. In section 4.5, we use the best result of parameter estimation method (Interlaced GA-EM) in four image segmentation methods. Four segmentation methods (MGLD, SA, SA-RCF, and MR) are applied to the images in this section and performances of the results are demonstrated. Section 4.6 provides a discussion and conclusion on proposed methods.

4.1 Background and Significance in μCT imaging of dental caries

Dental caries is a chronic disease that involves the destruction of tooth structure, which can lead to loss of masticatory function, pulpal abscess and tooth extraction [74]. Despite the significant progress in the prevention of dental caries, through the widespread use of fluoride and other oral hygiene measures, dental caries continue to be the most prevalent dental disease, afflicting 88% of the U.S. population [75]. According to the last national dental survey the average teenager has 8 decayed, missing or filled surfaces by

age 17 and young adults are afflicted with an average of 31 decayed or filled tooth surfaces by age 40-44. Likewise, older adults continue to be afflicted by dental caries with an attack rate only slightly less than that of young adults [76].

By the time dental caries can be detected in dental practice using tactile/visual methods, the lesions are well advanced and have progressed about halfway through the enamel. Such lesions reflect the fact that the disease had been in progress for 1-2 years before it has been detected. Lack of early detection and control of caries leads every year to significant expenses for treating caries and its consequences in the U.S. [74].

There is a continuous demand for development of new diagnostic technologies for detection of caries at an early stage, thus halting or reversing the course of its progress through non-invasive treatment procedures. Such techniques will also be able to determine the impact of novel treatment procedures that are designed to gradually reverse the course of demineralization.

To address the need for early detection and size of dental caries several clinical techniques have been introduced and examined. Examples of these techniques include:

- Laser- and Light-Induced Fluorescence [77-80],
- Electrical Conductance Measurement (ECM) [81-83],
- Digital Fiber Optic Trans-illumination (DIFOTI) [84,85],
- Digital Subtraction Radiography [86, 87],
- Optical Coherence Tomography (OCT) [88],
- Terahertz Pulse Imaging (TPI) [89],
- Ultrasound Imaging [90],
- Tuned Aperture Computed Tomography (TACT) [91-93].

To assess the diagnostic accuracy of these techniques and establish the ground truth, there is a need for a reference method (also known as gold standard). Commonly, the sensitivity and specificity of a newly introduced clinical method is tested through *in vitro* and *in situ* models, followed by clinical studies.

The questions of interest in enamel and dentin demineralization and remineralization studies are:

1) Mineral content,

2) Geometrical attributes such as depth, area and volume,

3) Rate of change (of both mineral content and geometrical attributes).

In the past three decades several reference methods have been introduced and examined in caries validation studies. In this section we briefly overview the most common techniques.

4.1.1. Microradiography

In this approach, specimens are sectioned into thin slices. In transverse microradiography (TMR) the sections are cut perpendicular to the enamel surface (thickness ranging from 90μ to 200μ), while in longitudinal microradiography (LMR) specimens are cut parallel to the anatomical tooth surface (thickness about 0.5 mm) [94, 95]. In TMR, the most frequently used microradiographic method; the slices are placed on a high-resolution radiographic film, along with a calibration step-wedge and irradiated by an x-ray source. The optical density of the developed film is used to calculate mineral content, which is expressed in vol% or in $g.m^{-3}$. From TMR images one can obtain lesion depth and the mineral loss value, ΔZ (in vol% or in $g.m^{-3}$). ΔZ refers to the integrated difference between microradiographic density of sample with mineral loss and that of sound enamel. The primary advantage of TMR is its relative quantitative accuracy. Disadvantages of TMR are [96]:

It is destructive due to sectioning. Also there is significant loss of tissue associated with the sectioning process, hence it is unable to provide an accurate view of the entire lesion.

Requires uniformly thick sections, which is practically difficult to achieve. Non-uniformity in thickness leads to error in density measurement.

Requires a flat surface for analysis purposes, hence it fails for analyzing interproximal and occlusal lesions.

Decision of sectioning location is commonly based on visual examination of the specimen, which may lead to incorrect judgment on the existence and severity of carious lesions.

4.1.2. Microhardness

Microhardness testing measures physical changes in the enamel. In this method, a Knoop or Vickers diamond is positioned on the sample with a given load for a given time [97]. The indentation length left by the diamond in the sample is determined microscopically and expressed in μm. This method provides indirect evidence of mineral loss or gain. There are two different types of hardness measurements (1) surface microhardness (SMH), and (2) cross-section microhardness (CSMH). The SMH test is considered a sensitive method of measuring mineral changes, which can be used to measure early enamel changes occurring at the enamel surface [96]. The CSMH method has also been shown to correlate closely with changes in mineral content as determined by microradiography [98]. However, due to fracturing this method cannot assess the outer 25μ of enamel. Also because of destructive indentation, the same surface cannot be studied longitudinally. It has been also noted that microhardness fails to assess dentinal lesions due to relaxation of dentin indentation with time and drying shrinkage effect [99].

4.1.3. Polarized Light Microscopy

In this approach thin sections (about 80μ) are prepared and observed under a polarizing microscope. The difference in birefringence is used to differentiate between sound and diseased areas [100, 101]. In addition to its destructive nature, it is difficult to interpret quantitatively [102].

4.1.4. Other methods

Other methods include light scattering [103, 104], wavelength-independent microradiography [105], confocal laser scanning microscopy [106], scanning light microscopy, chemical analysis [97], and x-ray tomographic microscopy (XTM). Kinney

et al. [107] employed XTM (based on synchrotron radiation at Stanford Synchrotron Radiation Facility) to assess mineral content of dentin. They analyzed a 3mm by 1mm section of a single maxillary molar with natural dentinal caries. XTM measurement for this specimen was shown to be in agreement with the average estimates of the mineral concentration in dentin.

To summarize, the drawbacks of the common techniques for demineralization/ remineralization analysis are:

Destructive approaches do not allow analysis of the dynamics of demineralization and remineralization. Also, due to destruction, a larger number of specimens are needed to accommodate different demineralization/remineralization regiments, as well as reducing inter-subject variations.

It is also extremely difficult to obtain uniformly thick lesions for techniques requiring thin sections. Obtaining such sections of the occlusal surfaces leads to a significant percentage of fractures. Also, the flat surface requirement excludes lesions on interproximal and occlusal surfaces.

Most of these techniques analyze the lesion for a small window provided by the thin section. For example, to obtain a 100μ thin section for TMR analysis, about 200μ of tissue will be lost on each side of the intended section during cutting. Hence, it is not possible to obtain a complete view of geometrical extent and mineral content of the lesion.

There is a need for objective and accurate methods to assess the diagnostic potential of such innovative technologies for detection of caries. There are several methods currently employed for analysis of demineralization and remineralization. The consensus in the dental research is that there is significant inconsistency among these methods [108, 109]. Furthermore, these destructive approaches do not permit the monitoring of lesions longitudinally to study dynamics of lesion progression and regression. The objective of thesis is to design and evaluate a quantitative and non-destructive method for caries analysis and validation.

In the past few years researchers at Indiana University School of Dentistry (IUSD) have been examining the sensitivity and accuracy of several novel techniques with potential for early detection of caries and evaluation of its size with non-destructive method. An example of such techniques is μCT imaging. To address this problem, it was shown feasibility of a high-resolution computed tomographic (μCT) unit for caries assessment. The primary interest in this approach stemmed from the facts that (1) it can provide full 3-D view of lesion, (2) it makes possible to study the dynamics of demineralization and remineralization due to its nondestructive nature, and more importantly, and (3) if calibrated, it can provide objective and quantitative assessment of tissue content [110-114].

4.2 Tooth images

We have selected 9 dental images, which have each been scanned under a different signal-to-noise ratio by a μCT scanner, (Scanco Medical μCT20). The scanner can provide an in-plane resolution of 8-14 μm, depending on the size of field of view and a slice thickness of 25μm. These images were selected from different sets of 3-D volumes. There was no preference to select the image slices from the 3-D volumes except for having a variety of lesion sizes.

Figure 4.1 shows a slice image that has caries on both the enamel and the dentin area. It is an occlusal view of a premolar tooth, which was scanned by a μCT scanner.

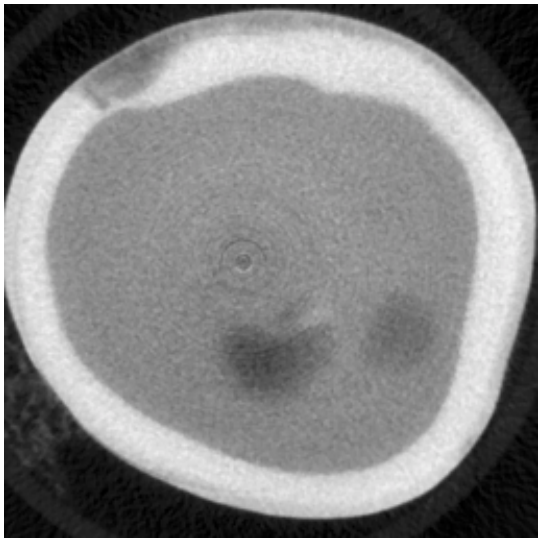


Figure 4.1. Tooth image, occlusal view, from a set of images taken by μCT.

When analyzing dental images, dentists look for regions that can be used for image interpretations. Those regions were marked as A1-A5 in figure 4.2. A1 shows the background, A2 shows the enamel, A3 shows the dentin area of tooth, A4 shows the caries on enamel, and A5 shows caries in the dentin area. It should be mentioned that however there are two marked A5 regions and apparently both of them belong to one class, caries; but those regions actually belong to two stages of caries development.

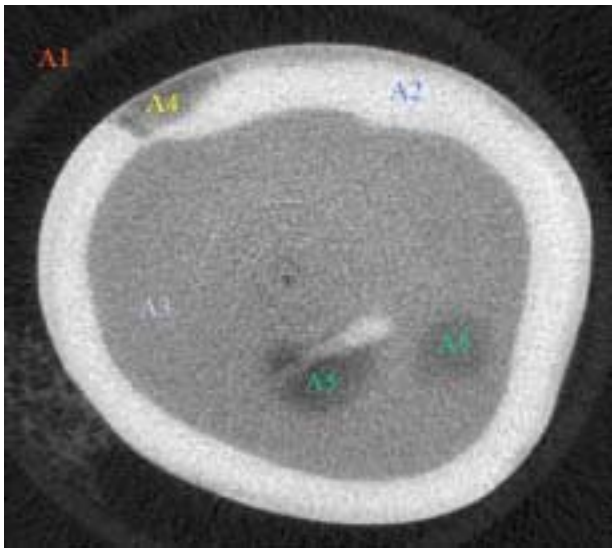
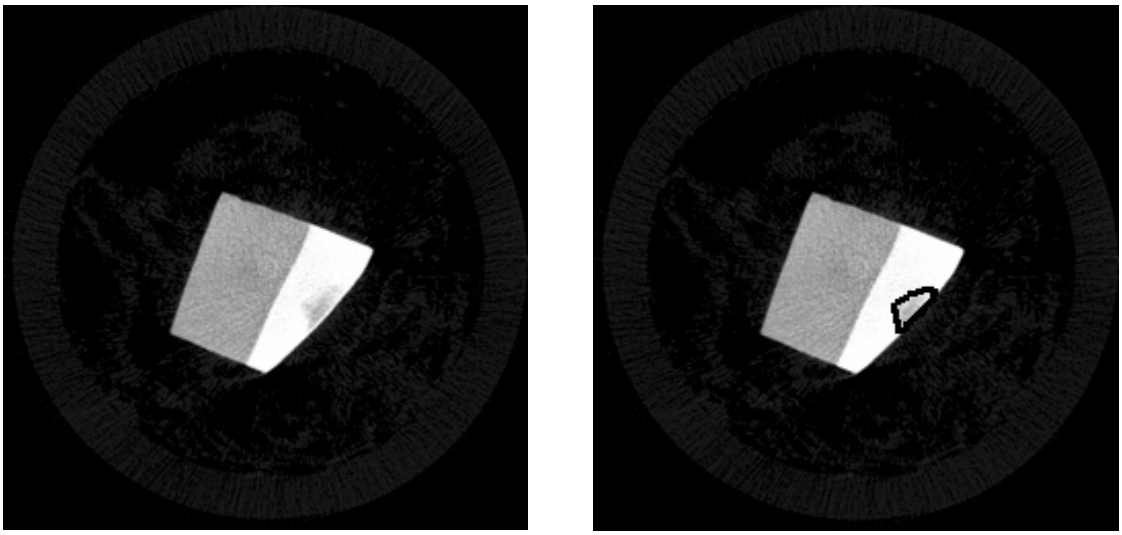


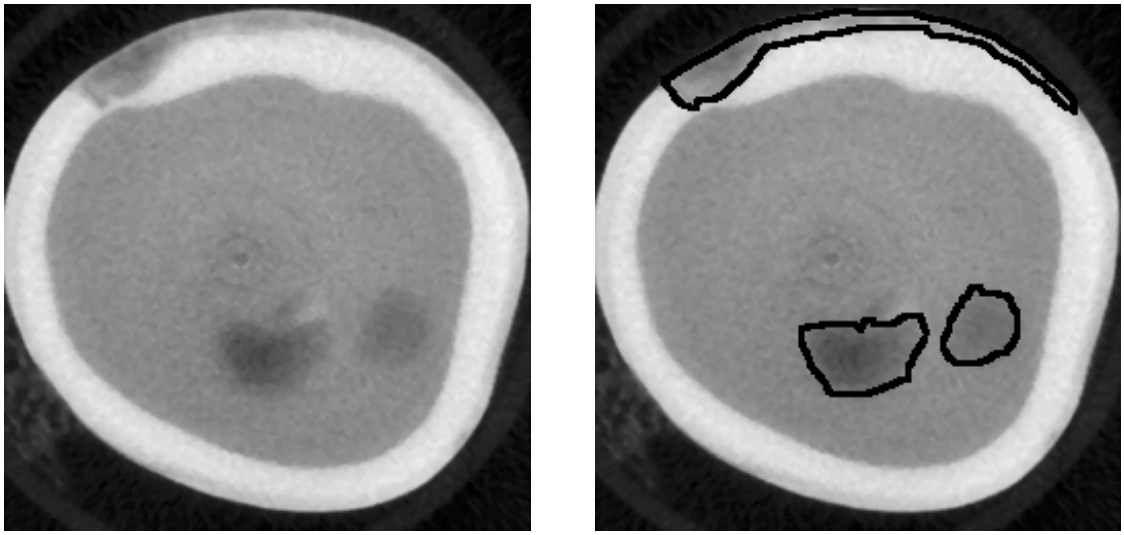
Figure 4.2. Tooth image, ROI on the image were marked from A1 to A5.

The overall purpose of this thesis is to detect caries on dental images unsupervised. In order to evaluate the image segmentation process, we need to have correct locations/sizes of caries on dental images. The caries were determined and marked by two faculty members, Dr. Masatoshi Ando and Dr. Hafsteinn Eggertsson, School of Dentistry, Indiana University. Throughout this chapter, the errors of the image segmentation methods were calculated as the average of segmentation error results based on both Dr. Ando and Dr. Eggertsson's caries detection. The teeth images are shown in figures 4.3 through 4.11, called Images A – K. The original image was labeled "a" and the marked caries labeled "b" for all images (A-K).



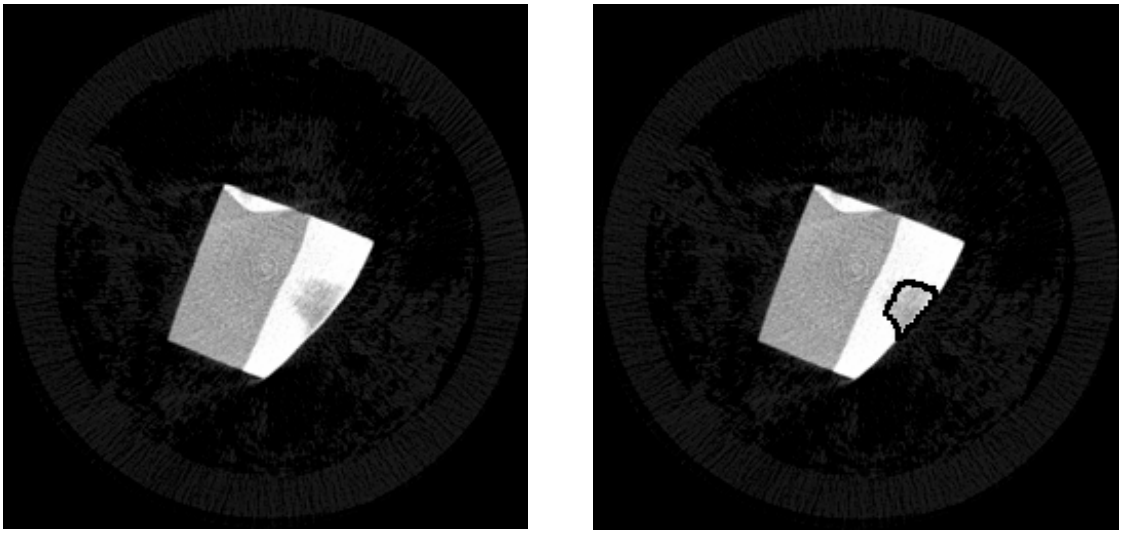
a b

Figure 4.3. a) Image A, b) Caries are distinguished on image with a thick black line.



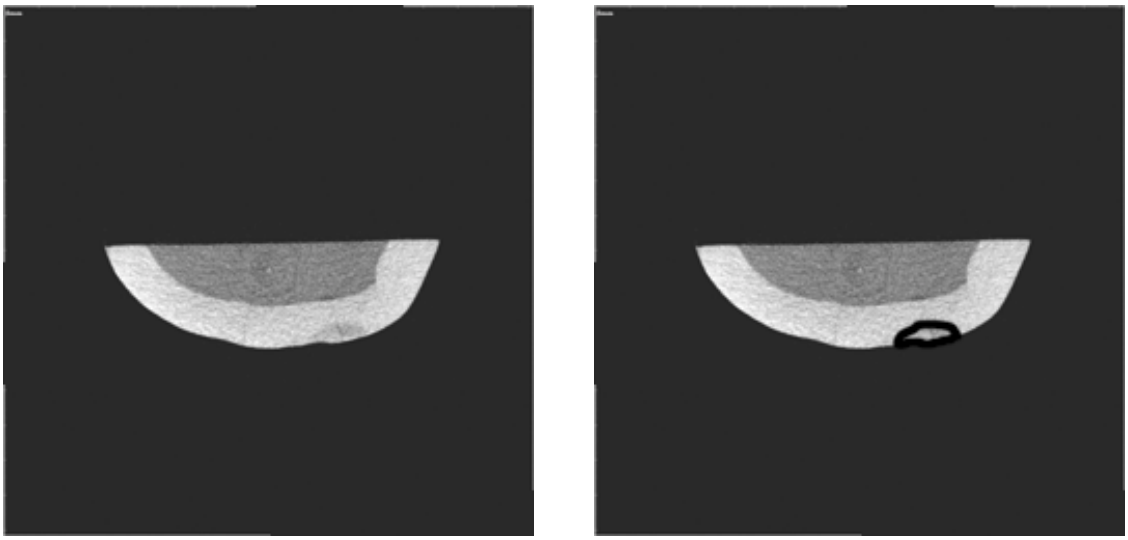
a b

Figure 4.4. a) Image B, b) Caries are distinguished on image with a thick black line.



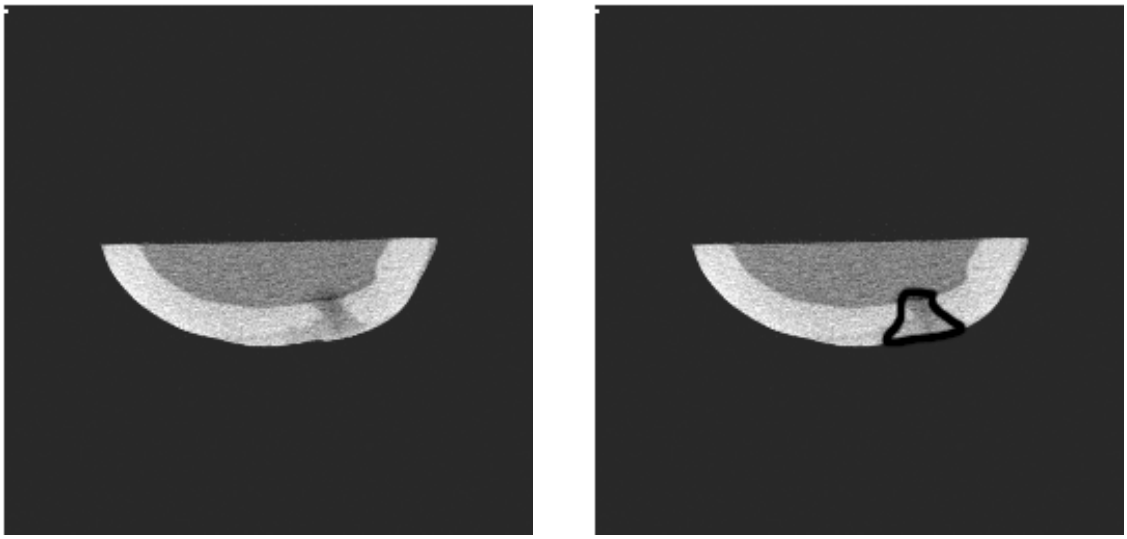
a b

Figure 4.5. a) Image C, b) Caries are distinguished on image with a thick black line.



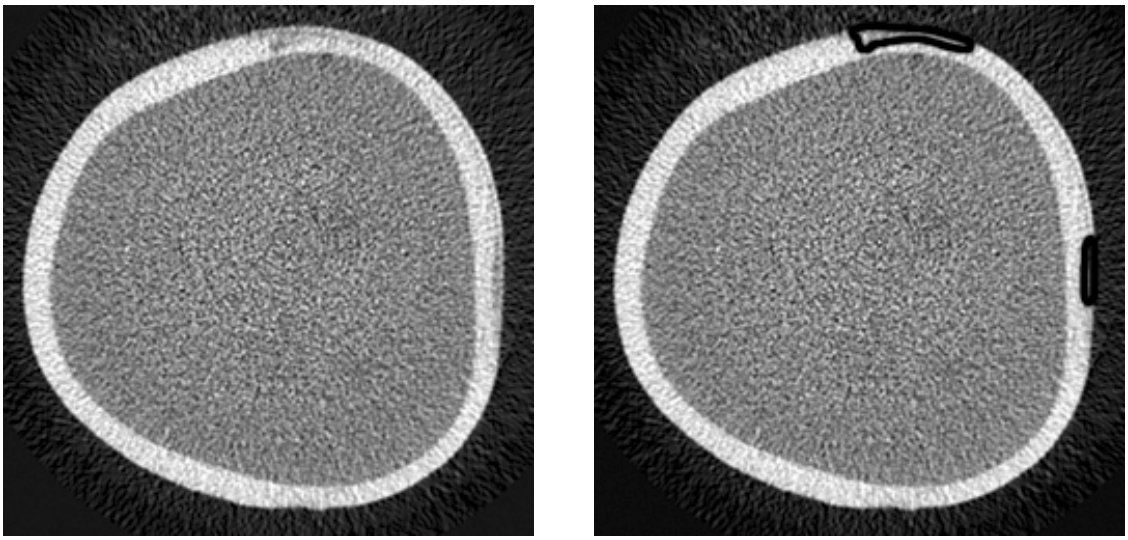
a b

Figure 4.6. a) Image D, b) Caries are distinguished on image with a thick black line.



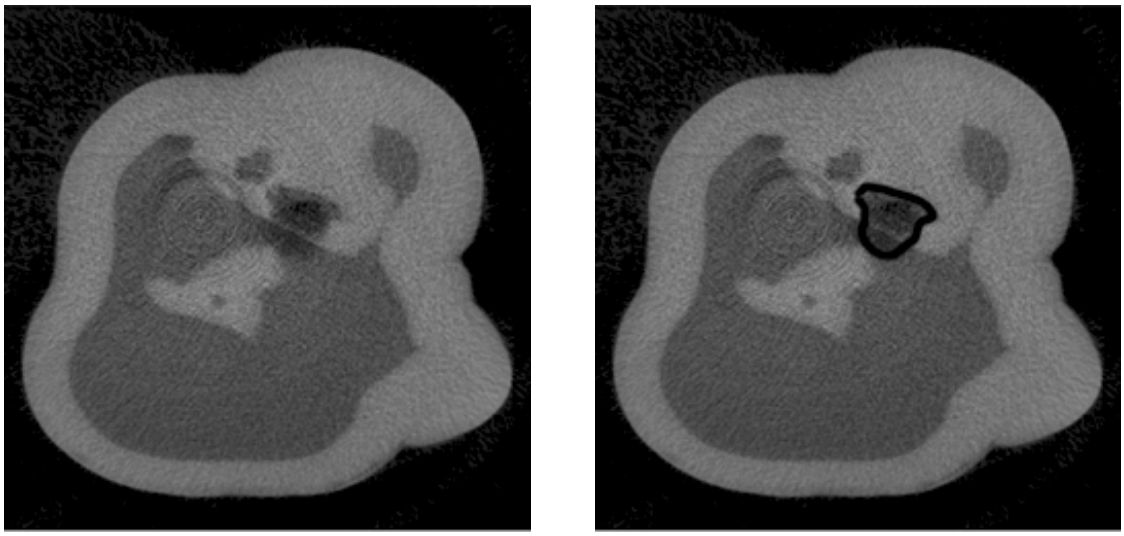
a b

Figure 4.7. a) Image E, b) Caries are distinguished on image with a thick black line.



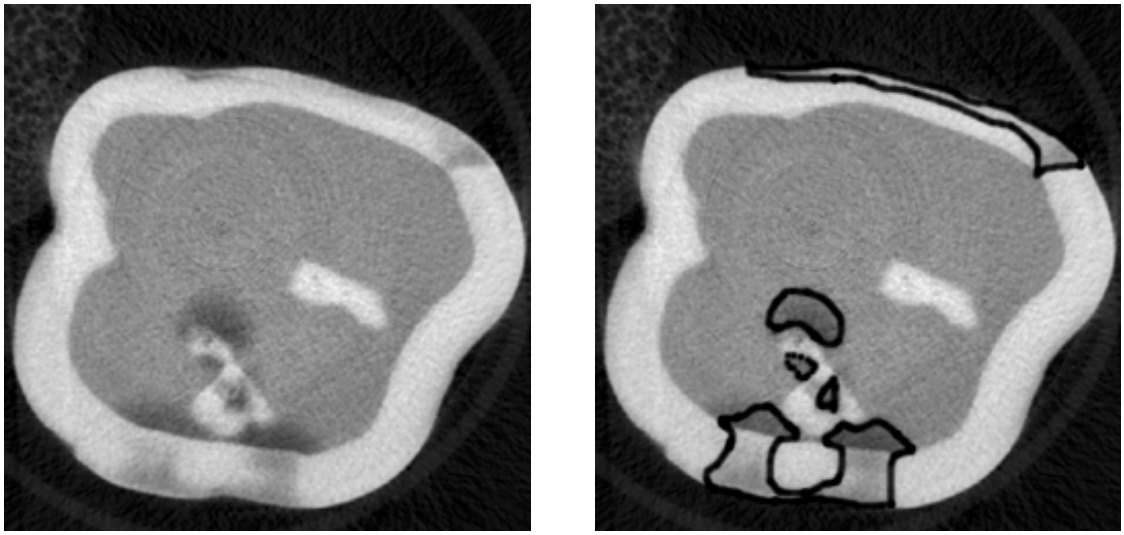
a b

Figure 4.8. a) Image F, b) Caries are distinguished on image with a thick black line.



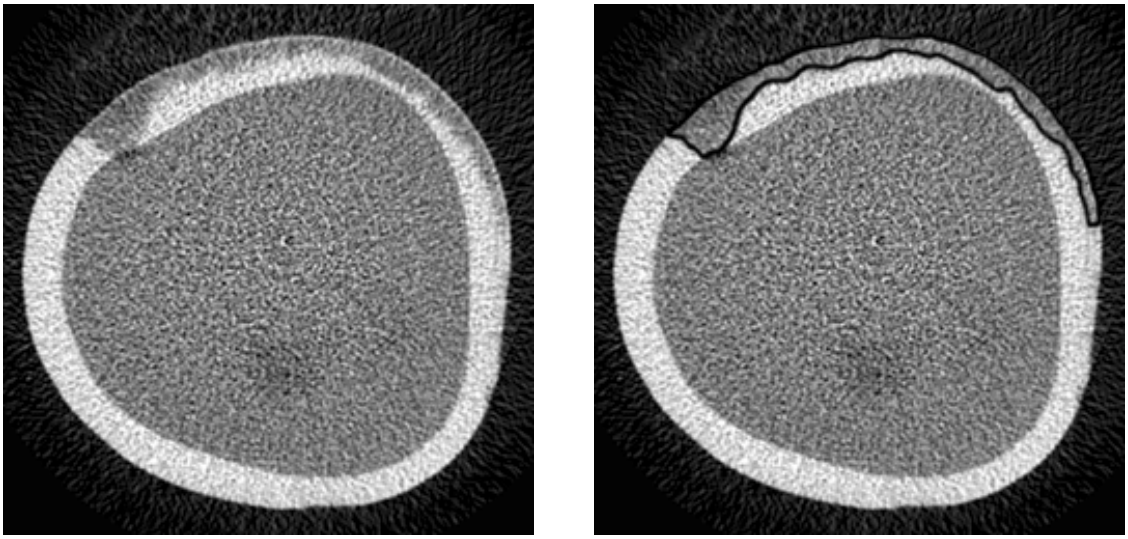
a b

Figure 4.9. a) Image G, b) Caries are distinguished on image with a thick black line.



a b

Figure 4.10. a) Image H, b) Caries are distinguished on image with a thick black line.



a b

Figure 4.11. a) Image K, b) Caries are distinguished on image with a thick black line.

4.3 Intra observer agreement

Even experienced eyes are not very consistent in identifying caries on the images. We have asked Dr. Masatoshi Ando to mark caries on the images. He was asked to determine the caries for each image at two different times. We would like to know whether or not the specialist could detect the lesion's exact size and location consistently. The ability to correctly detect the lesion location and size for each image was calculated

as a percent of the intersection of regions per total union of lesion sizes, which were obtained at two separate times. Figures 4.12 and 4.13 show this Intra Observer Agreement (IOA) and the difference between the two observations respectively. The average of the IOA over 9 images was 81.21%. The average difference between the two observations was 18.75%. The Intra class Correlation Coefficients (ICC) for these sets of images was calculated as follows [115]:

$$ICC = \frac{\frac{n\sum_{i=1}^{a}(\bar{Y}_{io} - \bar{Y}_{oo})^2}{a-1}}{\frac{n\sum_{i=1}^{a}(\bar{Y}_{io} - \bar{Y}_{oo})^2}{a-1} + \frac{\sum_{i=}^{a}\sum_{j=1}^{n}(Y_{ij} - \bar{Y}_{io})^2}{N-a}} \quad (4.1)$$

where a is number of images, n is numbers times for each image (2 times per image), $N = a \times n$, $\bar{Y}_{io}$ is the average size of caries over two times, and $\bar{Y}_{oo}$ is the overall average size of all caries.

Sizes of caries are shown in figure 4.14. Please note that Image B, Image F and Image H have more than one lesion. The caries sizes at the first observation time are plotted versus the sizes at the second observation time for each image.

A regression line was fit to this data and ICC and coefficient of determination (R^2) was calculated. R^2 was 0.9613 and the intercept of regression line was 29.86. The ICC was calculated as 0.8642.

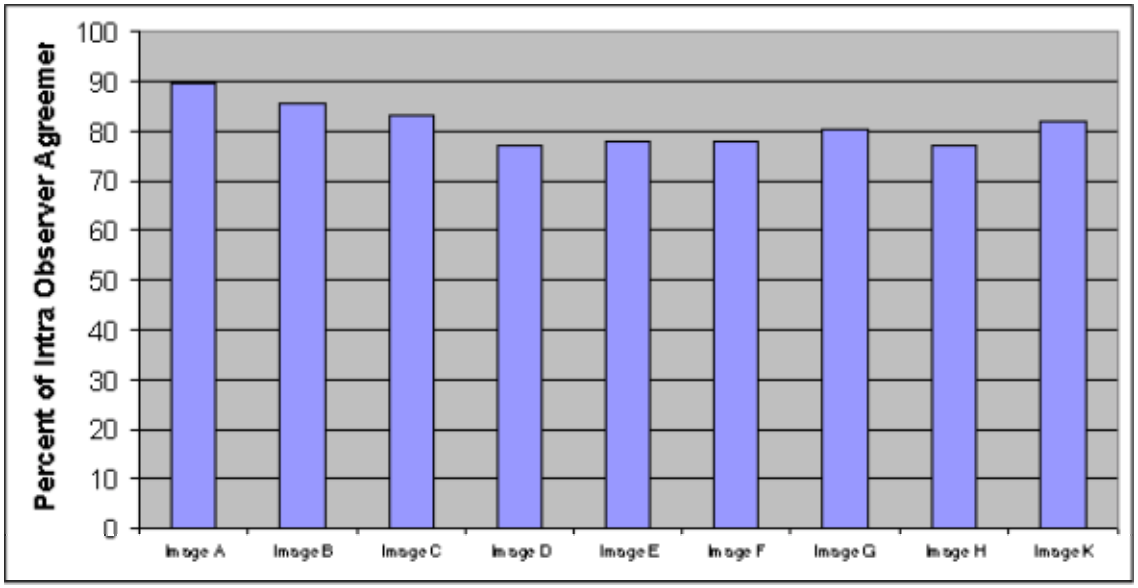


Figure 4.12. Percent of repeatability to detect caries for each image.

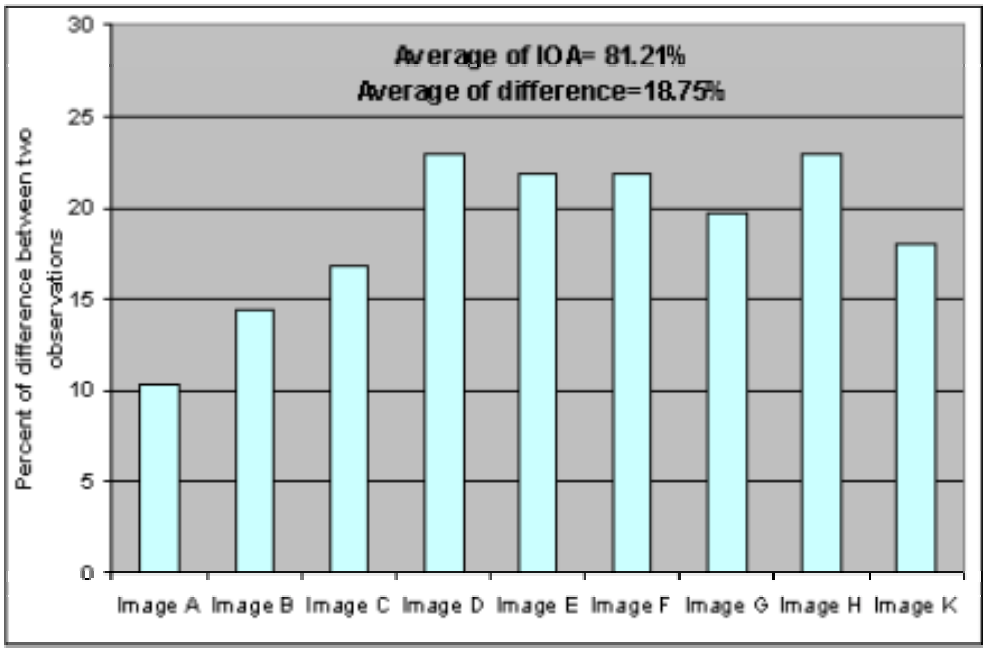


Figure 4.13. Percent of error to detect caries in teeth images at two observations. Please note that percent of error range on vertical axes is shown in a range from 0 to 30 to amplify error contrast among images.

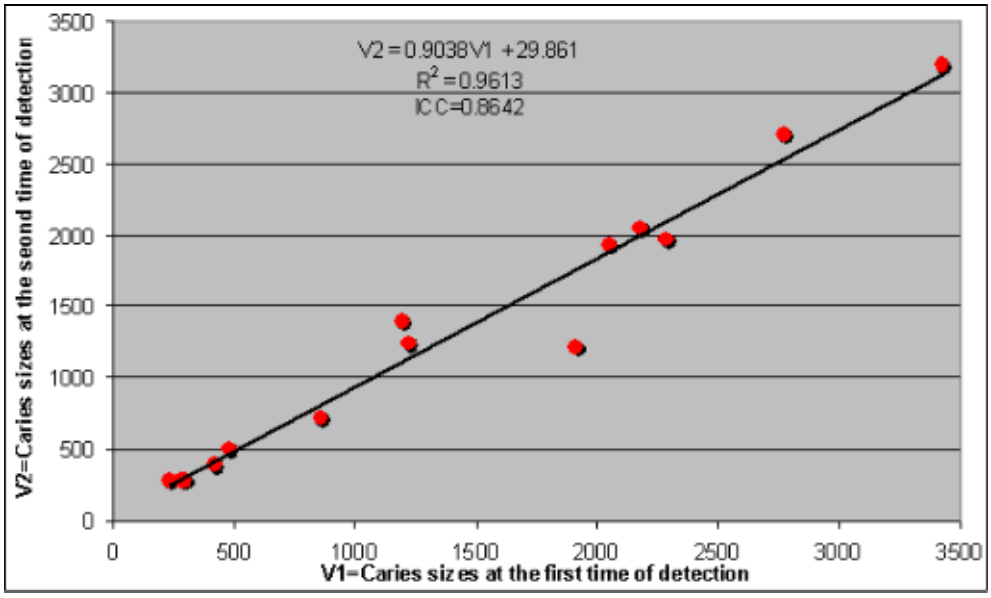


Figure 4.14. Caries sizes (in pixels) at two different times of detection.

The root-mean-square of signal-to-noise ratio for each image, denoted SNR_{rms}, is:

$$SNR_{rms} = \sqrt{\frac{\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}\hat{g}(x,y)^2}{\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}[\hat{g}(x,y)-g(x,y)]^2}} \tag{4.2}$$

where $\hat{g}(x,y)$ is considered a noiseless image, in which the gradation of enamel, dentin, caries and background has been adjusted, and $g(x,y)$ is the original image In the rectangular lattice $M \times N$. The caries sizes were calculated from $\hat{g}(x,y)$ and we have used them as ground truth. Average caries sizes of each image at two different times are shown in figure 4.15. The root mean square value of signal-to-noise ratio (SNR_{rms}) of each image, $g(x,y)$, is shown in figure 4.16.

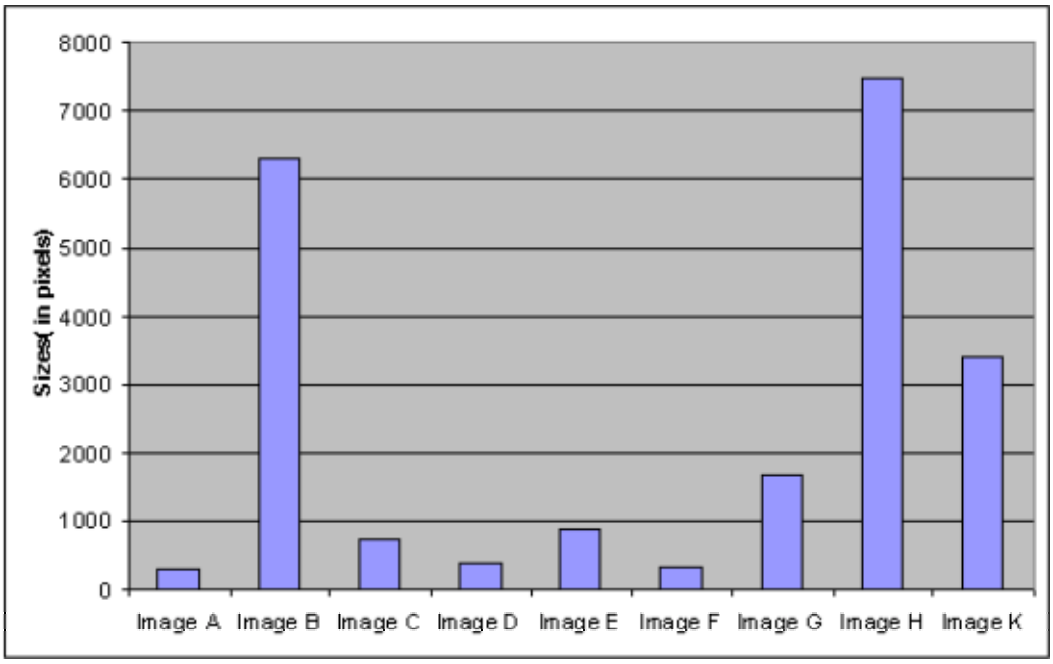


Figure 4.15. Caries sizes, in pixels, in each image

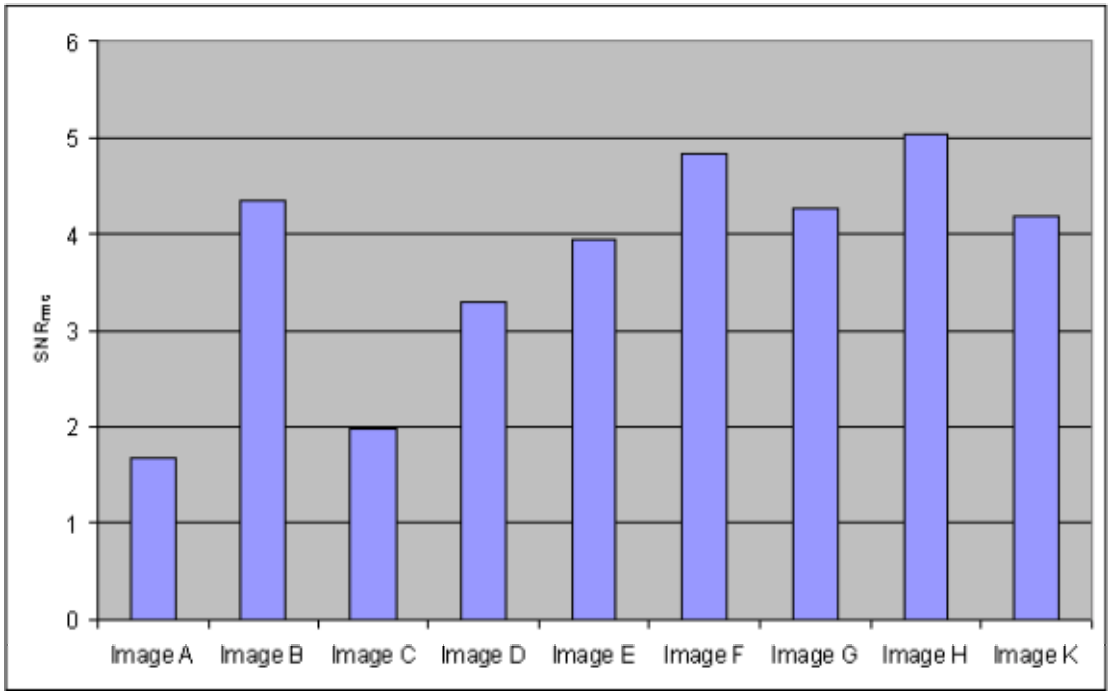


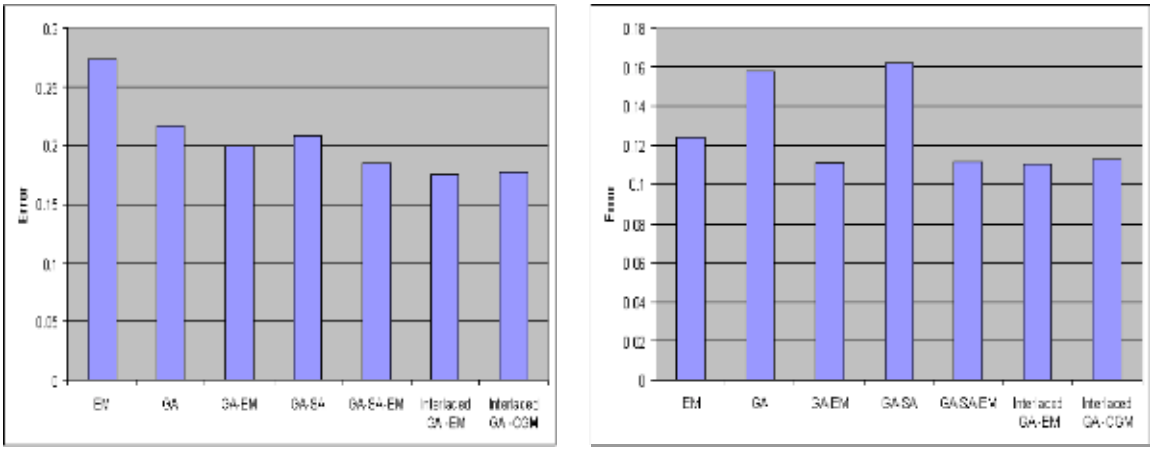
Figure 4.16. SNR_{rms} for nine dental images

4.4 Parameter estimation on tooth images, experimental results

The parameter estimation of the Gaussian mixture model is the first step of our unsupervised image segmentation procedure. We have introduced the parameter estimation methods in chapter two extensively and we apply them in this section. All methods have been applied to tooth images and errors are shown in figure 4.17 through 4.21. These methods are:

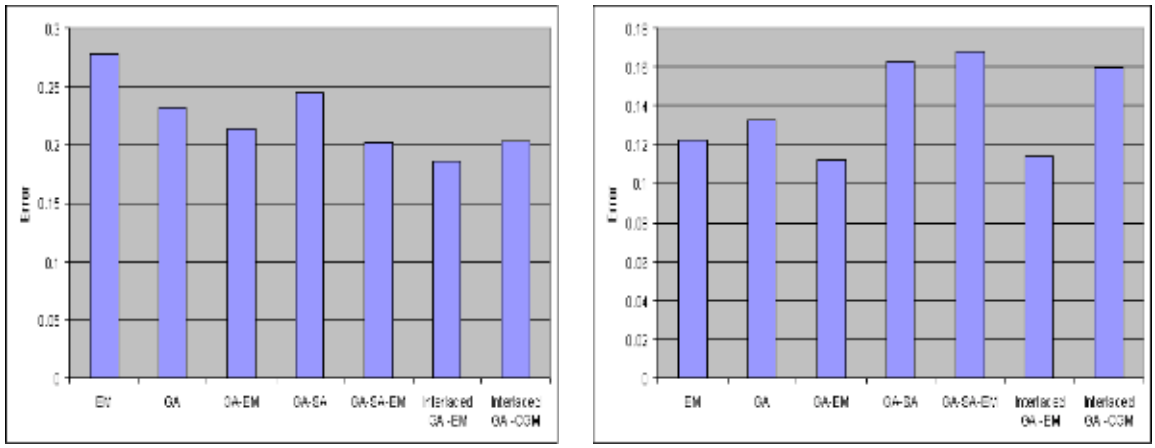
1) GA: Standard Genetic Algorithm,
2) EM: Standard Expectation Maximization,
3) GA-SA: A modification of standard GA with Simulated Annealing concept,
4) GA-EM: Uses final result of GA as initial input for EM,
5) GA-SA-EM: Final results of GA-SA are used as initial inputs for EM,
6) Interlaced GA-EM: Interactively parameter estimation on both GA and EM instead of sequential order, and
7) Interlaced GA-CGM: Interactively parameter estimation on both GA and Conjugate Gradient Method.

We compared the performance of each method using the minimum squared error between mixture models of each image as described in the section 2.9.



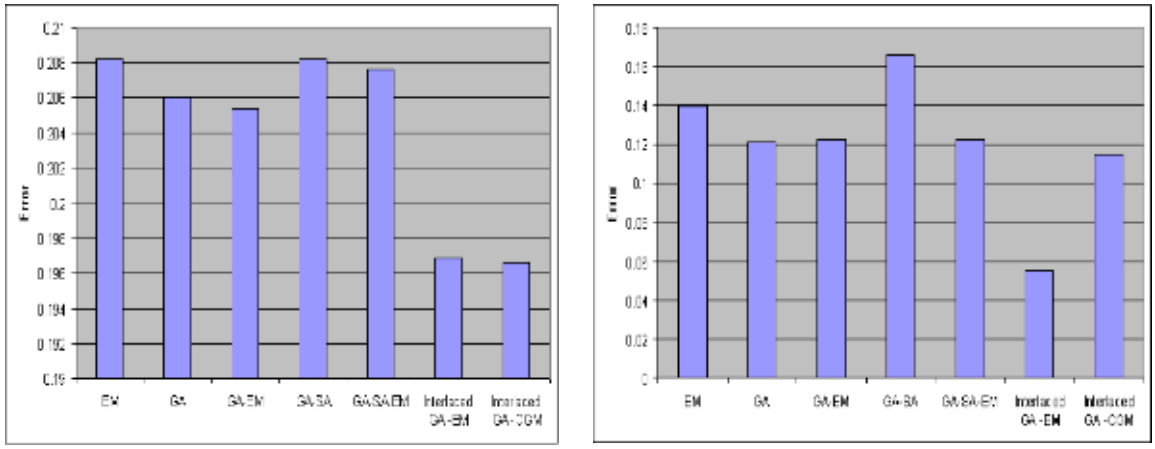
a b

Figure 4.17. Error of EM, GA, GA-EM, GA-SA, GA-SA-EM, Interlaced GA-EM, and Interlaced GA-CGM. a) Image A b) Image B.



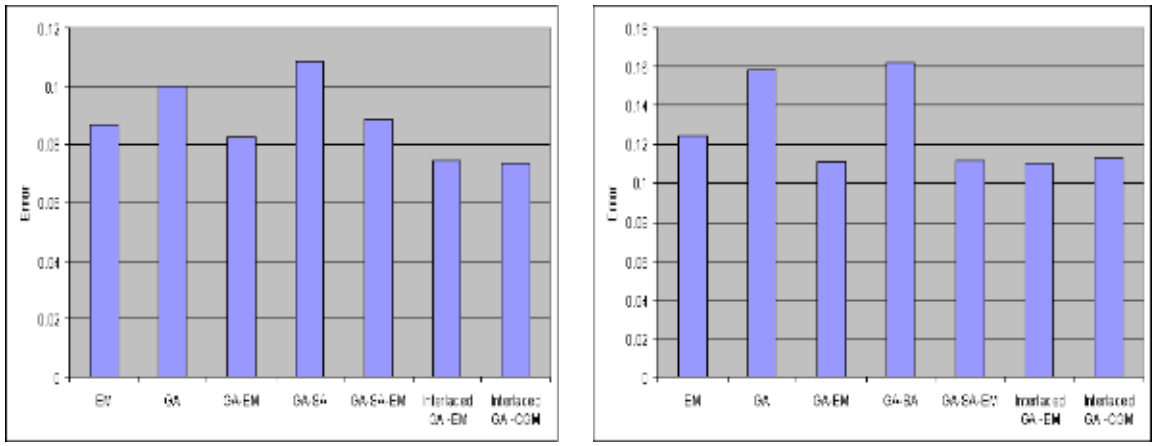
a b

Figure 4.18. Error of EM, GA, GA-EM, GA-SA, GA-SA-EM, Interlaced GA-EM, and Interlaced GA-CGM. a) Image C b) Image D.



a b

Figure 4.19. Error of EM, GA, GA-EM, GA-SA, GA-SA-EM, Interlaced GA-EM, and Interlaced GA-CGM. a) Image E b) Image F.



a b

Figure 4.20. Error of EM, GA, GA-EM, GA-SA, GA-SA-EM, Interlaced GA-EM, and Interlaced GA-CGM. a) Image G b) Image H.

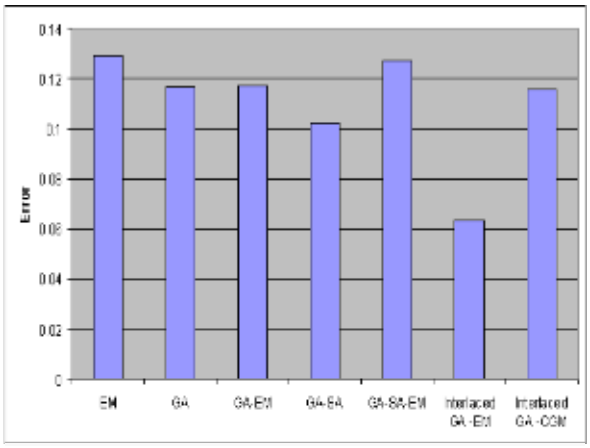


Figure 4.21. Error of EM, GA, GA-EM, GA-SA, GA-SA-EM, Interlaced GA-EM, and Interlaced GA-CGM for Image K.

4.5 Image segmentation of tooth images, experimental results

The image segmentation method was explained in the chapter 3. We have used those image segmentation methods (SA, MGLD, MR, and SA-RCF) to segment the 9 tooth images. As we have observed Interlaced GA-EM produced the minimum error in the parameter estimation, so we have used its values to segment images. The complete set of parameter estimation errors are shown in figure 4.22. The error of the image

segmentation was calculated as the percentage of pixels misclassified per total pixels in the image as denoted in section 3.4.

Results of image segmentation are shown in figures 4.23 to figure 4.31.

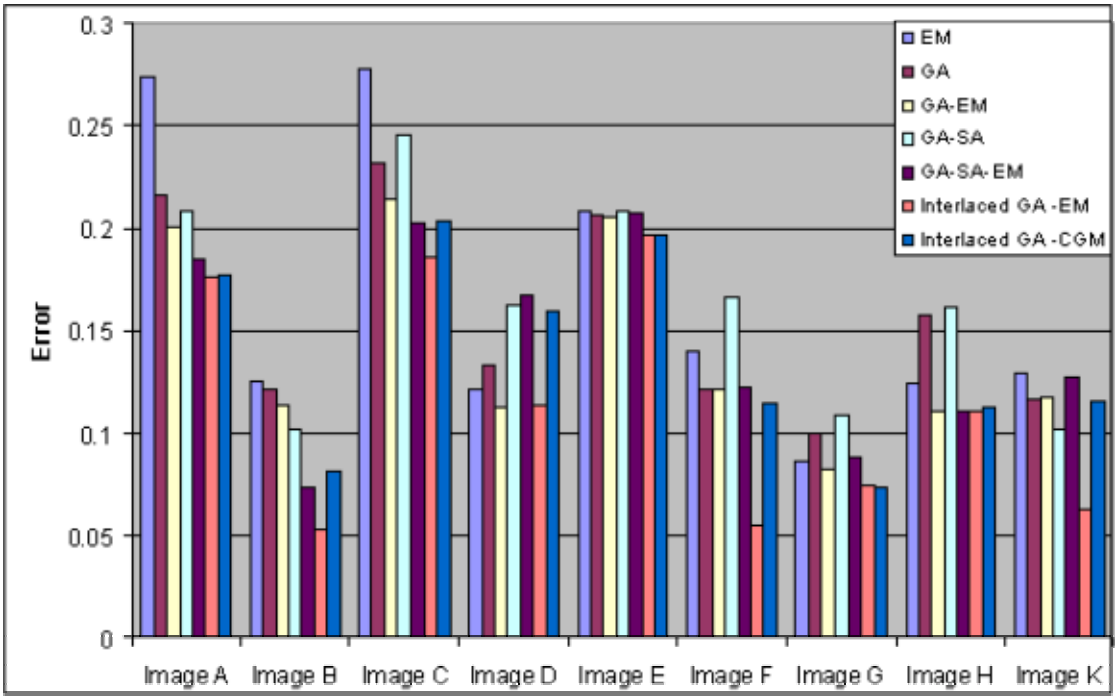
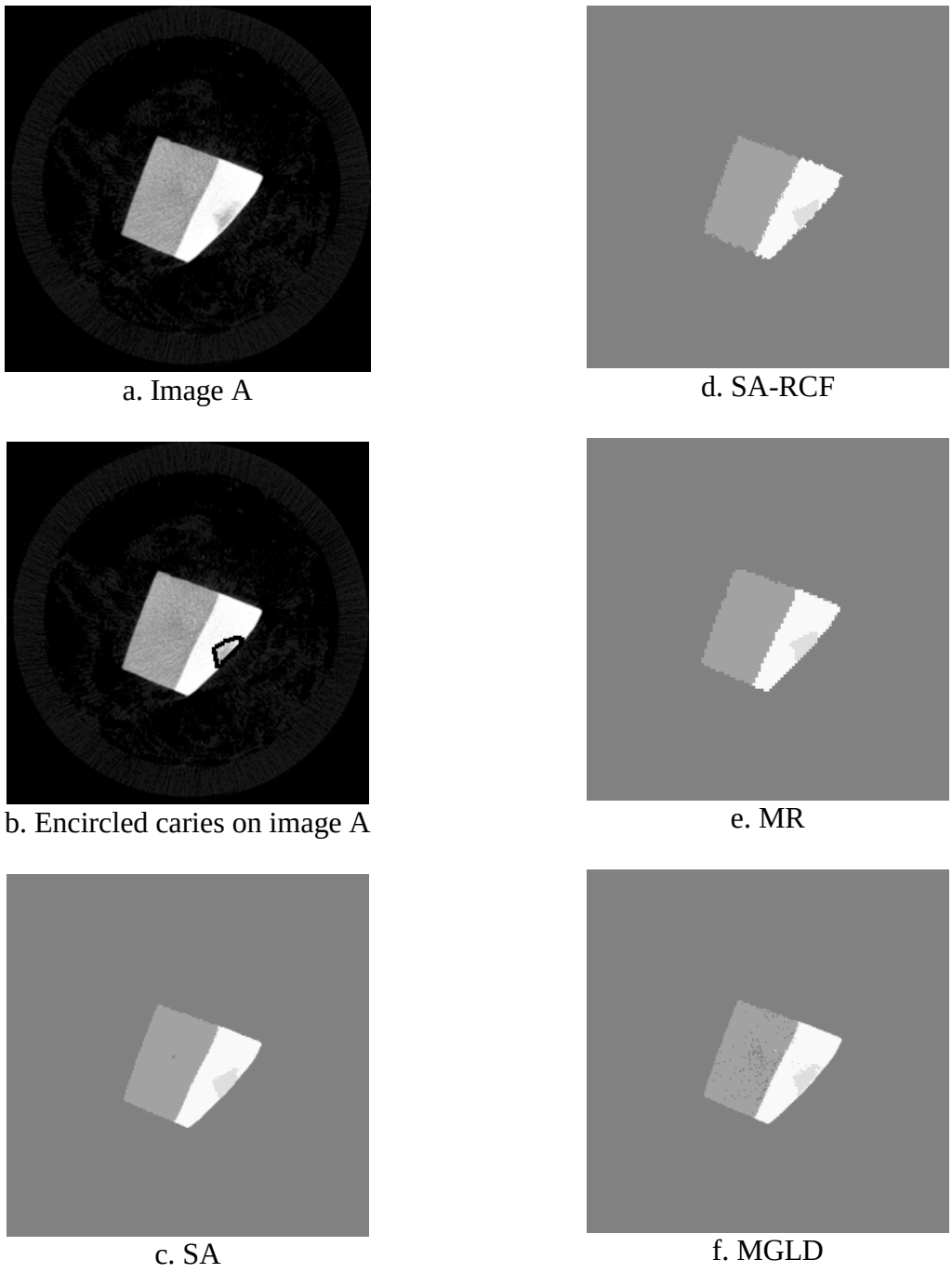


Figure 4.22. Error of EM, GA, GA-EM, GA-SA, GA-SA-EM, Interlaced GA-EM, and Interlaced GA-CGM for Images A through K.



a. Image A

b. Encircled caries on image A

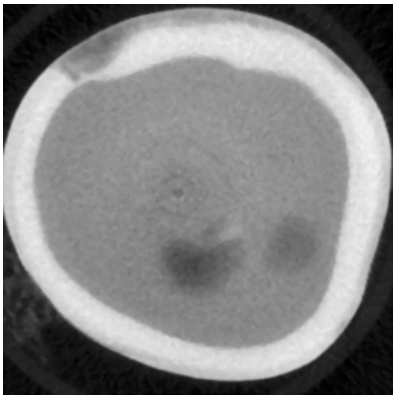
c. SA

d. SA-RCF

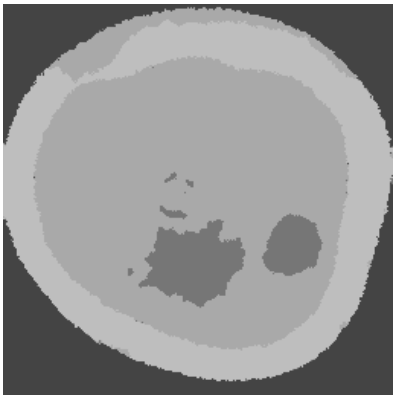
e. MR

f. MGLD

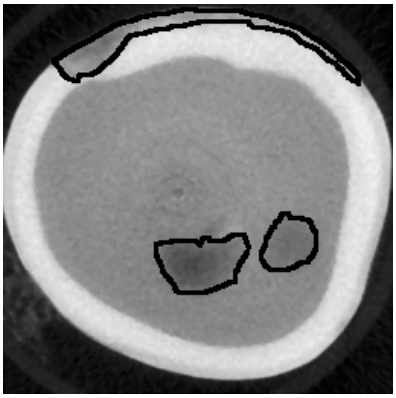
Figure 4.23. a) Image A b) Caries on the Image A. c) Result of SA. d) Result of SA-RCF. e) Result of MR. f) Result of MGLD. β =0.22 was selected for all these images.



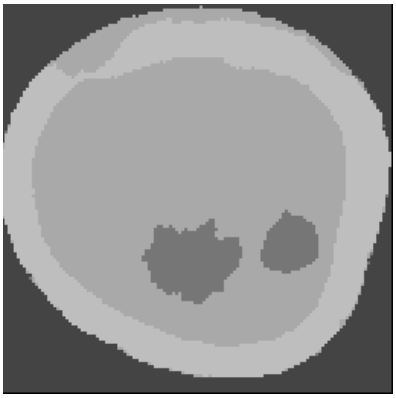
a. Image B



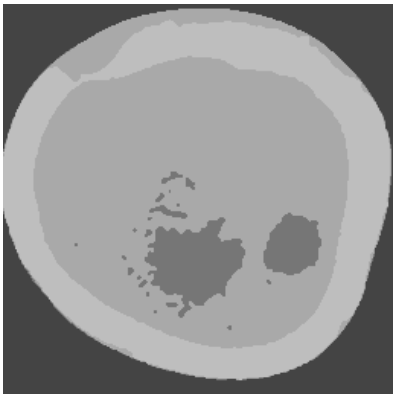
d. SA-RCF



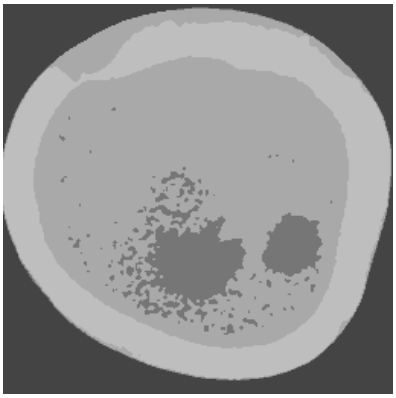
b. Encircled caries on image B



e. MR

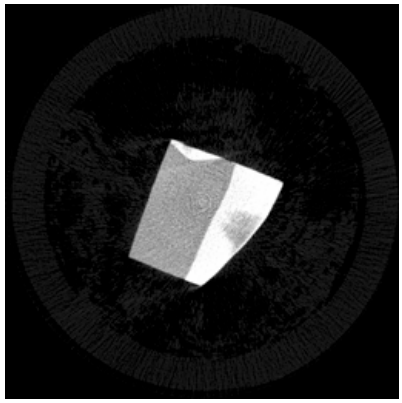


c. SA

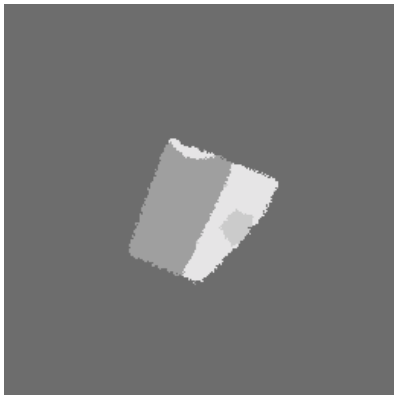


f. MGLD

Figure 4.24. a) Image B. b) Caries on the Image B. c) Result of SA. d) Result of SA-RCF. e) Result of MR. f) Result of MGLD. β =0.22 was selected for all these images.



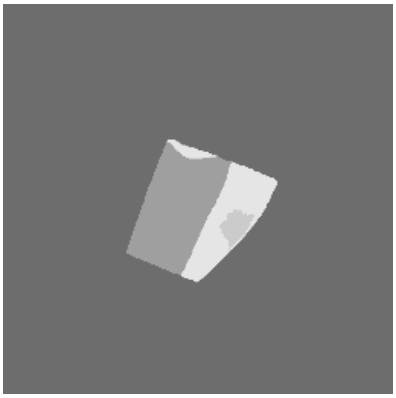
a. Image C



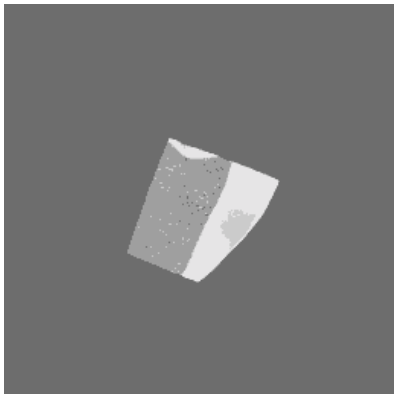
d. SA-RCF

b. Encircled caries on image C

e. MR

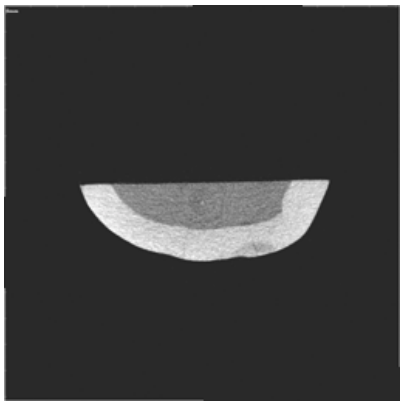


c. SA



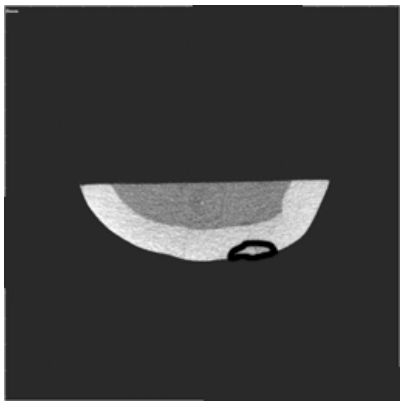
f. MGLD

Figure 4.25. a) Image C. b) Caries on the Image C. c) Result of SA. d) Result of SA-RCF. e) Result of MR. f) Result of MGLD. β =0.22 was selected for all these images.



a. Image D

d. SA-RCF



b. Encircled caries on image D

e. MR

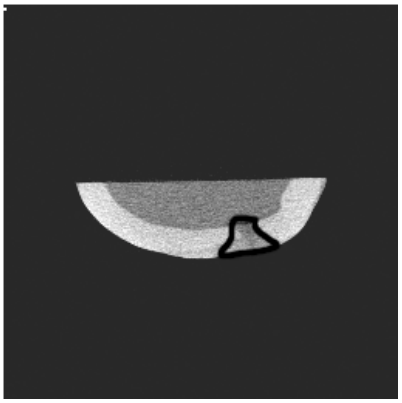
c. SA

f. MGLD

Figure 4.26. a) Image D. b) Caries on the Image D. c) Result of SA. d) Result of SA-RCF. e) Result of MR. f) Result of MGLD. β =0.22 was selected for all these images.

a. Image E

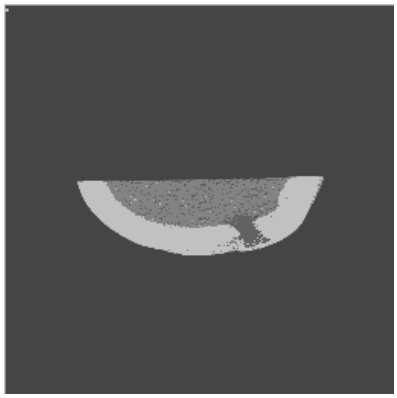
d. SA-RCF



b. Encircled caries on image E

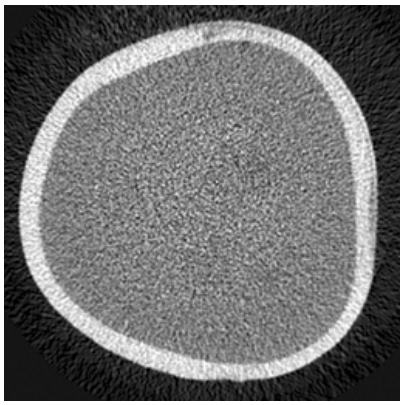
e. MR

c. SA

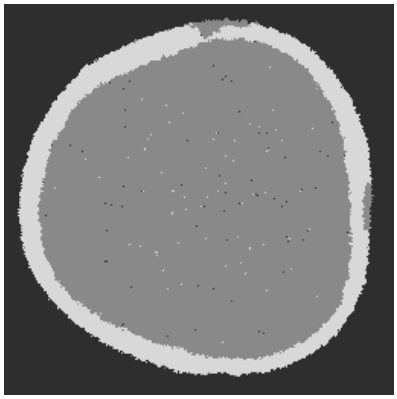


f. MGLD

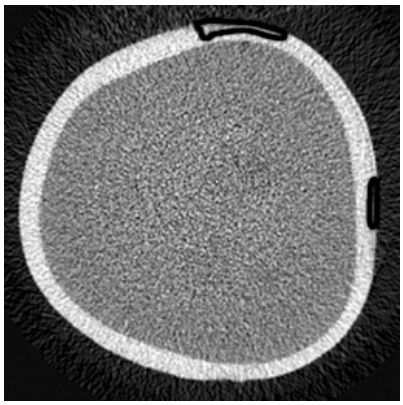
Figure 4.27. a) Image E. b) Caries on the Image E. c) Result of SA. d) Result of SA-RCF. e) Result of MR. f) Result of MGLD. β =0.22 was selected for all these images.



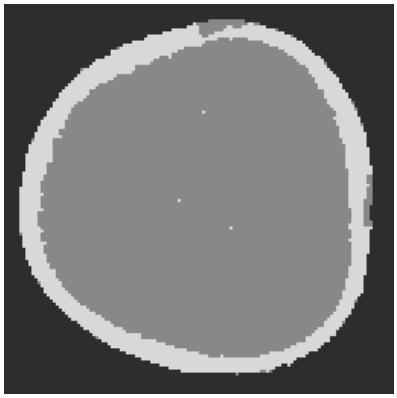
a. Image F



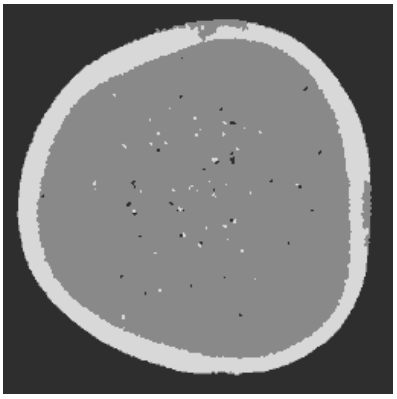
d. SA-RCF



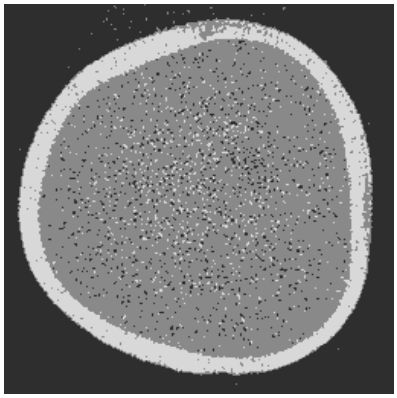
b. Encircled caries on image F



e. MR



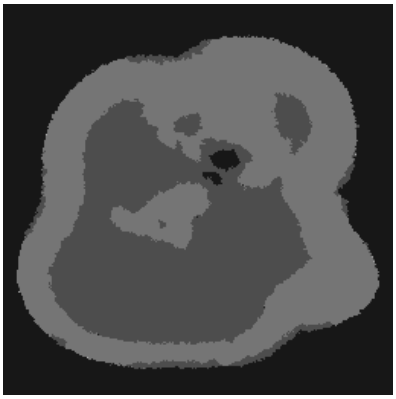
c. SA



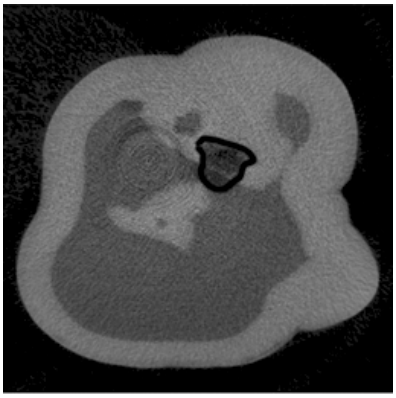
f. MGLD

Figure 4.28. a) Image F. b) Caries on the Image F. c) Result of SA. d) Result of SA-RCF. e) Result of MR. f) Result of MGLD. β =0.22 was selected for all these images.

a. Image G

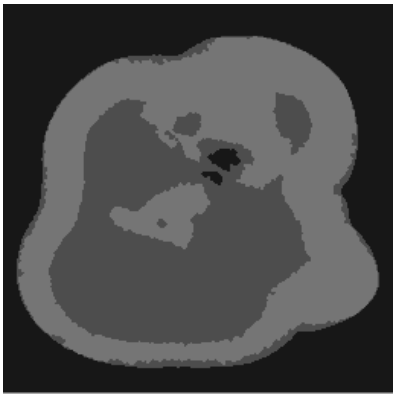


d. SA-RCF

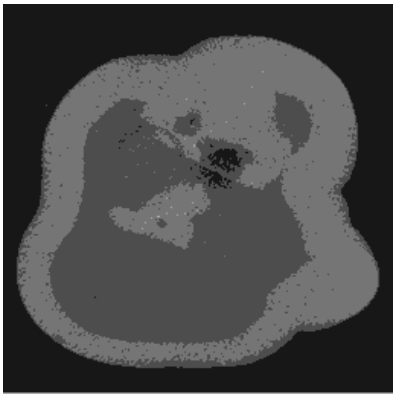


b. Encircled caries on image G

e. MR



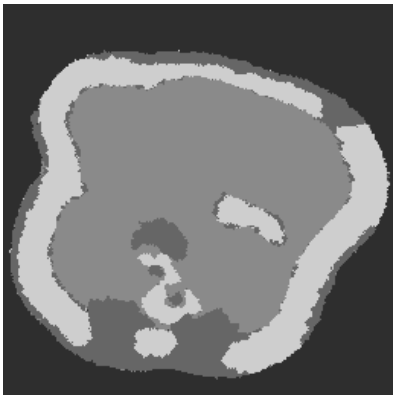
c. SA



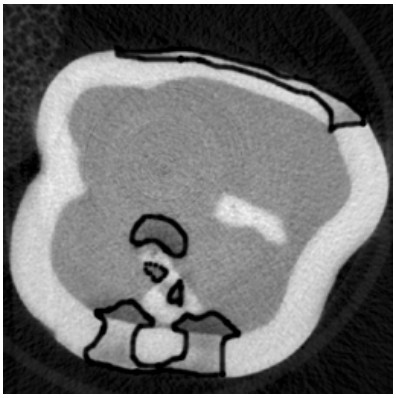
f. MGLD

Figure 4.29. a) Image G. b) Caries on the Image G. c) Result of SA. d) Result of SA-RCF. e) Result of MR. f) Result of MGLD. β =0.22 was selected for all these images.

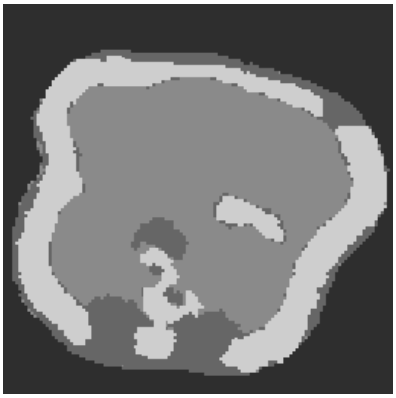
a. Image H



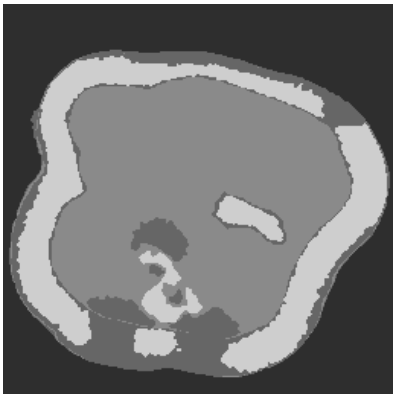
d. SA-RCF



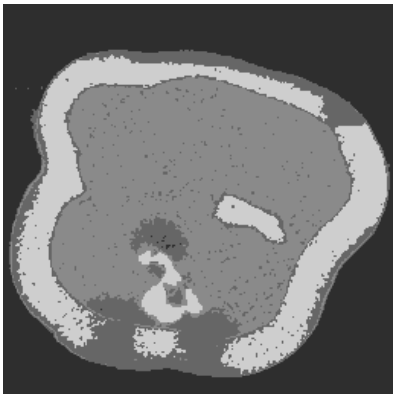
b. Encircled caries on image H



e. MR

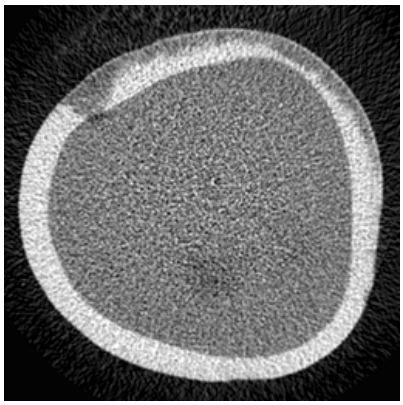


c. SA

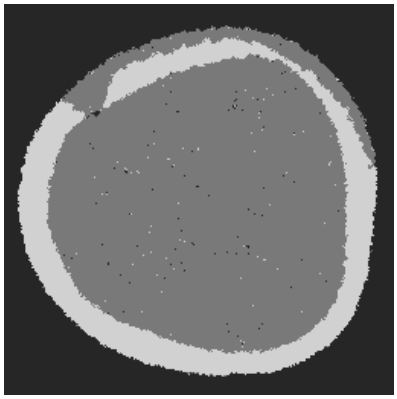


f. MGLD

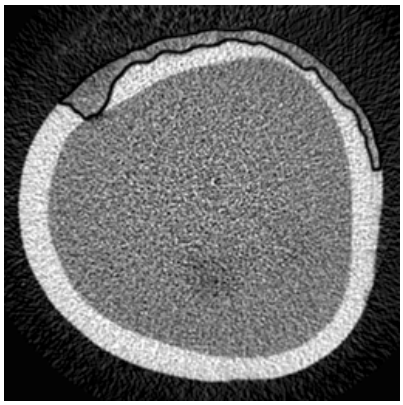
Figure 4.30. a) Image H. b) Caries on the Image H. c) Result of SA. d) Result of SA-RCF. e) Result of MR. f) Result of MGLD. β =0.22 was selected for all these images.



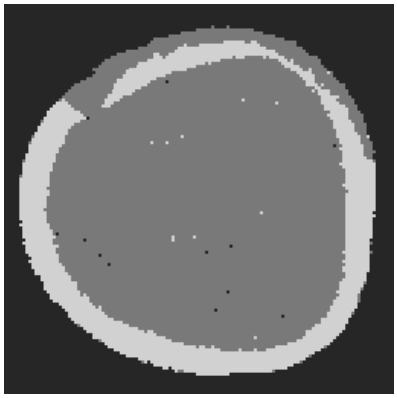
a. Image K



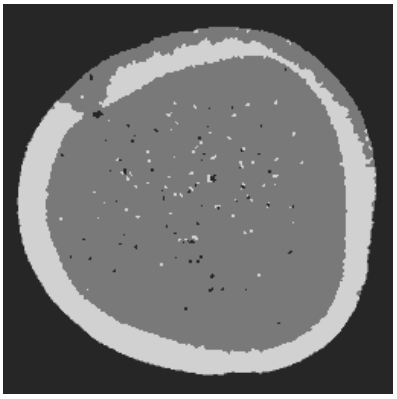
d. SA-RCF



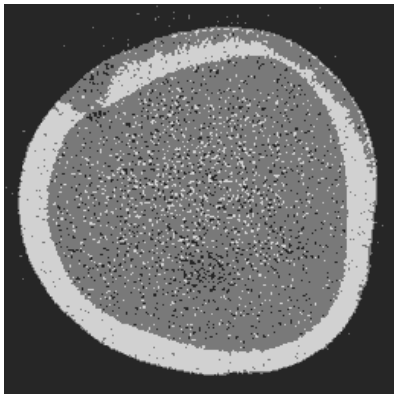
b. Encircled caries on image K



e. MR



c. SA



f. MGLD

Figure 4.31. a) Image K b) Caries on the Image K. c) Result of SA. d) Result of SA-RCF. e) Result of MR. f) Result of MGLD. β =0.22 was selected for all these images.

To assess image segmentation results, gradations on dental images (Enamel, Dentin, caries and background) were adjusted by region and the error of image segmentation was calculated. The results of image segmentation errors for the value of β=0.22 are shown in Figure 4.32.

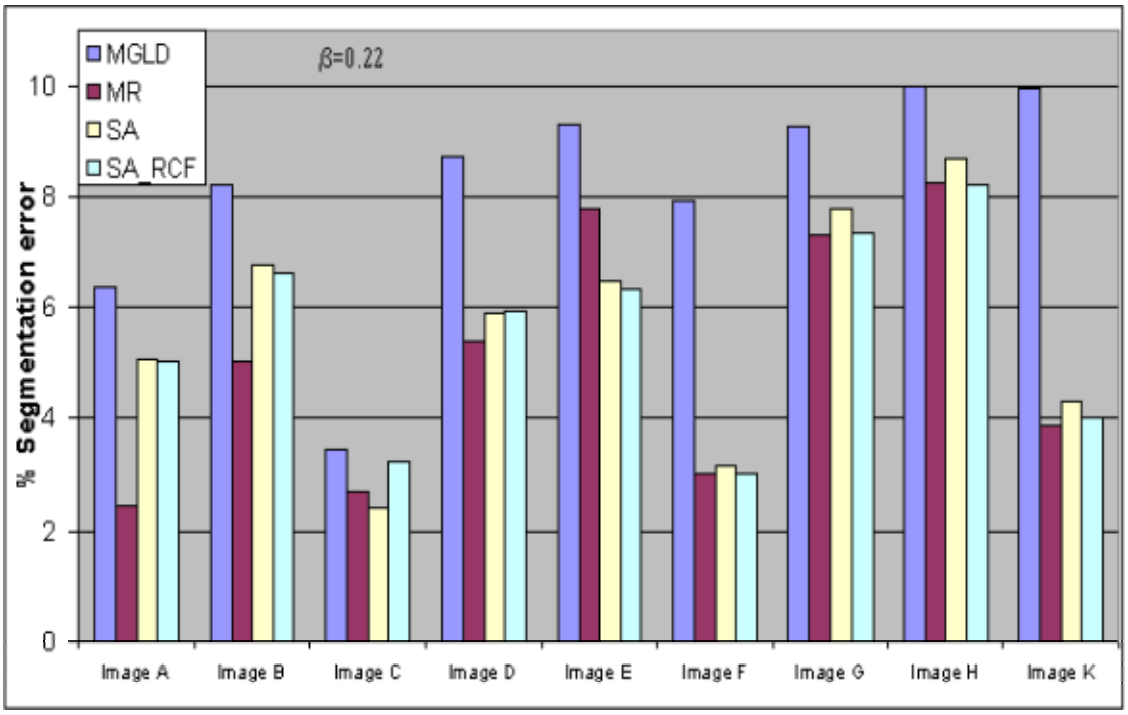


Figure 4.32. Error of image segmentation on tooth images.

The results show that MR has produced minimum error value for most of the images compared with the other methods. The performance rank to obtain minimum error among four methods is shown in figure 4.33 for the 9 dental images.

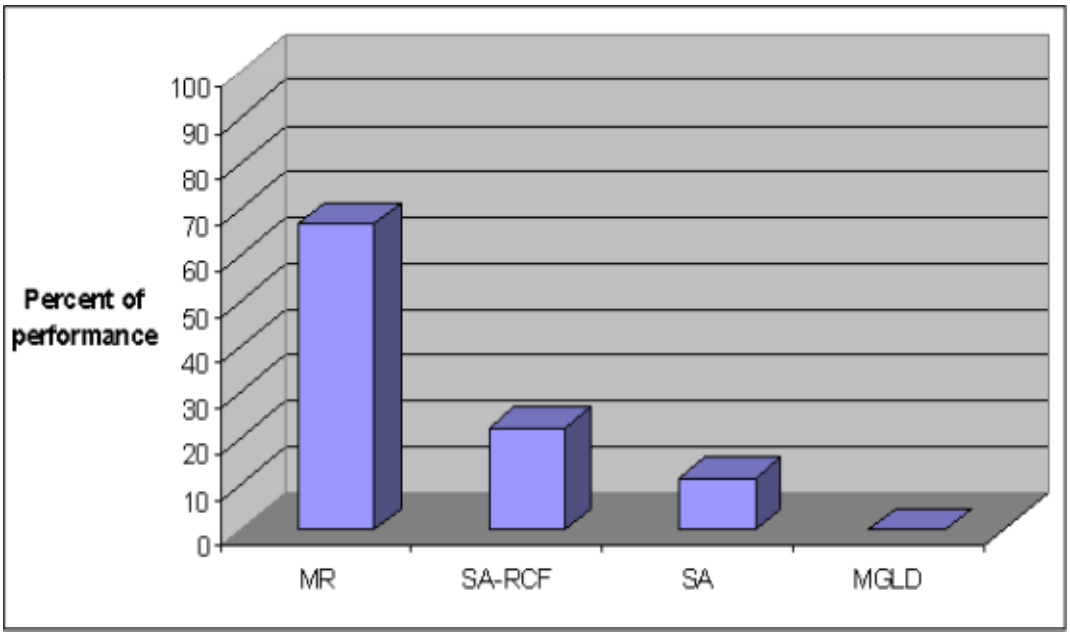


Figure 4.33. Performance rank for MGLD, MR, SA and SA-RCF image segmentation methods.

4.6 Conclusion

We have demonstrated the advantages of unsupervised image segmentation compared with manual segmentation in terms of accuracy of the results. This chapter illustrated the Intra Observer Agreement and the difference between the two observations at two different times. A regression line was fit to this data; ICC and R^2 were calculated. The value of "1" for R^2 and ICC shows the perfect agreement between the lesion sizes in two image sets, while "0" value proves no agreement in the results. R^2 was calculated as 0.9613 which the regression line has an intercept of 29.86. The desired intercept value is 0 while R^2 approaches "1." The ICC was calculated as 0.8642. Our data shows a high ICC for repeatability because the range of our data goes from approximately 200 to 3500 pixels. As compared to the large variability between caries, the differences within caries are relatively small. All of our pairs of data are distinct so again that will lead to a high ICC. Alternatively, we calculated the error of identifying dislocation of lesions. The dentist's average error when determining the caries sizes and locations at the two different times was 18.79%. We are certain that the average error will be in the interval of 15.49%-22.09% with a 95% Confidential Interval (CI).

In this chapter we used proposed methods of parameter estimation and segmentation proposed in chapters 2 and 3 respectively to segment without supervision nine μCT dental images.

In chapter 2, we observed that Interlaced GA-EM produced the minimum estimation error of the Gaussian mixture model of phantom images; and as we expected it was also correct for μCT dental images. Since the Interlaced GA-EM gave the lowest error value for the mixture model parameters, the estimated parameters of Interlaced GA-EM were used to segment dental images. All segmentation methods introduced in chapter 3 were applied to these images. Segmentation performance was evaluated based on the ratio between the number of misclassified pixels in each image and the total numbers of pixels in the image. MR was most successful in minimizing the segmentation error and computational time. MGLD yielded poor segmentation results since its performance is highly related to the noise level of the image.

5. CONCLUSIONS

This thesis proposed solutions and addressed two major problems in image analysis:

1) Parameter estimation on the Gaussian mixture model

2) Image segmentation using the Markov random field

GAs, SA, and a mix of them were introduced to estimate image parameters. These methods were mixed with EM to increase their abilities to find better solutions for parameters in the mixture the Gaussian model.

Previously, it was shown that GA is independent of initialization parameters and GAs are an efficient technique to optimize difficult functions in large search spaces. We have shown that using the EM algorithm might not always be successful because the true number of mixing components is unknown and the algorithm can get trapped in one of the many local maxima of the likelihood function; furthermore, there is no generally accepted method for parameter initialization.

We introduced GA to estimate the Gaussian mixture parameters; the performance of GAs is highly dependent on how the objective function of the system is defined. We tested several types of objective functions, and in this thesis we used the best of them for parameter estimation of medical images and CT phantoms. In that, we presented another new method to find out the Gaussian mixture model parameters by GAs. An alternative approach to this problem is to have a two-step process, first estimating the parameters by GA and getting crude *a priori* information. Second the *a priori* information is used as an initial guess for EM algorithm to find out the final parameters. So the problem of parameters can be solved using both GA and EM (GA-EM) algorithms.

We introduced a new algorithm (GA-SA) for the purpose of increasing accuracy in estimation parameters. Since both GA and SA are robust for finding the minimum global energy, a mix of them with EM (GA-SA-EM) increases the chance of finding global

minimum energy, thus providing yet another alternative solution to parameter estimation problems.

In the last step to boost parameter estimation methods, we have introduced Interlaced GA-EM methods. We proposed a new incorporated genetic algorithm with EM (Interlaced GA-EM) to improve the estimation of the Gaussian mixture parameters. The method used a population of mixture models, rather than a single mixture, interactively in both GA and EM to determine the Gaussian mixture parameters. Alternatively, we have used the same technique to introduce Interlaced GA-CGM (Conjugate Gradient Method).

Performance accuracy of the parameter estimation of EM, GA, GA-EM, GA-SA, GA-SA-EM, Interlaced GA-EM and Interlaced GA-CGM was calculated on both CT phantoms, based on a modified Shepp-Logan method, and a series of dental images. It has been shown that GAs are independent of initialization parameters and can provide an efficient technique to optimize functions in large search spaces while the solution obtained by EM is a function of initial parameters, hence the relatively high likelihood of achieving a sub-optimal solution, due to trapping in a local maxima. The results indicated that the EM algorithm, as expected, was heavily impacted by initial values and the EM failed to calculate the parameters of the image for one of the images. Coupling GA-SA with EM (GA-SA-EM) not only improves the overall accuracy, it also provides estimates that are independent of initial seed values. The best result was obtained from Interlaced GA-EM on both error value and accuracy of the obtained parameters. EM has a lower computational time while Interlaced GA-CGM has the highest. The rest of the methods have almost all the same computational time higher than the EM and much lower than the Interlaced GA-CGM.

The proposed methods offer an accurate and stable technique to estimation parameters of the Gaussian mixture models, with a higher likelihood of achieving the global optimal. Obtaining such accurate parameters is a key requirement for image analysis and signal processing approaches.

This thesis also proposed two improvements to Simulated Annealing for performing image segmentation using the Markov random field. The first improvement addressed

performance of the algorithm in noisy images by randomizing the pixel used in the cost function of the image (SA-RCF). The second improvement reduces the computational complexity and time by performing stochastic sampling within a multiresolution scheme (MR).

We compared MR, SA-RCF, SA and MGLD methods for segmenting images into a discrete set of classes in CT phantom images. The SA-RCF and MR algorithms achieve higher segmentation accuracy in the presence of noise. MGLD provides excellent segmentation accuracy for noise-free images, with comparable computational time. The MR method achieves higher segmentation accuracy at significantly lower computational cost compared to SA and MGLD. Both MR and SA-RCF are less dependent on β values (a fixed control parameter on the energy function) compared to SA and MGLD. Independence of small variations of the β value confirms that both MR and SA-RCF are more successful to find global minimum energy than SA and MGLD. However, due to its multiresolution decomposition of images, segmentation error is not distributed uniformly across regions of various sizes, *i.e.*, the error in smaller regions is greater than in larger regions. SA and SA-RCF have the same computational complexity, while SA-RCF offers a more robust segmentation for noisy images. Those segmentation methods have segments unsupervised a series of nine dental images that had different caries sizes and signal-to-noise ratios. Since Interlaced GA-EM had the best performances to estimate the mixture model parameters of dental images, image segmentation methods have used its estimated parameters to segment the images in which caries sizes were different. Performances of image segmentation methods are compared to a dentist's manual segmentation. The results showed that MR has better segmentation results in significant lower computation time as compared with SA and SA-RCF. MGLD obtained poor segmentation results since its performance is highly related to the noise level of an image.

In the statistical approach, we also showed that a dentist's error average to detect a lesion on the tooth image at two different times has an interval of 15.49%-22.09% with a 95% confidential interval.

This page deliberately left blank

REFERENCES

REFERENCES

[1] R. C. Gonzalez and P. Wintz, *Digital Image Processing*. Addison-Wesley Pub. Co., 1987.

[2] Jia Li, Robert M. Gray, *Image Segmentation and Compression Using Hidden Markov Models*, Kluwer Academic Publishers; 1st edition 2000.

[3] Blake and Zisserman, *Visual Reconstruction*, The MIT Press, Cambridge, USA, 1987.

[4] Nikhil R. Pal and Sankar K. Pal, "A review on image segmentation techniques," *Pattern Recognition*, vol. 26, no.9, pp.1277-1294, 1993.

[5] Robert M. Haralick and Linda G. Shapiro, "Image segmentation techniques," *Computer Vision Graphics, and Image Processing*, vol.29, pp.100-132, 1985.

[6] S. Z. Li, "Modeling image analysis problems using Markov random fields", In C.R. Rao and D.N. Shanbhag (Ed), *Stochastic Processes: Modeling and Simulation*, Elsevier Science. Volume 20 of Handbook of Statistics, pp. 1-43, 2000.

[7] J. M. Hammersley and P. Clifford, "Markov field on finite graphs and lattices," Unpublished 1971.

[8] J. Besag, "Spatial interaction and the statistical analysis of lattice systems," *J. of Royal statistical Society*, Series B, vol. 36, pp.192-236, 1974.

[9] U. Grenander, *Tutorial in Pattern Synthesis*, Brown University, Division of Applied Mathematics, 1983.

[10] Geman and D. Geman, "Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 6, no. 6, pp. 721-741, 1984.

[11] B. Gidas,"A renormalization group approach to image processing problems," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 11, no.2, pp. 164-180, 1989.

[12] R. Chellappa, "Two-Dimensional discrete Gaussian Markov random field models for image processing," In L. N. Kanal and A. Rosenfeld, editors, *Progress in Pattern Recognition 2*, pages 79-112, 1985.

[13] F. S. Cohen and D. B. Copper, “Simple parallel hierarchical and relaxation algorithms for segmentation non casual Markovian random fields,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.9, no. 2, pp. 195-218, March 1987.

[14] H. Dernin, H. Elliott, R. Cristi, and D. Geman, “Bayes smoothing algorithms for segmentation of binary images modeled by Markov random fields,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.6, no. 6, pp. 707-720, November 1984.

[15] N. S. Friedland and A. Rosenfeld, “Compact object recognition using energy function based optimization,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.14, no. 7, pp. 770-777, July 1992.

[16] F. R. Hansen and H. Elliott, “Image segmentation using simple Markov random field models,” *Computer Graphics Image Processing*, vol. 20, pp. 101-132, 1982.

[17] M. Hassner and J. Slansky, “Use of Markov random field as models of texture,” *Computer Graphics Image Processing*, vol.12, pp. 357-370, 1980.

[18] S. Z. Li, “Invariant surface segmentation through energy minimization with discontinuities,” *International Journal of Computer Vision*, vol. 5, no. 2, pp. 161-194, 1990.

[19] S. Z. Li, “Close-form solution and parameter selection for convex minimization based edge-preserving smoothing,” *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 20, no. 9, pp. 916-932, September 1998.

[20] S. Lakshmanan and H. Derin, “Gaussian Markov random field at multiple resolutions,” In R. Chellappa and A. Jain, editor, *Markov Random Fields- Theory and application*, pp.131-157. Academic Press, 1993.

[21] C. Bouman and M. Shapiro, “A multiscale random field model for Bayesian image segmentation,” *IEEE Transaction on Image Processing*, vol. 3, no. 2, pp. 162-177, 1994.

[22] H. Dernin and H. Elliot, “Modeling and segmentation of noisy and textured images using Gibbs random fields,” *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 9, pp. 39-55, Jan. 1987.

[23] Geoffrey J. McLachlan, Kaye E. Basford, *Mixture Model Inference and Application to Clustering*, Marcel Dekker, Inc. 1988.

[24] Richard A. Render and Homer F. Walker, “Mixture densities, maximum likelihood and the EM algorithm,” *SIAM review*, vol. 26, no. 2, April 1984.

[25] N. A. Vlassis, G. Papakonstantinou, and P. Tsanakas, “Mixture Density Estimation Based on Maximum Likelihood and Test Statistics,” *Neural Processing Letters Journal*, 1998.

[26] Tood K. Moon, “The expectation maximization algorithm,” *IEEE Signal Processing Magazine*, pp. 47-60, November 1996.

[27] M. Krzywinski, V. Sossi, T.J. Ruth, “Comparison of FORE, OSEM and SAGE algorithms to 3DRP in 3D PET using phantom and human subject data,” *IEEE Transactions on Nuclear Science*, vol. 46, no.4, pp. 1114 -1120, August 1999.

[28] Jr. R.S., Lynch, “Performance results of recognizing various class types using classifier decision fusion,” *Proceedings of the Fifth International Conference on Information Fusion*, vol. 1, pp. 266 -271, 2002.

[29] G. R. Dattatreya, “Gaussian mixture parameter estimation with known means and unknown class-dependent variances,” *Pattern Recognition*. vol.35, pp.1611-1616, 2002.

[30] D. E. Goldberg, *Genetic Algorithms in Search Optimization and Machine Learning*, Addison Wesley Publishing Company, 1989.

[31] Sankar K. Pal and Paul P. Wang, *Genetic Algorithms for Pattern Recognition*, CRC press Inc., 1996.

[32] N. Majdi Nasab, and M. Analoui “Mixture conditional estimation using genetic algorithms,” *IEEE ISSPA 2001: Statistical Signal Processing*, Kuala-Lampur, Malaysia, 13-16, August 2001.

[33] N. Majdi Nasab and M. Analoui “Genetic algorithm and expectation maximization for parameter estimation of mixture Gaussian model phantom,” *SPIE, Medical Imaging*, vol. 4684, pp. 864-871, 2002.

[34] C. Bouman and B. Liu, “Multiple resolution segmentation of texture images,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 2, pp. 99-113, February 1991.

[35] Z. Kato, M. Berthod, and J. Zerubia, “Multiscale Markov random field models for parallels image classification,” *Proceeding of the forth International Conference on Computer Vision*, Berlin, Germany, May 1993.

[36] S. Krishnamachari and R. Chellappa, “Multiresolution Gauss-Markov random Field model,” Technical Report, University of Maryland, College Park, Maryland, Jan. 1995.

[37] N. Majdi Nasab and M. Analoui, "An efficient multiresolution approach for image segmentation based on Markov random field," *Proc. SPIE, Medical Imaging 2001: Image Processing*, vol. 4322, pp.1066-1074, 2001.

[38] N. Majdi Nasab, M. Analoui, and Edward J. Delp, "Robust and efficient image segmentation approaches using Markov random field," *Journal of Electronic Imaging*, vol. 12, no.2, pp.50-58 Jan. 2003.

[39] Geoffrey J. McLachlan, Kaye E. Basford, *Mixture Model Inference and Application to Clustering*, Marcel Dekker, Inc. 1988.

[40] Richard A. Render and Homer F. Walker, "Mixture densities, maximum likelihood and the EM algorithm," *SIAM review*, **26**, No. 2, April 1984.

[41] N. A. Vlassis, G. Papakonstantinou, and P. Tsanakas, "Mixture density estimation based on maximum likelihood and test statistics" *Neural Processing Letters Journal*, vol. 9, No. 1, pp.63-76, 1999.

[42] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. of Royal statistical Society*, Series B, vol. 39, no. 1, pp.1-38, 1977.

[43] P. Masson and W. Pieczynski, "SEM Algorithm and unsupervised statistical segmentation of satellite images," *IEEE Geosciences and Remote Sensing*, vol. 31, no. 3, pp. 618-633, May 1993.

[44] D. Geman, "Bayesian image analysis by adaptive annealing" *Proc. IGARSS'85*, pp.269-277, Amherst, USA, Oct. 1985.

[45] M. L. Comer and E. J. Delp, "The EM/MPM Algorithm for segmentation of textured images: analysis and further experimental results," *IEEE Transactions on Image Processing*, vol. 9, no. 10, pp. 1731 -1744, October 2000.

[46] Davis, L., Ed., *Handbook of Genetic Algorithms*, Van Nostrand Reinhold, New York 1991.

[47] A. Jain, *Fundamental of Digital Image Processing*, Prentice Hall, NJ, 1989.

[48] D. G. Luenberger, *Linear and Nonlinear Programming*. Readings, MA: Addison-Wesley, 2nd Ed, 1984.

[49] F. Lin, C. Kao, and C, Hsu "Applying the Genetic approach to Simulated Annealing in solving some NP-hard problems," *IEEE Transaction on Systems, Man. and Cybernetics*, vol. 23, No. 6. Nov. 1993.

[50] N. Majdi Nasab, M. Analoui, and Ed. J. Delp, "Improved Methods for Parameter Estimation of Mixture Gaussian Model Using Genetic and Maximum Likelihood Algorithms," *SPIE, Medical Imaging 2004: Image Processing*, vol. 5370, 2004.

[51] R. Salakhutdinov, S. Roweis, and Z. Ghahramani, "Optimization with EM and Expectation-Conjugate-Gradient," *International Conference on Machine Learning 20*, pp.672-679, 2003.

[52] L. E. Ortiz, and L. P. Kaelbling. Notes on methods based on maximum-likelihood estimation for learning the parameters of the mixture of The Gaussians model. Technical Report CS-99-03, Brown University, Providence, IR, Feb., 1999.

[53] E. K. P. Chong, S. H. Zak, *An Introduction to Optimization*, 2nd Ed., John Wiley Sons Inc., 2001.

[54] R. Kinderman and J. L. Snell, *Markov Random Fields and Their Applications*. Providence, Rhode Island: American Mathematical Society, 1980.

[55] P. Pérez, F. Heitz, "Restriction of a Markov random field on a graph and multiresolution statistical image modeling," *IEEE Transaction on Information Theory*, vol. 42, no. 1, pp. 180-190, January 1996.

[56] M. Lafarté, P. Pérez, F. Heitz, "Discrete Markov modeling and inference on the quad-tree," *IEEE Transaction on Image Processing*, vol. 9, no. 3,pp. 390-404, March 2000.

[57] S. Z. Li, *Markov Random Field Modeling in Image Analysis*, Springer-Verlag, 2nd edition, Tokyo, 2001.

[58] J. M. Laferté, F. Heitz, P. Pérez, and E. Fabre, "Hierarchical statistical models for the fusion of multiresolution image data," *In Proc. Int. Conf. Computer Vision*, pp. 908-913, Cambrige, USA, June 1995.

[59] C. Graffigne, F. Heitz, P. Perez, F. Preteux, M. Sigelle, J. Zerubia, "Hierarchical Markov random field models applied to image analysis: a review," *Proc. SPIE*, vol. 2568, pp. 2-17, 1996.

[60] W. Hackbusch, *Multigrid Methods and Applications*, Spronger-Verlag, Berlin, 1985.

[61] D. Terzopoulos, "Image analysis using mutigrid relaxation methods," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 8, no. 2, pp. 129-139, 1986.

[62] F. C. Jeng, "Subsampling of Markov random fields" *J. of Visual Com. And Image Representation*, vol. 3, no. 3, pp. 225-229, 1992.

[63] G. K. Nicholls and M. Petrou, “Multiresolution representation of Markov random fields,” Technical Report VSSP-TR3/93, University of Surrey, 1993.

[64] P. Pérez, F. Heitz, “Global Bayesian estimation, constrained mutiscale Markov random fields and the analysis of visual motion,” *Proc. 12th International Maxent workshop*, pp. 383-388, Paris, July, 1992.

[65] J. Besag, “On the statistical analysis of dirty pictures,” *J. of Royal statistical Society*, Series B, vol. 48, no.3, pp. 259-302, 1986.

[66] M. Barlaud, L. Blanc-Féraud, and P. Charbonnier, “Image coding using multiresolution Markov random fields,” *Proc. SIPE*, vol.1657, pp.25-35, 1992.

[67] F. Marques and A. Gasull, “Stochastic image model for segmentation. Application to image coding,” *Proc. Scandinavian Conf. Image Analysis*, pp. 265-272, Tromso, Norway, 1993.

[68] Z. Kato, J. Zerubia, and M. Berthod, “Unsupervised parallel image classification using hierarchical Markovian model,” *Proc. Int. Conf. Computer Vision*, pp. 169-174, Cambridge, USA, June 1995.

[69] B. Gidas. “A multilevel-multiresolution technique for computer vision via renormalization group ideas,” *Proc. SPI*, vol. 880, pp. 214-218,1988.

[70] P. Pérez, F. Heitz, “Multiscale Markov random fields and constrained relaxation in low level image analysis,” *Proc. Int. Conf. Acoustic Speech Signal Proc*., vol. III, pp. 61-64, March, 1992.

[71] Athanasios Papoulis, *Probability, Random variables, and the Stochastic Process*, McGraw-Hill, Inc., 1991.

[72] X. Descombes, M. Moctezuma, H. Maître, J.P. Rudant, “Coastline detection by a Markovian segmentation of SAR images,” *Signal Processing*, vol. 55, pp. 123-132, 1996.

[73] A. Johnson and D. W. Wichern, *Applied Multivariate Statistical Analysis*. 4th Edition, Prentice-Hall Inc., 1998.

[74] DT. Zero, “Application of clinical models in remineralization research,” *The Journal of Clinical Dentistry X*:74-85, 1999.

[75] U.S. Department of health and Human Services, National Institute of Dental Research. Oral Health of United States Children, The National Survey of Dental Caries in U.S. School Children: 1986-1987. NIH Publication No. 89-2247. 1989.

[76] U.S. Department of health and Human Services, National Institute of Dental Research. Oral Health of United States Adults: The National Survey of Oral Health in U.S. Employed Adults and Seniors: 1985-1986. NIH Publication No. 87-2868, 1987.

[77] F. Sundstrom, K. Fredriksson, S. Montan, U. Hafstrom-Bjorkman, J. Strom, "Laser-induced fluorescence from sound and carious tooth substance: spectroscopic studies," *Swedish Dental Journal* 9:71-80, 1985.

[78] H. Bjelkhagen, F. Sundström, B. Angmar-Månsson, H. Rydé, "Early detection of enamel caries by the luminescence excited by visible laser light," *Swed Dent J* 6:1-7, 1982.

[79] A. Lussi, S. Imwinkelried, C. Longbottom, E. Reich, "Performance of a laser fluorescence system for detection of occlusal caries', *Caries Res* 34:297, 1998.

[80] H. Eggertsson, M. Ando, M. Analoui, and GK. Stookey, "Early caries detection using QLF and IR (Diagnodent), validated by TMR," *Intl National Association of Dental Research*, Washington, DC, 2000.

[81] DN, Ricketts, E. Kidd, R. Wilson, "Electronic diagnosis of occlusal caries in vitro: Adaptation of the technique for epidemiological purposes," *Community Dent Oral Epidemio*l, 238-241, 1997.

[82] WP. Rock, EAM Kidd, "The electronic detection of demineralization in occlusal fissures," *Br. Dent.J.* 164:243-247, 1998.

[83] EH. Verdonschot, P. Rondel, MC Huysmans, "Validity of electrical conductance measurements in evaluating the marginal integrity of sealant restorations," *Caries Research* 29:100-106, 1995.

[84] DT. Zero, "Caries detection using digital imaging fiber-optic trans-illumination," *Indiana Conference on Early Caries Detection*, Indianapolis, IN, Indiana University, 1999.

[85] A. Scheiderman, M. Elbaum, T. Schultz, "Assessment of dental caries with digital imaging, fiber-optic transillumination (DIFOTI): In vitro Study," *Caries Res* 31:103-110, 1997.

[86] M. Analoui, H. Eggertsson, and Stookey G.K, "Quantitative radiographic techniques for caries detection and analysis," *Indiana Conference on Caries Detection Indianapolis*, IN, Indiana University.1999.

[87] JB. Briley, SB. Dove, EJ. Mertz-Fairhurst, CB. Hermesch, "Computer-assisted densitometric image analysis (CADIA) of previously sealed carious teeth: a pilot study," *Oper.Dent.* 22:105-114, 1997.

[88] MJ. Everett, BW. Colston, US. Sathyam, LB. da Silva, D. Fried, and JD. Featherstone, "Noninvasive diagnosis of early caries with polarization-sensitive optical coherence tomography (PS-OCT)," Proc. SPIE Vol. 3593. Lasers in Dentistry, V 177-182. 1999.

[89] DD. Arnone, CM. Ciesla, A. Corchia, S. Egusa, "Applications of terahertz (THz) technology to medical imaging," *Proc. SPIE Terahertz Spectroscopy and Applications II*, vol. 3828, 209-219, 1999.

[90] F. Yanikoglu, F. Ozturk, O. Hayran, M. Analoui, "Detection of natural white spot lesions by an ultrasonic system," *Caries Res*. 2000.

[91] XQ Shi, P. Han, U. Welander, B. Angmar-Mansson, "Tuned-aperture computed tomography for detection of occlusal caries," *Dentomaxillofac.Radiol.* 30:45-49, 2001.

[92] JM. Abreu, DA. Tyndall, E. Platin, JB. Ludlow, C. Phillips, "Two- and three-dimensional imaging modalities for the detection of caries. A comparison between film, digital radiography and tuned aperture computed tomography (TACT)," *Dentomaxillofac.Radiol.* 28:152-157, 1999.

[93] MK. Nair, DA. Tyndall, JB. Ludlow, K. May, "Tuned aperture computed tomography and detection of recurrent caries," *Caries Res* 32:23-30, 1998.

[94] J. Arends, JL. Ruben, D. Inaba, "Major topics in quantitative microradiography of enamel and dentin: R parameter, mineral distribution visualization, and hyper- remineralization," *Adv.Dent Res* 11:403-414, 1997.

[95] E. de Josselin de Jong, JJ. ten Bosch, J. Noordmans, "Optimised microcomputer-guided quantitative microradiography on dental mineralised tissue slices," *Phys.Med Biol* 32:887-899, 1987.

[96] J. Arends, JJ. ten Bosch, "Demineralization and remineralization evaluation techniques," *J Dent Res* 71 Spec No:924-928, 1992.

[97] T. Koulourides, "Dynamics of tooth surface-oral fluid equilibrium," *Adv.Oral Biol* 2:149-171, 1966.

[98] JD. Featherstone, JM. ten Cate, M. Shariati, J. Arends, "Comparison of artificial caries-like lesions by quantitative microradiography and microhardness profiles," *Caries Res* 17:385-391, 1983.

[99] FM. Herkstroter, M. Witjes, J. Ruben, J. Arends, "Time dependency of microhardness indentations in human and bovine dentine compared with human enamel," Caries Res 23:342-344, 1989.

[100] LM. Silverstone, JS. Wefel, BF. Zimmerman, BH. Clarkson, MJ Featherstone, "Remineralization of natural and artificial lesions in human dental enamel in vitro," *Effect of calcium concentration of the calcifying fluid. Caries Res* 15:138-157, 1981.

[101] L. Holmen, A. Thylstrup, J. Artun, "Clinical and histological features observed during arrestment of active enamel carious lesions in vivo," Caries *Res* 21:546-554, 1987.

[102] S. Kashket, "Historical review of remineralization research," *The Journal of Clinical Dentistry X*:56-64, 1999.

[103] JJ. ten Bosch, H. van der Mei, P. Borsboom, "Optical monitor of in vitro caries," *Caries Res* 18:540-547, 1984.

[104] J. Brinkman, J. ten Bosch, P. Brosboom, "Optical quantification of natural caries in smooth surfaces of extracted teeth," *Caries Res* 22:262, 1996.

[105] F. Herkstroter, J. Noordmans, J. ten Bosch, "Wavelength-independent microradiography: Its use to measure mineral changes in curved and thick surfaces," *Caries Research* 24:399, 1990.

[106] AGF. Zandona, M. Analoui, AH Kafrawy, and GK. Stookey, "Comparison of histology and confocal laser microscopy on analyses of occlusal fissure demineralization," *International Association of Dento MaxilloFacial Radiology*, Louisville, KY,1997.

[107] J. Kinney, Jr G Marshal, S. Marshal, "Three-dimensional mapping of mineral densities in carious dentin: Theory and Method," *Scanning Microscopy* 8:197-205, 1994.

[108] A. Wenzel, H. Hintze, "The choice of gold standard for evaluating tests for caries diagnosis," *Dentomaxillofac.Radiol*. 28:132-136, 1999.

[109] J. M. ten Cate, M. D. lagerweij, JM. Wefel, GK. Stookey, and et.al., "In Vitro Validation Studies of Quantitative Light Fluorescence," Indianapolis, IN, Indiana University. Indiana Conference on Caries Detection, 1999.

[110] M. Analoui, M. Ando, and GK. Stookey, "Comparison of reflectance spectra of sound and carious enamel," *Lasers in Dentistry VI*, John D. Featherstone; Peter Rechmann; Daniel Fried; Eds. *SPIE* vol. 3910, p. 227-234, 2000.

[111] H. Eggertsson, M. Analoui, RL. Isaacs, GJ. Eckert, and GK. Stookey, "Comparison of caries lesion detection using radiographic films and DDR," *J Dent Res 77*(713 (abstr # 651), 1998.

[112] L. Skartviet, M. Analoui, N. Majdi Nasab, J. Spire, C. Brown, “Correlation between Quantitative Bone Analysis and Viewer’s Examination for Detection of Apical Lesions,” IADR, June 25-28, 2003.

[113] AGF. Zandona, RL. Isaacs, M.H. van der Veen, and GK. Stookey, “Indiana pilot study of quantitative light fluorescence,” *Early Detection of Dental Caries II,* Indianapolis, IN, Indiana University, 219-230. 2001.

[114] M. Analoui, H. Eggertsson, A. Laib, and GK. Stookey, “High-resolution computed tomography for caries validation: a pilot study,” *J Dent Res 80* (Special Issue (AADR Abstract)). 2001.

[115] J. Witmer and M. Samuels, *Statistics for the life science*. Prentice Hall, 3rd edition, 2002.

This page deliberately left blank

VITA

VITA

Name: Nariman Majdi Nasab

Address:

Purdue University
Department of Biomedical Engineering
500 Central Drive
West Lafayette, Indiana 47907-2022
Phone: +1 765 49-47054
Fax: +1 765 49-41193
E -mail: nmajdina@ecn.purdue.edu, majdi_nasab@hotmail.com

Education:

Degree	*Date*	*School*
Ph.D.	2004	Purdue University

Dissertation: Stochastic and Biological metaphor parameter estimation on the Gaussian mixture model and image segmentation by Markov random field.

Advisors: Edward J. Delp and Mostafa Analoui

MSBME	1996	Tehran University, Tehran, Iran.
BSEE	1993	K.N. Toosi Industrial University, Tehran, Iran

Professional Affiliations:

IEEE, Institute of Electrical and Electronics Engineering
SPIE, The International Society For Optical Engineering

Publications:

N. Majdi Nasab, M. Analoui, and Edward J. Delp, "Improved methods for parameter estimation of mixture Gaussian model using genetic and maximum likelihood algorithms," SPIE Medical Imaging, vol. 5370, San Diego, CA, USA.

L. Skartviet, M. Analoui, N. Majdi Nasab, J. Spire, C. Brown, "Correlation between Quantitative Bone Analysis and Viewer's Examination for Detection of Apical Lesions," IADR, June 25-28, 2003.

N. Majdi Nasab, M. Analoui, and Edward J. Delp, "Robust and efficient image segmentation approaches using Markov random field," Journal of Electronic Imaging, vol. 12, no.2, pp.50-58 Jan. 2003.

N. Majdi Nasab and M. Analoui, "Genetic Algorithm and Expectation Maximization for Parameter of Phantom with Mixture Gaussian Model," Proc. SPIE, Medical Imaging 2002: Image Processing, Feb. 23-28, 2002, San Diego, CA, USA.

N. Majdi Nasab and M. Analoui, "Mixture Conditional Estimation Using Genetic Algorithms," IEEE ISSPA 2001: Statistical Signal Processing, 13-16 August 2001, Kuala-Lampur, Malaysia.

N. Majdi Nasab, and M. Analoui, "An efficient multiresolution approach for image segmentation based on Markov random field," Proc. SPIE Vol. 4322, Medical Imaging 2001: Image Processing, San Diego, CA, USA, Feb. 17-23, 2001.

N. Majdi Nasab, F. Arámbula, "Modelos Jerárquicos para Segmentación de Imágenes utilizando Campos Aleatorios de Markov- Hierarchical Models for Image Segmentation Using Markov Random Field", SOMI XV, Oct. 16-20, 2000, Guadalajara, Jal., Mexico.

F. Arámbula Cosío and N. Majdi Nasab, "Analysis of genetic algorithms for automatic recognition of prostate in ultrasound images," Mexican International Conference on Artificial Intelligence 2000, April 10-14, 2000, Acapulco, Mexico.

N. Majdi Nasab, S. Shirani, and M. Analoui, "Generation of Bessel limited diffraction beams with hexagonal sparse arrays," Proc. SPIE Vol. 3982, p. 234-242, Medical Imaging 2000: Ultrasonic Imaging and Signal Processing, Feb. 14-17, 2000, San Diego, CA, USA.

N. Majdi Nasab, I. Herrera, and F. Arámbula, "Optimización de un algoritmo genético para reconocimiento automático de próstata utilizado MatLab," SOMI XIV, Oct. 4-8 1999, Tonantzintla, Puebla, México.

N. Majdi Nasab, H. Soltanian-Zadeh, “Producing of Bessel limited diffraction beam with 41×41 elements by sparse arrays,” 1st International conference on biomedical engineering, 11-14 Nov. 1998, Mazatlan, Mexico.

A. R. Zapata F., A. Gaona, N. Majdi Nasab**,** “Biomedical engineering standards and useful equipment for the global economy in Latin America,” 20th Annual International Conference of the IEEE in Medicine & Biology Society, 29 Oct. – 1 Nov. 1998, Hong Kong.

N. Majdi Nasab, H. Soltanian-Zadeh, “Producing of X wave with sparse arrays,” CIMAT, 27Jun-3 July 1998, Guanajuato, Gjo, Mexico.

N. Majdi Nasab, M. Moctezuma, “Segmentation in "CT" images,” Visual98, Second Symposium, Apr. 20-24,1998, Mexico, D.F., Mexico.

N. Majdi Nasab, H. Sharifi, H. Soltanian-Zadeh, M. Oghabian, “Comparing Electrical Impedance Tomography in Circle and Ellipse Surface Model,” 8th International Symposium in Biomedical Eng., Amirkabir Univ. Tehran, Iran.

N. Majdi Nasab, S. Shirani, H. Soltanian-Zadeh, “Generation of Limited Diffraction Beams By Sparse arrays,” 1996 IEEE International Ultrasonic Symposium, Nov. 1996, San Antonio, Texas, U.S.A.