PARARMETER ESTIMATION IN GIBBS-MARKOV IMAGE MODELS

Ayman El-Baz and Aly A. Farag
Computer Vision and Image Processing Laboratory
University of Louisville, Louisville, KY 40292
E-mail: {elbaz, farag}@cvip.Louisville.edu, URL: www.cvip.louisville.edu

***Abstract* -** This paper introduces two novel approaches to estimate the clique potentials in discrete and multilevel realizations of Gibbs Markov random field (GMRF) models. The first approach employs a genetic algorithm (GA) in order to arrive at the closest synthesized image that resembles the original "observed" image. The Second approach is used to estimate the parameters of Gaussian Markov random field. Given an image formed of a number of classes, an initial class density is assumed and the parameters of the densities are estimated using the EM approach. Convergence to the true distribution is tested using the Levy distance. The segmentation of classes is performed iteratively using the ICM algorithm and a genetic algorithm (GA) search approach that provides the maximum a posteriori probability of pixel classification. During the iterations, the GA approach is used to select the clique potentials of the Gibbs-Markov models used for the observed image. The algorithm stops when a fitness function, equivalent to the maximum a posteriori probability, does not change. The approach has been tested on synthetic data and is shown to provide satisfactory results.

1 Introduction

The subject of image modeling involves the construction of models or procedures of the specification of images. These models serve a dual role in that they can describe images that are observed and also can serve to generate synthetic images from the model parameters. We will be concerned with a specific type of image model, the class of texture models. There are important areas of image processing in which texture plays an important area of image processing in which texture plays important role: for example, classification, image segmentation, and image encoding.

Several schemes have been proposed in the computer vision literature to estimate the parameters of an MRF. For MRFs defined on pixel sites (e.g. texture modeling), these schemes have been applied with considerable success. For MRFs defined in edge sites (line variable used to denote discontinuity between adjacent pixels), however, the available parameter estimation techniques [1]-[5] are difficult to apply because of the lack of true edge labels. Also the Least squares (LS) method is not accurate [6].

In this paper we introduce two novel unsupervised approaches to estimate GMRFs parameters. In the first approach we use a genetic algorithm to minimize the error between the original image and regenerated image. In the second approach we estimate the model parameters that maximize posteriori probability of each pixel in given image. The MAP estimate is obtained using a combination of genetic search and deterministic estimation using the iterated conditional mode (ICM) approach of Besag (e.g., [7]). The desired estimate of the GMRF parameters is those corresponding to the MAP estimate.

2. Statistical Framework

We first define some basic notation. We will use uppercase letters for random quantities and lowercase letters for their deterministic realization. Throughout this paper, we will assume that the observed image G is considered as a composite of two random process: a high level process G^h, which represents the regions (or classes) that form the observed image; and a low level process G^l, which describes the statistical characteristics of each region (or class). The representation $G = (G^h, G^l)$ has been widely used in the image processing literature in the past two decades.

The high level process (G^h) is a random filed defined on a rectangular grid S of N points, and the value of G^h will be written as G_s^h. Points in G^h will take values in the set $\{1,\ldots,M\}$, where M is the number of regions (or classes). Further, the conditional density function of G^l given G^h, is assumed to exist and to be strictly positive and is denoted by $p(G^l|G^h)$.

Finally, an image is a square array of pixels, or sites, $\{ (i, j): 1\leq i \leq L,\ 1\leq j \leq L\}$. We adopt a simple numbering of sites by assigning sequence number $t = j + L(i - 1)$ to site s. This scheme numbers the sites row by row from 1 to L^2, starting in the upper left.

3. Markov Random Field

The study of Markov random fields has had a long history, beginning with Ising thesis on ferromagnetism [8]. Although it did not prove to be to be a realistic model for magnetic domains, it is approximately correct for phase-separated alloys, idealized gases, and some crystals. The model has traditionally been applied to the case of either Gaussian or binary variables on lattice. Besag [1] allows a natural extension to the case of variables that have integer ranges, either bounded or unbounded. These extensions, coupled with estimation procedures, permit the application of the Markov random field to texture modeling. In this paper, we build on the models described in Geman and Geman 1984 [9], Dubes and Jain 1988 [10], Derrin and Elliot [6], and Farag and Delp [11].

The structure of the neighborhood system determines the order of the MRF. For a first order MRF the neighborhood of a site consists of its four nearest neighbors. In a second order MRF the neighborhood of a site consists of the eight nearest neighbors. The clique structures for a 2nd MRF are illustrated in figure 1. The order coding of the neighborhood up to order five is shown in figure 2

Definition: G^h is a Gibbs random field (GRF) with respect to the neighborhood system $\eta = \{\eta_s: s \in S\}$ if and only if

$$P(G^h = g^h) = \frac{1}{Z} e^{-U(g^h)/T}, \tag{1}$$

where Z is a normalizing constant called the partition function, T is a control parameter called temperature and U is the energy function of the form:

$$U(g^h) = \sum_{all\ cliques\ c} V_C(g^h), \tag{2}$$

where V_C is called a potential and is a function depending only on g_s^h, $s \in C$. Only cliques of size 2 are involved in a pairwise interaction model. The energy function for a pairwise interaction model can be written in the form [10]:

$$U(g^h) = \sum_{t=1}^{L} F(g_t^h) + \sum_{t=1}^{L}\sum_{r=1}^{N} H(g_t^h, g_{t+r}^h), \tag{3}$$

where H(a, b) = H(b, a), H(a, a) = 0, and N depend on the size of the neighborhood around each site. Function F (.) is the potential function for single-pixel cliques, and H(.,.) is the potential function for all cliques of size 2. For example, in the Derin-Elliott [10] model F(.), and H(.,.) are expressed as follows:

$$F(g_t^h) = \alpha\, g_t^h \quad \text{and} \quad H(g_t^h, g_{t+r}^h) = \theta_r I(g_t^h, g_{t+r}^h),$$

where I(a, b) = -1 if a = b and 1 if a ≠ b.

The simulations in this paper are based on the Gibbs-sampler approach of Geman and Geman [9](see also Derrin and Elliot [6] and Dubes and Jain [10]). A realization is generated as follows

Algorithm I

(1) Initialize the L by L image by assigning a color randomly from {0, 1, 2…, M-1} to each site. Call this initial coloring g^h.

(2) for s from 1 to L^2

  (a) Compute probabilities $p(g^h)$ for $g^h = 0, 1, M-1$ where $p(g^h) = p(G_s^h = g_s^h \mid G_r^h = g_r^h, r \in \eta)$, for all $s \in S$ and $\{G_s, s \in S\} \in \Omega$

  (b) Set the color of site s to g^h with probability $p(g^h)$.

(3) Repeat (2) N_{iter} times.

The convergence of this algorithm is assured if N_{iter} is large enough. Figure 3 shows realization of Derin-Elliot model generated with $N_{iter} = 50$, and using ten parameters coefficient corresponding to the second order neighborhood system. All are 64 X 64 images with $\alpha = 1$, and each image consist of two classes. The images shown in figure 3 contain two binary classes. One can map each class to take gray level from 0 to 255 according to certain distribution (e.g., normal distribution N (μ, σ^2).) as shown in figure 4.

4 Model-based parameters estimation

4.1 Parameter estimation for Discrete GMRF

In this section we introduce an approach to estimate the parameters for discrete GMRF such as the images shown in figure 3. The main idea of our approach is to use the GA to generate coefficient of $U(g^h)$, and evaluate these coefficient through the fitness function. We will focus on first and second-order GMRF models, the approach can be readily generalized for higher order models as well. To build the genetic algorithm we define the following parameters (see Goldberg [12] for the terminologies of the genetic algorithm).

1) ***Chromosome****:* A chromosome is represented in binary digits and consists of representations for model order and clique coefficients. Each chromosome has 41 bits. The first bit represents the order of the system. The remaining bits represent the clique coefficients, where each clique coefficient is represented by four bits (note that for first order system we estimate only five parameters, and the remaining cliques coefficient will be zero, but for the second order system we will estimate ten parameters).

2) ***Fitness****:* We defined the fitness of the individual as follow:

$$\text{Error} = |\, G\_image(i, j) - O\_image(i, j) \,|,$$

$$fitness = \begin{cases} 1/\text{Error} & \text{Error} > 0 \\ 10^{10} & \text{Error} = 0 \end{cases}, \tag{4}$$

where G_image is the regenerated image using the estimated parameters for the cliques, O_image is the original image

3) ***GA Process.*** First, the initial population was randomly generated from a sequence of zeroes and ones. Next, the fitness of each individual was calculated. Then crossover and mutation is applied to generate the next population [12].

Algorithm II

(1) Generate the first generation which consists of 30 chromosomes.

(2) Use *Algorithm I* to generate image corresponding to each chromosomes, and use original image as an initial image

(3) If the Fitness value is equal to 10^{10}, then stop and the chromosome, which gives this value, is the desired solution (If there are two chromosomes give the same fitness value we select the chromosomes which represent lower order system).

(4) If the fitness values for all chromosomes in the first generation do not satisfy the required condition, go to 2.

The above algorithm was applied on the six textures shown in figure 3. Table 1 shows the original parameters and estimated parameters for each image. Figure 5 shows the original image and the regenerated image using the estimated parameters. From these results we note that the estimated parameters are very close to original parameters, also the regenerated images close to the original images. These results are more favorable than the least square approach results in [6] and Markov Chain Monte Carlo (MCMC) in [13]. We also note that our method estimates 10 parameters while the results in [6] and [13] were based on four parameters only. Table 2 shows comparison of our results with the results shown in [13]. On the first texture shown in figure 3, the Genetic algorithm converged to the solution after 15 to 20 generation, which took two-three minutes on a PC with an Intel 2.2 GHz processor.

Table 1: Original and estimated parameters.

NO.	Original Parameters	Estimated Parameters
1	[1 0.3 0.3 0.3 0.3 0 0 0 0 0]	[1 0.27 0.29 0.2 0.29 0.01 0.09 0 0 0]
2	[1 1 1 -1 1 0 0 0 0 0]	[0.9 0.9 -0.99 0.9 0.9 0 0 0 0 0]
3	[1 2 2 -1 -1 0 0 0 0 0]	[1 1.99 1.97 -0.95 -0.94 0 0 0 0 0]
4	[1 1 1 -1 1 -1 1 -1 1]	[1 1 0.9 0.9 -0.8 0.99 0.98 1 0.93 0.9]
5	[1 1 1 -1 1 -1 1 -1 1 1]	[0.9 0.7 0.9 -0.8 1 -0.9 0.8 -0.9 0.9 1]
6	[1 -1 1 1 1 1 1 1 1 -1]	[1 -0.99 1 1 1 1 1 1 1 -0.98]
7	[1 1 -1 1 1 1 1 1 1 -1]	[1 1 -0.95 1 1 1 1 1 1 -0.97]
8	[1 1 1 1 1 1 1 1 1 -1]	[1 1 1 1 1 1 1 1 1 -0.91]

Table 2: MRF estimation using different methods

Textures	Methods	θ_1	θ_2	θ_3	θ_4
Texture1	Original	0.3	0.3	0.3	0.3
	LS	0.1152	0.152	0.1867	0.1444
	MCMC	0.3478	0.2762	0.296	0.2877
	σ	0.0266	0.0165	0.0130	0.0086
	Our results	**0.2780**	**0.2900**	**0.289**	**0.291**

4.2 Parameter estimation for Gaussian Markov Random Field

A typical outline for statistical-based image parameters estimation is as follows (e.g., [11], [14], [15]): The observed image process **G** is modeled as a composite of two random processes, a high level process $\mathbf{G}^h$ and a low level process $\mathbf{G}^l$, that is, $\mathbf{G} = (\mathbf{G}^h , \mathbf{G}^l)$. Each of the three processes is a random field defined on the same lattice S.
The maximum a posteriori (MAP) parameters estimation involves the determination of g^h that maximizes $P(G^h = g^h |G^l = g^l)$ with respect to g^h By Bayes' rule,

$$P(G^h = g^h |G^l = g^l) = [P(G^l = g^l|G^h= g^h)P(G^h= g^h)]/P(G^l= g^l). \qquad (5)$$

Since the denominator of Eq. 5 does not affect the optimization, the MAP parameters estimation can be obtained, equivalently, by maximizing the numerator of Eq. 5 or its natural logarithm; that is, we need to find g^h which maximizes the following criterion:

$$\Gamma(G^l,G^h)= \ln P(G^l= g^l|G^h= g^h)+ \ln P(G^h= g^h). \qquad (6)$$

The first term of Eq. 6 is the likelihood due to the low level process and the second term is due to the high level process. Based on the models of the high level and low level processes, the MAP estimate can be obtained.
In order to carry out the MAP parameters estimation in Eq. 6, one needs to specify the parameters in the two processes. A popular model for the high level process is the Gibbs-Markov model, with the probability density function as shown in Eq. 1. In this paper we will assume the model for the low level process to a mixture of normal distributions which follow the following equation:

$$p(g^l) = \sum_{i=1}^{M} \pi_i p_i (g^l), \qquad (7)$$

where M is the number of normal mixture(classes), and π is the mixing proportion.

In order to determine the number of classes for high level image and estimate the mean and variance for each class for low level image we will consider the low level process is a mixture of normal distributions and we will use the Expectation-Maximization (EM) algorithm to estimate the mean, the variance, and the proportion for each distribution (see for example [15] on application of EM algorithm for image classification).
We run the algorithm shown [15] in the first image in figure 4. The results are shown in figure 6. As can seen from figure 6 at m = 2 maximize the conditional expectation so we will select m = 2 for the mixture model. Also it is clear from Table 3 that the original parameters of the mixture are close to the estimated parameters.
In order to determine the convergence between the empirical distribution (which is computed from the low level image), and the reference distribution (mixture of normal distribution, which their parameters are estimated using EM algorithm) we will compute the Levy distance between the empirical distribution (P_{emp}) and the reference distribution (P_{ref}) for each class. The Levy $\rho (P_{emp} , P_{ref})$ distance is defined as:

$$\rho (P_{emp} , P_{ref}) = \inf\{\xi >0: \forall g^l \; P_{emp} (g^l-\xi) - \xi\leq P_{ref} (g^l) \leq P_{emp} (g^l + \xi) + \xi\}. \qquad (8)$$

For example, figure 7 shows the empirical density function of the mixture and reference density function of the mixture. Figure 8 shows the empirical distribution of the mixture and reference distribution for the mixture. For these distributions $\rho (P_{emp} , P_{ref}) = 0.0001$. Since as $\rho (P_{emp} , P_{ref})$ approach to zero $P_{emp}(g^l)$ converge weekly to $P_{ref} (g^l)$, then above result indicate that the $P_{emp} (g^l)$ converge to $P_{ref} (g^l)$, and this proves that the choosing mixture of normal distributions to be the model for the low level process is correct.

Table 3 shows The Estimated, and the original parameters

Parameter	μ_1	μ_2	σ_1^2	σ_2^2
Estimated	31.4570	153.0938	153.4138	157.5943
original	31.9027	152.8616	149.8957	158.09

In order to estimate the parameters of the model of GMRF that maximizes Eq. 6, we will use the iterative conditional mode (ICM) approach [7], and genetic algorithm (GA). The ICM is a relaxation algorithm to find a global maximum. The algorithm assumes that the classes of all neighbors of a pixel g are known. The high level process is assumed to be formed of *M*-independent processes each of the *M* processes is modeled by Gibbs-Markov random field which follow Eq. 1. Then g^l can be classified using the fact that $p(g_i^h | g_i^l)$ is proportional to $p(g_i^l | g_i^h) p(g_i^h | \eta_s)$; i.e.,

$$p(g_i^h | g_i^l) \alpha \; p(g_i^l | g_i^h) p(g_i^h | \eta_s), \tag{9}$$

where η_s is the neighbor set to site S belonging to class g_i^h, $p(g_i^h | \eta_s)$ is computed from Eq. 1

To get initial labeling image (high level process), we will use Bayes classifier (see [16] for more details about Bayes classifiers).

In order to run ICM, first we must know the coefficient of potential function $U(g^h)$, so we will use GA to path the coefficient of $U(g^h)$, and evaluate these coefficients through the fitness function.

We will use the same structure of GAs that described in the section 4.1 and using Eq. 6 to be the fitness function of each chromosome.

Algorithm III

(1) Generate the first generation which consists of 30 chromosomes
(2) Apply the ICM algorithm for each chromosome on each image and then compute the fitness function for each chromosome
(3) If the fitness value for all chromosomes do not change from one population to another population, then stop and select the chromosome, which gives maximum fitness value (If there are two chromosomes give the same fitness value we select the chromosomes which represent lower order system). Otherwise go to step 2.

We run the previous algorithm on the six textures had shown in figure 4. Table 4 shows the original parameters and estimated parameters for each image. Figure 9 shows the original image and the regenerated image using the estimated parameters. From these results we note that the estimated parameters are very close to original parameters, also the regenerated images close to the original images.

5 Conclusion

In this paper we introduced two novel approaches to estimate the clique potentials in Gibbs-Markov image models. First approach is used to estimate the clique potentials for discrete Gibbs Markov random field (GMRF) by using genetic algorithms (GAs). Second approach is used to estimate the parameters of Gaussian Markov random field. The outline steps of the second algorithm are as follow. Given an image formed of a number of classes, an initial class density is assumed and the parameters of the densities are estimated using the EM approach. Convergence to the true distribution is tested using the Levy distance.

References

[1] J. Besag, "Spatial interaction and the statistical analysis of lattice system," J. Royal Statistical Society, Ser. B, vol. 36, pp. 192-236, 1974

[2] J. Besag, " Efficiency of pseudolikelihood estimation for simple Guassian fields, "Biometrika, vol. 64, pp. 616-618, 1977.

[3] R. Kashyap and R. Chellappa, "Estimation and choice of neighbors in spatial interaction models of images," IEEE Trans. Information Theory, vol. 29, pp. 60-72, 1983.

[4] S. Lakshmanan and H. Derin, "Simultaneous parameters estimation and segmentation of Gibbs random field using simulated annealing," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 11, pp. 799-813, 1989.

[5] Sateesha G. Nadabar and Anil K. Jain, " Parameter Estimation in Markov Random Field Contextual Models Using Geometric Models of objects," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 18, pp. 326-329, 1996.

[6] Haluk Derin, Howard Elliott, "Modeling and Segmentation of Noisy and Textured Images Using Gibbs Random Fields", IEEE Tran. On pattern Analysis and Machine intelligence, vol. Pami-9, No. 1, Januarry 1987.

[7] J. Besag, "On the statistical analysis of dirty pictures", Journal of the Royal Statistical Society, London, Series B, Vol. B-48, pp. 259-302, 1986.

[8] E. Ising, Zetischrift Physik, vol. 31, p. 253, 1925.

[9] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images," IEEE trans. Pattern Analysis and Machine intelligence, vol. 6, pp. 721-741, 1984.

[10] Richard C. Dubes, and Anil K. Jain, " Random field models in image analysis," Journal of applied statistics, vol. 16, No. 2, pp. 131-164, 1989

[11] A. A. Farag and E. J. Delp, ``Image Segmentation Based on Composite Random Field Models,'' *Journal of Optical Engineering,* Vol. 12, pp. 2594-2607, December 1992.

[12] D. E. GoldBerg, "Genetic Algorithms in Search, Optimization and Machine Learning ," Addison-Wesley Publishing Company, INC, 1989.

[13] Lei Wang, Jun Liu, Stan Z. Li, "MRF parameter estimation by MCMC method", Pattern Recognition,33, PP. 1919-1925, 2000

[14] C. A. Bouman and B. Liu, "Multiple resolution segmentation of textured images", IEEE, Trans. on Pattern Analysis and Machine Intelligence, 13:99-113, Feb. 1991.

[15] Ayman El-Baz, Aly A. Farag," Image segmentation using GMRF models: parameters estimation and applications", International Conference on Image Processing, Barcelona, Spain Sept. 2003

[16] R. O. Duda, P. E. Hart and D. Stork., *Pattern Classification,* John Wiley & Sons, 2nd edition, 2001.

Table 4: Original and estimated parameters.

NO.	Original Parameters	Estimated Parameters
1	[1 0.3 0.3 0.3 0.3 0 0 0 0 0]	[0.9 0.25 0.27 0.28 0.21 0.01 0.07 0 0 0]
2	[1 1 1 -1 1 0 0 0 0 0]	[0.86 0.96 -0.91 0.9 0.8 0 0 0 0 0]
3	[1 2 2 -1 -1 0 0 0 0 0]	[1 1.97 1.96 -0.93 -0.92 0 0 0 0 0]
4	[1 1 1 -1 1 -1 1 -1 1 1]]	[0.9 0.8 0.9 0.8 -0.8 0.8 0.88 1 0.81 0.99]
5	[1 1 1 -1 1 -1 1 -1 1 1]	[0.91 0.9 0.8 -0.92 -0.98 0.8 -0.9 0.99 1]
6	[1 -1 1 1 1 1 1 1 1 -1]	[1 -0.9 0.8 0.82 0.79 0.9 0.8 0.9 0.9 -0.9]
7	[1 1 -1 1 1 1 1 1 1 -1]	[0.9 0.8 -0.9 0.9 0.8 0.8 0.9 0.9 0.9 -0.9]
8	[1 1 1 1 1 1 1 1 1 -1]	[1 0.8 0.88 0.98 0.99 0.96 0.95 1 1 -0.91]

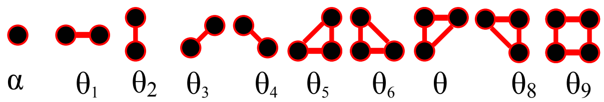

Figure 1: Clique structure for the 2[nd] order neighborhoods

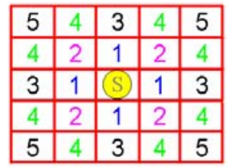

Figure 2: Neighborhood System for GMRF

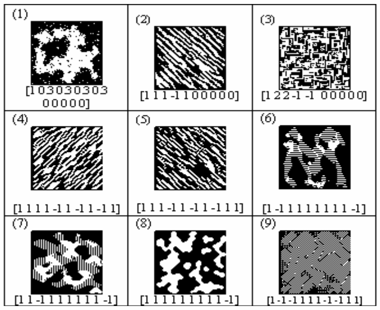

Figure 3: Examples of 64 X 64 realizations of Derin-Elliott model for various values of (α, θ_1, θ_2, θ_3, θ_4, θ_5, θ_6, θ_7, θ_8, θ_9).

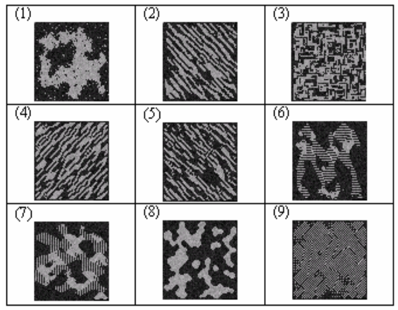

Figure 4: Examples of 64 X 64 realizations of Auto-normal random process for various values of (α, θ_1, θ_2, θ_3, θ_4, θ_5, θ_6, θ_7, θ_8, θ_9).

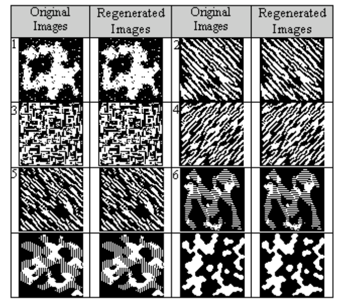

Figure 5: Original and synthesized images using the parameters in Table 1.

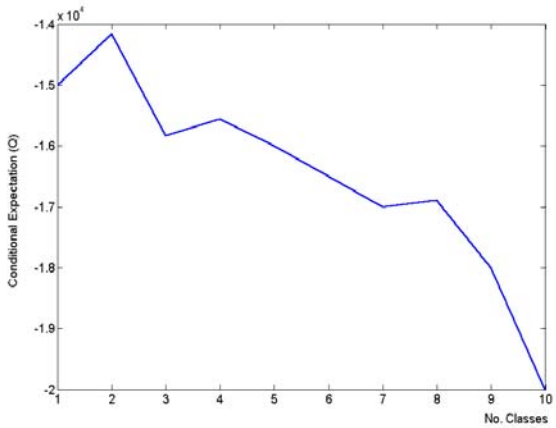

Figure 6 Evolution of Conditional Expectation with the number of mixture assumed in the scene for the absolute error shown in figure

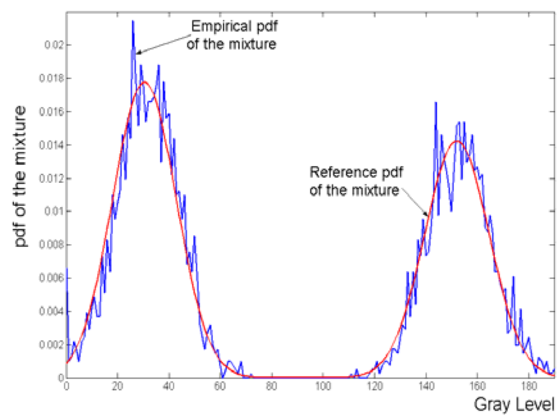

Figure 7 Empirical density function of the mixture $p_{emp}(g)$, and Reference density function of the mixture p_{ref}

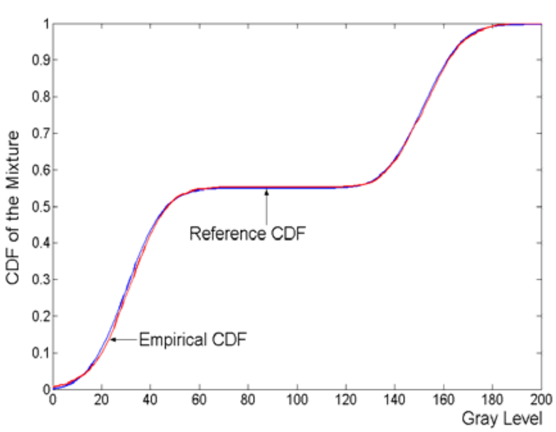

Figure 8 Empirical Distribution $P_{emp}(g)$, and Reference Distribution $P_{ref}(g)$.

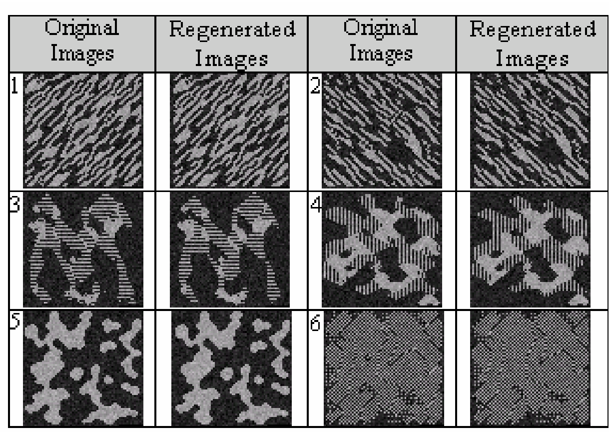

Figure 9 Original and synthesized images using the parameters in Table 5.