Chapter 2

Image Model and Parameter Estimation

2.1 INTRODUCTION

In general, most of the image processing and image analysis are performed under any one of the following techniques: (i) Transform based and (ii) model based.

The transform based techniques are widely used in image processing such as image coding and restoration. Various transform techniques and their efficiencies have been reported in the literature. Among the existing transform techniques, the most widely used are Discrete Fourier Transform (DFT), Discrete Cosine Transform (DCT), Discrete Sine Transform (DST), Karhunen-Loeve Transform (KLT), Walsh-Hadamard Transform, Harr Transform and Orthogonal Polynomial Transform. Each transform technique has its own specific features. In recent years, the Wavelets transform based Wavelets and Fractal image coding techniques becoming popular due to its efficiency. Most of the transform based techniques demand more computations and some of them require large amounts of memory. Also, applications of the transform based techniques are limited at low-level image processing and analysis when compared to the model based techniques.
The model based techniques are most appropriate to effectively handle the problems involved with large amount of data like image filtering, object recognition, etc. To handle this volume of data, it would be preferable to have an underlying model that explains the dominant statistical characteristics of the given data. The different classes of models that have been suggested in the literature, exploit the statistical properties among the neighboring pixels for low-level image processing and analysis. The statistical models attracted many researchers, due to its wide range of applications at low-level image processing such as texture analysis, image smoothing, enhancement, restoration, segmentation, edge detection, image data compression.

The statistical models are becoming increasingly important because of their role in the development of useful algorithms for image processing and analysis. It is observed that most of the applications of image processing use some sort of statistical models. Generally, the models are not usually made explicit, but are made implicit by the adoption of assumptions that incorporate certain model assumptions within them. Most of the algorithms use the assumption that the image can be treated as a random process with wide sense of stationary properties, linear dependency, white uncorrelated noise, etc. In that sense, Markov Random Field (MRF) and Autoregressive (AR) models are most appropriate for almost all the low-level image processing. Many researchers have explored the efficiency of the MRF and AR models for low-level image processing such as image smoothing, object recognition, classification, segmentation, texture representation, texture synthesis, compression and reconstruction. Most of the images satisfy the properties of MRF and AR models. The pixels in a two-dimensional image are spatially
equal interval of distance in row wise and column wise as in time series (equal interval of time) and the pixels in the images satisfy the sampling properties and satisfy stationarity, linear dependency and white noise uncorrelation. Hence the MRF and AR models have drawn the attention of the many researchers in different fields of image processing and analysis. As discussed in Chapter 1, different types of stochastic models that are used for image processing and analysis include, Autoregressive (AR), Moving Average (MA), Autoregressive Moving Average (ARMA) and Autoregressive Integrated Moving Average (ARIMA) with various assumptions about the image. The assumptions, Random Field, Markov Random Field, Gibbs Field and $\sigma$-Field etc. are made on the basis of the nature of the images. A brief review of the related literature, under the different approaches, is given below.

Generally, the main advantage of the AR model over the other models is that it is regenerative, that is, it represents all the information in an $N \times N$ image by two sets of parameters, one set containing a minimum number of parameters having most of the information while another set containing $N^2$ parameters, the so-called residuals, having the remaining information. The residuals can be stored with minimum number of bits than the original image pixels without sacrificing any accuracy [Kash80]. With the use of stored parameters and the residual values, the original image can be reconstructed with good quality. The textured images can be generated with the use of stored parameters of the model only and without any compromise in the quality of the image.
Several authors have shown a considerable attention on MRF and AR models, due to its simplicity, i.e., less computational complexity, finite memory or memory loss and wide range of applications, especially in texture analysis, segmentation, inpainting, reconstruction and in data mining, which searches and retrieve the images from the volume of database that contains images.

In order to analyse the content of an image, many researchers have investigated various statistical models such as co-occurrence matrix method, MRF, AR models, Principal Component Analysis (PCA), etc. It is observed that the MRF and AR models play a significant role almost in all the low-level image processing. For instance, Wong [Wong68] analysed the random field model for image coding applications and Kashyap [Kash80] has used the same for image segmentation. Many authors [Kash80, Chel85, Boum93, Panj95, Benn98, Li00, Elia03] have discussed the image synthesis and analysis of colour and monochrome images with MRF model. Autoregressive model has been examined in many application areas, that includes image compression [Delp79], image segmentation [Lu95] and restoration [Kada98]. A common choice in image applications is a homogeneous Gibbs distribution that corresponds to a homogeneous Markov Random Field. This, however, is made with little comment, and model checking is rarely considered. This suggests an edge-site model for image reconstruction such as that proposed by [Gema84, Chal89]. Derin and Elliott [Deri87] have investigated Gibbs field models for the segmentation of noisy and textured images.

Moreover, the degree of accuracy of parameter estimation also plays a significant role in obtaining the satisfactory results in image
processing. The parameter estimation techniques are generally
grouped into three classes, viz., sampling theory (or classical
approach), Bayesian approach and iterative procedure. The classical
methods, depending on the degree of accuracy and starting with the
most precise method and continuing in order of precision, are the
following [(Prie81, Kend90]: Maximum Likelihood method, Least
Square method, appropriate least square method and Yule-Walker
method.

In earlier works, the classical approach such as Maximum
Likelihood Estimation (MLE) method was adopted to estimate the
parameters of the MRF model for various low-level image analysis
[Panj95, Balg98, Kada98] and for reconstruction [Shep82, Vard85].
The Least Square (LS) method was chosen in [Delp79, Benn98] for
image compression and texture synthesis. It is observed from the
literature that the Yule-Walker method [Harv84] was chosen in the
rare case in the context of image analysis. In the case of classical
approach, the parameter estimation is found to be unsatisfactory for
reconstructing the images after compression, and for segmentation
[Delp79, Aykr99] or it requires some post processing [Wils83, Chen95]
or it needs iterative procedures [Lu95, Come99, Clau00]. The main
drawback of this method is that it produces excessive noise in the
final reconstruction.

The prior knowledge about the parameters of the model to be
estimated is used in the Bayesian approach. This procedure, however,
usually assume that any prior parameters are known. First, Besag,
and Geman and Geman [Besa83, Gema84] introduced the concept, in
the reconstruction of images. Subsequently, many authors [Gema84,
Gree90, Higd97 Elia03] have used the Bayesian approach and they reported that it gives high precision and expected results.

Over the recent years, Markov Chain Monte Carlo (MCMC), Expectation Maximization (EM) approach, combined MCMC and EM approach and some other iterative procedures are used either alone or with the concept of neural network. The combined MCMC and EM approach was proposed by Geman and McClure [Gema87] and used by Chalmond [Chal89] for homogeneous prior models. Besag et al. [Besa95] used MCMC with Bayesian approach for image restoration. Comer et al. [Come99] used the modified version of the EM algorithm to estimate the parameters of the Multiresolution Gaussian Autoregressive model for textured image segmentation. Over the recent years, many authors have studied the Neural Network approach [Lu95, Clau00]. For example, Lu and Xu [Lu95] used for texture description and textured image segmentation based on an AR model with neural networks, and Clausen and Wechler [Clau00] used for colour image compression based on Principal Component Analysis (PCA) with neural networks. Most of the authors suggested that even though these methods give better results, they demand high computational effort for the estimation procedure when compared to classical approaches. This is the main drawback of these methods. Another important point that can be observed from the literature is that the emphasis is more on the parameter estimation and identification of the order of the model is not given any importance [Anja97]. In practice, in the context of image analysis, almost all the researchers have fixed the order of the models as two or maximum of three. There can be two reasons for not concentrating on the order determination: first one is the computational complexity arises as the
order increases and the second one, is the authors may have thought that the sufficient information can be captured with order three. This problem is avoided in the proposed method, since the proposed model has a feature, viz., infinite structure with a finite number of parameters and so completely avoids the problem of order determination of the model. The features of the proposed model are discussed in detail in the next section.

Kashyap and Chellappa [Kash83] proved that for a model, appropriate order should be fitted to obtain good results in low-level image processing, like, image reconstruction, classification, segmentation and image data compression. The order selection problem becomes more difficult due to the numerous varieties of model structures. Thus the quality of the processed image varies considerably depending on how similar the underlying model is to the true model, and so the use of the appropriate neighbor set is important. It is further reported that the selection of the model of the appropriate order based on visual inspection, that is, qualitative measure is unreliable. It is possible to choose an appropriate model on a quantitative basis without visual inspection, from a family of such models for any given image. But this task is more tedious.

It is observed from the literature that, several authors have proposed numerous algorithms for model selection. For instance, the machine learning algorithms for model selection was first proposed by [Arin94], then adopted by others [Chu94, Venk99, Prud02], by treating the model selection as a classification problem. The Box-Jenkins models [Box70], family of exponential smoothing models [Mont90] and many Artificial Neural Networks models [Dorf96] have
been proposed. Collopy and Armstrong [Coll92] have introduced the landmark Rule-Based Forecasting system with 99 rules to configure and combine widespread time series models. Recently, the suitable versions of the Model Symbolic classifier are employed for image processing [De01] and for information filtering [Beze02, De03, Prud04] and they produce the results that are successful both in terms of accuracy and execution time. Despite the diversity of models, empirical research has shown that there is no single model that performs better than the others in all types of images (time series) [Coll92]. Selecting the most adequate model for describing individual time series can strongly improve the performance of the image analysis. The main drawback of this model selection method is that it is time consuming and computationally expensive.

With the consideration of all the points aforesaid, a statistical model, called a new Family of Autoregressive models is proposed here. The model selection problem is overcome in our proposed method to some extent, since the proposed model is completely free from order determination problem. So the proposed model will automatically fix the order according to the nature of the images. The main advantage of the proposed model is that it works on both textured and untextured images, unlike the existing ones.

Generally, a model with statistical properties, which describes the probability structure of a time series and in general any sequence of observations, is called stochastic process. The image to be analysed may be thought of as one particular realization, produced by the underlying probability mechanism of the image under study. That is,
in analysing an image we regard it as a realization of a stochastic process.

**Definition:** If a stochastic process, that is, a family of time dependent random variables \( \{X(s)\} \) satisfies

\[
E(X(s)/X(s-j); j=1,2, \ldots) = E(X(s)/X(s-1), \ldots, X(s-p))
\]

then \( \{X(s)\} \) is said to satisfy the Markov property.

On the left hand side (LHS) the expectation is conditional on the infinite history of \( X(s) \). On the right hand side (RHS) it is conditional on only part of the history. From the definition, an AR(p) model is seen to satisfy the Markov property. In that sense, time series models, Markov Random models and Stochastic Process, all are interrelated and are more associated with each other in image processing. In the next section, the proposed, a new Family of Full Range Autoregressive (FRAR) models is introduced.

### 2.2 Proposed Models

Let \( X \) be a random variable that represents the intensity value of a pixel at location \((k,l)\) in an image of size \( N \times N \). We also assume that \( X \) may have noise and is considered as independently and identically distributed Gaussian random variable with discrete time space and continuous state space with mean zero and variance \( \sigma^2 \) and is denoted as \( \varepsilon(k,l) \) i.e. \( \varepsilon(k,l) \sim N(0, \sigma^2) \).

Since \( \{X(s); s \in S\} \) is a stochastic process, where \( S = \{s: (k,l); 1 \leq k, l \leq N\} \), \( \{X(s)\} \) can be considered as a Markov process because we have the conditional probability

\[
P\{X(s_n) = i_n \mid X(s_k) = i_k ; k=0,1,2, \ldots, n-1\}
\]
\[ P(X(s_n) = i_k \mid X(s_{n-1}) = i_{n-1}) \]

\[ \forall i_k, k = 0, 1, 2, \ldots, n-1 \text{ and } s_k \text{ belonging to the state space } S \text{ and } s_0 < s_1 < \ldots < s_n. \]

Here, the two-dimensional images \( \{X(k,l), 1 \leq k, l \leq N\} \) are modelled by means of discrete spatial time series of random fields [Kash80]. Thus, we propose a model as in equation (2.1), a family of models by a discrete-time stochastic process \( \{X(s)\}, t = 0, \pm 1, \pm 2, \pm 3, \ldots \), called the family of Full Range Autoregressive (FRAR) model by the difference equation

\[
X(k,l) = \sum_{r=-1}^{M} \sum_{q=-1}^{M} \frac{K \sin(r \theta) \cos(r \phi)}{\alpha^r} X(k+r, 1+q) + \varepsilon(k,l)
\]

\[ = \sum_{r=-1}^{M} \sum_{q=-1}^{M} \Gamma_{rq} X(k+r, 1+q) + \varepsilon(k,l) \quad \text{ ...(2.1)} \]

where \( \Gamma_{rq} = \Gamma_r = \frac{K \sin(r \theta) \cos(r \phi)}{\alpha^r}, \forall q \)

In the above equation (2.1), \( X(k+r, l+q) \) accounts for the spatial variation owing to image properties and \( \varepsilon(k,l) \) is the spatial variation owing to additive noise and \( \Gamma_{rq} = \Gamma_r = \frac{K \sin(r \theta) \cos(r \phi)}{\alpha^r}, \forall q \) is the \( r^{th} \) coefficient of variation among the low level primitives in the small image region. The coefficients are interrelated. The relationship is
established through the parameters $K$, $\alpha$, $\theta$, and $\phi$ which are real. The coefficients in (2.1) are functions of $K$, $\alpha$, $\theta$, and $\phi$ as well as $n$. That is,

$$\Gamma_r = \Gamma_r(K, \alpha, \theta, \phi) = \frac{K}{\alpha^r} \sin(r\theta) \cos(r\phi)$$

where $K \in \mathbb{R}$: set of real numbers; $\alpha > 1$; $\theta, \phi \in [0, 2\pi]$ and $n \in \{1, 2, \ldots\}$.

The proposed model is employed to analyse a two-dimensional discrete gray-scale image. The image is partitioned into various subimages of size $M \times M$, $(M<N)$, to locally characterise the nature of the image. With the Markovian assumption, the conditional probability of $X(s)$ given all other values only depends upon the nearest neighbourhood values, i.e., $P(X(s)/X(s-j)); j=1,2, \ldots, M$.

The initial assumptions about the parameters are $K \in \mathbb{R}$, $\alpha>1$, and $\theta, \phi\in[0, 2\pi]$. Further restriction on the range of the parameters is placed by examining the identifiability of the model. It is interesting to note that some of the models used in the previous works, such as, white noise, autoregressive finite order and infinite order autoregressive models can be regarded as special cases of the proposed model.

Thus

(i) if we set $\theta=0$, then the FRAR model reduces to the white noise process.

(ii) when $\alpha$ is large, the coefficients $\Gamma_r$s become negligible as $r$ increases. So the FRAR model will be reduced to AR($p$) model approximately, for a suitable value of $p$. 
(iii) when $\alpha$ is chosen less than one, then the FRAR model becomes an explosive infinite order AR model.

The fact that $X(s)$ has regression on its neighbourhood pixels gives rise to the terminology of autoregressive process. However, in this case, the dependence of $X(s)$ on neighbourhood values may be true to some extent. In fact, the process is Gaussian under the assumption that the $\varepsilon(k,l)$s are Gaussian and in this case its probabilistic structure is completely determined by its second order properties. Finally the range of the parameters of the model are set as with the constraints $K \in \mathbb{R}$, $\alpha > 1$, $0 < \theta < \pi$, $0 < \phi < \pi/2$ and $\varepsilon(k,l)$ are independent and identically distributed Gaussian random variables with mean zero and variance $\sigma^2$.

The noncausal model given in equation (2.1) represents the pixel $X(s)$ as a linear combination of nearest neighbourhood values on each side as shown in Figure 1.1(a) and the influence of the horizontal and vertical pixels on the centre is higher than the diagonal pixels. The order of influence of neighbourhood pixels on its centre is shown in Figure 1.1(b). The pixel $X(s)$ is predicted by using the closest neighbourhood pixels on each side.

![Figure 2.1: (a) Relationship among neighbourhood pixels;](image-url)
(b) Order of influence of neighbourhood pixel on its centre.

The model considered here is very important from theoretical point of view since it has infinite structure with a finite number of parameters and hence completely avoids the problem of order determination. It can solve systematically many important problems, such as, testing whether an image is homogeneous or not with the choice of the relevant set of neighbours. It is also proved in this thesis that it gives fruitful results for many practical problems such as texture representation, classification, edge detection and image compression.

2.3 Parameter Estimation

In order to implement the proposed FRAR model, we must estimate the parameters. The parameters $K$, $\alpha$, $\theta$, and $\phi$ are estimated, by taking the suitable prior information for the hyper parameters $\beta$, $\nu$, and $\delta$, based on numerical integration technique and Bayesian methodology and the posterior density of the parameters of the model is obtained. The FRAR model, discussed in the previous section, is defined by a discrete time stochastic process $\{X(t)\}$ as

$$X(k, l) = \sum_{r=-1}^{M} \sum_{q=-1}^{N} \Gamma_{rq} X(k + r, l + q) + \varepsilon(k, l) \quad \text{...}(2.2)$$

where $\Gamma_r = \frac{K \sin(r\theta) \cos(r\phi)}{\alpha}$.

$K$, $\alpha$, $\theta$, and $\phi$ are parameters, $\varepsilon(k, l)$s are independent and identically distributed normal random variables with mean zero and
variance \( \sigma^2 > 0 \). It is assumed that the domain of the parameters is given by \( K \in \mathbb{R}, \alpha \in (1, \infty), \theta \in (0, \pi) \) and \( \phi \in (0, \pi/2) \), which ensures that the model is identifiable.

Suppose that there are \( N \) observations \( X_1, X_2, \ldots, X_N \). The problem is to estimate the unknown parameters \( K, \alpha, \theta, \phi \) and \( \sigma^2 \), using the Bayesian methodology.

Only for the computational purpose, the pixel values of each subimage are arranged as one-dimensional vectors \( X(t), t = 1,2,3, \ldots, N \) \((M \times M=M^2 = N)\). Since the error term \( \varepsilon(k, l) \) in (2.1) are independent and identically distributed Gaussian random variable, the joint probability density function of the stochastic process \( \{X(t)\} \) is given by

\[
p(X/H) \propto \left(\sigma^2\right)^{-N/2} \exp\left[-\frac{1}{2\sigma^2} \sum_{t=1}^{N} \left( X_t - K \sum_{r=1}^{\infty} S_r X_{t-r} \right)^2 \right] \quad \ldots(2.3)
\]

where \( X = (X_1, X_2, \ldots, X_N) \)

\[
H = (K, \alpha, \theta, \phi, \sigma^2) \text{ and}
\]

\[
S_r = \frac{\sin(r\theta)\cos(r\phi)}{\alpha^r}.
\]

When we analyze the real data with finite number of \( N \) observations, the range for the index \( r \) viz., \( 1 \) to \( \infty \), reduces to \( 1 \) to \( N \) and so in the joint probability density function of the observations
given by (2.3) the summation \( \sum_{r=1}^{\infty} \) can be replaced by \( \sum_{r=1}^{N} \) which gives

\[
P(X/\Theta) \propto (\sigma^2)^{-N/2} \exp\left[ -\frac{1}{2\sigma^2} \sum_{t=1}^{N} \left\{ X_t - K \sum_{r=1}^{N} S_r X_{t-r} \right\}^2 \right] \quad \ldots (2.4)
\]

By expanding the square in the exponent, we get

\[
P(X/\Theta) \propto (\sigma^2)^{-N/2} \exp\left[ -\frac{1}{2\sigma^2} \left\{ T_{00} + K^2 \sum_{r=1}^{N} S_r^2 T_{rr} + 2K^2 \sum_{r,s=1, r < s}^{N} S_r S_s T_{rs} - 2K \sum_{r=1}^{N} S_r T_{0r} \right\} \right]
\]

where \( T_{rs} = \sum_{t=1}^{N} X_{t-r} X_{t-s}, \quad r, s = 0, 1, 2, \ldots, N \)

The above joint probability density function can be written as

\[
P(X/\Theta) \propto (\sigma^2)^{-N/2} \exp\left[ -\frac{Q}{2\sigma^2} \right] \quad \ldots (2.5)
\]

where \( Q = T_{00} + K^2 \sum_{r=1}^{N} S_r^2 T_{rr} + 2K^2 \sum_{r,s=1, r < s}^{N} S_r S_s T_{rs} - 2K \sum_{r=1}^{N} S_r T_{0r} \)

\[ K \in \mathbb{R}, \alpha > 1, \quad 0 < \theta < \pi, \quad 0 < \phi < \pi/2 \quad \text{and} \quad \sigma^2 > 0. \]
The prior distribution for the parameters is assigned as follows:

1. \( \alpha \) is distributed as the displaced exponential distribution with parameter \( \beta \), i.e.

\[
P(\alpha) = \beta \exp(-\beta(\alpha - 1)) ; \quad \alpha > 1 ; \quad \beta > 0
\]

2. \( \sigma^2 \) has the inverted gamma distribution with parameter \( \nu \) and \( \delta \), i.e.

\[
P(\sigma^2) \propto \exp(-\nu / \sigma^2) (\sigma^2)^{-\left(\delta + 1\right)} ; \quad \sigma^2 > 0, \quad \nu, \delta > 0
\]

3. \( K, \theta \) and \( \phi \) are uniformly distributed over their domain, i.e.

\[
P(K, \theta, \phi) = C, \text{ a constant}
\]

\[K \in \mathbb{R}, \quad 0 < \theta < \pi, \quad 0 < \phi < \pi/2\]

So, the joint prior density function of \( \Theta \) is given by

\[
P(\Theta) \propto \beta \exp(-\beta(\alpha - 1) - \nu / \sigma^2) (\sigma^2)^{-\left(\delta + 1\right)} ; \quad \sigma^2 > 0, \quad \alpha > 1, \quad 0 < \theta < \pi, \quad 0 < \phi < \pi/2.
\] ... (2.6)

where \( P \) is used as a general notation for the probability density function of the random variables given within the parentheses following \( P \).

Using (2.5), (2.6) and Bayes theorem, the joint posterior density of \( K, \alpha, \theta, \phi, \) and \( \sigma^2 \) is obtained as

\[
P\left(\frac{X}{\Theta}\right) \propto \sigma^2^{-\frac{N}{2}} \exp\left(-\frac{Q}{2\sigma^2}\right) \exp\left(-\beta(\alpha - 1) - \gamma / \sigma^2\right) (\sigma^2)^{-\left(\delta + 1\right)} ;
\]
\[ \propto \exp(-\beta(\alpha - 1)) \exp(-1/2\sigma^2) (Q + 2\nu)^{\left(\frac{N}{2} + \delta + 1\right)} \] 

\[ \text{K} \in \mathbb{R}, \; \alpha > 0, \; 0 < \theta < \pi, \; 0 < \phi < \frac{\pi}{2} \; \text{and} \; \sigma^2 > 0. \]

Integrating (2.7) with respect to \(\sigma^2\), the posterior density of \(K, \alpha, \theta\) and \(\phi\) is obtained as

\[ P(K, \alpha, \theta, \phi/X) \propto \exp(-\beta(\alpha - 1))(Q + 2\nu)^{\left(\frac{N}{2} + \delta\right)} \] 

where

\[ [Q + 2\nu] = \begin{bmatrix}
K^2 \sum_{r=1}^{N} S_r^2 T_{rr} + 2K^2 \sum_{r<s}^{N} S_r S_s T_{rs} \\
-2K \sum_{r=1}^{N} S_r T_{or} \\
+ T_{oo} + 2\nu
\end{bmatrix} \]

That is,

\[(Q+2\nu) = aK^2 - 2Kb + T_{oo} + 2\nu, \]

\[= C \left[ 1 + a \left( K - b_{1} \right)^2 \right] \]

where
\[
C = T_{00} - \frac{b^2}{a} + 2\nu
\]
\[
a = \sum_{r = 1}^{N} S^2_{rr} + 2 \sum_{r, s = 1}^{N} S_{rs} T_{rs}
\]
\[
b = \sum_{r = 1}^{N} S_r T_{0r}
\]
\[
a_1 = \frac{a}{C}, \quad b_1 = \frac{b}{C}
\]

Thus, the above joint posterior density of \(K, \alpha, \theta, \phi\) can be rewritten as

\[
P(K, \alpha, \theta, \phi / X) \propto \exp(-\beta(\alpha - 1)) \left[ \frac{1}{C} \left(1 + a_1 (K - b_1)^2\right)\right]^{d - 1} \]

\[\text{where} \quad d = \frac{N}{2} + \delta\]

This shows that, given \(\alpha, \theta, \phi\) the conditional distribution of \(K\) is ‘t’ distribution located at \(b_1\) with \((2d - 1)\) degrees of freedom.

The proper Bayesian inference on \(K, \alpha, \theta, \phi\) can be obtained from their respective posterior densities. The joint posterior density of \(\alpha, \theta, \phi\), namely, \(P(\alpha, \theta, \phi / X)\), can be obtained by integrating (2.9) with respect to \(K\). Thus, the joint posterior density of \(\alpha, \theta, \phi\) is obtained as

\[
P(\alpha, \theta, \phi / X) \propto \exp(-\beta(\alpha - 1)) C^{-d} a_1^{-\frac{1}{2}} \quad ...(2.10)
\]
\[ \alpha > 1, \ O < \theta < \pi, \ O < \phi < \frac{\pi}{2} \]

The above posterior density of \( \alpha, \theta \) and \( \phi \) in (2.10) is a complicated function and is analytically not solvable. Therefore, we can find the original posterior density of \( \alpha, \theta \) and \( \phi \) numerically from the joint density (2.10).

That is,

\[ P(\alpha) \propto \iint P(\alpha, \theta, \phi, /X) \, d\theta \, d\phi \]

Similarly,

\[ P(\theta) \propto \iint P(\alpha, \theta, \phi, /X) \, d\alpha \, d\phi \quad \text{and} \]

\[ P(\phi) \propto \iint P(\alpha, \theta, \phi, /X) \, d\alpha \, d\theta \]

The point estimates of the parameters \( \alpha, \theta \) and \( \phi \) may be taken as the means of the respective marginal posterior distribution i.e. posterior means. With a view to minimize the computations, we first obtain the posterior mean of \( \alpha \) numerically. Then fix \( \alpha \) at its posterior mean and evaluate the conditional means of \( \theta \) and \( \phi \) fixing \( \alpha \) at its mean. We fix \( \alpha, \theta \) and \( \phi \) at their posterior means respectively and then evaluate the conditional mean of \( K \).

Thus, the estimates are
\[ \hat{\alpha} = E(\alpha) \]
\[ (\hat{\theta}, \hat{\phi}) = E(\theta, \phi/\alpha = \hat{\alpha}) \text{ and} \]
\[ \hat{K} = E(K/\hat{\alpha}, \theta = \hat{\theta}, \phi = \hat{\phi}). \]

The estimated parameters \( \hat{K}, \hat{\alpha}, \hat{\theta} \) and \( \hat{\phi} \) are adopted to compute the coefficients \( \Gamma_i \)'s of the model in equation (2.1).

The advantage of the present approach is that the number of parameters are fixed as only four and the order does not increase the computational complexity, since the estimation of these parameters is the same irrespective of the order of the model, and hence it increases efficiency of the model. The usage of the proposed model with parameter estimation is highlighted in texture analysis, edge detection and image compression in the subsequent chapters.