National Institute of Electricity and Electronics

INELEC-BOUMERDES

DEPARTEMENT OF RESEARCH

THESIS

Presented in partial fulfilment of the requirement of the

DEGREE OF MAGISTER

in Electronic System Engineering

by

Djamel-Eddine FADLI

Nonparametric Approach To Digital Image Restoration

Defended on: 1995 before the jury:

President: Dr. A. MAAFI

Members: Dr. M. DJEDDI

Dr. M. BENTARZI

Dr. K. HARICHE

Mr. A. BOUKLACHI

Registration Number 02/95.

DEDICATION

I dedicate this work specially to my mother, my father and my sister Madjda for their valuable help and moral support without their motivations I would have never achieved such work.

I dedicate this work also to my brother Adib and my little sister Assma who have encouraged me continuously to hold on.

I dedicate this work to my supervisor Dr. M. DJEDDI for his intelligent help and technical support without his advises I would never finish this work.

I dedicate this work also to all my friends and particularly to Boukhalfa Ben-Abdallah.

I dedicate this work to all the technical staff of INELEC.

ACKNOWLEDGEMENTS

I like to express my most sincere gratitude to my technical advisor Dr. M. DJEDDI for his continued research guidance that without his advises I would not have finished this research work.

I am specially grateful to Dr. K. HARICHE, head of the research department, for providing research facilities to make our project a success.

My special thanks to Mrs. BERRI and Miss. BOUZIDI.

ABSTRACT

Image Analysis is concerned with restoration and interpretation of images that have been contaminated by noise and possibly some (nonlinear transformation). The purpose of our research work is to implement the statistical nonparametric methods to the restoration of noisy images which are characterized by the range of possible values for the intensity function $\vartheta(s)$ evaluated at different regular points of a lattice plane D. The nonparametric approach is more easier to be mastered by Engineers than the parametric one, and it does not require a priori strong assumptions concerning the unknown model of the image.

The first contribution consists of regarding the restoration of a given degraded image as the estimation of the intensity function from the observed degraded colors. This task was accomplished by the use of the Kernel method of regression. It was noticed that the implemented method does not perform satisfactorily in the regions where the intensity function fluctuates rapidly (sharp changes in the coloring). This drawback can be reduced to a certain extent by using the variable Kernel method.

In order to overcome the drawback of the curve fitting approach, a stochastic nonparametric approach was developed. In this approach, the value of the intensity function at each pixel was regarded as an observation generated from a random variable which is locally dependent on the neighboring random variables. Therefore, the restoration of images reduces to the estimation of an unknown stochastic process using one of the suggested nonparametric methods. The first method considers the observed image as a realization of a spatial process, whereas the second one is based on the estimation of the density function of the true unknown color of a pixel using the neighboring principle.

The two developed methods have been found to be very efficient. Furthermore, their performances were found to be better than that of the parametric Weiner approach.

TABLE OF CONTENTS

CHAPTER 5

CHAPTER 1

GENERAL INTRODUCTION TO THE PROBLEM

Human beings are faced to assess and interpret quite large variety of data most of which are coming under the form of imaging information. Our visual system perceives more than 75% of such information. Imagery was a subject of research for many scientists and engineers. It has continued to grow significantly in parallel with the advances in other fields such as remote sensing by satellites, optical astronomy, radar, electron microscope imaging, photography, computer vision and graphics, ultrasound, magnetic resonance imaging, photon emission tomography, etc.

The last four decades have been characterized by a rapid parallel improvement of low cost digital processing hardware. The structure of the imagery area is quite large, owing to the fact that the above listed disciplines rely too much on image processing. Fig 1.1 depicts the different fields of digital image processing.

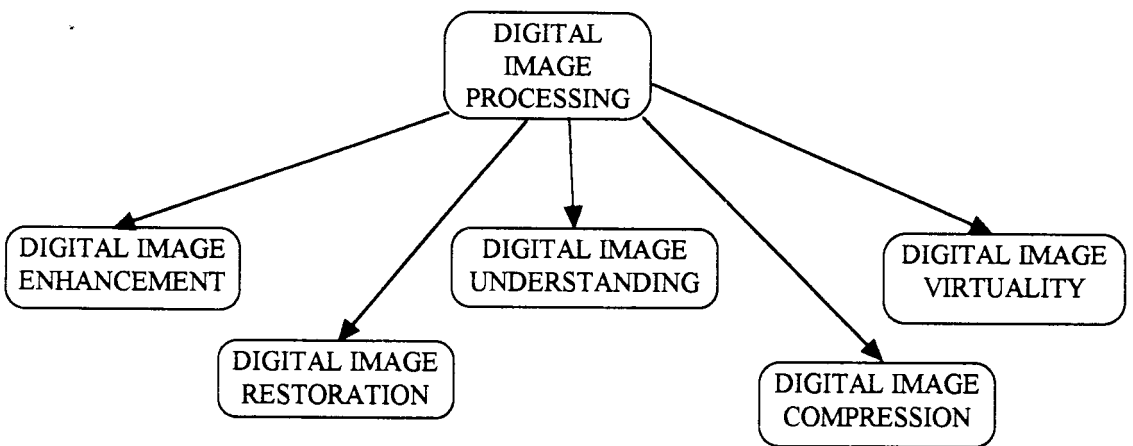


Figure 1.1-Showing the diffrent desciplines of digital image processing

Basically, any digital image is a two dimensional bounded set of regularly spaced sites called pixels. Each pixel possesses a certain intensity color corresponding to a specific sampled gray level. There are always situations where the sampled image will be distorted so that the gray level shade looses its original value which should be recovered in order that the image will be correctly interpreted. The restoration process is required, for example, when images are coming from a space shuttle and have been seriously degraded by a noise or a linear/non-linear degradation process, especially when we are dealing with important data sets.

Most of the sophisticated techniques of image restoration are computationally expensive due to their reliance on the model of the degrading process which should be mathematically formulated. Fig 1.2 shows a typical model based on a parametric restoration scheme which requires the knowledge of the model. This approach has the advantage of constraining the solution to belong to a specified family. Usually, once the parametric family of the model is specified, the restoration can be carried out and assessed through rigorous criteria functions.

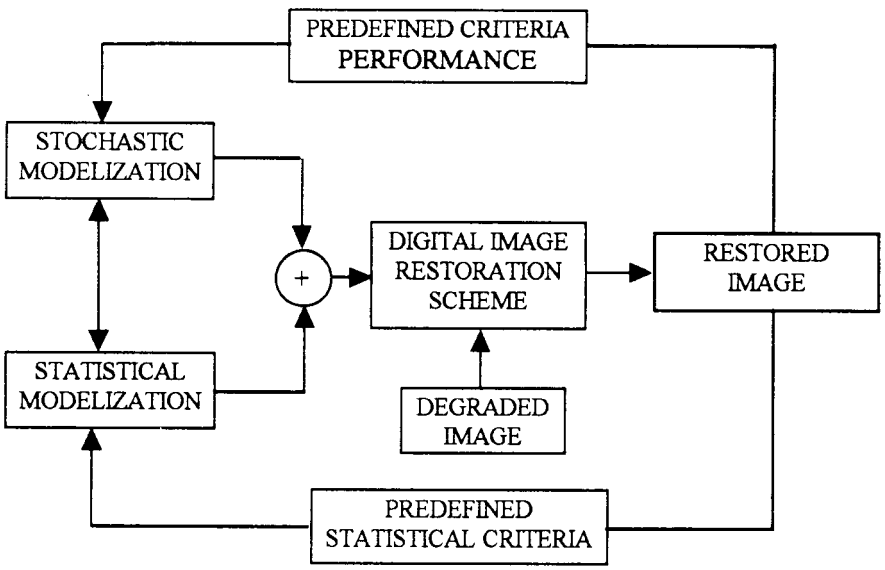


Figure 1.2 -Showing the model based scheme of restoration.

The solution is usually described by a distribution which has a mean that would be the optimal solution and a variance which represents the uncertainty of this solution. If the model is quite

reliable, then it can always be shown that the solution is optimal in the sense that it converges to this mean.

However, there are many cases where the choice of the parametric family is not justified or erroneous.

On the other hand, if the model is not reliable, neither this mean nor its variance could have an appealing sense because the parametric approach, in this case, will constrain the solution to belong to an invalid family of solutions.

Therefore, it is interesting to estimate the optimal solution using a wider family of models when the functional form of the members of the family can not be specified. These methods are called non model based or nonparametric methods. Fig 1.3 depicts a typical nonmodel based approach of restoration.

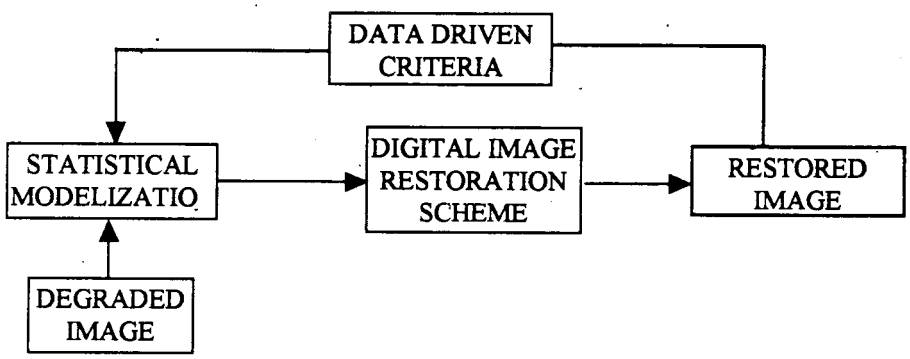


Figure 1.3-Showing the nonmodel based scheme of restoration.

In this approach the restoration is carried out using only the information contained in the observed data i.e. the noisy image. This is the reason for which nonparametric methods are usually called data driven restoration methods. A quick comparison between Figs 1.2 and 1.3 reveals that the data driven methods would be expected to be not computationally demanding as do their parametric counterparts. However, their use is not without price since these methods have bias problems and low convergence rates.

Our research work will be dealing with an eventual potential application of these data driven methods to the problem of noisy digital image restoration problem. Their statistical properties have been extensively studied by many statisticians. The nonparametric method to be used is the well known kernel method of regression and density estimation. The study will cover two approaches of restoration. One is termed as the regression or curve fitting approach. The other will be called the nonparametric stochastic approach to image restoration.

Basically, the regression approach consists of fitting a curve to a sample of observed data which would have been contaminated by a noise as depicted in Fig 1.4.

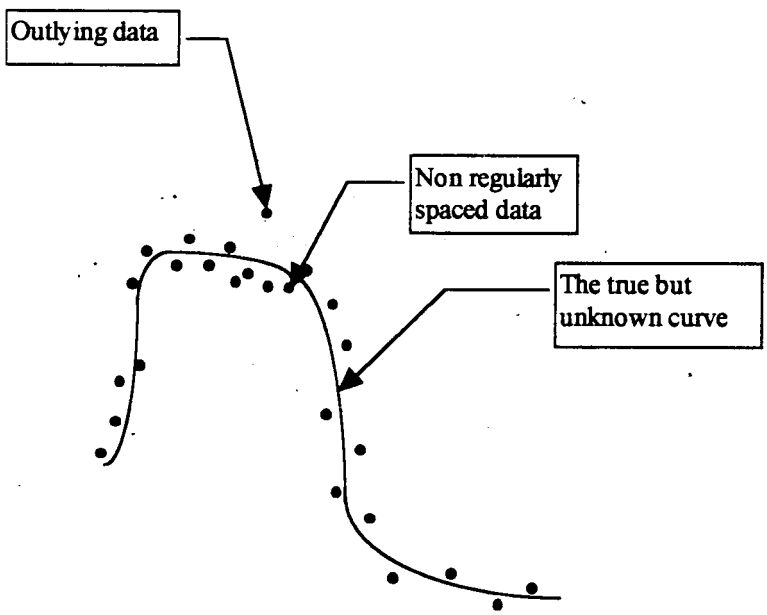


Figure 1.4-Regression aspect of estimation

The model based regression approach takes advantage of the fact that the functional form of the curve is known in advance apart from some unknown parameters to be estimated whereas in the nonparametric approach no such prior knowledge is required. Most of the nonparametric regression methods can be regarded as simple local weighted averaging techniques. Since this approach is data driven, it is sensitive to outlying observed data. Consequently, this approach of estimation behaves also as a variance reduction estimation, and the true peaks that may actually belong to the data could be smoothed out with the noise. In

this approach, the problem of the restoration of degraded noisy images can be formulated as the estimation of the unknown intensity function of the true image from the observed data.

The stochastic approach to image restoration relies on the statistical amplitudes that are existing near the point of estimation. In other words, the restoration is carried out by regarding the value of the intensity function at a given point (pixel) as a random variable whose distribution depends on those of the neighboring pixels. In this approach, the true value of the intensity function will be estimated by the mean of its distribution. Since the stochastic estimation methods are based on the computation of the likelihood of these distribution functions, the nonparametric approach requires that the distribution functions means be estimated only from the data without assuming a parametric family.

This thesis will be organized as follows:

In chapter 2, we will review the fundamentals of imagery systems, formulate the image restoration problem, and discuss the familiar parametric Weiner method of image restoration. In chapter 3, we will present nonparametric methods of regression estimation and will stress, in detail, the kernel method of regression estimation and its properties. In chapter 4, we will discuss the extension of the kernel method to the bivariate data which permits us to construct a simple method of restoration of noisy images. Chapter 5 deals with the stochastic approach. It will describe two methods which are based on the local conditional Markov models. In the first method, we will propose two ways of modeling the data by a local spatial interaction via the unilateral conditional or bilateral conditional specification of the image lattice structure, and propose a simple nonparametric conditional mode smoother. The second method will be concerned with an iterative approach based on the estimation of the conditional mode by combining the information contained in the observed records and the local dominant intensity color. Finally we will resume our work by concluding remarks.

CHAPTER 2

IMAGERY PROBLEM FORMULATION

2.1-INTRODUCTION

In the introduction we have stressed on the impact impressed by imagery systems in the information science and technology. Some times, it is required to map any other information into an image information for the seek of more interpretability. Engineers and scientists are so often faced with the interpretation and the processing of many images of different complexities. One field which has grasped much of the research interest is the restoration of badly recorded images in a noisy environment. This consists of building cheap methods that recover the original image by removing all kinds of distortion that may have corrupted the image. Statistical methods are regarded as a familiar approach to image restoration problems due to their ease of implementation and their flexibility of use. Model specification of all possible degradations is an essential step in image restoration.

However, if a complete formulation which takes into account most of the susceptible degradation processes is used, problems of computational environment arise. Many models have been proposed for handling the restoration of degraded images, but their usefulness are being conditioned by the induced computational costs. Thus, most of the restoration proposals fail as a limit of the currently available computational hardware technology although much of the algorithmic implementation is being relegated now by parallel processing. Therefore, one has to make a compromise between which model to choose among the ones which approximate the degradation process, and the computational facilities that are allowed through the available computing hardware.

Images can be modeled by continuous or discrete models. Although, the signal at its acquisition level is analog (continuous), the practical recording, storage and further processing make the discrete model formulation more appealing. This formulation is expected to become more dominant due to the extensive drifts in the digital processing hardware and the developments of fast discrete transforms and convolution algorithms. Consequently, the

matrix or vector representation of images seem to be more adequate. Two major raisons have been identified for such representation. The first reason is the availability of sophisticated linear algebra tools which can be of great usefulness in analyzing the eventual problems that are related to matrix operations. The second raison is justified by the target hardware on which the numerical computations are carried out. Subsequently, the digital hardware is mandatory for implementing the algorithms that would result from an image restoration scheme. Therefore, any model that would be suggested for a given restoration scheme should be conditioned by the motivation of these reasons.

In this chapter, we will review some of the familiar methodologies used for formulating the image restoration problem. The formulation is related to the degradation process and the type of the image to be restored. For general purposes, we will be interested mainly in colored images which are defined as bounded sets of sites called pixels having gray level shades or intensities. In statistical terminology, the image is referred to as spatial data of sites having continuous intensities.

2.2-IMAGE FORMATION

An image is whatever can be interpreted visually via artificial or natural imaging systems. Unlike the human visual system which can sense only the visible spectrum light (green, blue, dark, light ...), there are artificial physical systems that can also sense the invisible light spectrum. The x-ray radiographic systems of physicists is one of them. The human vision relies on the electromagnetic radiation energy to produce an impression of an image. On the other hand, artificial imaging systems are described by a relatively small number of physical concepts and an associated set of mathematical equations.

The basic concept of a typical image recording system is depicted in Fig 2.1 where it is assumed that an object is radiating an electromagnetic energy $f(\xi,\eta)$ through the coordinate object (ξ,η).

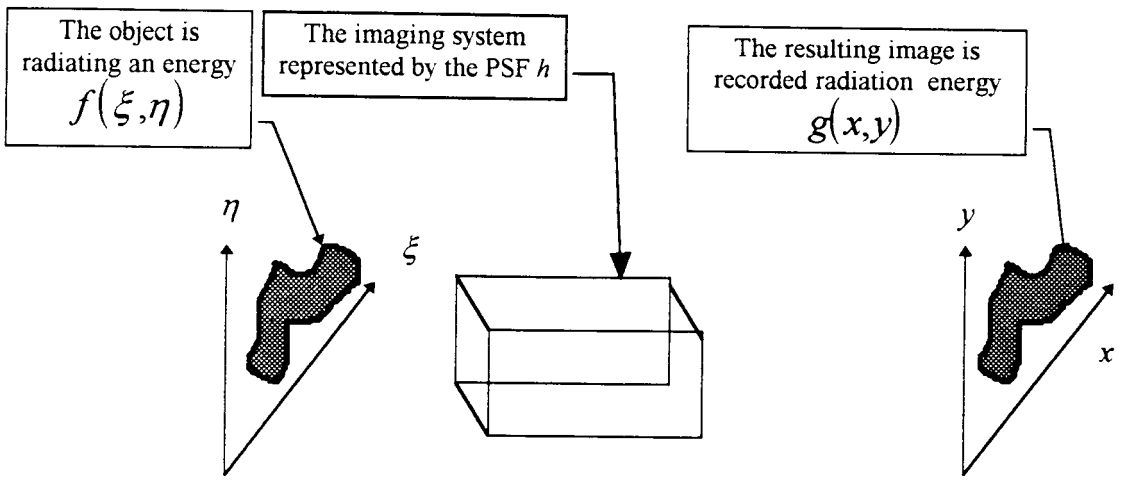


Figure 2.1-Typical imaging system

The radiated energy by the object image $f(\xi,\eta)$ is sensed by the imaging system which in turn produces a recorded image $g(x,y)$ through a recording coordinate media (x,y). The system is said to perform a recording of what is called commonly an image $g(x,y)$. Thus, images are representations of an object or a collection of objects that are indirectly sensed, and the various forms of radiant energy transport are known as the mechanisms by which the sensing is carried out. In this regard, there are three physical principles upon which image formation can be achieved:

(I) The Neighboring principle: It states that any recorded point on the image plane is the action of many neighboring points from the object plane.

(2) The Nonegativity principle: It states that the smallest energy quantum which can contribute for recording a point on the image plane is zero. That is:

$$f(\xi,\eta) \geq 0$$

and,

$$g(x,y) \geq 0$$

(3) The Superposition principle: It states that any imagery system is described by a transfer function.

If h denotes the transfer function of the sensing system, this principle states that:

$$g(x,y) = \iint h(\xi,\eta,x,y,f(\xi,\eta))d\xi d\eta \qquad (2.1)$$

Furthermore, if $h(\xi,\eta,x,y,f(\xi,\eta)) = h(\xi,\eta,x,y)f(\xi,\eta)$, the system is said to be linear.

In this case, using the neighboring principle, the image is then being formed according to:

$$g(x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(\xi,\eta,x,y) f(\xi,\eta) d\xi d\eta \tag{2.2}$$

The transfer function h is commonly called the Point Spread Function (PSF). It is called a Space Invariant PSF and denoted by SIPSF if:

$$h(\xi,\eta,x,y) \equiv h(\xi - x, \eta - y)$$

This equation is a parallel form of the impulse response function of a linear time invariant system. In this case, (2.2) becomes the convolution of the PSF and the object image:

$$g(x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(\xi - x, \eta - y) f(\xi,\eta) d\xi d\eta \tag{2.3}$$

Furthermore, when the PSF is separable i.e. $h(\xi - x, \eta - y) = h_1(\xi - x) h_2(\eta - y)$, one can integrate (2.3) sequentially into two scans to reduce computational complexity for an eventual deconvolution. The above equations are essential in the modeling concepts of radiant energy for image formation.

More details of image formation systems which convey the above equations can be found in the literature such as in Goodman (1968) and Van Bladel (1964).

The sensing systems are divided into two major technologies namely the photoelectronic and the photochemical systems which are exemplified by the CCD camera and the photographic films respectively.

The photographic film is characterized by an inherent randomness of the grain deposition during the fabrication process. The randomness of the deposition is the raison for noise generation signal. The grain noise is a signal dependent noise, constituting the most troublesome kind of noise, Mees (1954). Hung (1966) proposed the following approximation to characterize this noise:

$$\begin{aligned} d_r(x,y) &= d(x,y) + \alpha \mathrm{E}(d)^{\beta} n(x,y) \\ &= d(x,y) + n_1(x,y) \end{aligned} \tag{2.4}$$

where $d(x,y)$ is the local free noise and $\mathrm{E}(d)$ is the global mean density of the whole image. The parameters α and β are chosen in order that n_1 be a random Gaussian noise with zero mean and standard deviation $\sigma = \alpha \mathrm{E}(d)^{\beta}$. This approximation has the effect of bringing the noise dependent signal amplitude statistics to satisfy additive properties.

The noise generated in the photoelectronic imagery is not simple. Two processes are likely to contribute to the generation of the noise during the sensing and the recording mechanisms.

One comes under the form of random fluctuations in the number of photons and photoelectrons. The other one is due to the randomness which occurs in the electronic circuitry. This noise is always signal dependent as opposed to the film grain noise.

If $e(x,y)$ denotes the local mean electron emission rate at position (x,y), and $n(x,y)$ is a zero mean Gaussian noise with unity standard deviation, then Andrews and Hunt (1977) have expressed the noisy photoelectron image $e_r(x,y)$ as:

$$e_r(x,y) = e(x,y) + e(x,y)^{1/2} n(x,y) \tag{2.5}$$

The photoelectronic technology turned out to be suitable for the digital recording and processing despite the fact that the photochemical technology has the advantage of incorporating simultaneously the sensing and the recording mechanisms, as opposed to the CCD camera which includes a serial process of sensing and recording.

The above discussion and the associated equation settings are depicted in Fig 2.2 which shows the process of recording an image for the two mentioned technologies. It also includes the degradation of the system under the form of the PSF.

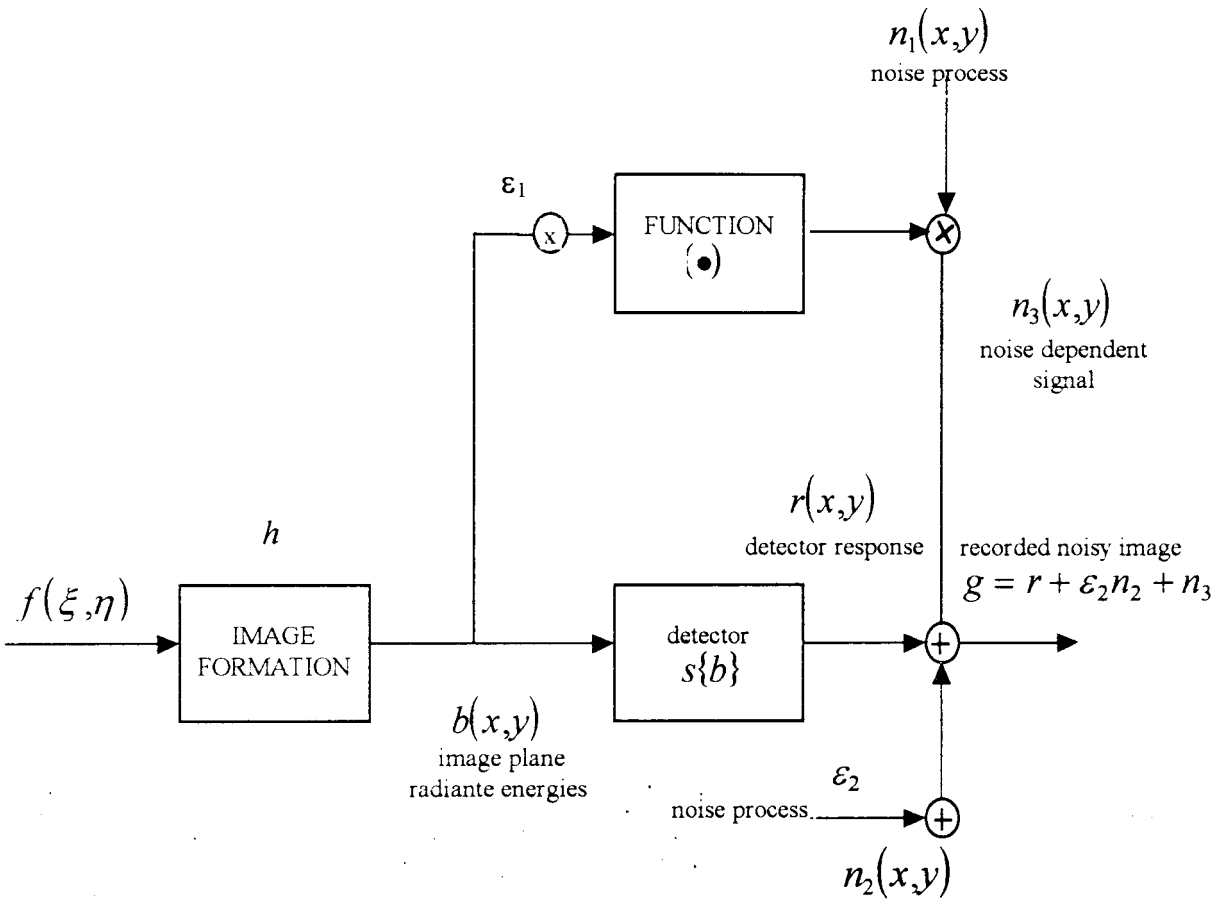


Figure 2.2-Complete model of image formation, sensing and recording

2.3-IMAGE RECORDING

It has been pointed out, in section 2.1 that image restoration is suitably carried out through discrete formulation. We need to discuss the mechanisms by which the digital image recording is achieved in order to understand model representation of restoration scheme. In as much as we are familiar with the sampling in time domain signals, the sampling of an image is carried according to a lattice structure. The digital image record is obtained by spatial sampling of the radiated electromagnetic energy of the sensed image *g(x,y)*. This can be described as:

$$g_s(i\Delta x, j\Delta y) = \sum_i \sum_j g(x,y)\delta(x - i\Delta x, y - j\Delta y) \quad \text{for} \quad i,j = 1,\ldots,N. \qquad 2.6$$

where g_s is the sampled image record at pixel $(i\Delta x, j\Delta y)$, and δ is the two dimensional impulse function or Dirac function. Δx and Δy are the lattice steps of which may be taken, for

simplicity, as unity. The sampled recorded values g_s for $i,j=1,...,N$ can be summarized in a matrix form $G=[G_{ij}]$, where $G_{ij} = g_s(i\Delta x, j\Delta y)$. In other words, one specific color or gray level intensity corresponds to each pixel.

We recall from sampling theory in time domain that for a possible signal reconstruction, the sampling speed should be at least twice the highest frequency content of the signal. Similarly, the sampling of the image should be at least spatially twice finer than the smallest detail of the image for a possible image reconstruction. More details about sampling techniques can be found in Hunt and Bree-Love (1974). However, there is a limit beyond which problems of storage and processing complexity become important. Thus, a trade off must be imposed between achieving high resolution image quality and the inherent complexities. For instance, an image of 1000x1000 pixels, each with 8 bits of bins (e.g. 2^8 gray levels) results in 1 Mb of computer memory storage. It is left, therefore, for the user to weight the computational complexity related to image processing

On the other hand, if the sampling lattice step is larger than some practical limits, then it may be regarded as a degradation source since insufficient sampling speed would result in the loss of fine information. A loss of information or imperfect reconstruction is thus expected for lower resolution images. Hence, the sampling can be considered as a part of the PSF imperfectness. Furthermore, the PSF may also include, for example, the superposition of a blurring effect as a result of a bad focus of the image recording and sensing system. As a consequence, the PSF is considered as the transfer function which sums up all the imperfectness of the imagery system.

As for linear time domain systems, the PSF of the image formation system can be mathematically represented through a simple set of equations. It is well known from linear system theory that this transfer function is the response of the system to the impulse function

(in our case, this will be the two dimensional Dirac function). Thus, the recorded image $g(x,y)$ can then be represented as the convolution of the Point Spread Function h imperfectness with the sampled recorded image:

$$G_{ij} \equiv g_s(x,y) = h(x,y) * g(x,y) \tag{2.7}$$

where $x=i\Delta x$ and $y=i\Delta y$.

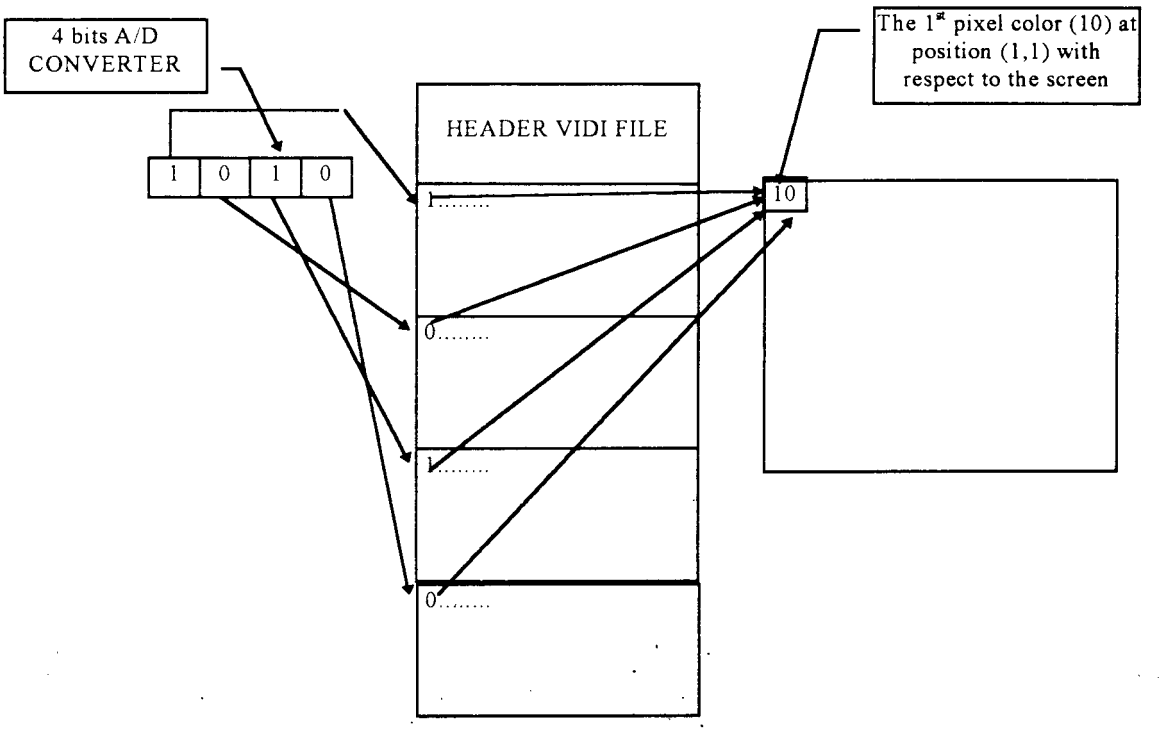
In this research work, we will not deal with the problem due to the imperfectness effects of the PSF but we will be interested in the development of statistical techniques to noise removal. However, the restoration that would be carried out to remedy for the imperfectness effects of the PSF on the recorded image, can be done separately either after or before removing the noise. Moreover, the above discussion suggests that the noise is characterized by signal dependent noise assumptions. However, there are two reasons for which such characterization should be simplified by signal independent noise assumption. The first being that there is no real mathematical gain on dealing with signal dependent noise. The second one being that the solution will not, significantly, have a better accuracy if it is based on the signal dependent noise model, Walkup (1974). Hence, the whole objective of this work is termed as the smoothing or the regularization of noisy images which have been corrupted by additive noise.

2.4-THE DIGITALIZATION OF REAL IMAGES

The above grabbed image in Fig 4.11 was obtained from a video acquisition card that samples the video signal coming from the CCD camera. The VIDI-PC video acquisition card software driver generates a file format called FILE_NAME.VID. The structure of this file comprises first the header called VIDI header which includes all the information of the image data, such as the number of bits of the A/D (Analog to Digital) video converter, the size of the grabbed image and so forth.

The file then comprises sequentially each bit plane by its turn depending on the number of bits per grabbed pixel color. For example if the image is captured under 16 gray levels acquisition

mode which corresponds to 4 bits of A/D video digitizer the rest of the VIDI file will include 4 planes of data as indicated by figure shown below



Hence, the data of the image under VIDI file structure format is a bit mapped and therefore does not permit the direct access file to read at once the whole information color of a given pixel.

Of course since the header file of the VIDI file format is known we have constructed a small program that reads the data under VIDI file format and converts it to a raw data file in order that the user can have an easy access to read and manipulate the pixel colors according to the logical order of occurrence with respect to the displaying screen. For example if one is interested to read the color of the 1st pixel he just have to read the first integer data that would correspond to a 10 according to the previous figure.

That is the way we followed to bring examples of real images within the MATLAB environment and simulates our methods on digital image restoration since the MATLAB software can read easily a raw file of data and coverts it to its MATLAB matrix format.

For more details on the hardware setup of the VIDI-PC video acquisition card and the MATLAB software environment is provided in appendices A and B at the end of this thesis.

2.5-GENERAL CONCEPTS IN IMAGE RESTORATION PROBLEM

Various methods have been proposed for image restoration problem. It is not an easy task to keep track of all the available techniques through a unified benchmark because image restoration is an ill-conditioned problem. We will give a general overview of this problem.

In general, independently on whether the restoration problem is within a statistical framework or not, the restoration is carried out by applying a reverse process to the sources of degradation on the degraded record image g. On the other hand, the characterization of the solution to the above problem is desired to be in the strict sense requiring a particular attention to the existence and uniqueness of an inverse process transformation.

Suppose that the process of recording an image is described by:

$$g(x,y) = \int_a^b \int_a^b h(x,y,\xi,\eta) f(\xi,\eta) d\xi d\eta \tag{2.10}$$

A reverse process transformation of recovering f from g is to look for a transformation T such that:

$$T^{-1}: g \to f \tag{2.11}$$

In the case where the PSF h is orthogonal to the original object f over a region on which the recording process is defined i.e.:

$$\int_a^b \int_a^b h(x,y,\xi,\eta) f(\xi,\eta) d\xi d\eta = 0 \tag{2.12}$$

the restoration of f is obviously impossible. In this case h is termed singular with respect to $f(\xi,\eta)$.

When h is not singular with respect to f, and the system is characterized, for example, by a linear transfer function of a blurring process with the addition of the noise, the recorded image can be modeled as:

$$g(x,y) = \int_a^b \int_a^b h(x,y,\xi,\eta) f(\xi,\eta) d\xi d\eta + n(x,y) \tag{2.13}$$

or in matrix form:

$$G = HF + n \tag{2.14}$$

where $G_{ij}=g(i\Delta x, j\Delta y)$ for $i,j=1,...,N$ and $F_{ij}=f(i\Delta x, j\Delta y)$ is the object image matrix.

In this case the inverse transformation T^{-1} exists due to the assumptions made about h. However, if the noise is not zero there can be no unique mapping between f and g. Every possible random process n is potentially generating a solution f from g. Consequently, there is an infinite family of candidate solutions from which one must select the optimal one.

The deterministic approach to image restoration consists of viewing (2.14) as a problem of finding a deterministic function f satisfying a set of linear equations.

In other words, the deterministic approach assumes that the PSF H is a square nonsingular matrix, and a candidate solution of f can be generated by:

$$\tilde{f} = H^{-1}g = f + H^{-1}n \tag{2.15}$$

which consists of the original object and a term involving the noise.

If H has a large condition number so that it is too ill-conditioned, H^{-1} will have some very large entries with respect to others. Therefore, the term $H^{-1}n$ can dominate the solution f and $\tilde{f}$ will be a bad estimate of f. Consequently, one may argue that the presence of an inherent error into (2.15) when H is ill-conditioned will cause a large inherent uncertainty in the solution.

For example, if the relative signal to noise ratio (SNR) is defined as:

$$\alpha = \frac{\|f\|}{\|n\|}$$

then the solution $\tilde{f}$ possesses an SNR of:

$$\beta = \frac{\|f\|}{\|H^{-1}\|\|n\|}$$

where || || denotes the norm of the matrix.

Thus, if H is ill conditioned, the deterioration factor in SNR from data to solution is $\frac{\alpha}{\beta}$ which could be considerably greater than one. When H is ill-conditioned this factor can reach up to 1000 or more.

Although, the ill-conditioned behavior of such solution can be relegated through the so-called generalized inverse (or pseudo inverse) matrices, this observation illustrates one of the difficulties involved in image restoration using the deterministic approach.

The statistical approach to image restoration using (2.14) implies an estimation of a set of values (vector) of f subject to random disturbance caused by noise. One way to do this, is by assuming that the solution is a random (regression curve estimation) function f but the nonrandom product Hf is corrupted by a random noise η. Furthermore, if the probability density function of this disturbance is known, then it can be used to estimate the function f given the recorded image g, at each site by:

$$\tilde{f} = \mathrm{E}\{P_n(f|g)\} \tag{2.16}$$

where P_n is the density function of η. The density function of f given g, at each pixel is simply the density of n centered at the recorded value g.

Moreover, any rigorous restoration technique should be based on the more plausible representation of an image recording process given by:

$$g = s\{Hf\} + n \tag{2.17}$$

where the nonlinear degradation function $s\{\bullet\}$ is included in the degradation process in contrast to (2.14).

In theory, the formulation (2.17) of g is the one which provides a complete reliable characterization about what might actually happen during real recording processes.

Let's specify a linear system represented by the matrix L which acts upon g to produce a restoration $\tilde{f}$ such that:

$$\tilde{f} = Lg = L\left(s\{Hf\} + n\right) \tag{2.18}$$

If $n = 0$, then $\tilde{f} = f$ would be true only if $L = s^{-1}H^{-1}$ and

$$s\{Hf\} = Hs\{f\} \tag{2.19}$$

Since linear and nonlinear operators do not commute, then (2.19) cannot be usually fulfilled. Therefore, it is doubtful that a linear L can exist, specially when $n \neq 0$, where L should satisfy:

$$L = H^{-1}s^{-1}\{s\{Hf\} + n\} \tag{2.20}$$

Thus, theoretically speaking, there would be no promising reconstruction due to the existence of random disturbances which cannot be known in advance. However, if one attempts to draw the behavior of (2.20) under small variations there would be some practical remedy. If we assume that the nonlinear function s, defined as above, anticipates for perturbations about a constant $\bar{g}$, then, using the Taylor series expansion we will obtain:

$$s\{\bar{g} + \Delta g\} \approx \left\{ s\{g\} + \frac{\partial s}{\partial g} \right\}\Bigg|_{\bar{g}} \Delta g \tag{2.21}$$

so that an approximation to (2.20) becomes as:

$$s^{-1}\{s\{Hf\}\} + \frac{\partial s^{-1}}{\partial g}\Bigg|_{g=s\{Hf\}n} \tag{2.22}$$

Therefore, (2.18) can be approximated by:

$$\tilde{f} \cong f + H^{-1}kn \tag{2.23}$$

where k is a constant that accounts for the first derivative in the Taylor series terms.

This shows that linear restoration still be devised when the recording process is subjected to nonlinear degradation. There is indeed a sufficient match of experimental witness to suggest

that processing using a linear approximation to (2.18) is relevant, independently on whether (2.22) is valid or not.

Consequently, from a parametric model view of restoration, there is no explicit proper solution, and the natural solution is obtained by assuming prior knowledge on f and η using desired criterion performance by involving many simplifications. The restoration using this approach is termed as a parametric model based estimation problem.

In the next section, we will present the Weiner approach of carrying out a restoration. It is well known in the digital parametric image restoration literature and requires a previous knowledge of the covariance matrices of f and η.

2.6- APPLICATION OF THE WEINER METHOD TO THE RESTORATION OF A REAL IMAGE

In this section, we will illustrate one of the most popular approaches of digital image restoration. This approach is based on the classical Fourier minimum mean squared error criterion. The estimator is known to give good results as long as the SNR is high.

Assume a linear, signal-independent-noise model of image formation and recording:

$$g = Hf + n \tag{2.24}$$

When the PSF H is the identity matrix, the method is referred as the Weiner filter.

We wish to restore directly f from g by using the mean of the square error (MSE) criterion that explicitly measures how close the result of the restoration is to the original object intensity distribution. Let's call the estimate $\tilde{f}$ of the object unknown image f obtained by using this criterion, and write the error of estimation as:

$$e = f - \tilde{f} \tag{2.25}$$

The MSE criterion requires that the expected total error of estimation to be minimum over the entire ensemble of all possible images. The fluctuation of the error (2.25) can be both positive and negative. Consider the minimum of the positive quantity as:

$$\min_{\tilde{f}} = \mathrm{E}\{e'e\} \tag{2.26}$$

The Weiner filter estimates a linear transformation L such that $\tilde{f} = Lg$ which minimizes (2.26). In this case, the error (2.26) can be written as:

$$\mathrm{E}\left\{ Tr\left[(f - Lg)(f - Lg)' \right] \right\} = 0 \tag{2.27}$$

It can be shown that it is equivalent to:

$$\mathrm{E}\left\{ Tr\begin{bmatrix} (ff' - L(ff' + nf') - (ff' + fn')L') + \\ L(ff' + nf' + fn' + nn')L' \end{bmatrix} \right\} = 0 \tag{2.28}$$

Using the fact that the noise is uncorrelated with the image object and assuming that the covariance matrices of the noise and that of the object image are available, the solution for L in (2.28) is given by:

$$L = Q_f \left(Q_f + Q_n \right)^{-1} \tag{2.29}.$$

where Q_f and Q_n are respectively the covariance matrices of the object image and the noise.

Thus, the use of the Mean Square Error criterion (MSE) requires the knowledge of some coarse parameters such as the covariance matrices of f and n.

Details on how to estimate these covariance matrices and their approximation by the circulant block matrices using tensor operations are given in Andrew and Hunt (1977).

The Fourier transform of the covariance matrices Q_f and Q_n are the power spectrum densities P_f and P_n respectively. A Fourier form of (2.29) produces the following estimator:

$$\tilde{f}(x,y) = \mathcal{F}^{-1}\left\{ \frac{P_f(w_x, w_y)}{P_f(w_x, w_y) + P_n(w_x, w_y)} \mathcal{F}[g(x,y)] \right\} \tag{2.30}$$

where $\mathcal{F}^{-1}$ is the inverse Fourier transform, and P_f, P_n are respectively the power spectral densities of the object image and noise and w_x, w_y are the frequency components.

We have used the Weiner approach to filter the noisy image given in Fig 2.4 which was obtained by the addition of white Gaussian noise whose standard deviation $\sigma=7$ and mean $\mu=0$ to the original image shown in Fig 2.3. The construction of the Weiner filter was done by using the exact covariance matrices rather than their estimates. The restored image is shown in Fig 2.5 . The performance error rate is defined as:

$$\frac{\sum_{i=1}^{N}\sum_{j=1}^{N}\left|f(i,j)-\tilde{f}(i,j)\right|}{\sum_{i=1}^{N}\sum_{j=1}^{N}\left|n(i,j)\right|}$$

Assuming that the total noise summation is 100%, the computed error rate value was found to be equal to 46%.

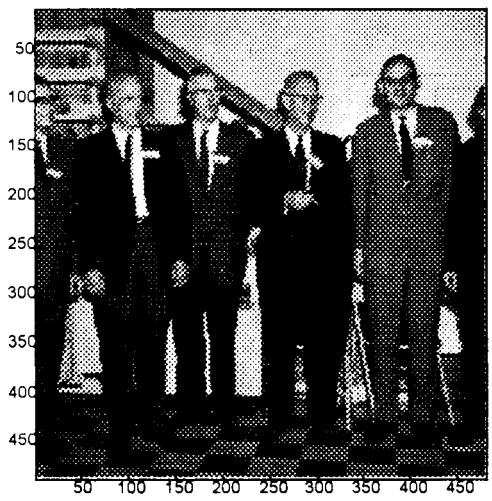


Figure2.3-Original object image.

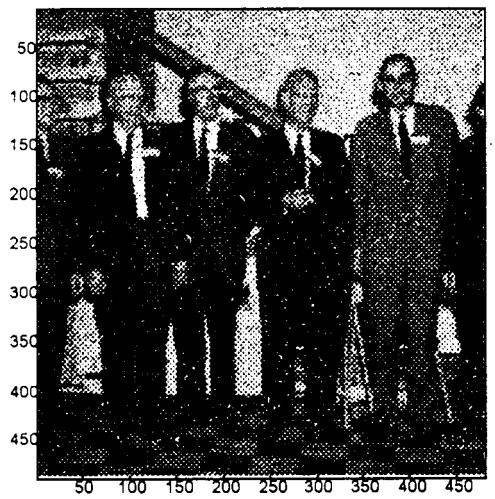


Figure2.4-Noisy image with low SNR.

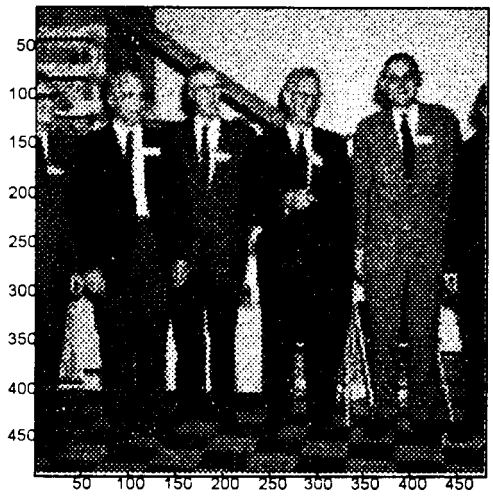


Figure 2.5-Weiner filtering of the noisy image (low SNR).

CHAPTER 3

KERNEL REGRESSION METHOD OF ESTIMATION

3.1-INTRODUCTION

In this chapter, we discuss in detail the nonparametric statistical methods of regression estimation which will be used to restore noisy images in chapter 4. We assume that the available observations/data are of the form (t_j, y_j) for $j=1,\ldots,n$, generated from a continuous random variable Y, using a specified sampling scheme at n predefined values of a variable t, from the regression model of the form:

$$y_j = \mu(t_j)+\varepsilon_j \tag{3.1}$$

where the ε_j are zero mean uncorrelated random variables with common variance σ^2 and $\mu(t)$ is an unknown function to be estimated.

The regression estimation problem consists of finding one function, amongst a specified set of functions, which explains adequately the variable Y in function of the variable t. In other words, the regression analysis refers to the estimation of the function $\mu(.)$ of the model (3.1) from the data at hand (t_j, y_j), $j=1,\ldots,n$ using one of the two statistical approaches known as the parametric or the nonparametric approaches of regression estimation.

The parametric approach assumes that the functional form of the function μ is known apart from a certain number of parameters which we denote by $\beta=(\beta_1, \ldots, \beta_p)'$. In this case, the model (3.1) reduces to :

$$y_j = \mu(x_1(t_j),\ldots,x_p(t_j),\beta)+\varepsilon_j \tag{3.1a}$$

Since the functional form $\mu(x_i(t_j),\beta)$ is known a priori, the problem will consist of estimating the vector β from the observed data. Usually, when $\mu(t)$ is assumed to be linear in the explicative functions $x_1,...,x_p$, (3.1a) can be written as:

$$\mu(t)= \sum_{j=1}^{n} \beta_i x_i(t)+\varepsilon_j \quad for \ i=1,...,p \tag{3.1b}$$

where $x_1,...,x_p$ are known functions called explanatory variables.

The most popular method of estimating $\mu(x_i,\beta)$ of the model (3.1b) from the data is the well known least square method which consists of choosing the vector β which minimizes the sum of squared residuals (called commonly the risk criterion) defined by:

$$RSS = \sum_{j=1}^{n}\left(y_j - \sum_{i=1}^{p}\beta_i x_i(t_j)\right)^2$$

It can easily be seen that if $(x^t x)^{-1}$ exists, the solution is of the form:

$$\tilde{\beta} = (x^t x)^{-1} x^t y$$

where x^t is the transpose matrix of the matrix $x=[x_1,...,x_p]$, and $x_j=(x_j(t_1),...,x_j(t_n))'$ for $j=1,...,p$.

When the function $\mu(.,\beta)$ is not linear in its parameters β, the regression estimation problem becomes, in general, complex specially when it cannot be linearized using one of the existing transformations

In the parametric approach, the estimated model rely more heavily on the information contained in the functional form of $\mu(x_i,\beta)$ since the data are mainly used only for selecting one function amongst the assumed parametric family. This allows it to possess good properties of convergence in the sense that when the sample size n tends to infinity the error criterion defined by $E\left(\|\beta-\tilde{\beta}\|^2\right)$ tends to zero at typical rates of order n^{-1}. Furthermore, the

parameters β_i's exhibit usually practical interpretations which allow one to judge the performance of the estimator. However, if the assumed parametric family is chosen using weak arguments, the approach may lead to very low performances and unwanted properties. Usually, the poor selection of the parametric family occurs when the process is hard to be parametrized through a mathematical representation or it is a new process. In this case, the nonparametric approach can be devised to overcome such difficulties.

The nonparametric approach of regression estimation is a collection of simple model free methods that can be used for the estimation of the model (3.1) without making strong assumptions (the knowledge of the functional form) on the unknown function μ. They offer wider convenience to estimate μ because these estimation procedures rely mainly on the information contained in the observed data rather than on a specific parametric family. For example, one has to make only few weak assumptions on the unknown function. The most common assumptions are related to the smoothness properties of μ such as differentiability and integrability of its second derivative that are usually satisfied in many practical situations.

There are three main popular nonparametric regression estimation methods the spline method of curve fitting, the orthogonal series estimators, and the kernel method of regression estimation. The first two methods will be briefly discussed in the next section whereas the kernel method will be reviewed in detail in section 3.3, since it is the method which we will use to restore noisy images.

3.2-NONPARAMETRIC REGRESSION PROBLEM

3.2.1-INTRODUCTION AND DEFINITION

The nonparametric approach of regression estimation is based on using the information contained in the observed data by taking advantage of the smoothness property of μ. When μ is actually smooth, then a valuable information conveying μ at t is contained in the observations obtained at, the neighbors t_j of t. In other words, the construction of an estimate of μ at t may be obtained by averaging locally the data in the neighborhood of t.

Most of the nonparametric estimators belong to the family of linear estimators which have the following form:

$$\mu_{\lambda}(t)= \sum_{j=1}^{n} K\left(t,t_j;\lambda\right) y_j \qquad (3.2)$$

where $K(.,t_j;\lambda)$ $j=1,...,n.$ is a collection of weight functions that depends only on a parameter λ, called the smoothing parameter or the smoothing bandwidth.

Although, one cannot expect unbiasedness for a nonparametric estimator of μ when the size of the data is finite and assuming that μ is to be an arbitrary function in a set W, an asymptotically unbiased estimator for large n over W is seeked. This is the reason why the performance of nonparametric estimators is usually assessed by an asymptotic consistency analysis. The performance criterion which is commonly used is called the risk or the average risk, and defined by:

$$R(\tilde{\mu}) = n^{-1} \sum_{j=1}^{n} \mathrm{E}\left(\mu\left(t_j\right) - \tilde{\mu}\left(t_j\right)\right)^2 \qquad (3.3)$$

where $\tilde{\mu}$ is the estimated function of μ.

The rate with which $R(.)$ decays to zero as n becomes large may be used to compare nonparametric estimators. It is well established that it is of order $n^{-\delta}$ where $0<\delta<1$. The fact

that δ is strictly smaller than 1 is one of the prices that one should pay for making fewer assumptions on μ than those used in the parametric approach.

3.2.2-SMOOTHING SPLINE ESTIMATORS

Assume that we have univariate observations Y_j at design points t_j; $j=1,...,n$ and the observations are assumed to satisfy the following model:

$$y_j = \mu(t_j) + \varepsilon_j$$

where the errors ε_j are uncorrelated with zero mean and constant variance σ^2.

On the other hand, we wish to estimate the function $\mu(t)$ without assuming any parametric functional form for $\mu(t)$. It can be seen that if we put no restrictions on the curve μ, we can reduce the sum of the squared residuals (or the risk) which is defined by:

$$RSS = \sum_{j=1}^{n} \left(y_j - \mu(t_j)\right)^2$$

to zero by choosing μ to be any curve which interpolates the data (t_j, y_j). The obtained trivial solution is obviously unsatisfactory because of its artificial rapid fluctuations.

The spline approach is a method which balances between the production of a good fit to the data and to avoid too much rapid local variations which can be measured by the roughness penalty defined by:

$$R(\mu) = \int \mu''^2(t)dt$$

Using this penalty measure, the *RSS* criterion is replaced by the modified sum of squared residuals defined by:

$$\begin{cases} n^{-1}\sum_{j=1}^{n}\left(\frac{y_j - \mu(t_j)}{\delta_j}\right)^2 + \lambda\int_a^b \mu''(t)^2 dt \\ \lambda \geq 0 \end{cases} \qquad (3.5)$$

where δ_j controls locally the overall extent of smoothing which can be replaced (if available) by the standard deviation of the random error at t_j. The parameter λ represents the weight that should be placed on the influence of the roughness penalty $\int_a^b \mu''(t)^2 dt$ where [a,b] is the interval to which the observation belongs to.

The solution to the constrained problem (3.5) was shown by Reinsch (1967) and Woodford (1970) to be a cubic spline. The form of the estimator $\tilde{\mu}_\lambda$ will therefore be a piecewise polynomial of the third degree:

$$\tilde{\mu}_\lambda(t) = a_j + b_j(t-t_j) + c_j(t-t_j)^2 + d_j(t-t_j)^3 \quad for \quad t_j \le t < t_{j+1}$$

where $c_1 = c_n = d_1 = d_n = 0$, $d_j = \frac{c_{j+1} - c_j}{3h_j}$, and, $b_j = \frac{a_{j+1} - a_j}{h_j} - c_j h_j - d_j h_j^2$ *for* $j = 1,...,n$.

The coefficients $c = (c_2,...,c_{n-1})'$ and $a = (a_1,...,a_n)'$ are known to satisfy the condition:

$$Tc = Q'a \tag{3.6}$$

where T and Q are matrices defined by:

$$T \equiv \begin{cases} T(j,j) = \frac{2(h_{j-1} + h_j)}{3} \\ T(j,j+1) = T(j+1,j) = \frac{h_j}{3} \end{cases}$$

$$Q \equiv \begin{cases} Q(j-1,j) = \frac{1}{h_{j-1}} \\ Q(j,j) = -\frac{1}{h_{j-1}} - \frac{1}{h_j} \\ Q(j+1,j) = \frac{1}{h_j} \end{cases}$$

and $h_j = t_{j+1} - t_j$

Using (3.5) and lengthy manipulations to cancel out a and c, it can easily be shown that:

$$\tilde{\mu}_\lambda = \left(I - Q\left(Q^t Q + \lambda T\right)^{-1} Q^t\right) y \tag{3.7}$$

3.2.3-THE ORTHOGONAL SERIES ESTIMATORS

The spirit of the orthogonal series estimators $\tilde{\mu}(t)$ is an extension of that of parametric linear model estimators. It is represented in the form of the linear regression model with λ functions $x_1,...,x_\lambda$ which spann the space of μ, and it is given by:

$$\tilde{\mu}(t) = \sum_{j=1}^{\lambda} \beta_j x_j(t) \tag{3.8}$$

Usually, the functions $x_1,...,x_\lambda$ are selected so that they form a complete orthonormal system (CONS). The familiar complex exponentials provide an important CONS of functions. They are defined in the interval $[a,b]$ by:

$$x_j(t) = (b-a)^{-1/2} e^{2\pi ijt/(b-a)};\ for\ j=0,\pm1,... \tag{3.9}$$

The corresponding estimator $\tilde{\mu}(t)$ is known as the classical Fourier estimator.

In this setting the estimation of μ reduces to the estimation of the regression parameters β_j's, once λ had been specified, using one of the standard methods amongst which the least square method.

The theoritical properties of this estimator is based on the following proposition.

Proposition 3.2.1: Luenberger (1969)

If $\{x_j\}_{j=1}^{\infty}$ is a CONS for $L_2[a,b]$, and $\mu \in L_2[a,b]$ then $\sum_{j=1}^{\lambda}\beta_j x_j$ is the best approximation to μ, amongst the set of estimators, in the sense that:

$$\left\|\mu-\sum_{j=1}^{\lambda}\beta_j x_j\right\| \le \left\|\mu-\sum_{j=1}^{\lambda}\alpha_j x_j\right\| \tag{3.10}$$

where β_j is the inner product defined by:

$$\beta_j = \langle \mu, x_j \rangle, j = 1,\ldots \text{ for all } \alpha = (\alpha_1,\ldots,\alpha_\lambda) \in \Re^{\lambda}. \tag{3.11}$$

Moreover as $\lambda \to \infty$ we have:

$$\mu(t) \xrightarrow{\text{converges}} \sum_{j=1}^{\infty}\alpha_j x_j(t) \tag{3.12}$$

Therefore, using (3.8) we deduce that:

$$\mu(t) \equiv (b-a)^{-1}\sum_{j=-\infty}^{\infty}\beta_j e^{2\pi ijt/(b-a)} \tag{3.13}$$

where $\beta_j = \int_a^b \mu(s)e^{-2\pi ijt/(b-a)}ds$

Furthermore, since μ is a real function, then the conjugate $\beta_j^* = \beta_{-j}$. Hence:

$$(b-a)\mu(t) \approx \beta_0 + \sum_{j=1}^{\infty}\left(\beta_j e^{2\pi ijt/(b-a)} + \beta_{-j}e^{2\pi ijt/(b-a)}\right)$$

If we fix $\lambda \ge 0$, the fitting of the function $\sum_{j=-\lambda}^{\lambda}\beta_j x_j(t)$ to the data by the method of least squares leads to the estimates of β_j's defined by:

$$\begin{aligned}\beta_\lambda &= (\beta_{\lambda-\lambda},\ldots,\beta_{\lambda\lambda})' = (X_\lambda^* X_\lambda)^{-1}X_\lambda^* Y \\ X_\lambda &= \left\{e^{2\pi ijr/n}\right\}_{\substack{r=0,n-1. \\ j=-\lambda,\lambda}}\end{aligned} \tag{3.14}$$

where X^* denotes the conjugate transpose of X

β_λ can be simplified to:

$$\beta_{\lambda j} = n^{-1}\sum_{r=1}^{n} y_r e^{-2\pi i j(r-1)/n}; \quad \text{for} \quad j = -\lambda,\ldots,\lambda. \tag{3.15}$$

On the other hand, it can easily be shown that $\tilde{\mu}_\lambda(t)$ can be expressed as:

$$\tilde{\mu}_\lambda(t) = \sum_{j=-\lambda}^{\lambda} \beta_{\lambda j} e^{2\pi i j t} \tag{3.16}$$

which can be rewritten as:

$$\tilde{\mu}_\lambda(t) = n^{-1} \sum_{r=1}^{n} y_r K_\lambda(t - t_r) \tag{3.17}$$

where $K_\lambda(x) = \sin(\pi(2\lambda+1)x)/\sin(\pi x)$

The function $K_\lambda(x)$ is known as the Dirichlet kernel. It is a periodic function even for noninteger argument λ and it satisfies the property:

$$\sum_{r=1}^{n} K_\lambda(t - t_r) = n. \tag{3.18}$$

As a consequence, the estimator $\tilde{\mu}_\lambda(t)$ can be computed , once the number λ is selected in an optimal sense, easily without assuming the knowledge of the parametric family of the unkown function $\mu(t)$.

3.3-KERNEL ESTIMATORS (CURVE FITTING APPROACH)

In this section we will review in detail the kernel method of regression estimation. It is based on a simple idea of assigning a local set of weights and averaging data. This method is assessed using asymptotic properties. If it is evaluated from these asymptotic properties, the kernel estimator cannot be argued against other nonparametric estimators such as the smoothing splines. Advantages of one nonparametric method with respect to another may come from a matter of convenience. The estimator performances rely on an appropriate choice of a smoothing parameter λ. Due to its importance, the concern of this choice has dominated on its own a quite wide nonparametric statistical literature.

We will review the kernel method in the univariate case. Its generalization to the bivariate case will be straightforward. Furthermore, the bivariate kernel method will be shown in the next chapter to be useful for the description of a very simple alternative way of smoothing noisy images.

3.3.1-GENERALITIES ON THE KERNEL METHOD

We assume that n observations $\{(t_1,y_1),...,(t_n,y_n)\}$ are generated from the model:

$$y_j = \mu(t_j)+\varepsilon_j$$

where the ε_j are zero mean, uncorrelated random variables and the t_j are deterministic ordered design points that satisfy the condition:

$$a \le t_1 < t_2, ...,< t_n \le b$$

where a and b are finite constants.

The idea behind kernel estimators goes back to the classical Fourier series estimators. We have seen that the Fourier series estimator can be expressed as:

$$\mu_\lambda(t) = \sum_{j=1}^{n} y_j k_\lambda(t - t_j) \tag{3.19a}$$

where k_λ is the Direchlet kernel defined by:

$$K_\lambda(u) = \sin(2\pi u/\lambda)/\sin(\pi u) \tag{3.19b}$$

which is a symmetric periodic function, and its maximum occurs at zero and dies out in magnitude away from it. Although these characteristics are desired, the estimator (3.19) can be well defined as long as λ is a nonnegative integer valued scalar.

The development of the kernel estimators was based on whether μ_λ given in (3.19a) is still an estimator of μ when λ is a non integer scalar. Indeed, due to the properties previously discussed of $K_\lambda(u)$, $\mu_\lambda(t)$ will tend to be always a type of weighted local averager of the observations near t even if λ is not an integer. Generating the weights, rather than solving a set of equations as in the orthogonal series estimators, by an appropriate kernel (function) is a more simple and an appealing approach since we don't need the solution of equations systems as required for instance, in the construction of Fourier series estimators. This approach led to the suggestion of the kernel estimator which is defined by:

$$\mu_\lambda(t) = \sum_{j=1}^{n} y_j K\left(t; t_j, \lambda\right) \tag{3.23a}$$

PROPERTIES OF THE KERNEL FUNCTION $K(t,\lambda)$:

First, since the estimator is based on local weighted averages, it is reasonable to assume that the support of the kernel $K_\lambda(.)$ is finite. The reason why we assume that $K(u)$ should have a finite support is due to the fact that when an infinite support kernel $K(u)$ is used in the estimation, the global bias is considerable at the boundaries of the observed data interval. Thus, kernels which have finite support are more desirable. Nonetheless, kernels which do not have finite supports are some times used since any kernel of infinite support can be rescaled so

that $K(u)$ is negligible outside the support, say, $[-1,1]$. Therefore, there is no loss of generality to assume the support of the kernel to be finite.

The kernel function $K(u)$ is a weighting function and the sum of the weights should be taken to be equal to 1, i.e.:

$$\int_{-1}^{1} K(u)du = 1 \tag{3.20a}$$

On the other hand, the contribution of the observations y's to the left and the right of t is independent on whether they are smaller or larger than t. Therefore, the kernel has to be symmetric on [-1,1]:

$$\int_{-1}^{1} uK(u)du = 0 \tag{3.20b}$$

Two other conditions are required for the asymptotic consistency properties of the kernel estimators:

$$\int_{-1}^{1} u^2 K(u)du = \alpha \neq 0 \tag{3.20c}$$

$$\int_{-1}^{1} K(u)^2 du < \infty \tag{3.20d}$$

The kernels which satisfy the above conditions are known as kernels of order 2. Kernels of higher order are preferred if one is interested in the estimation of the derivatives of $\mu(t)$. Other higher order kernels result from the imposition of additional moment conditions on the function K in (3.20c) and (3.20d), see Hall and Marron (1988).

If we want to generate the weights according to a kernel satisfying conditions (3.20), without losing the distributional properties of the response variable y, we may use the following:

$$K(t,t_j;\lambda) = \frac{1}{\lambda} K(\lambda^{-1}(t-t_j)) \text{ where } \lambda > 0. \tag{3.21}$$

Thus, the estimator $\tilde{\mu}(t)$ given in (3.23a) will be written as:

$$\tilde{\mu}(t) = \sum_j \frac{1}{\lambda} K\left(\lambda^{-1}\left(t - t_j\right)\right) y_j \tag{3.22}$$

The expression (3.22) looks like that of the classical Fourier series estimator given in the previous section 3.2.4. The main difference between the kernel estimator and classical Fourier series estimators is that the Direchlet kernel $K_\lambda(u)$ cannot be written as a function involving the argument u/λ since the term $\sin(\pi u)$ does not contain the parameter $1/\lambda$. However, if we approximate $K_\lambda(u)$ using the first term of Taylor series expansion, we will have:

$$K_\lambda(u) \approx 1/\lambda \frac{\sin(\pi u/\lambda)}{(\pi u/\lambda)}$$

This shows the strong similarities between the two nonparametric estimators.

The parameter λ, as it has been described earlier, is called the bandwidth or the smoothing parameter. Moreover, if $K(u)$ has a support interval on [-1, 1] then $K(u/\lambda)$ will have support on the interval $[-\lambda,\lambda]$. Consequently, the parameter λ in $K(u)$ will determine how many of the observations will be taken into account for estimating $\mu(t)$ at t where t lies in the middle of the support. The smoothing parameter also controls the variability of the kernel and hence the degree of the dependence of the kernel estimator on the information near t. In other words, it will be a wiggly estimator if small values of λ are used, and it will be a very smoothed estimator when large values of λ are used. Furthermore, if $K(u)$ is nonnegative then (3.21) can be used in probability density estimation. This is one of the reasons which motivates the choice of kernel estimation method among others nonparametric methods of estimation in the restoration of images problem.

The commonly used kernels with finite support in both density and regression estimation are summarized in table 3.1. The Gaussian kernel has also been extensively used in the literature

although it has an infinite support and defined by:

$$K(u)\left(\frac{1}{2\pi}\right)^{1/2} e^{-1/2u^2}$$

Kernel	$K(u)$	$V(K)=\int K(u)^2 du$	$M(K)=\int u^m K(u) du$	$\left(V^m M^2\right)^{\frac{2}{2m+1}}$
Uniform	$\begin{cases}\frac{1}{2}, \lvert u\rvert \le 1\\ 0, \lvert u\rvert > 1\end{cases}$	$\frac{1}{2}$	$\frac{1}{3}$	.3701
Triangular	$\begin{cases}1-\lvert u\rvert, \lvert u\rvert \le 1\\ 0, \lvert u\rvert > 1\end{cases}$	$\frac{2}{3}$	$\frac{1}{6}$	.3531
Quadratic (Optimal) (m=2)	$\begin{cases}\frac{3}{4}\left(1-u^2\right), \lvert u\rvert \le 1\\ 0, \lvert u\rvert > 1\end{cases}$	.6	.2	.3491
Minimum Variance (m=4)	$\begin{cases}\frac{3}{8}\left(3-5u^2\right), \lvert u\rvert \le 1\\ 0, \lvert u\rvert > 1\end{cases}$	$\frac{9}{8}$	−.08571	.6432

Table 3.1: Various popular kernel functions

Apart from the gaussian kernel and the kernel termed as « the minimum variance » which violates condition (3.20c), all other kernels satisfy the conditions (3.20) simultaneously. For more details concerning the kernel form (3.22) see Priestly and Chao (1972). However, as we will see later, the choice of the kernel to be used is not crucial.

There are other several variants of the kernel estimator (3.22) or the crude form (3.23a) which have been investigated and used in the literature, Eubank (1988), among of which we give:

$$\mu_\lambda(t) = \lambda^{-1}\sum_{j=1}^{n}\left(t_j - t_{j-1}\right)K\left(\lambda^{-1}(t-t_j)\right)y_j \qquad (3.23b)$$

3.3.2-ASYMPTOTIC THEORY

In this section, we will discuss the asymptotic properties of the kernel regression estimator. Without loss of generalities, these properties will be discussed for the estimator (3.23d) given by:

$$\mu_\lambda(t) = \sum_{j=1}^{n} y_j \int_{s_{j-1}}^{s_j} \frac{1}{\lambda} K\left(\lambda^{-1}(t-s)\right) ds$$
$$s_0 = a; s_{j-1} \le t_j \le s_j; j = 1,\ldots,n-1; s_n = b \qquad (3.24)$$

where for simplicity the support $[a,b]$ is assumed to be $[0,1]$, $t_j = j/n$; $j=1,\ldots,n$ and that $s_j = t_j + t_{j+1/2}$.

In asymptotic consistency analysis, a natural objective is to obtain an asymptotic expression for the risk $R_n(\lambda)$ defined by:

$$R_n(\lambda) = n^{-1} \sum_{r=1}^{n} \mathrm{E}\left(\mu(r/n) - \mu_\lambda(r/n)\right)^2 \qquad (3.25)$$

using the fact that the point expected mean square error is defined by:

$$R(t,\lambda) = \mathrm{E}\left\{(\mu(t) - \mu_\lambda(t))^2\right\} \qquad (3.26a)$$

which can be expressed as:

$$R(t;\lambda) = \mathrm{E}\left(\mu(t) - \mathrm{E}\,\mu_\lambda(t)\right)^2 + Var\left[\mu_\lambda(t)\right] \qquad (3.26b)$$

It can be shown, e.g. Eubank (1988), that:

$$Var\left[\mu_\lambda(t)\right] = \frac{\sigma^2}{n\lambda^2} \int_0^1 K\left(\lambda^{-1}(t-u)\right) du + O\left((n\lambda)^{-2}\right) \qquad (3.27)$$

where σ^2 is the common variance of the errors ε_j's.

The variance $Var\left(\mu_\lambda(t)\right)$ is desired to converge to zero as $n \to \infty$. This is valid when the following conditions are satisfied:

1. $\lambda \to 0$ as $n \to \infty$

2. $\lambda n \to \infty$ as $n \to \infty$

Priestly and Chao (1972) have shown that the expected mean of the estimator $\mu_\lambda(t)$ is given by:

$$E\{\mu_\lambda(t)\} = \frac{1}{\lambda}\int_0^1 K\left(\lambda^{-1}(t-s)\right)\mu(s)\,ds + O\left(n^{-1}\right) \tag{3.28a}$$

and

$$E\{\mu_\lambda(t)\} - \mu(t) = \frac{\lambda^2}{2}\mu''(t)\int_0^1 u^2 K(u)\,ds + o\left(\lambda^2\right) O\left(n^{-1}\right) \tag{3.28b}$$

Expression (3.28b) suggests that the bias is larger when the slope of $\mu(t)$ is changing rapidly, and it is moderate when $\mu(t)$ is smooth. Furthermore, we can notice that the estimator μ_λ will tend to overestimate the valleys of μ, and underestimate the peaks of the function. Therefore, the inspection of the data may give the user an idea on the regions where the estimator may not behave as good as other estimators.

The substitution of the expressions (3.28a) and (3.28b) into (3.26a) leads to an approximation of the point risk expression by:

$$R_n(t;\lambda) \approx \frac{s^2}{n\lambda}\int_{-1}^1 K(u)^2\,du + \frac{\lambda^4\left(\mu''(t)\right)^2}{4}\left[\int_{-1}^1 u^2 K(u)^2\,du\right]^2 \tag{3.29}$$

This expression shows clearly that the choice of the smoothing parameter λ is very crucial.

Using our objective, one may choose the optimal value λ_{opt} as the one which minimizes the right hand side of (3.29).

The minimization leads to:

$$\lambda_{opt} = \left\{\frac{\sigma^2 n^{-1}\int_{-1}^1 K(u)^2\,du}{\mu''(t)^2\left[\int_{-1}^1 u^2 K(u)^2\,du\right]^2}\right\}^{\frac{1}{5}} \tag{3.30}$$

When $\mu''(t) \neq 0$, *for* $t \in [-1,1]$; the corresponding risk is approximately given by:

$$R_n\left(t,\lambda_{opt}\right) \approx \frac{5}{4}n^{-\frac{5}{4}}\left\{\left|\mu''(t)\right|\sigma^4\left[\int_{-1}^{1}u^2K(u)du\right]\left[\int_{-1}^{1}K(u)^2du\right]\right\}^{\frac{2}{5}} \quad (3.31)$$

It is worth noting that these expressions give the optimal bandwidth and the risk for a specified point rather than the global optimal values of λ and the risk. However, these global values can be computed using (3.26) to give similar expressions as those of (3.30) and (3.31), except that $\mu''(t)^2$ is replaced by $\int \mu''(t)^2 dt$.

3.3.3-SELECTION OF THE OPTIMAL KERNEL

In this subsection we will give an answer to whether there is a best kernel to choose. The results from the previous subsection may be used in this regard. This is merely the minimization of the risk given in (3.31) with respect to K. This reduces to the minimization of the term involving the kernel K i.e.:

$$R(K) = \left\{\left[\int_{-1}^{1}u^2K(u)du\right]\left[\int_{-1}^{1}K(u)^2du\right]^2\right\}^{2/5} \quad (3.32)$$

Gasser and Müller (1979) have shown that, the minimization solution to (3.32) is a quadratic kernel with two sign changes on the support $[-1,1]$. To avoid such oscillations, they have imposed the restriction that the kernel should be positive on $[-1,1]$. This has resulted merely in the functional form of the kernel given in table 3.1 of section 3.3.2 under the designation « optimal quadratic kernel » or Epanichnikove kernel, Epanichnikove(1969).

Another intuitive criterion for choosing the optimal kernel is based on the smoothness property of the kernel estimator stating that $\mu_\lambda\left(t_j\right)$ can be thought of as an estimator which should have a smaller variance than the observed data by tolerating the existence of a criterion bias. In other words, the kernel method is a variance reduction method. Thus, one is led to select K which minimizes the variance of the estimator μ_λ.

The asymptotic analysis shows that the selection of a suitable kernel can be based on the minimization of (3.32), subject to the condition 3.20d (section 3.3.1), i.e.:

$$V(K)\int_{-1}^{1} K(u)^2 du \tag{3.33}$$

One possible solution to (3.32) subject to the constraint (3.33) is given by a linear polynomial $\theta_0 + \theta_1 u$ satisfying the conditions $\int_{-1}^{1}(\theta_0 + \theta_1 u)du = 1$ and $\int_{-1}^{1} u(\theta_0 + \theta_1 u)du = 0$.

The particular solution which corresponds to $\theta_0 = 1/2$ and $\theta_1 = 0$ corresponds to the rectangular kernel given in table 3.1, which we denote by $K_r(u)$.

Any other kernel on $[-1,1]$, can be generated by writing $K_r + \delta_K$ where δ_K is a functional form satisfying:

$$\begin{cases} \int_{-1}^{1} \delta_K(u)du = 0 \\ \int_{-1}^{1} u\delta_K(u)du = 0 \end{cases} \tag{3.34}$$

Since

$$V(K) = \int_{-1}^{1}\left(K_r(u) + \delta_K(u)\right)^2 du \text{ and } V(K_r) = \int_{-1}^{1} K_r(u)^2 du$$

Then

$$V(K_r) < \int_{-1}^{1} K_r(u)^2 du + \int_{-1}^{1} \delta_K(u)^2 du$$

Thus, the rectangular kernel is the one which possesses asymptotically the minimum variance. However, we can see from table 3.1 that $R(K)$ is minimum for the quadratic kernel whereas the $V(K)$ is quite similar for both kernels. Consequently, the choice of the kernel is not of so much concern, once the optimal selection of λ has been made.

3.3.4-SELECTION OF λ (Smoothing Bandwidth)

We have pointed out, in the previous section that, the important problem associated with the use of a kernel estimator is the selection of a good value of λ. On the other hand, in section 3.3.2, the optimal smoothing parameter, which is obtained from the asymptotic expression of the risk depends on the unknown function $\mu(t)$ whose knowledge is not available. Several methods have been suggested to overcome this problem. The natural approach consists of replacing $\mu(t)$ by a crude parametric estimate $\tilde{\mu}\left(t,\tilde{\beta}\right)$. Other more appealing methods known as the data driven techniques have been suggested for the estimation of λ. One of these methods is the method of Generalized Cross-Validation (GCV) due to Craven and Wahba (1982).

The GCV is a method which attempts to obtain an asymptotic minimizer of the expected risk expression without requiring the knowledge of the observations noise variance and the knowledge of $\mu(t)$. Basically, the method is a weighted version of the expected risk by utilizing the trace property of the estimator and the fact that the kernel estimator is linear.

Let's assume the kernel estimator 3.23d, expressed as:

$$\mu_\lambda = H(\lambda)y$$

where the entries of $H(\lambda)$ are occupied by the kernel function elements:

$$h_{jk}(\lambda) = \frac{1}{\lambda}\int_{s_{k-1}}^{s_k} K\left(\lambda^{-1}\left(t_j - s\right)\right)ds$$

Thus:

$$tr[H(\lambda)] = \frac{1}{\lambda}\sum_{j=1}^{n}\int_{s_{k-1}}^{s_k} K\left(\lambda^{-1}\left(t_j - s\right)\right)ds$$

The GCV criterion to be minimized is defined by:

$$\mathrm{GCV}(\lambda) = \frac{n^{-1}\sum_{j=1}^{n}\left(y_j - \mu_\lambda(t_j)\right)^2}{\left[(1-n\lambda)^{-1}\sum_{j=1}^{n}\int_{s_{k-1}}^{s_k} K\left(\lambda^{-1}(t_j - s)\right)ds\right]^2} \tag{3.35}$$

which can be minimized to give an asymptotically optimal value of the smoothing parameter λ. The trace $Tr[H(\lambda)]$ has the property of downwarding the bias due to the missing knowledge of σ^2.

There is no loss for generalizing this usage to other estimators (3.23). Actually, the GCV can be used in any other nonparametric method. It has been used the first time in smoothing splines estimators.

Rice (1984), Hardle, Hall and Marron (1985) have asserted that no smoothing bandwidth selector can be expected to give a very good estimate of λ_{opt}. They have shown a somewhat discouraging convergence rates of the order $n^{-3/10}$.

The use of a global λ_{opt} may increase the bias when the data is not homogeneous. This is an important remark for our work concerned with the smoothing of noisy images since arbitrary images are unlikely to be generated by a smooth intensity function but they possess many sharp regions separating the different local scenes. When sharp regions occur most frequently, a more rigorous choice of λ_{opt} would be one which is locally based on the scenes of the observed data. One possible way to do this is by taking advantage of an additional information concerning the smoothness of μ.

Assume that $\mu \in C^m[0,1]$ for $m>2$, where C^m is the set of all differentiable functions up to the m^{th} derivative. A more accurate expression for the bias rather than (3.28) of section 3.3.2 is given by:

$$\mathrm{E}\{\mu_\lambda(t)\} - \mu(t) = (-1)^m \frac{\lambda^m}{m!} \mu^{(m)}(t) \int_{-1}^{1} u^m K(u) du \tag{3.36}$$

This means that the kernel should instead satisfy the following moment conditions:

1. $\int_{-1}^{1} K(u)du = 1,$

2. $\int_{-1}^{1} u^j K(u)du = 0,$ *for* $j=1,\ldots,m-1,$

3. $\int_{-1}^{1} u^m K(u)du = \alpha \neq 0,$

A kernel of this type is called a kernel of order m. The previous kernels satisfying conditions 3.20 are called kernels of order 2. In this case the variance of the estimator is given by:

$$\mathrm{Var}(\mu_\lambda(t)) = \frac{\sigma^2}{n\lambda} \int_{-1} K(u)^2 du \tag{3.37}$$

and the total expected risk is:

$$R_n(t,\lambda) = \frac{\sigma^2}{n\lambda} \int_{-1} K(u)^2 du + \frac{\lambda^{2m}}{m!^2} \mu^{(m)}(t)^2 \int_{-1} u^m K(u) du \tag{3.38}$$

An asymptotically optimal bandwidth λ_{opt} at t can be given, Eubank (1988), by:

$$\lambda_{opt}(t) = \left[\frac{\sigma^2}{n} \frac{m!^2}{2m} \frac{V/M^2}{\mu^{(m)}(t)^2} \right]^{1/(2m+1)} \tag{3.39}$$

where $M = \int u^m K(u)du$, $V = \int K(u)^2 du$, and σ^2 is the variance of the noise.

In this case, the rate of convergence is of the order $n^{-2m/(2m+1)}$ and that the order n^{-1} can be attained when $m \to \infty$.

Estimators for which $\lambda = \lambda(t)$ varies with t are called variable bandwidth kernel estimators. In practice $\lambda_{opt}(t)$ still also cannot be obtained since it involves some unknown terms as does the

global optimal λ_{opt} given in section 3.3.3. However, this difficulty may be overcome by substituting an estimated $\mu^{(m)}(t)$ in the expression of $\lambda_{opt}(t)$.

One may estimate $\mu^{(m)}$ the derivative of order m of $\mu(t)$ using the kernel method by:

$$\mu_{\lambda}^{(m)}(t) = \lambda^{-(m+1)} \sum_{j=1}^{n} y_j \int_{s_{j-1}}^{s_j} K^{(m)}\left(\lambda^{-1}(t-s)\right) ds \tag{3.40}$$

where the kernel used must satisfy the following conditions:

$$\int_{-1}^{1} u^j K(u) = \begin{cases} 0; & j=0,\ldots,m-1;\ j \neq d \\ (-1)^d d!; & j = d \\ a \neq 0; & j = m \end{cases} \tag{3.41}$$

For instance, to estimate the first derivative one appropriate kernel is given by:

$$K(u) = 105/32\left(-9u^5 + 14u^3 + 5u\right)$$

Thus, an estimate of $\lambda_{opt}(t)$ can be chosen locally depending on the additional smoothness properties of $\mu^{(m)}(t)$ to reduce the bias. For further details, see Müller (1985 a,b); Gasser et al (1984); Gasser and Müller (1984) and Gasser, Müller and Mammitsch (1985).

The known bootstrap method provides another way to estimate $\lambda_{opt}(t)$ in a more rigorous probabilistic way. This will not be discussed here in details. It is basically based on a resampling technique of the residuals of the estimator during the estimation. A very rich review of this method is provided in Hârdle and Bowman (1988) and Taylor (1989).

3.3.5-ROBUST KERNEL ESTIMATORS

The regression kernel estimation is capable of fitting a general curve to data as long as the unknown function μ is smooth. However, we can easily see that the kernel estimators are not resistant to the effects of outliers because they are data driven methods. This is due to the fact that the kernel estimators are weighted types of least square estimators not well suited to

outliers. This lack of robustness motivated the development of some robust alternative estimators to 3.23, known as the M-type kernel estimators, see Eubank (1988).

An M-type estimator $\mu_\lambda(t)$ is defined as the minimizer of:

$$\sum_{j=1}^{n}\left(\int_{s_{j-1}}^{s_j} K\left(\lambda^{-1}(t-s)\right)ds\right)\rho\left(y_j-\theta\right) \tag{3.42}$$

with respect to some reference variable θ and ρ is a nonnegative function.

Provided that $\rho' = \psi$ is well defined, (3.42) is equivalent to finding a solution to:

$$\sum_{j=1}^{n}\left(\int_{s_{j-1}}^{s_j} K\left(\lambda^{-1}(t-s)\right)ds\right)\psi\left(y_j-\theta\right)=0 \tag{3.43}$$

The estimator $\mu_\lambda(t)$ derived from this way, satisfies distribution robustness.

For example, it can be shown that when $\rho(u)=u^2$, (3.43) results in the estimator given in (3.23d). Other functions of ρ that would generate more robust M-estimators are for example:

$$\rho(u)=|u| \tag{3.44}$$

and

$$\rho(u)=\begin{cases} u^2; & |u|\le c \\ c|u|-c^2/2; & |u|>c \end{cases} \tag{3.45}$$

where c is an appropriately chosen nonegative constant.

Conceptually, the use of the nondifferentiable function $\rho(u)=|u|$ leads to an estimator that is similar to a median action. This can be viewed as the median of the n values $x_1,\ldots,x_n$ which is the minimizer of $\sum_{i=1}^{n}|x_i-\theta|$. Function (3.45) improves a little more this view by combining u^2 with $|u|$ so that the influence of an extreme observation is reduced by the use of c instead of u for large deviations $|u|$.

The M-type estimator may be more appealing than the conventional kernel method of regression in image smoothing problems. The latter induces a blurring effect since an arbitrary digital image have usually close scenes of sharply changing colors commonly known as high frequency components.

3.4- SIMULATED EXAMPLES AND DISCUSSION

In this section we will simulate the three previously discussed nonparametric methods to illustrate their ability to smooth a univariate sample noisy observed data. In the simulation study we have used a simple periodic function on 100 regularly spaced samples defined as:

$$y_j = \sin(2\pi j/100)^2 + \varepsilon_j, \ for \ j=1,...100 \tag{3.46}$$

where ε_j are random Gaussian errors of zero mean and variance $\sigma^2 = .05$.

The original regression curve and its noisy degraded data are shown in Fig 3.1 The curve possesses low μ_j'' when μ_j is maximum or minimum, at which the smoothness of the original curve is reduced.

We have estimated the function $\mu_j(t)$ using the three nonparametric methods is shown in Fig 3.2.

The cubic spline and the kernel estimator have the following linear form:

$$\mu_\lambda = H(\lambda)y$$

where $H(\lambda) = \left(I - Q(Q'Q + \lambda T)^{-1}Q'\right)$, with Q and T are as defined before in subsection 3.2.2 for the cubic spline estimators whereas, and for the kernel estimator $H(\lambda)$ has typical entries:

$$h_{i,j}(\lambda) = \frac{1}{\lambda}K\left(\frac{i-j}{\lambda}\right), \ for \ \ j=1,...100$$

The corresponding matrices are depicted in Fig 3.2 and 3.3. We notice that the cubic spline weight matrix has weight entries that goes slightly under zero whereas the weight kernel matrix entries are all positive.

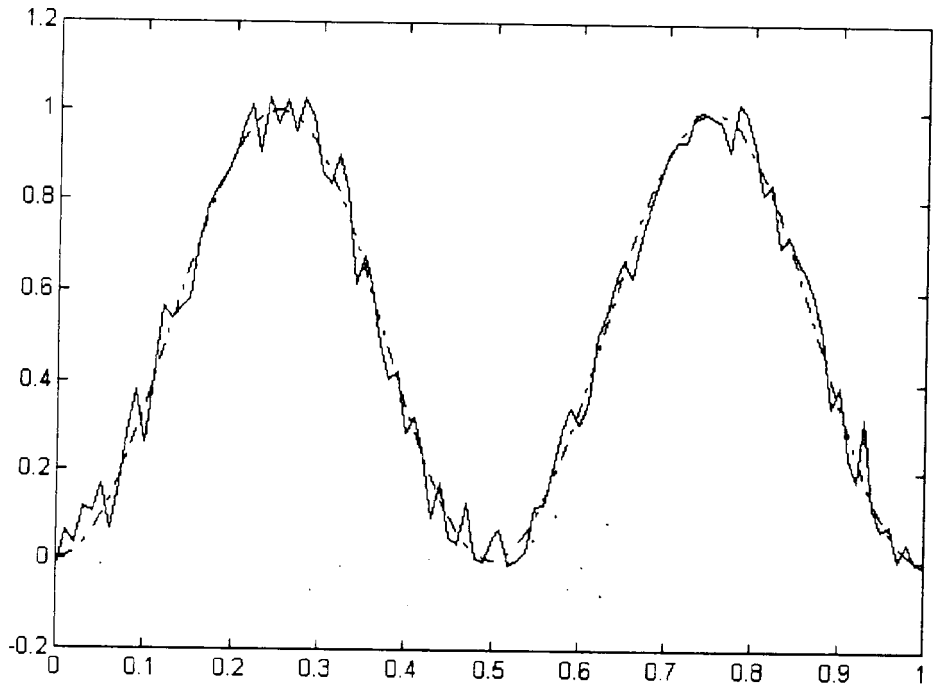


Figure-3.1 The rgression curve and its noisy observed record.

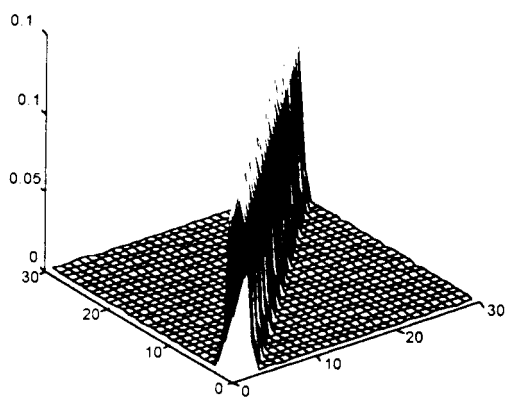


Figure-3.2 The kernel weight matrix $H(0.094)$.

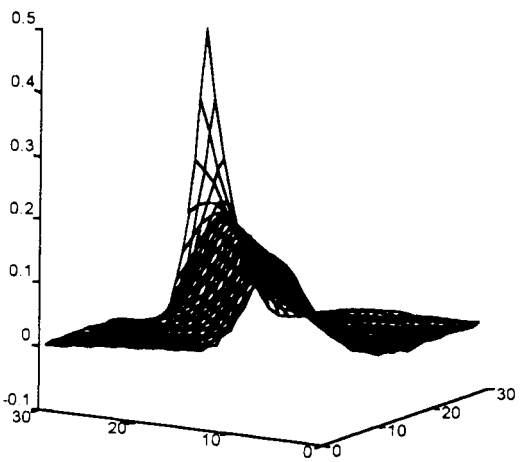


Figure-3.4 The cubic spline weight matrix λ=0.1.

The generalized Fourier series estimator was computed using (3.14) which gives a closed form for the estimator. The Generalized Fourier coefficients β_j's which have been computed for λ_{opt} =2 are given in table 3.2.

The coefficients $\beta_{-2,\ldots,2}$ for λ =2,
-0.2478 + 0.0161i
-0.0027 + 0.0006i
0.4984 - 0.0000i
-0.0027 - 0.0006i
-0.2478 - 0.0161i

Table 3 . 2: The estimates of Fourier coefficients

The application of the three nonparametric estimators to smooth the noisy curve of Fig 3.1 is given in Fig 3.4 showing that there is no significant diffrence between the three estimation.

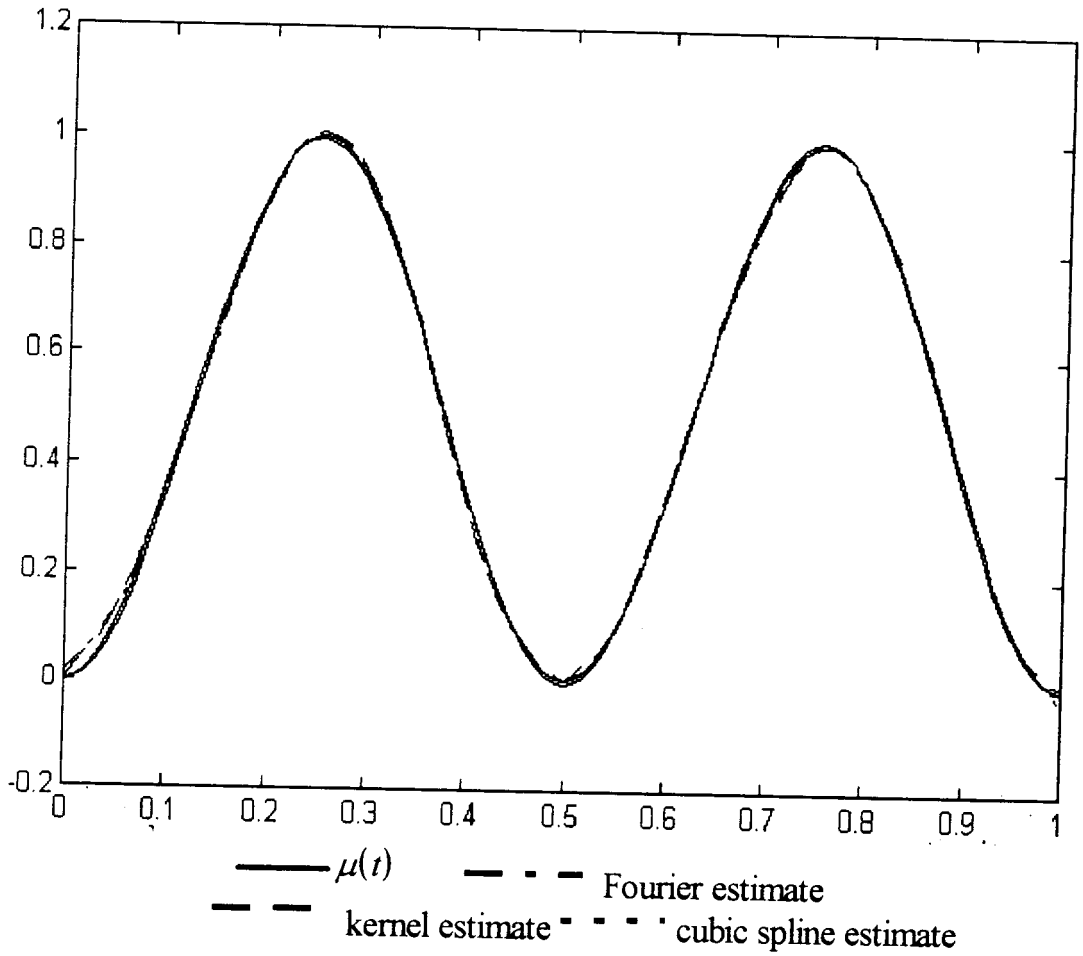


Figure-3.4 The three nonparametric estimates and the original rgression curve.

The three nonparametric estimators were compared using the error rate criterion which is defined by:

$$\frac{\sum_{j=1}^{100}\left|\mu_j - \tilde{\mu}_j\right|}{\sum_{j=1}^{100}\left|\varepsilon_j\right|}$$

	Error rate	λ_{opt}
Orthogonal series	0.1165 - 0.0000i	2
Cubic splines	0.1165	0.03
Kernel	0.1173	0.094

Table 3 . 3: Optimal error rates of the three estimators

The kernel estimator does not require the resolution of a linear system of equations as do the orthogonal series estimators and neither it does involve the inverse of matrix as do the cubic splines. Therefore, it is the easiest nonparametric method which requires, as compared to the others ones, a simple choice of the weight matrix computed from a single kernel.

It is interesting also to note that, the weight matrix $H(\lambda)$ for either kernel or cubic splines is nearly the same. The maximum weight for both matrices, occurs at the observed values (diagonal of the matrix) which indicates that the observed values contribute considerably to $\tilde{\mu}(t_j)$. Moreover, one can easily see the effect of the smoothing parameter λ on the assignment of the other diagonal weight entries. For the orthogonal series estimates the complex part of all the coefficients is close to zero and the greater the index j of the function term x_j is, the faster the corresponding coefficient vanishes to zero. This indicates the smaller the smoothing parameter is, the more higher order frequency components of the regression are attenuated.

The kernel estimator can be easily generalized to handle bivariate data samples. For example, the weight matrix would look like an umbrella centered at the site of estimation giving it a maximum weight. The construction follows easily from the multiplication of each row by its transpose. The simulation results show that the kernel method of regression is computationally

faster and more stable than the two others. This is clear, since the orthogonal series use the least square method which is computationally unstable (in the presence of noise). Moreover, since the kernel estimators have a convolution forms, they can be implemented using fast Fourier transforms to increase the computational speed, Silverman (1982).

CHAPTER 4

KERNEL SMOOTHING OF NOISY IMAGES

4.1-INTRODUCTION

In this chapter, we will propose the kernel method of regression estimation to smooth noisy images. The approach is based on regression modeling of the observed data which consist of the values of the intensity object image function discussed in chapter 2.

In nonparametric regression analysis, the prediction values of one response variable is based on only an observed sample of data i.e. prior knowledge of the functional form of the degradation process is not required. In image reconstruction (smoothing), the observed data is the color of pixels (or the gray level). From regression concept, these data are sustained to be generated by the same intensity function which is a two dimensional smooth curve i.e. the object image. The observed data is a sample image corrupted with an additive noise. Consequently, the problem of nonparametric image restoration reduces to the prediction of the two dimensional curve from just its observed records where it is assumed that the original object image was smooth before it has been recorded in the noisy environment. Our aim is to recover the original intensity function using the regression kernel smoother which has many attractive properties.

Among these properties, we list its asymptotic consistency. The estimator is both locally and globally asymptotically consistent. This means that as the sample size n becomes large, $n \to \infty$, the curve kernel estimator $\mu_\lambda \to \mu$ if the smoothing parameter λ tends to 0. The rate of

convergence is strictly related to the proper choice of λ, and inherently to the smoothness of the original image.

In image restoration, the data is so abundant and therefore the asymptotic convergence is guaranteed (good results are expected). Moreover, the kernel regression estimator is built on neighboring property which is possessed by many images in practice. This property is interpreted as that two neighboring sites (pixels) tend to have the same intensity values, see chapter 2. Therefore, the kernel method is an appealing method for smoothing noisy images.

4.2-IMAGE SMOOTHNESS ASSUMPTIONS

The noisy image is composed of pixels having a false color of gray level intensities. If the true image is believed to be smooth in a certain sense except in few regions, then the kernel estimator can be used to estimate the true pixel intensities. This estimation is carried out by using the information contained in the neighboring pixels of each pixel.

To begin with, let's assume that the observed image has been generated by the following process:

$$g_s = f_s + n_s \tag{4.1a}$$

or in matrix form

$$G = F + n \tag{4.1b}$$

where s is the spatial sites occupied by the equally sampled pixels, defined as $s=(i,j)$ for $i,j=1,...,N$. n is some random errors belonging to an i.d.d. random process having zero mean and constant covariance $\sigma^2 I$ matrix. g_s and f_s are the observed and the true colors at site s respectively.

The curve fitting approach or regression approach to image smoothing assumes the continuous discrete model of image recording process see Andrew (1972), in which the original object F is

believed to be continuous possessing a certain degree of smoothness, and the observed noisy record G is discrete. The aim is therefore to trace out a curve or a regression (matrix) function $\widetilde{F}$, from G, which should fit the sampled observed pixels intensities while minimizing the higher order perturbations. When these perturbations are believed to be due to the noise exclusively, the derivation of $\widetilde{F}$ is equivalent to minimize the local variance of the data. This argument suggests that the kernel method which is a curve fitting method can be considered as a method which restores the original image from the degraded image.

The smoothness assumption of an image is defined as that the original object does not contain sharp regions which are the scenes of pixels that possess a sharp change of gray level intensities. However, in practice, it is unlikely that an arbitrary image could not contain such components. Thus, the kernel estimator will attenuate these sharp regions because it assumes that the unknown F is smooth. The role of the smoothing parameter λ used in the kernel estimator is for guarding an acceptable compromise between the fit to the observed pixels and the smoothness of the unknown true image. If λ is taken small enough, it would keep as much as possible resolution but the resulting estimator is a wigglier estimator depending on the local amplitude statistics of the data and the noise. On the other hand, if the smoothing parameter λ has a large value, then the over smoothing leads to the creation of patches of the same color.

We have pointed out earlier in chapter 2 that the nature of noise process is essentially dependent upon the signal of the true image. This is true since most of the available technologies are usually based on a two recording type systems. Namely, these are the photoelectronic and the photochemical processes of sensing and recording systems. Both of them generate the film grain-noise. However, there is always a practical evidence that the digital samples are taken at greater

correlation distances than the average film-grain correlation distance. Therefore, film-grain noise is a white process when viewed from practical view of point. In this case, the power spectrum of the observed image should be composed of the sum of a constant term (due to noise) and the term involving the power spectrum of the true but unknown image. It is well established, in the literature, that the spectral power of the object image function becomes approximately flat beyond some frequency. The mean level about which the fluctuations occur can be identified as the spectral power associated with the noise. The other term is the associated « hump » at the lower frequencies. The PSF function usually accelerates the decay of these humps toward zero as shown in Fig 4.1.

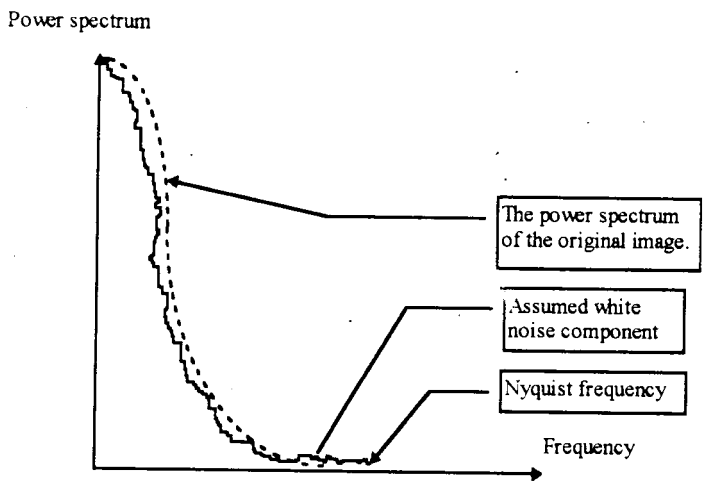


Figure 4-1 The power spectral analysis of the recorded image.

4.3 SOME PREVIOUS NONPARAMETRIC STATISTICAL ATTEMPTS

There have been an attempt to use the same approach of smoothing using a different type of nonparametric estimator, see Peyrovian and Sawchuck (1977). They have used a criterion based on the locally variable statistics and minimization of the second derivative. The resulting scheme is merely a smoothing cubic spline estimator where the smoothing parameter controls the tradeoff between resolution and over smoothness. The estimator has been able to restore a degraded image

by film grain noise. The method consist of finding a spline f which minimizes the following criterion:

$$\int_{s_1}^{s_N}\left[f''(s)\right]^2 ds+\lambda\left\{\sum_{i=s_1}^{s_N}\left[\frac{g(s_i)-f(s_i)}{\delta_i}\right]^2\right\} \tag{4.2}$$

where $s_1, s_2, \ldots, s_N$ are the pixel sites. In the cubic spline approach, the role played by the smoothing parameter λ is the inverse of the one played by the smoothing parameter in kernel estimator.

Titterington (1985a, 1985b) gave a general frame work under the goal of image restoration. The problem was posed as a least square estimation problem. He proposed a generalized framework of such estimators to obtain an estimate of F from G. A roughly based least square estimator for F can be obtained by finding $\tilde{F}_0$ that minimizes the following quadratic equation:

$$(G-F)'(G-F) \tag{4.3}$$

However, the least square solution gives unstable results if the PSF function (chapter 2) is not identity, particularly in the presence of the noise. This led him to propose a regularization method of noisy image which is based mainly on finding a tradeoff between two distance measures Δ_1 and Δ_2 respectively. One measures the distance between the $\tilde{f}_\infty$ and f, and, the other measures the distance between $\tilde{f}_0$ and f where $\tilde{f}_\infty$ is the ultra smooth version of f and f_0 is the under smoothed version of f. This is equivalent to finding an appropriate p such that:

$$\inf_{f\in\Phi}\left[\Delta_1(f,\tilde{f}_0)+p\Delta_2(f,\tilde{f}_\infty)\right] \tag{4.4}$$

where Φ is the set of all possible realizations of f. Hence a complete fidelity to the data can be obtained by making $p=0$. As p increases, $\tilde{f}_p$ becomes smoother and smoother toward the ultra smooth estimator $\tilde{f}_\infty$. He has listed some of the available candidates for Δ. These include:

1- Quadratic distance:

$$\Delta_I(f_1,f_2) \equiv (f_1 - f_2)'(f_1 - f_2) \tag{4.5}$$

2-Weighted quadratic distance:

$$\Delta_W(f_1,f_2) \equiv (f_1 - f_2)' W(f_1 - f_2) \tag{4.6}$$

2-Kullback-Leibler direct divergence:

$$\Delta_{KL}(f_1 - f_2) = \sum_{s\in D} q_1(s) \log\left(\frac{q_1(s)}{q_2(s)}\right) \tag{4.7}$$

where $q_1(s)$ and $q_2(s)$ are obtained from f_1 and f_2 by:

$$q(s) = \frac{f(s)}{\sum_{u\in D} f(u)}, \qquad s \in D = \{1,\ldots,n\}$$

If $\tilde{f}_\infty$ represents a prior image of specific interest, then Δ_2 is often taken to be a Kullback-Liebler measure and typically Δ_1 is taken to be either quadratic distance or weighted quadratic distance. Therefore the two distances can be not the same. If (4.4) has quadratic form then explicit solutions are available The smoothing parameter p can be chosen by the Cross-Validation data method. He has also established a relation between this estimator and the Maximum A priori Probability estimator.

4.4-GENERALIZATION OF THE KERNEL METHOD TO BIVARIATE DATA

We have seen that the kernel method of estimation provides a rigorous, and a simple way to construct asymptotically consistent nonparametric estimator that we will use to smooth noisy images next in this chapter. The kernel estimators are more flexible and require less computational complexity. Moreover, they can be extended easily to handle a vector valued t . In particular, for the case of bidimensional set of values of observations y_j's, in which j becomes a pair valued scalar integer called a site in spatial statistics. The extension is forward since one can use, in the multivariate case, a p-vectored kernel Cacoulos (1964) of the form:

$$K\left(u_1,\ldots,u_p\right)=\prod_{k=1}^{p}K_k\left(u_p\right) \tag{4.8}$$

where $K_{p.}$'s are univariate kernels which should individually satisfy the moment conditions given in (3.20).

Thus, from (3.22) (section 3.3.1), an estimator of μ may be expressed as:

$$\mu_\lambda(t)=\frac{1}{n}\sum_{j=1}^{n}y_j\prod_{r=1}^{p}\lambda_r^{-1}K_r\left(\lambda^{-1}\left(t_r-t_{jr}\right)\right) \tag{4.9}$$

where the vector $t_j=\left(t_{j1},\ldots,t_{jp}\right)'$ is the point where y_j is observed.

All the results that have been obtained for the univariate kernel hold also for the multivariate p-kernel Eubank (1988). Moreover, the Generalized Cross Validation method (GCV) can be used to make a selection of the smoothing parameter λ in the same way as described in section 3.3.4. The extension of the univariate GCV requires the use of matrix tensor operations of linear algebra. For more details see Craven and Wahba (1979).

4.5-SOME RELATED ASPECTS OF RESTORATION

The two dimensional kernel can be used for the purpose of regression smoothing of noisy images. Actually, in the p-variate estimator (4.9), one may select the same kernel i.e. $K_1(.)=K_2(.)$. Moreover, since the noise has constant covariance matrix $\sigma^2 I$, then the same value of the smoothing parameter λ can be used unless there is some convenience about some dominant image direction. A dominant image direction is the direction on which the image has its maximum power. In terms of what has been revealed in Fig 4.1, the maximum power is usually located in the lower frequencies. If we have a dominant image direction, then most of the smooth scenes are located on this direction in which the use of different values for the smoothing parameter would be more appropriate.

Since the two dimensional kernel estimator is a separable function, the problem of under smoothing or over smoothing sharp regions can be avoided. This requires an automatic scheme to detect the principle dominating axis of the image and, in this case, the kernel estimator can be used into a raster scan. In Fig 4.2 the black sharp scenes have forms that are oriented on the axis of the image plane whereas the gray scenes have forms which are not situated on the mair image plane.

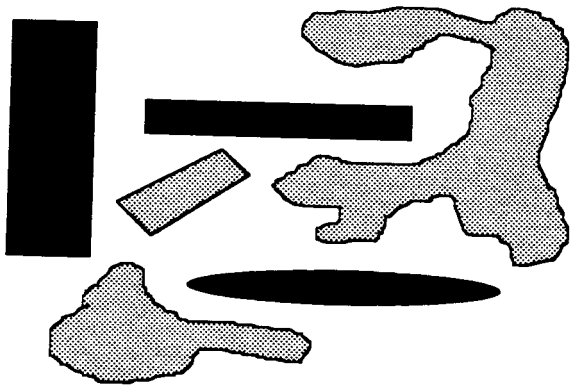


Figure 4.2-Showing objects which are situated on their principal components

4.6-APPLICATION OF THE DEVELOPED METHODS TO REAL IMAGES

We have applied the regression kernel method to smooth the image given in chapter 2 and only the main results are shown here, see appendix A and B for details of the practical implementation. We have used the bivariate kernel defined by:

$$K_\lambda(u,v)= K_\lambda(u) \cdot K_\lambda(v)$$

where $K_\lambda(u)$ is the minimum variance kernel defined by:

$$K(u) = \begin{cases} \frac{4}{3} - 8u^2 + 8|u|^3, & |u| < \frac{1}{2} \\ \frac{8}{3}\left(1-|u|^3\right)^3, & \frac{1}{2} \le |u| \le 1 \\ 0, & |u| > 1 \end{cases}$$

Fig 4.3 shows the graph of the univariate version of this kernel. The bivariate can be computed in a straightforward manner, by first computing the univariate weight vector $K_\lambda(u)$ and then the

bivariate is computed from $K_\lambda(u,v)= K_\lambda(u) \cdot K_\lambda(v)'$. Its graph is shown in Fig 4.4, whose form looks like a bell.

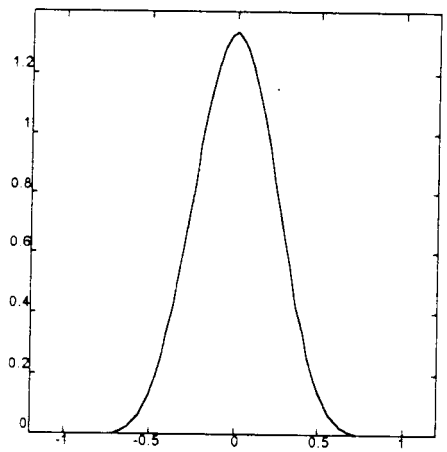


Figure-4.3 Functional form of the univariate kernel $K_\lambda(u)$ for λ=0.8.

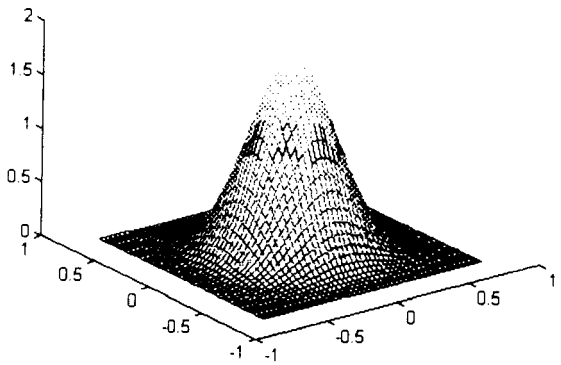


Figure-4.4 The bivariate kernel functional form $K_\lambda(u,v)= K_\lambda(u) \cdot K_\lambda(v)'$

The result of applying the approach of the kernel method to smooth noisy images is shown in Figs 4.4 to 4.9. These figures give respectively, the original, the noisy, and the smoothed image for various values of the smoothing parameter λ. For illustration, the smoothing was carried out for different values of smoothing parameters ranging from $\lambda=4.2\text{x}10^{-3}$ to $\lambda=15\text{x}10^{-3}$. The optimal smoothing parameter $\lambda_{opt}= 9\text{x}10^{-3}$ was obtained by minimizing the loss function defined by::

$$L = \sum_{i}^{N}\sum_{j}^{N}\left|f(i,j) - \tilde{f}_\lambda(i,j)\right|$$

where f is the original object image, $\tilde{f}_\lambda$ is kernel regression smoothed noisy image, and N is the number of pixels.

the optimal smoothed noisy image is given in Fig 4.7

The error rate is computed by the percentage of the total absolute value error of pixels that have not been recovered, defined previously in chapter 2, as:

$$\frac{\sum_{i=1}^{N}\sum_{j=1}^{N}\left|f(i,j)-\tilde{f}(i,j)\right|}{\sum_{i=1}^{N}\sum_{j=1}^{N}\left|n(i,j)\right|}$$

assuming that the noisy image has 100% of error rate. The percentage of this error rate in function of λ is shown in Fig 4.3.

It is worth noting that the graph of the error rate, given in Fig 4.3, suggests that choice of λ is not very crucial since the rate is nearly flat in the interval $\lambda \in \left[7\text{x}10^{-3}, 11\text{x}10^{-3}\right]$.

Fig 4.6 illustrates clearly what we argued earlier i.e.: when a small value of λ is used, the resolution is better preserved at the price that the noise has not been totally smoothed out. On the other hand Fig 4.8 which corresponds to the use of $\lambda=15\text{x}10^{-3}$ shows that the resolution has been sacrificed for the elimination of the noise.

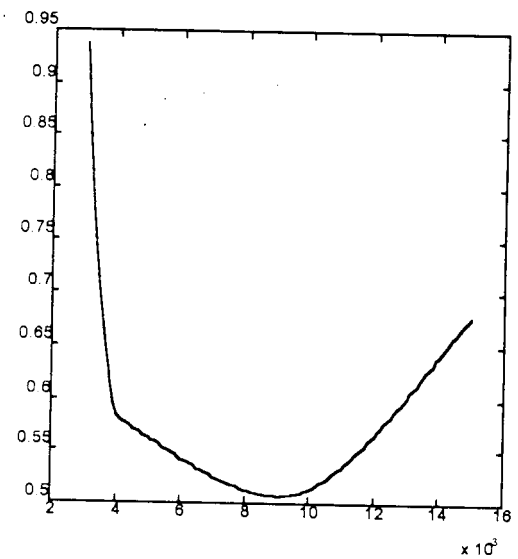


Figure 4.3 error rate versus λ in percentage.

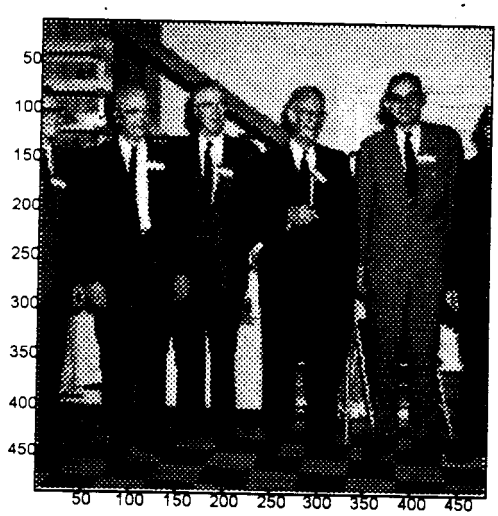


Figure 4.4- Original image 64 grey levels 480x480 pixels.

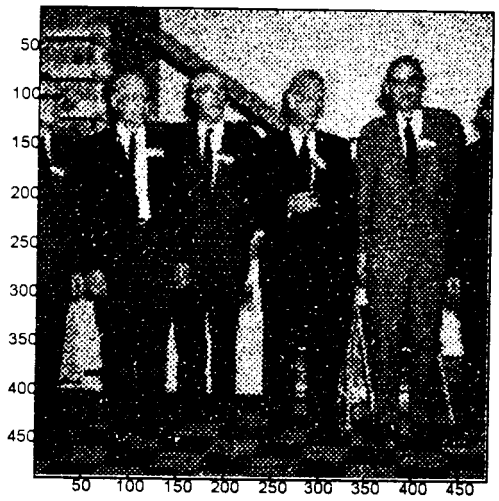


Figure 4.5 Noisy image with gaussian noise, μ=0 and σ=7.(SNR is low).

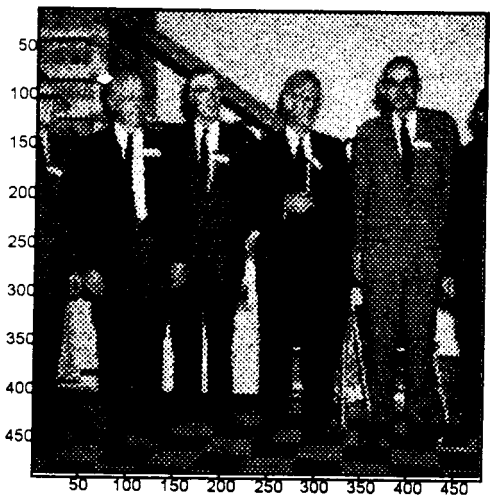


Figure 4.6-under smoothed image with $\lambda=4.2\text{x}10^{-3}$

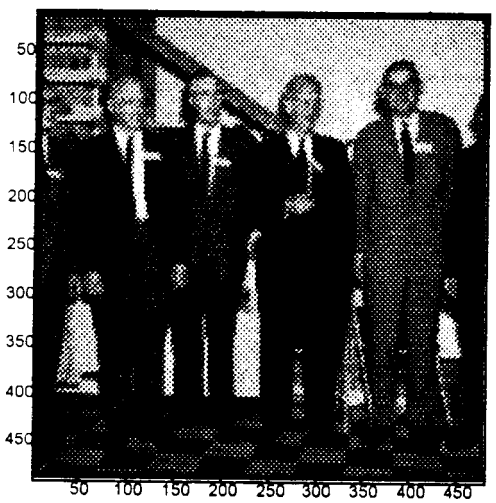


Figure 4.7- Optimally smoothed image with $\lambda=9\text{x}10^{-3}$

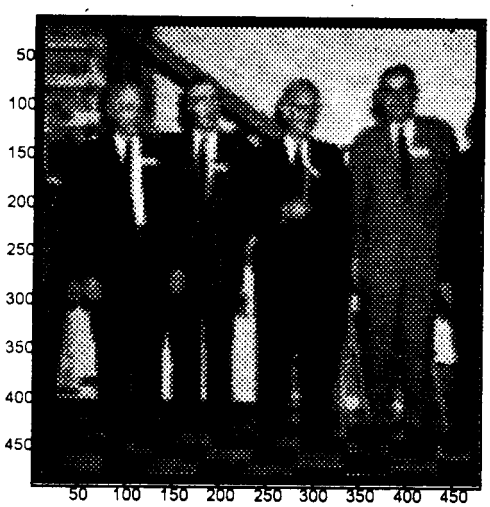


Figure 4.8-over smoothed image with $\lambda=15\text{x}10^{-3}$

The second illustrative example shows the same original image but with an additive noise of smaller standard deviation σ=2.5 rather than σ=7. The noisy and the optimal smoothed images are shown in Fig 4.9 and Fig 4.10 respectively. The recovery is better with higher SNR (high Signal

to Noise Ratio). In this case the error rate was found to be 39.74% whereas it was 46% in the previous example. Since the noise tends to increase the variability of the image, this example confirms that the optimal rate performances are dependent on the original smoothness of the image.

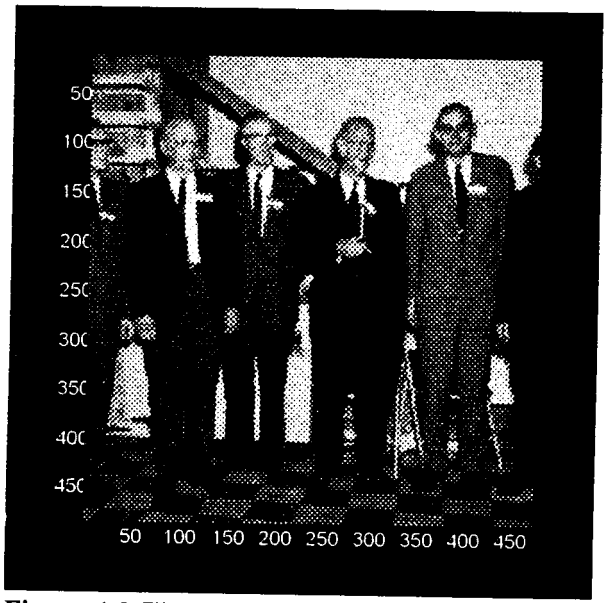


Figure 4.9-The noisy image with σ =2.5.

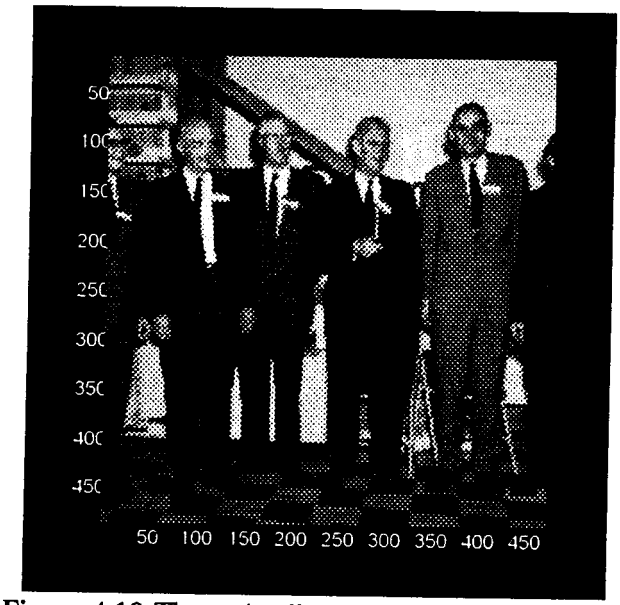


Figure 4.10-The optimally smoothed image with λ_{opt}=0.0035 and the error rate recovery is 39%

We have also applied the regression kernel approach to another image subdivided into 89x89 pixels and consisting of 16 gray levels, see Fig 4.11, to which we have added a Gaussian noise of mean zero and variance 1. The noisy image is given in Fig 4.12. The results of applying the regression kernel estimator for three different values of the smoothing parameter λ are shown in Figs 4.13 to 4.15. The optimal value of the smoothing parameter was found to be λ_{opt}=0.25, and the error rate was equal to 61%, see Fig 4.14 the corresponding optimal smoothed image. Fig 4.15 shows the over smoothed image which has been blurred whereas Fig 4.13 shows the under smoothed image.

Figure 4.11-The original image 89x89 pixels and 16 gray levels.

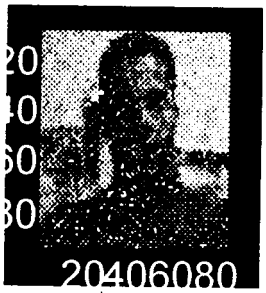


Figure 4.12-The noise image with Gaussian noise $\sigma^2 I=1$.

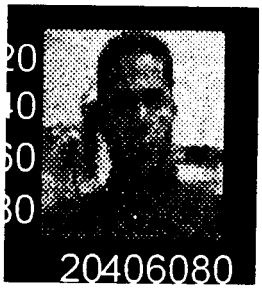


Figure 4.13-The under smoothed image with $\lambda=0.2$

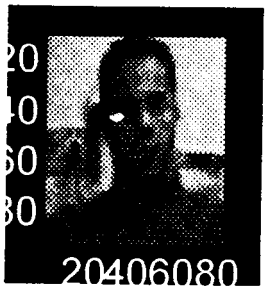


Figure 4.14-The optimally smoothed image with $\lambda_{opt}=0.25$ and 61% error rate.

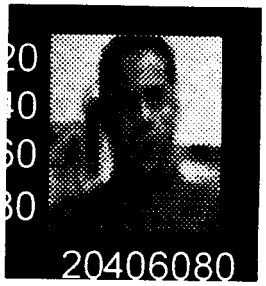


Figure 4.15-The over smoothed image with $\lambda=0.35$.

CHAPTER 5

CONDITIONAL KERNEL DENSITY ESTIMATION APPROACH

5.1-INTRODUCTION AND DEFINITION

The first part of our work has dealt with the application of the nonparametric regression estimation to image smoothing. The approach has shown flexibility since no prior knowledge on the model is required whereas parametric approaches require the knowledge of the degrading process. From the illustrative examples of the last chapter, we notice that the nonparametric regression approach generates « blurring effect » since it assumes that the intensity gray level function is smooth. This holds as long as the variation among the pixel colors is smooth. However, this is not always true in practice since there are many abrupt changes which occur in the colors of the scene pixels forming the complete image.

Most of the image restoration methods using classical approaches have their inferential structure based on a low level statistics analysis and modeling. In other words, they operate on a small correlation distance among pixels without taking into account the nature of the scenes to which they belong because independent and identically distributed models (i.i.d.) are usually assumed in the restoration. This is valid while it is believed that the image structure conveys large scale variations. Some times this may be relegated through nonhomogenous data models. Lack of homogeneity in data is usually accounted for in statistical models by a nonconstant mean assumptions. However, when nonhomogeneous small-scale variations are suspected, models involving statistical dependence are often more realistic.

In this chapter we should view the smoothing problem from a conditional distribution estimation point of view rather than assuming smoothness of the process. In this approach, the construction of a restoration scheme takes into account a higher level of statistical dependence. These are

characterized by a complement knowledge on the conditional interdependence among their low level statistical structures. The statistical area which deals with this concept is known as the spatial lattice statistics, see for example Cressie (1991). Unlike with the time series conditional models, in this approach dependence is present in all directions and becomes weaker as data locations become more dispersed. Data which are close together in space are likely to be correlated. This is a natural notion and it has been successfully used by statisticians to model spatial lattice processes.

Data analysis and statistical modeling of spatial data can be described using space geometry. In this type of analysis, the data are given by the spatial locations $D=\{s_1,...,s_n\}$ and the intensities $\{Z(s_1),...,Z(s_n)\}$ observed at those locations. The locations are in general considered as nonrandom because imprecise positioning is generally not modeled. The tractability of this approach to image smoothing is its ability to characterize procedures using extra information in the pixels neighborhood.

We will discuss in section 5.2 procedures where the spatial lattice could be made ordered as a time series by the use of the spatial interaction Conditional Markov (CM) models. The approach will be used in section 5.3 to construct unilateral and bilateral conditional nonparametric mode estimators of the true but unknown intensities at the given pixels using the kernel method of density estimation. In section 5.4 we will review the Markov random fields (M.r.f's) and the iterated method of restoration of Besag (1986) and, we will use the M.r.f's modeling of the spatial interaction pixels to propose a nonparametric iterative conditional estimates for the unknown pixel colors (NICM).

5.2-DESCRIPTION OF THE CONDITIONAL MODE ESTIMATION APPROACH

The conditional approach that we will see is based on the paradigm used in the estimation of time series generated by Markov processes, where past observed realizations are used to predict the present value. The idea is that time series generated by Markov process have their conditional interdependence characterized via the flow of time. In this case, the process is considered as a conditional Markov process of some order where the correlation function schemes, for instance, can be used to construct estimates based on standard criterion. This section will deal with the construction of unilateral and bilateral Markov models for the image smoothing problem. The words « unilateral » or « bilateral » stand for the unidirectional propagation with respect to the spatial lattice which enables us to construct a nonparametric unilateral or bilateral conditional mode estimator of the estimate of the pixel colors using the kernel method of density estimation.

By analogy to time series, an image obeying a unilateral or bilateral Markov models can be partitioned in two parts. The « past » and the « future » parts, so that the definition of Markovianity used in time series will remain valid for the spatial lattice data structures. For the description of the approach, the determination of the past part is sufficient.

We assume that the image consists of the observed records of a finite set of gray level pixels $\{z(s), s=(i,j) \in D\}$, $\{s=(i,j), 1 \le i,j \le M\}$ sampled on a grid lattice D, where $z(s)$, is the gray level of the cell s. Let N be a neighborhood of the site $s=(i,j)$ in the image lattice Das shown in Fig 5.1 for the site (0,0). If the pixel of interest lies on site s=(0,0) in a prescribed neighborhood N, then the pixel $(0,1)$ is at the east of $(0,0)$and $(1,0)$is at the south of $(0,0)$ which can be viewed

as the future pixels whereas (-1,0) and (0,-1) which are respectively located at the west and at the north of site (0,0) are viewed as the past pixels.

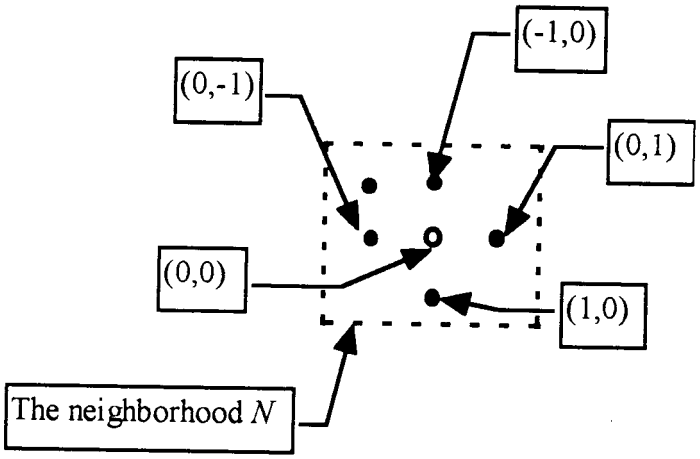


Figure 5.1 image CM based models

In the unilateral Conditional Markov (CM), the set N is defined as the nonsymmetric subset of the upper half plane S^+satisfying the following properties:

$$\begin{aligned} &s_1 \in S^+, s_2 \in S^+ \Rightarrow s_1 + s_2 \in S^+ \\ &s \in S^+ \Rightarrow -s \notin S^+, \quad (0,0) \notin S^+ \end{aligned} \tag{5.1}$$

In the case of image lattice structure, the definition of S^+ in (5.1) is not unique. One way (out of many) to make it unique, can be achieved by ignoring all the future parts of S^+ near $(0,0)$. In the situation which is depicted in Fig 5.2 (another way is available in Fig 5.3), the estimator may be constructed using only the past gray sites. The white ones that correspond to the future sites are not taken into account. The word « future » here, is left to be interpreted for the convenience of someone with respect to another.

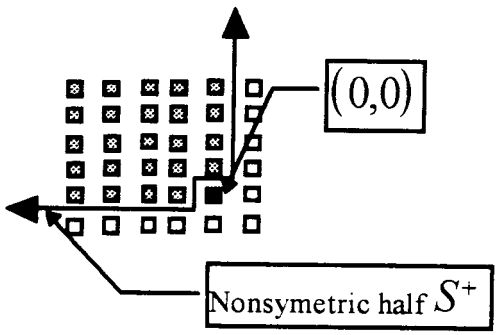


Figure 5.2- Showing S^+ as causal unilateral model

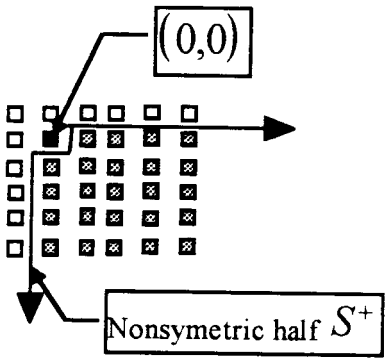


Figure 5.3- Showing another possibility of S^+.

Therefore, for any pixel s in D, we can divide an arbitrary finite image lattice D in two parts, namely $D_{s,N}$ (The upper left quadrant) and its complement in N. $D_{s,N}$ can be interpreted as the « past » and the remaining part excluding s as the « future ».

The unilateral Markov model possesses the one dimensional Markov property defined as Chellappa and Kayshap (1982):

$$P\big(\theta(s) \mid \theta(r),\ \forall r \in D\big) = P\big(\theta(s) \mid \theta(s+r), \forall r \in D_{s,N}\big) \tag{5.2}$$

The popular class of causal models which consists of the three neighbor model with N={(0,-1), (-1,0), (-1,-1)}, (i.e. only the west, north and, north-east pixels) is a special class of unilateral models so that N is a subset of the top left quadrant square centered at the pixel of interest.

In regard to this definition of N, an image obeying the unilateral model can be synthesized recursively using appropriate initial conditions. The appropriate initial conditions are meant for the availability of the topmost rows and the leftmost columns for the neighbor of N. These are conditioned upon the properties of convergence rates of the nonparametric estimator which is require more data for asymptotic consistency estimation.

However, there are valuable color information in the future pixels for the unilateral specified model, that are closer to pixel s than in the so-called past S^-. In this case, a better specification of an arbitrary image lattice structure can be achieved by what is called a bilateral specification in which a better local neighborhood symmetry is achieved.

An image is said to have a bilateral conditional Markov property with respect to the symmetric neighbor N, see Chellappa and Kayshap (1982), if the following is satisfied:

$$p\big(\theta(s)|\ all\ \theta(r),\ except\ r=s\big)=p\big(\theta(s)|\ all\ \theta(r+s), r\in N\big)$$

and we will assume the subset N_s of N, to satisfy the condition:

$$s\in N_s \Rightarrow -s\notin N_s\ \ where\ \ N=\left[s: s\in N_s \text{ or } -s\in N_s\right]$$

This is illustrated in Fig 5.4.

It is worth noting that any model satisfying the bilateral CM, does not necessarily satisfy the unilateral CM model. This is due to the difference in the geometrical adjacencies of the neighboring pixels from all possible directions.

This method of ordering neighboring pixels as past and future information enables us to formulate the problem of image restoration as the one which consists of predicting the intensity value z_{s+1} at pixel $s+1$ from the past values. In other words, an estimate of the z_{s+1} will be expressed as:

$$\theta_{s+1}=\mathbf{E}[z_{s+1}\mid z_u\ \ -s\le u<-1]$$

In this section our objective is to estimate Θ_{s+1} using the kernel method of nonparametric density estimation.

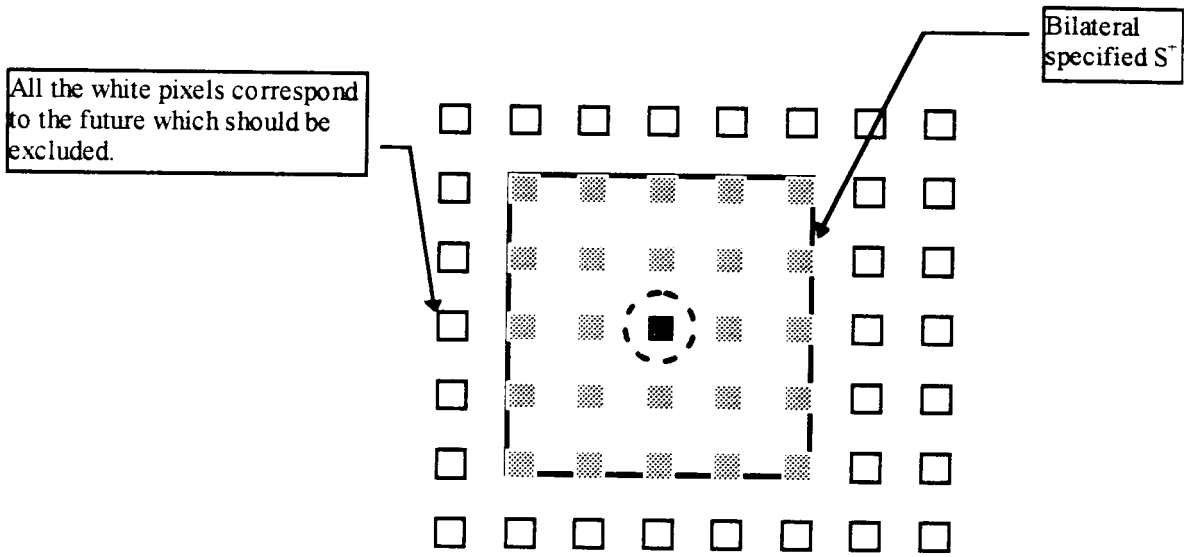


Figure 5.4-The Bilateral specified CM models. Notice the local symmetry provided by S^+ around s than in the unilateral CM.

5.3 KERNEL METHOD OF DENSITY ESTIMATION

In this section we will review the form of the kernel density estimator which will be used to smooth noisy images when modeled by the CM (Conditional Markov) process.

Let $(t_1,y_1),\ldots,(t_n,y_n)$ be independent and identically distributed random variables having a common continuous bivariate probability distribution (T,Y). We assume once again our original model:

$$y_j = \mu(t_j) + \varepsilon_j; j = 1,\ldots,n$$

where ε_j are assumed to be independent and not necessarily identically distributed with common variance σ^2, and t_j are random design points, rather than deterministic, admitting a certain distribution.

One way to estimate the unknown value of $\mu(t)$ is by the conditional mean given by:

$$\tilde{\mu}(t) = \mathrm{E}\{Y|T = t\} \tag{5.3}$$

If we assume that (T,Y) have a joint continuous density function $f(t,y)$, the marginal density function of T is given by:

$$f_T(t) = \int_{-\infty}^{\infty} f(t,y)dy$$

Then, the conditional density function for Y, given T is:

$$f_{Y|T}(y|t) = \frac{f(t,y)}{f_T(t)} = \frac{f(t,y)}{\int_{-\infty}^{\infty} f(t,y)dy} \tag{5.4}$$

Using (5.3), we see that $\tilde{\mu}$ becomes:

$$\begin{aligned}\tilde{\mu}(t) &= \int_{-\infty}^{\infty} y f_{Y|T}(y|t)dy \\ &= \int_{-\infty}^{\infty} y \frac{f(t,y)}{\int_{-\infty}^{\infty} f(y,t)dy}dy\end{aligned} \tag{5.5}$$

Therefore, an estimate for $\mu(t)$ can be obtained by plugging kernel density estimate of $f_{Y|T}$ into (5.5).

The kernel method of density estimation traces its origins back to the works of Rosenblatt (1956) and Parzen (1962). Its extension to the estimation of multivariate density functions can be found in Cacoullos (1964).

Assume that there are n independent and identically distributed p-vectored observations $x_1,...,x_n$ where $x_i = \left(x_{i1},...,x_{in}\right)'$ obtained from a distribution with density function $f(x)$ where $x = \left(x_{p1},...,x_{pn}\right)'$ then, Cacoullos (1964):

$$f_\lambda(x) = \frac{1}{n\lambda^p}\sum_{r=1}^{n}\prod_{j=1}^{p} K_j\left(\lambda^{-1}\left(x_j - x_{rj}\right)\right) \tag{5.6}$$

where x_{rj} is the j^{th} component of the vector x_r, and the kernel functions K_j are assumed to have finite supports. The kernels K_j should not necessarily have the same functional forms, see Epanichnikove (1969), and λ is a smoothing parameter to be chosen.

For bivariate density functions (5.6) becomes:

$$f_\lambda(t,y) = \frac{1}{n\lambda^2}\sum_{j=1}^{n} K_1\left(\lambda^{-1}\left(t - t_r\right)\right)K_2\left(\lambda^{-1}\left(y - y_r\right)\right) \tag{5.7}$$

where the observations are $(t_1,y_1),...,(t_n,y_n)$.

Therefore, the marginal density f_T can be estimated by:

$$\begin{aligned} f_{\lambda T}(t) &= \int_{-\infty}^{\infty} f_\lambda(t,y)dy \\ &= \frac{1}{n\lambda}\sum_{r=1}^{n} K_1\left(\lambda^{-1}\left(t - t_r\right)\right) \end{aligned} \tag{5.8}$$

Using the definition of the conditional density $f_{Y|T}$ given by:

$$f_\lambda(y|t) = \frac{f_\lambda(t,y)}{f_{\lambda T}(t)} \tag{5.9}$$

an asymptotically consistent nonparametric estimator $\widetilde{\mu}$ of μ can be derived as:

$$\begin{aligned}\mu_\lambda(t) &= \frac{1}{f_{\lambda T}(t)}\int_{-\infty}^{\infty} y f_\lambda(t,y)dy \\ &= \frac{1}{f_{\lambda T}(t)}\frac{1}{n\lambda^2}\sum_{r=1}^{n} K_1\left(\lambda^{-1}(t-t_r)\right)\int_{-\infty}^{\infty} y\sum_{r=1}^{n} K_2\left(\lambda^{-1}(y-y_r)\right)dy \\ &= \frac{1}{f_{\lambda T}(t)}\frac{1}{n\lambda}\sum_{r=1}^{n} K_1\left(\lambda^{-1}(t-t_r)\right)\int_{-\infty}^{\infty}(y_r+\lambda u)K_2(u)du \\ &= \frac{\sum_{r=1}^{n} K_1\left(\lambda^{-1}(t-t_r)\right)y_r}{\sum_{r=1}^{n} K_1\left(\lambda^{-1}(t-t_r)\right)}\end{aligned} \qquad (5.10)$$

Estimator (5.10) is obtained by making use of the conditions (3.20) seen in chapter 3 and knowing that the choice of one kernel with respect to another is not of much concern.

Estimator (5.10) has a long history which can be related back to the works of Nadraya (1964) and Watson (1964), and referred as « Nadraya Watson estimator ». Beneditti (1974) has considered the case where t is deterministic design points, and have independently derived the estimator (3.23c), the same functional form of estimator (5.10).

Under certain conditions, asymptotic consistency properties have been established some of which can be found in Noda (1976), Rosenblatt (1969), Schuster and Yakowitz (1979), Coulomb (1981). Devroye (1978a,1980 a) has shown that when $n \to \infty$, $\lambda \to 0$,and $n\lambda \to \infty$:

$$Sup_{a \le t \le b}\left|\mu_\lambda(t) - \mu(t)\right| \overset{\text{asymptotically}}{\longrightarrow} 0$$

Schuster (1979) have shown that μ_λ is asymptotically normal. If K_1 is continuous and $n \to \infty$, $\lambda \to 0$, in such a way $n\lambda^3 \to \infty$ and $n\lambda^5 \to 0$ then $(n\lambda)^{1/2}\left(\mu_\lambda(t) - \mu(t)\right)$ has an asymptotically a zero mean normal distribution with variance:

$$Var(Y|T=t) = \frac{\int_{-1}^{1} K^2(u)du}{f_T(t)}$$

Hardl and Marron (1985a) considered the problem of bandwidth selection in the case where t 's are random and f_T is known. If the GCV (Generalized Cross Validation) is used, it must be modified by including a weighting function and other ways are available, see Eubank (1988).

Let $\mu_\lambda(j)$ denote the estimator of μ for the case when (t_j, y_j) has been omitted from the data such that:

$$\mu_{\lambda(j)}(t) = \frac{1}{(\lambda(n-1))} \frac{\sum_{i \neq j} K\left(\lambda^{-1}(t - t_j)\right) y_j}{f_T(t)}$$

then the modified cross-validation (weighted CV) criterion to be minimized will be:

$$..CVW(\lambda) = \frac{1}{n} \sum_{j=1}^{n} \left(y_j - \mu_{\lambda(j)}(t)\right)^2 w(t_j) \qquad (5.11)$$

where $w(t_j)$ is some nonnegative weight function, see Hardle and Marron (1985a).

It can be established that the $CVW(\lambda)$ always gives an asymptotically optimal selection rule, see Hardel and Marron (1985 a).

Yakowitz (1985) has shown that the same results hold when $Y_1,...,Y_n$ are random variables generated by a random Markov process i.e. applicable to estimate $\mathrm{E}[Y_{i+1} | Y_i = y_i]$.

Let's assume that the image can be modeled by the unilateral conditional Markov additive process as:

$$z(s) = \theta(s) + \eta(s)$$

where s is defined in section 5.2, $\eta(.)$ is noise, and $\theta(s+1)$ is a linear combination of the past observed intensities $z(s)$. The aim is that based on observations S^+ of a neighborhood N of pixel $s+1$, it is desired to obtain a conditional predictor defined by:

$$\tilde{\theta}_{s+1} = \mathrm{E}[z_{s+1} | z_s]$$

for all $z(s)$ for $s \in D$ and having each S_s^+ of size w^2.

The typical choice of the set S_s^+ and S_{s+1}^+ when it is specified by the unilateral CM model is depicted in Fig 5.5.

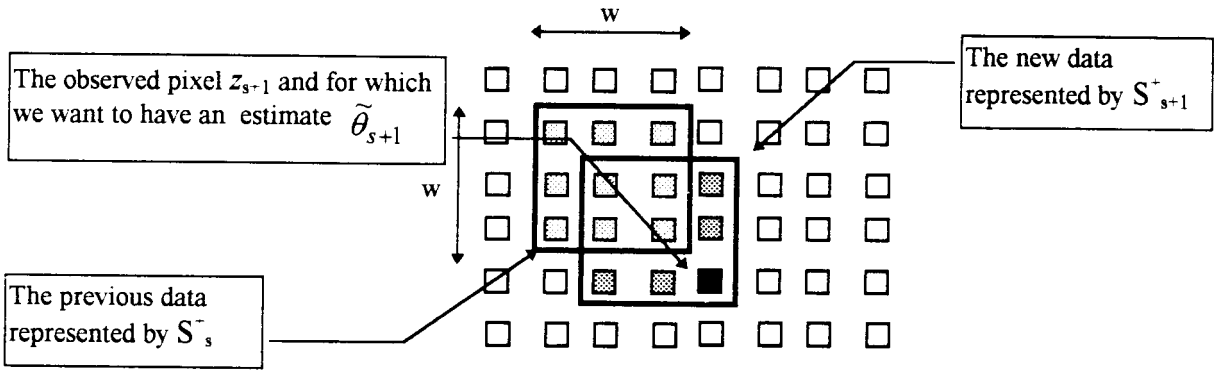


Figure 5.5-Showing the simulation of the unilateral conditional mode predictor.

Let $(z(s),z(s+1))$ have an associated continuous joint density function $f(z_s,z_{s+1})$ where the marginal density function of z_s is:

$$f(z_s)=\int_{-\infty}^{\infty} f(z_s,z_{s+1})\partial z_{s+1} \tag{5.12}$$

The conditional density function of z_{s+1} given z_s, is :

$$f(z_{s+1}|z_s)=\frac{f(z_s,z_{s+1})}{f(z_s)}$$

Therefore the estimate of pixel at site $s+1$, $\tilde{\theta}_{s+1}$ is:

$$\tilde{\theta}_{s+1}=\int_{-\infty}^{\infty} z_{s+1} f(z_{s+1}|z_s)\partial z_{s+1} \tag{5.13}$$

Choosing a size w (window) of S^+ and using the above results of the nonparametric density function of the kernel method of estimation we obtain:

$$\begin{cases} \tilde{\theta}_{s+1} = \dfrac{\sum_{j=1}^{w^2} K\left(\lambda^{-1}\left(z_s - z_j\right)\right) z_{j+1}}{\sum_{j=1}^{w^2} K\left(\lambda^{-1}\left(z_s - z_j\right)\right)} \\ z_{j+1} \in S_{s+1}^+ \text{ and } z_j \in S_s^+ \end{cases} \tag{5.14}$$

The typical choice of the set S_s^+ and S_{s+1}^+ when it is specified by the bilateral CM model is depicted in Fig 5.6. For either unilateral or bilateral the data is simulated by grouping the past data S^+ into a single array according to the nearest up to the farthest pixels.

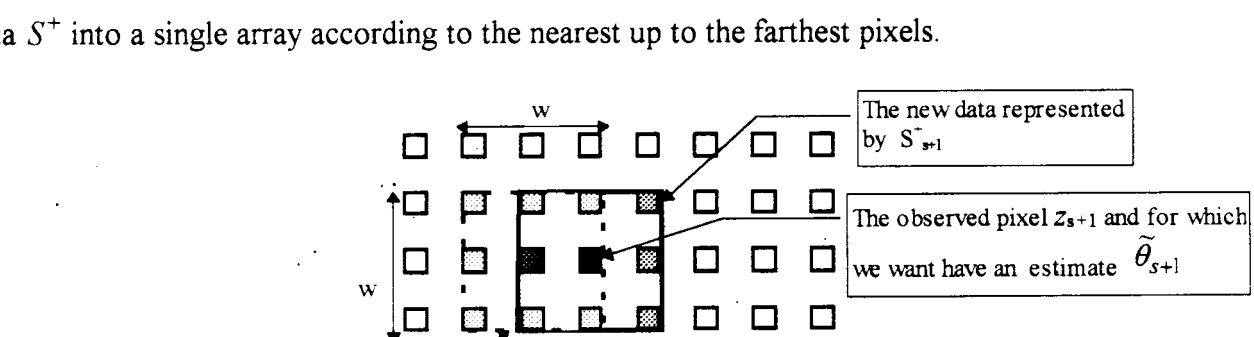


Figure 5.6-Showing the simulation of the bilateral conditional mode predictor.

The unilateral approach has been applied to the noisy image shown in Fig 5.8 which has been obtained by adding Gaussian noise with zero mean and σ=7 to the original image shown in Fig 5.7. The optimal smoothed image, shown in Fig 5.9, was obtained through unilateral conditional Markov specification model by using w=3 and λ =3 for which the relative error rate was 51.89% (see chapter 2 and 4 for definition of the error rate). Fig 5.10 shows, for the same unilateral specification, the case where an over smoothing, i.e. a blurring effect, occurs which was obtained when w=5 and λ =5. This effect is commonly known, in the literature of digital image restoration as a low pass filtering effect.

During the simulation results the choice of the smoothing parameter λ was not found crucial. This is intuitively true since it intervenes only for obtaining a local smooth density function estimates. The functional form of the estimator indicates that the estimated local density function of the « past window » of data is used to force the « future window » of data to have the same distribution. Therefore, the window size w is mostly the one for which the smoothing effect is due whereas the parameter λ has just the effect of smoothing the local density function of the observed noisy pixels z_s's.

We have applied as well the bilateral approach to the same noisy image (Fig 5.8) for the same size of window w and the smoothing parameter λ as in the previous example. The result is shown in Fig 5.11 for which, the relative error rate was found to be 40.31% which is smaller than the one obtained using the unilateral CM model.

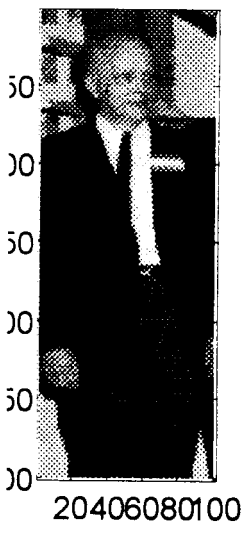


Figure 5.7- Original image 301x101 and 64 gray levels.

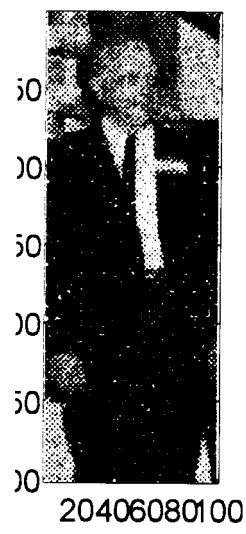


Figure 5.8-Noisy.

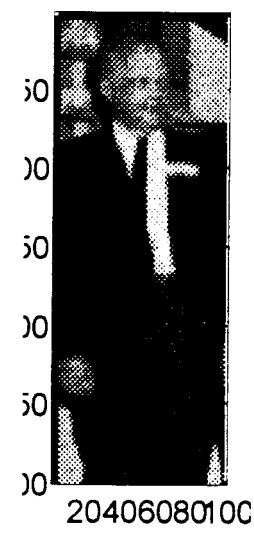


Figure 5.9-Unilateral smoothing.

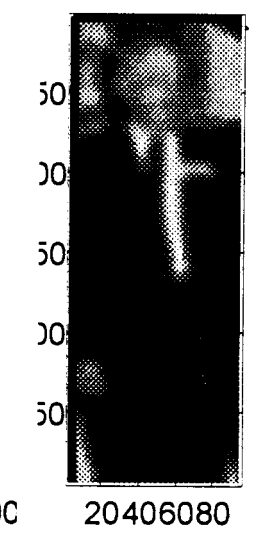


Figure 5.10- Unilateral over smoothed w=5, λ=9.

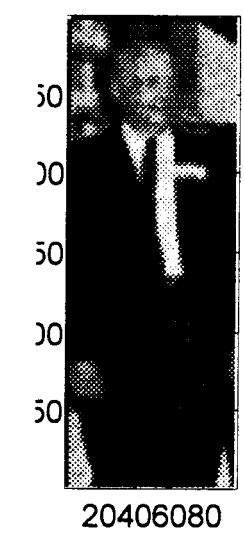


Figure 5.11- Bilateral smoothing

It is worth noting through a visual comparison between Fig 5.9 and Fig 5.11 that the unilateral CM generates a small local patches which are sharper than the ones that are generated by the bilateral CM. For the bilateral CM, these are smoothed and therefore cannot be distinguished.

This shows that it is better to use a bilateral CM model to specify the pixel color neighborhood dependence.

To illustrate the smoothing advantages using the CM specified models, the same settings have been applied to the image shown in Fig 5.12 which represents a bigger scene from which it has been has cut Fig 5.8 with the same characteristics of the noise. The simulation result is shown only for the case of bilateral CM in Fig 5.13 and Fig 5.14 . Fig 5.14 represents a case of over smoothing, with λ=9 and w=5, while the former is the one which has been optimally obtained e.i. λ=3 and w=3. The nonparametric smoothing of noisy images by the CM models has resulted in better performances than the parametric Weiner and the nonparametric regression approaches under the same conditions of noise and original data record statistics.

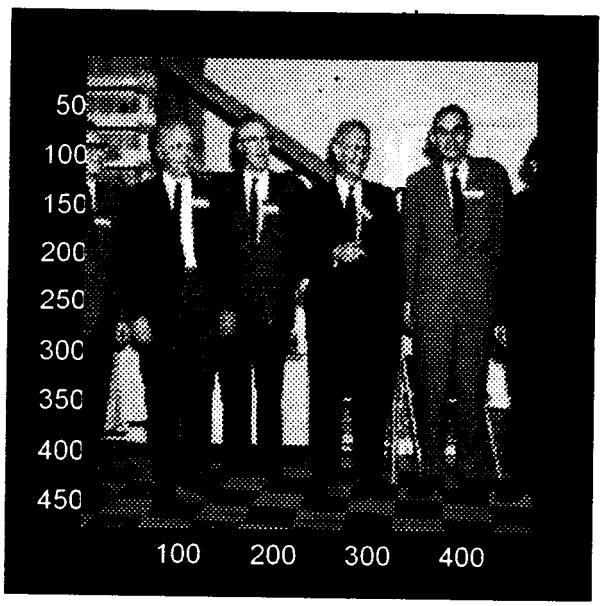


Figure-5.12 The original image.

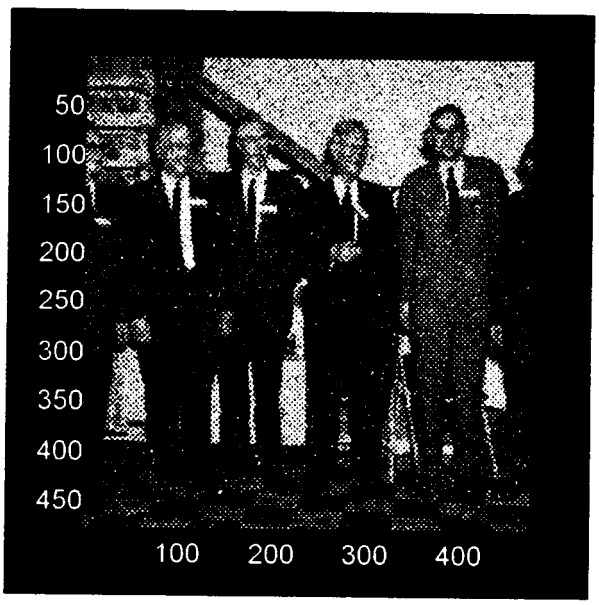


Figure-5.13 The noisy image (a bigger scene of Fig 5.8).

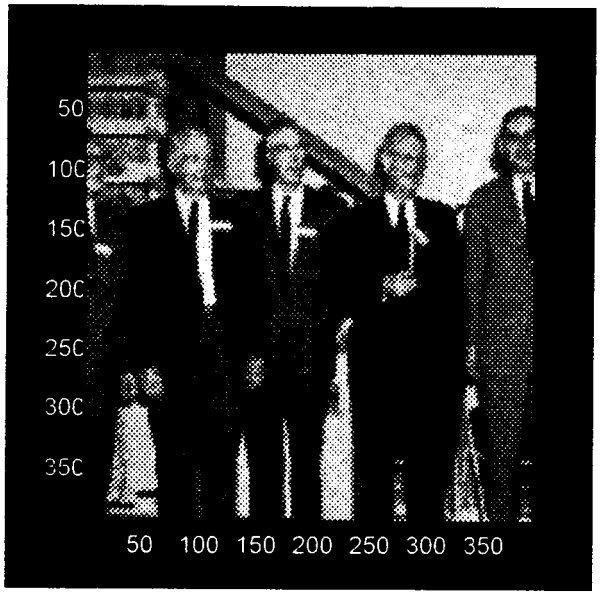


Figure 5.14 Optimally smoothed using the bilateral CM specification (w=3, λ=3).

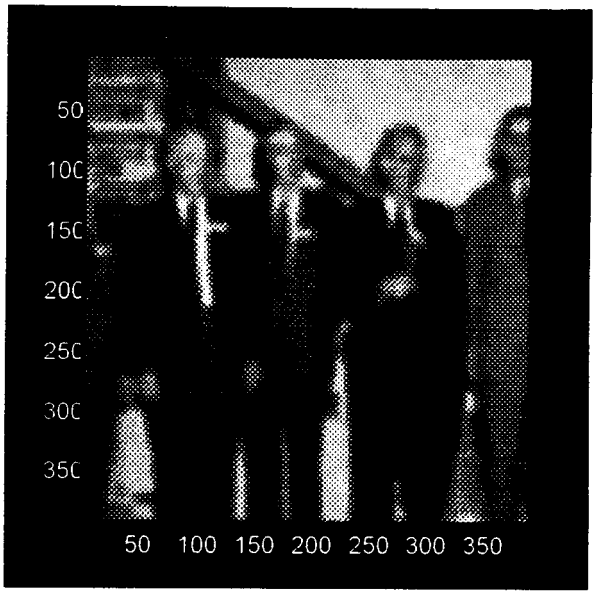


Figure-5.15 Over smoothed image w=5, λ=9.

5.4-NONPARAMETRIC ITERATIVE APPROACH

5.4.1-INTRODUCTION

The previous sections 5.2 and 5.3 have dealt with the specification of an image lattice structure through a unilateral and bilateral conditional Markov models. However, as it was outlined, the lattice structure of an arbitrary image does not meet such specification when the pixel colors do not support the unidirectional propagation via the unilateral or the bilateral modeling. Therefore, another way of departing from a parallel view of time ordering scheme is necessary. Furthermore, if one uses the nonparametric approach since the observed noisy data are the only available source of knowledge, the local spatial smoothing should be mostly constrained by such knowledge.

A more realistic inference (from a nonparametric view of the problem) would involve the elaboration of estimates constructed on two basis: using a Markov specified chain models and including an arrangement that removes the forgetting aspect related to the estimator of the former sections i.e. the CM estimators.

Therefore, the required specification should be constructed through a combination of a structure of Markov chain models with a complement of the observed records via the Baysian framework.

Besag (1986) has proposed a Baysian method to restore noisy images using a combination of nondegenerate Markov random fields of the local representation of the true but unknown scene and a scheme which keeps track of the observed records. The iterative method which he has called the ICM (Iterated Conditional Mode estimator), departs slightly from the simulated annealing, Gemman and Gemman (1984) paradigm by using a synchronous approximation to the conditional probability distribution.

5.4.2-MARKOV RANDOM FIELDS

In the next subsection we discuss Besag(1986) approach to image restoration based on Markov random fields specification of image lattice structure. The true image $\Theta=\{\theta_1,...,\theta_n\}$ is said to be a realization of a locally dependent Markov random fields if the distribution of the coloring at pixel s conditioned on the coloring of the other pixels is the same as that conditioned on a local neighborhood N. In other words, if $p_s(\theta)$ is the probability distribution which assigns colorings to the pixel s of D, then the conditional probability distribution which assigns a color θ_s to pixel s, $p(\theta_s|\theta_{D/s})$ is equal to $p(\theta_s|\theta_{N/s})$.

5.4.3-THE ITERATED CONDITIONAL MODE ESTIMATOR

Given an observed image $Z=\{z_1,...,z_n\}$, the first step consists of restoring the unknown image by using one of the existing method such as the maximum likelyhood classifier. Then the obtained estimate $\Theta^0=\{\theta_1^0,...,\theta_n^0\}$ can be improved iteratively by taking into account the local dependence. In each iteration, the new updated estimate of the color θ_s at pixel s is computed, and it is defined as the color which maximizes the conditional probability given by:

$$p(\theta_s|z_s,\tilde{\theta}_{N/s}) \tag{5.15}$$

where z_s is the observed color and $\tilde{\theta}_{N/s}$ is the estimated color of the neighboring pixels, which had been obtained from the previous iteration.

Using the property of Markov random fields, and Bayes theorem (5.15) can be shown to be proportional to:

$$p(\theta_s|z_s,\tilde{\theta}_{N/s}^t) \propto f(z_s|\theta_s)p_s(\theta_s|\tilde{\theta}_{N/s}^{t-1}) \tag{5.16}$$

where $f(z_s|\theta_s)$ is the conditional density of the color z_s given that the true color is θ_s and $p(\theta_s|\tilde{\theta}_{N/s})$ is the posterior probability that the color at pixel s is θ_s given the previous estimated colors in the neighborhood of s.

Therefore, at each iteration, the estimate $\tilde{\theta}_{N/s}$ is the color k which maximizes:

$$L(s)=f(z_s|k)p_s(k|\tilde{\theta}_{N/s}) \tag{5.17}$$

The iterations are repeated a certain number of times or until convergence.

5.4.4-NONPARAMETRIC KERNEL ITERATED METHOD

The iterative method of Besag (1986) is a parametric method requiring the knowledge of the prior and posterior distributions of the coloring process. Moreover, the updated estimate is based on the information contained in the four immediate neighbors of the pixel of concern.

The ideas behind the Baysian method of Besag (1986) motivated us to tackle the problem of image restoration nonparametricaly as follows:

Assume that $z(s)$ and $\theta(s)$ are the observed, and the true but unknown colors at pixel s respectively. Furthermore, we assume that the colors of the neighboring pixels are identically distributed.

The proposed method is based on the neighboring principle which states that the color $\theta(s)$ is more plausible to be the same color for the remaining neighboring pixels. This suggests that an estimate of the common color θ in a given neighborhood N can be given by the empirical average of the observed colors to cancel out the various noises i.e.:

$$\tilde{\theta}_o = \sum_{s_i \in N} z(s) w_i \tag{5.18}$$

where $w_i = \Pr\left[\theta(s) = z(s) \mid z_{N/s}\right]$.

The weights w_i can be estimated nonparametricaly by the relative frequency of color *z(s)* in the neighborhood N i.e.:

$$w_i = \frac{Nbr\ of\ pixels\ having\ color\ z(s)}{Total\ number\ of\ pixels\ in\ N} \tag{5.19}$$

A more consistent estimator of w_i can be obtained using the kernel method of density estimation as follows:

Using the neighboring principle and the fact that the errors ε_i are i.i.d. in the neighborhood N, then if the color z_s is regarded as a continuos random variable, an estimate of its density can be obtained using the kernel method by:

$$\hat{f}_\lambda(z) = \frac{\sum_{j=1}^{n^2} K\left(\lambda^{-1}\left(z_j - z\right)\right)}{n^2\lambda} \tag{5.20}$$

where n^2 is the number of pixels in N.

This estimator can be refined to take into consideration the situations where the coloring is regarded as a discrete random variable. In this situations, an estimate of w_i will be:

$$\tilde{w}_i = \int_{z(s_i)-1/2}^{z(s_i)+1/2} f_\lambda(z)dz \tag{5.21}$$

It is worth noting that w_i is proportional to $f_\lambda(z)$, when it is unimodal density, which suggests the following estimate:

$$\tilde{\theta}(s) = c \sum_{s_i \in N} z(s_i) f_\lambda\left(z(s_i)\right) \tag{5.22}$$

where c is a normalizing constant given by:

$$c = \frac{1}{\sum_{s_i \in N} f_\lambda\left(z(s_i)\right)}$$

Furthermore, the neighborhood N of pixel s can be chosen to be centered at s and possessing as few as 9 pixels because it is well known that the kernel method of density estimation produces satisfactory estimators when the size of the sample is as small as 6.

We have used this naive estimator, given in (5.22), to smooth the noisy image shown in Fig 5.16. The noisy image has been obtained by adding noise to the original image shown in Fig 5.15 with the same characteristics as in the image of Fig 5.8 of section 5.3. The results are shown from Fig 5.17 to 5.24 for different window size w and smoothing bandwidth λ in the nonparametric density estimator. However, one can notice easily the patching effect in Fig 5.19 (wide patches of the same color). Therefore, the inclusion of the records is essential. Note, if the window size is large the blurring effect increases as illustrated by Fig 5.17. The optimal rate recovery was obtained in Fig 5.24 with error rate of 68% corresponding to λ=3, w=3.

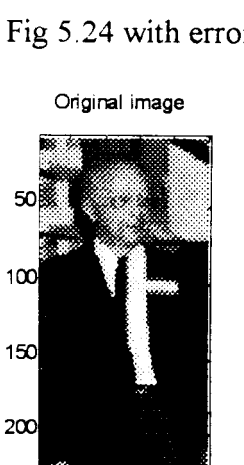


Figure 5.15

Figure 5.16

Figure 5.17

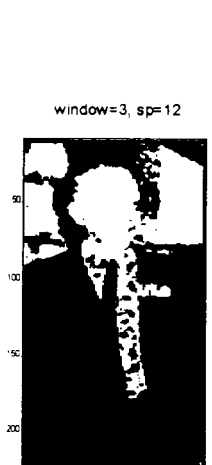


Figure 5.18

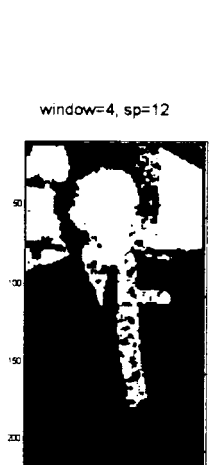


Figure 5.19

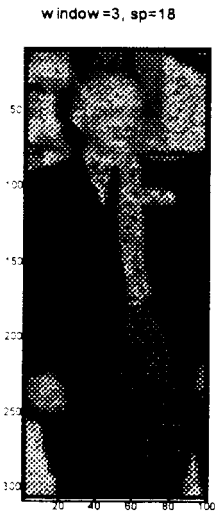


Figure 5.20

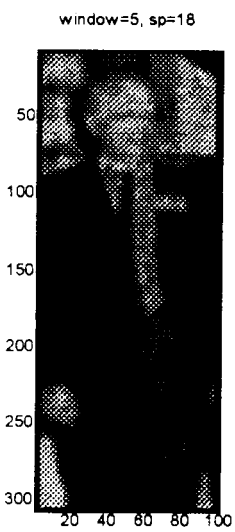


Figure 5.21

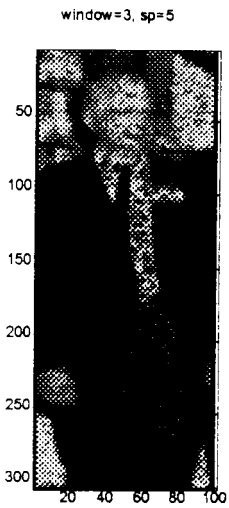


Figure 5.22

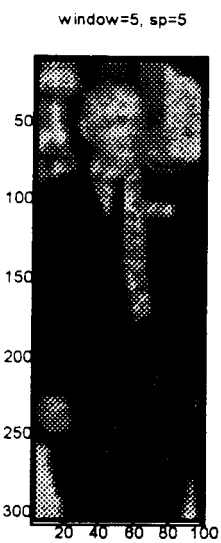


Figure 5.23

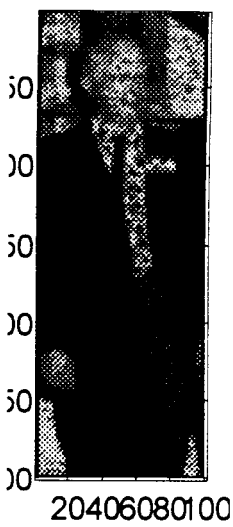


Figure 5.24-Optimal smoothing with error rate of 68%.

The relatively low performance of the estimator $\tilde{\theta}_s$ given by (5.22) is due to the fact that it relies mainly on the observed colors $z(1),...,z(n^2)$ which may be seriously degraded colors.

Following the spirit of the Baysian method of Besag (1986), the estimator $\tilde{\theta}_s$ can be improved iteratively as follows:

The straight forward method to overcome the considerable influence of the observed colors on $\tilde{\theta}_s$ is to update the estimator by replacing $z(s)$, at the new iteration, by the obtained estimate i.e.:

$$\tilde{\theta}_s^t = \sum_{s_i \in N} \tilde{\theta}(s) w_i^t \tag{5.23}$$

where $w_i^t = c \dfrac{\sum_{j=1}^{n^2} K\left(\lambda^{-1}\left(\tilde{\theta}(s_j) - \tilde{\theta}(s_i)\right)\right)}{n^2 \lambda}$

Due to the lack of the convergence property of this iterative approach since the records (colors) are not kept fixed, we use the familiar technique which consists of using the following estimator:

$$\tilde{\theta}_s^{(*)t} = \frac{1}{\gamma+1}\Psi_1^t + \frac{\gamma}{\gamma+1}\Psi_2^t \tag{5.24}$$

where Ψ_1^t and Ψ_2^t are the estimators defined by :

$$\Psi_1^t = \sum_{s_i \in N} z(s_i) w_i^t \text{ and } \Psi_2^t = \sum_{s_i \in N} \theta^{t-1}(s_i) w_i^{(*)t}$$

where w_i^t and $w_i^{(*)t}$ are determined, as described earlier, using the kernel density estimator computed using the samples $\left\{\theta_{s_i}^{t-2},\ s_i \in N\right\}$ and $\left\{\theta_{s_i}^{t-1},\ s_i \in N\right\}$ respectively.

The parameter γ is a control parameter which can be interpreted similarly as the control parameter β of Besag (1986). If it is taken greater than one it indicates that there is a large uncertainty about the observed records *z(s)*. One may garantee the termination of the algorithm by considering the control parameter γ as a decreasing function *γ(t)*.

We illustrated this iterative method on the noisy image given in Fig 5.15. The method has been applied for the same size of window, and smoothing parameter i.e. λ=3 and *w*=3, and γ has been changed from a pass to the next starting from an initial value of 9 to 5. Despite that we have started with the noisy image, the results were very interesting since the method has lead smaller error rates, 30%, than the previous methods of section 5.2 and 5.3 and the one of chapter 4 (Fig 5.32 at the 6th iteration). However, the error started to increase after the 6th iteration. The history of the error rate is depicted in Fig 5.35.

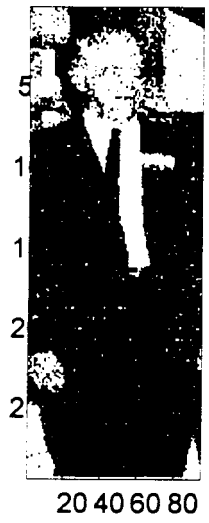


Figure 5.25 Noisy s=7, m=0, 100%.

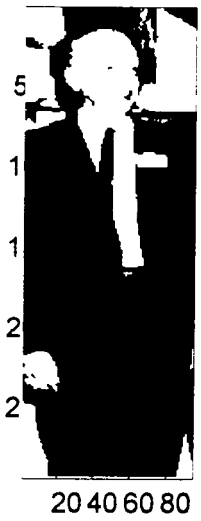


Figure 5.26 Original

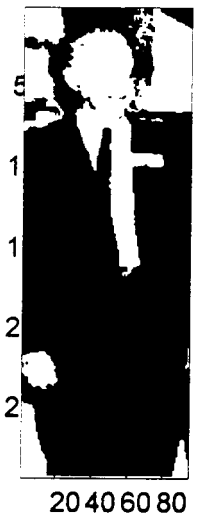


Figure 5.27 1st iteration, 38.54%.

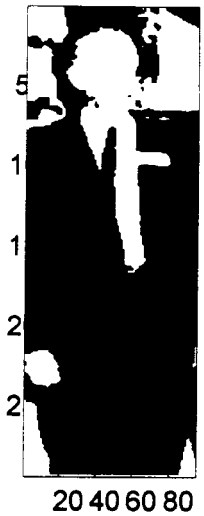


Figure 5.28 2nd iteration, 32.38%.

Figure 5.29 3rd iteration, 31.32%.

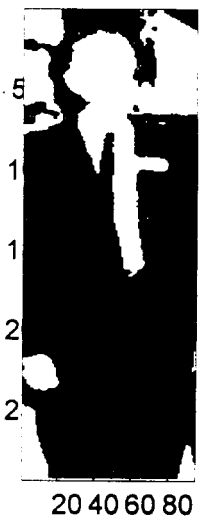


Figure 5.30 4th iteration 30.92%.

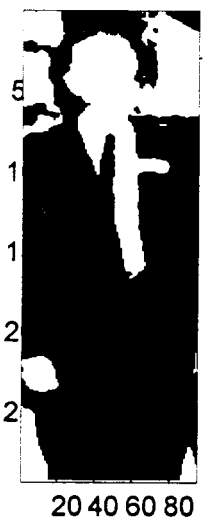


Figure 5.31 5th iteration 30.89%.

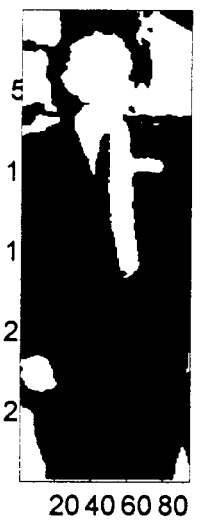


Figure 5.32 6th iteration 30.88%.

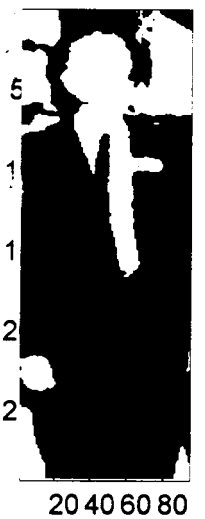


Figure 5.33 7th iteration 31.10%.

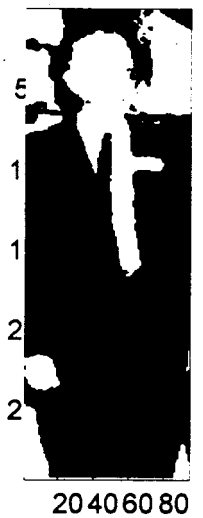


Figure 5.34 8th iteration 36.00%.

The application of the nonparametric iterative method on a wider noisy image from which Fig 5.25 has been cut is shown in Fig 5.36 which has been obtained at the 6th iteration. The obtained

rate of error was found to be equal to 29.58% which is nearly the same as the 30% obtained in Fig 5.32 in account to the edging effect which is smaller for the bigger scene of Fig 5.36.

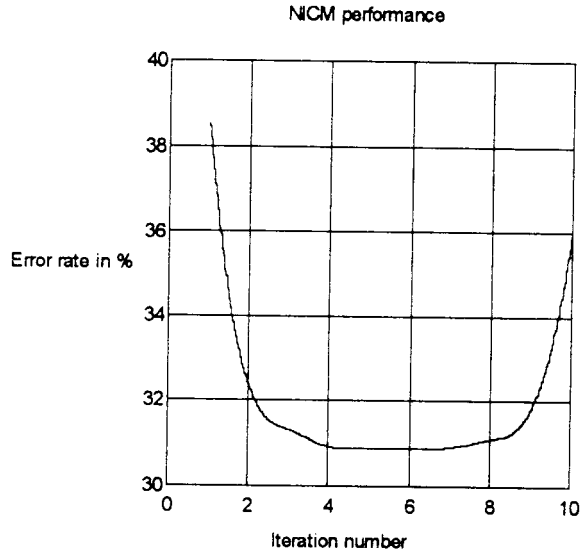


Figure 5.35- Nonparametric iterative method convergence rate performance with respect to iterations.

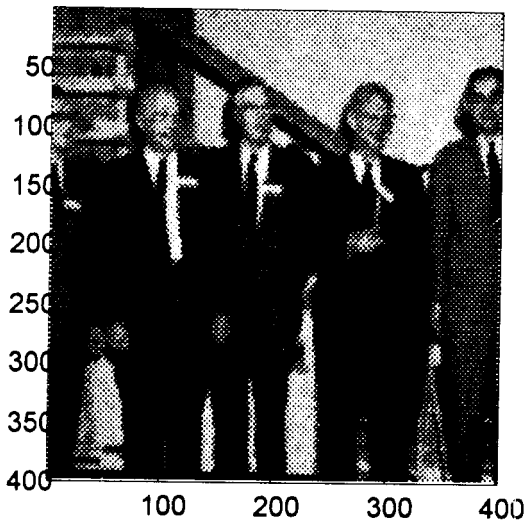


Figure 5.36- A more larger scene of Fig 5.26 at the 6th iteration with error rate equal 29.54%.

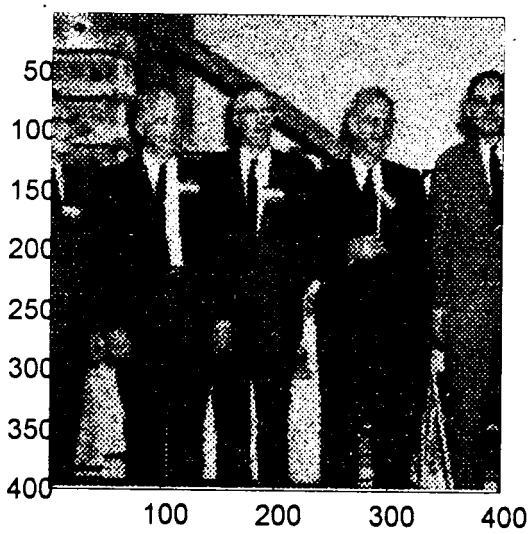


Figure-5.37 The noisy image (a bigger scene of Fig 5.25).

CONCLUSION

Image restoration problem has been extensively treated using sophisticated parametric statistical methods which require a prior knowledge of the model of the degrading process. These methods are optimal when the assumed models are reliable. However, there are many situations where no sufficient information concerning the degrading process is available or it is hard to be modeled. In these situations, the nonparametric approach can be devised to restore noisy images.

In our research work, we have investigated the potential application of the nonparametric methods of estimation to the problem of image restoration. The problem was tackled using the kernel method of regression estimation and the nonparametric kernel method of density estimation.

The regression approach assumes that the unknown intensity function of the original image is a bivariate smooth curve whereas the stochastic approach assumes that the original colors are random variables satisfying the spatial conditional Markov chain property. This led us to suggest two nonparametric restoration methods that are similar to the existing parametric likelihood conditional mode methods.

The application of the three proposed methods on simulated noisy images suggests that they can be regarded as alternative appealing methods when the model of degrading process is unknown. The simulated images show that the assumption of the smoothness of the intensity function was crucial for the nonparametric regression approach which does not perform as good as the nonparametric stochastic approach. However the nonparametric iterative method was found to be the best method but it is time consuming, and involves the choice of a control parameter.

As a further scope, we will investigate the following problems:

- The use of local adaptive optimal smoothing parameter in the regression method to reduce the effect of the smoothness of the estimator in sharp regions.
- Development of data based methods for objective choice of the control parameter to improve the speed of convergence to the optimal solution.

REFRENCES

- **Abramson, I. S.** (1982). *On bandwidth variation in kernel estimates-A square root law.* Ann. Statist.;4, 1217-1223.

- **Andrew, H. C. and Hunt, B. R.** (1977). *Digital image restoration.* TA1632.A5 ISBN 0-13-214213-9. Ed Prentice-Hall, Inc.

- **Atkinson, A. C.** (1980). *A note on the gneralized information criterion for choice of a model.* Biometrika 67, 413-418.

- **Besag, J.** (1986). *On the statistical analysis of dirty pictures.* J. R. Statist. Soc. 48. No. 3 259-302.

- **Besag. J.** (1974). *Spatial interaction and the statistical analysis of lattice systems.* J. Roy. Statist. Soc. B, 36, 192-225.

- **Cacoullos, T.** (1964). *Estimation of multivariate density.* Ann. Math. Statist. 179-187.

- **Chellappa, R. and Kashyap, R. L.** (1982). *Digital restoration using spatial interaction models.* IEEE Trans. Acoust. Speech. Signal. Proc. 3, 461-471.

- **Cheng, K. F. and Lin, P. E.** (1981). *Nonparametric estimation of a regression function.* Z. Wahrsch. Verw. Gebiet 57, 223-233.

- **Coullomb, G.** (1981). *Estimation non-parametric de la regression: revue bibliographique. Internat.* Statist. Rev. 49, 75-93.

- **Craven, P. and Wahba, G.** (1979). *Smoothing noisy data with spline functions: estimating the correct degree of smoothing by method of generalized cross-validation.* Numer. Math. 31,377-403.

- **Cressie. N.**(1991). *Statistics for Spatial data.* QA278.2.C75, ISBN 0-471-84336-9.

- **Deveroye, L. P.** (1978a). *The uniform convergence of the Nadaraya-Watson regression function estimate.* Canada. J. Statist. 6 179-191.

- **Deveroye, L. P.** (1978b). *The uniform convergence of nearest neighbhor regression function estimators and their application in optimization.* IEEE Trans. Inform. Theory IT 24, 124-151.

- **Eubank, R. L.** (1988). *Spline smoothing and nonparametric regression* vol. 90 QA 278.2.E93.

• **FADLI. D. E. and DJEDDI. M.**(1994*).Nonparametric regularization of noise degraded image.* Advances in Modelling & Analysis, B, AMSE Press, Vol.31, N4, 7-13

• **Gasser, Th. and Muller, H. G.** (1979). *Kernel estimation of regression functions. In Smoothing Techniques for Curve Estimatoin* (Th. Gasser and M. Rosenblatt, eds), 23-68.

• **GEMAN. S. and GEMAN. D.**(1984). Stochastic relaxation, Gibbs distributions and the Baysian restoration of images. IEEE Transactions on pattern analysis and machine intellegence, PAMI-6, 721-741.

• **Graybill, F. A.** (1976). *Theory and application of the linear model.* North Scituate, Mass: Duxbury Press.

• **Goodman, J. W.**(1968). *Introduction to Fourier optics.* McGraw-Hill., New York.

• **Hall, P. and Titterington, D. M.** (1986). *On some smoothing techniques used in image restoration.* J. R. Statist. Soc. B 48, No. 3 330-343.

• **Hall, P. and Marron, J. S.** (1988). *Choice of kernel order in density estimation.* Ann. Statist., vol 16, No.1, 161-173.

• **Hardle, W. and Bowman, A. W.** (1988). *Bootstrapping in nonparametric regression: Local adaptive smoothing and confidence bands.* Amm. Statist. Ass. 83, 103-110.

• **Hardle, W. and Marron, J. S. (1985a).** *Asymptotic nonequivalence of some bandwidth selectors in nonparametric regression.* Biometrika 72, 481-484.

• **Hardle, W., Hall, P. and Marron, S.** (1985). How far are automatically chosen regression smoothing parameters from their optimum. Mimeo Series #1589, Departement of statistics, North carolina State Univ.--Chapel Hill.

• **Heikkinen, J. and Hogmander, H.** (1994). *Fully Baysian approach to image restoration with an application in biography.* Appl. Statist. 43, No. 4, 569-582.

• **HRMMERSLEY. J. M. and CLIFFORD. P.**(1971). *Markov fields on finite graphs and lattices.*

• **Huber, P. J.** (1964). *Robust estimation of a location parameter.* Ann. Math. Statist. 35, 73-101.

• **Huang, T. S.** (1966). *Some notes on Film-Grain-Noise.* NSF Summer Study Report, Woods Hole, Massachusetts, 105-109.

• **Hunt, B. R. and Breedlove, J.**(1975). *Scan and display consideration in processing images by digital computer.* IEEE Trans. Computers, Vol. C-24., 848-853.

• **Lunberger, D. G.** (1969). *Optimization by vector space methods.* John Wiley: New York.

- **Mallows, C. L.** (1973). *Some comments on C_p. Technometrics* 15, 661-675.

- **Mees, C. E. K.** (1954). *The theory of photographic process.*, McMillan Co, New York.

- **Muller, H. G. and Stadmuller, U.** (1987*). Variable bandwidth kernel estimators of regression curves.* Ann. Statist.; 1, 182-201.

- **Nadaraya, E. A.** (1964). *On estimating regression.* Theor. Probab. Appl. 9, 141-142.

- **Nussbaum, M.** (1985). Spline smoothing in regression models and asymptotic efficiency in L_2. Ann. Statist. 13, 984-997.

- **Parzen, E.** (1962). *On estimation of a probability density function and mode.* Ann. Math. Statist. 33, 1065-1076.

- **Peyrovian, M. J. and Sawchuk, A. A.** (1977). *Restoration of noisy blurred images by smoothing spline filter.* Appl. Optics. Vol. 16, No. 12. 3147-3153.

- **Priestly, M. B. and Chao, M. T.** (1972). *Nonparametric function fitting.* J. Roy. Statist. Soc. B 34, 385-392.

- **Reinsch, C. H.** (1967). *Smoothing by splinefunctions.* Numer. Math. 10, 177-183.

- **Rosenblatt, M. (1956).** *Remarks on some nonparametric estimates of density function.* Ann. Math. Statist. 27, 832-835.

- **Rosenblatt, M.** (1969). *Conditional probability density and regression estimators.* In multivariate Analysis II (P. Krishnaih, ed.), 25-31. New york: Acadimic Press.

- **Rosenblatt, M.** (1970). *Density estimates and Markov sequences.* In nonparametric techniques in statistical inference (M. L. Puri, ed.), 199-210. Oxford: Cambridge University Press.

- **Roussas G. G.** (1969). *Nonparametric estimation of the transition function of a Markov process.* Ann. Math. Statist. Vol 40, no. 4, 1386-1400.

- **Schuster, E. F.**(1972). Joint asymptotic distribution of the estimated regression function at a finite number of distinct points. Ann. Math. Statist. 43, 84-88.

- **Searle, S. R.** (1971). *Linear models.* New York: John Wiley.

- **Silverman, B. W.** (1982). *Kernel density estimation using the fast Fourier transform.* Jorn. Roy. Statist. Soc. AS176.

- **Silverman, B. W. and Young G. A.** (1987). *The bootstrap: to smooth or not to smooth?.* Biometrica, 74, 469-479.

- **Speckman, P.** (1985). *Spline smoothing and optimal rates of convergence in nonparametric regression models*. Ann. Statist. 13, 970-983.

- **Stone, C. J.** (1975). *Nearest neighbor estimators of a nonlinear regression function*. In Proc. Computer Sci. Statist. 8th Ann. Symp. Interface, 413-418.

- **Swanepoel J. H.** (1986). *On the construction of nonparametric density function estimators using the bootstrap*. Commun. -Theor. Meth., 15(5), 1399-1415.

- **Taylor C. C.** (1989). *Bootstrap choice of the smoothing parameter in kernel density estimation*. Biometrika. 76, 705-712.

- **Titterington, D. M.** (1985). *Common structure of smoothing techniques in statistics*. Internat. Statist. Rev. 53, 141-170.

- **Van Bladel, J.** (1964). *Electromagnetic fields*. McGraw-Hill Co., New York.

- **Wahba, G.** (1984a). *Cross-validated spline methods for the estimation of mutivariate functions from data on functionals*. In Statist, Ames: Iowa State Univ Press; 205-235.

- **Walkup, J. F. And Cohens, R. F. (1974).** *Image Processing in signal dependent noise*. Optical Engineering, vol.13, 258-266.

- **Watson, G. S.** (1964). *Smooth regression analysis*. Sankhya A 26, 359-372.

- **Wegman, E. J.** (1972). *Nonparametric probability density estimation: I. A summary of available methods*. Technometrics 14, 513-546.

- **Woodford, C. H.** (1970). *An algorithm for data smoothing using splines functions*. BIT, 10, 501-510.

- **Yakowitz, S. J.** (1985). *Nonparametric density estimation, prediction, and regression for markov sequences*. J. Amer. Statist. Assoc. 80, 215-221.

APPENDIX A

As discussed in chapter 2, we will give in this appendix a short description of the imagery system and the computing environment that we have used to carry on the simulation study of the previous chapters. In the next appendix we will see the software environment of the algorithmic implementation of the nonparametric estimators discussed in the previous chapters.

The computer representation and storage of a digital image are handled in the hardware environment of the host computer respectively by the graphic card and the hard disk. The storage ability of the host computer is related to the storage of the hard disk. The maximum resolution in the number of pixels and the possible maximum number of the gray level shades are dependent on the installed graphic card. The processing power of the host computer is related to the type of the Central Processing Unit or CPU and to the size of the available size of the base memory, and if it exists, to the extended memory.

Our computing environment is a PC-IBM compatible computer with the main characteristics listed in table a.1

Type	CPU	base memory	extended memory	speed	Hard disk capacity	video RAM	Graphic card resolution	Max. number of colors
IBM PS/1	INTEL 80446DX2	1024 Kb	7Mb	50MHz	340Mb	2Mb	1024x840 pixels	2^{24} colors

Table a.1

The host computer includes as well extension slots that are useful for incorporating extra needed digital modules such as data acquisition cards. Among such cards that we have used to capture real gray level images is the simple card called VIDI-PC card. The card is provided with a software driver which controls many parameters during the acquisition of the video signal that may come from either video cassette recorder or a CCD camera. The video signal that comes to the video acquisition card input is simply the video signal that gets out of the video IF stage of the commercial TV sets.

The video acquisition card is essential because the video data should be sampled first into digital data before it can be handled by the computer. Basically, the video acquisition card is constructed by three main electronic circuitry parts as depicted in Fig a.1. The composite video separation part

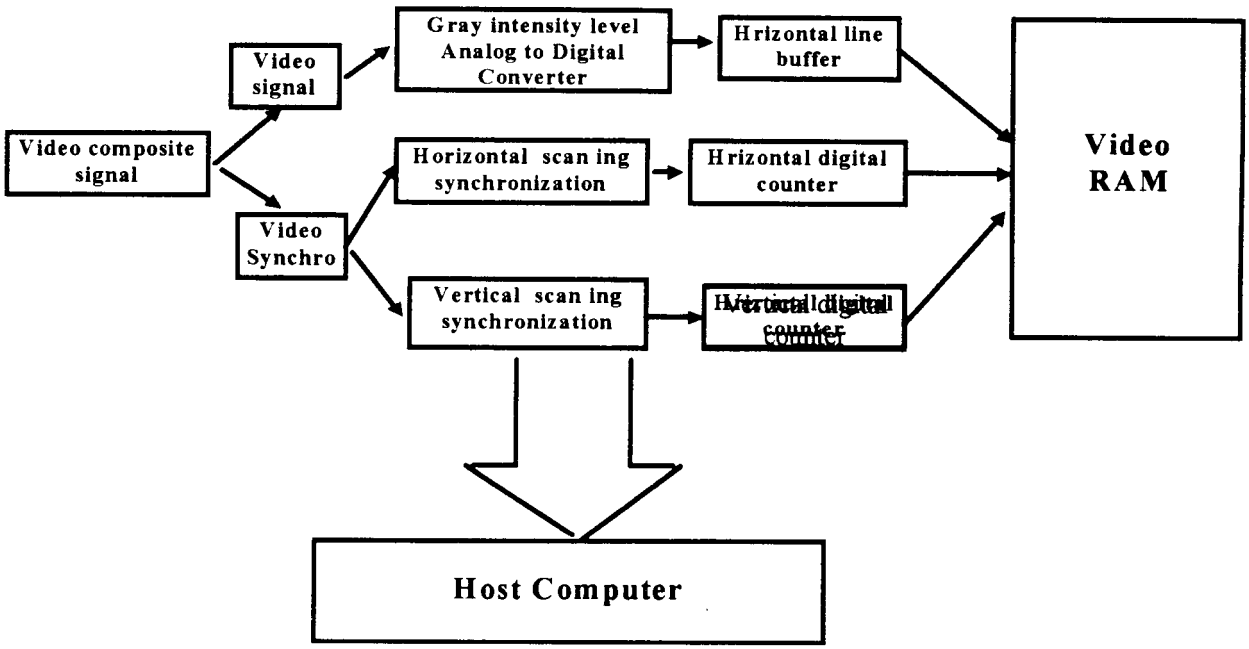


Fig-a.1 A typical computer based digital image acquisition system.

isolates the synchro signal from the video, the video Digital to Analog Converter, and the video memory. The VIDI-PC card is able to capture a digital image of up to 256x256 pixels with 16 gray levels. Therefore, the type of images that we have dealt with belong to the continuous gray level class images and correspond to the visible light, commonly called « black and white » images. The possible gray level tones that can be represented by the VIDI-PC video acquisition card are shown in Fig-a.2.

A real image that have been grabbed using this acquisition card is shown in Fig-a.3. Typical gray level values (which are not corrupted with noise) which correspond to the region bounded by the square in Fig-a3 are listed in table-a.2.

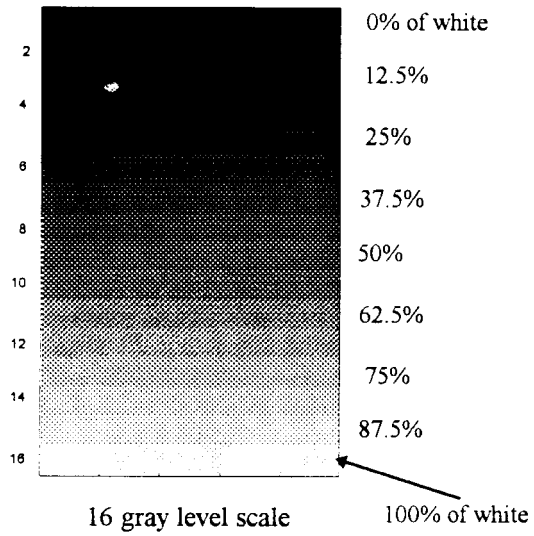


Fig a.2- Gray level scale of the VIDI-PC card

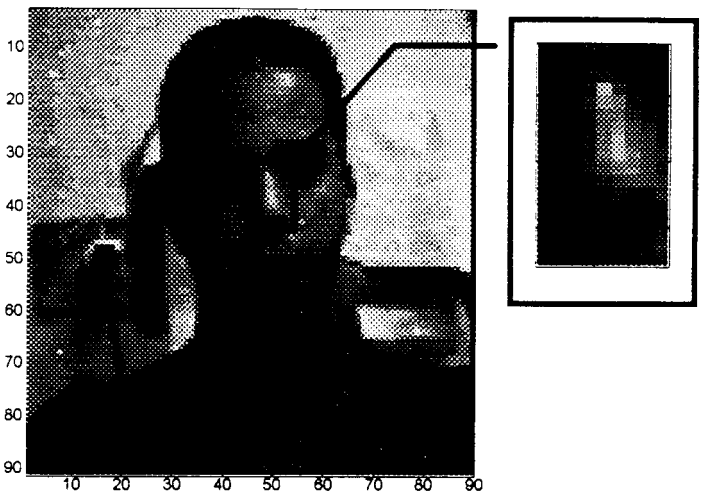


Fig-a.3 A real image of 89x89 pixels and 16 gray levels.

When the digital image acquisition environment is noisy the values which are listed in Table-a.2 will be modified. Table a.3 lists these values when they are corrupted with an additive Gaussian noise of zero mean and unity variance.

	44	45	46	47	48	49	50	51	52
29	2	3	3	6	5	4	3	2	3
30	2	2	3	5	5	5	4	3	3
31	2	3	4	7	7	6	5	3	2
32	3	2	4	7	13	8	6	4	3
33	3	2	5	8	12	11	7	5	5
34	3	3	4	7	11	11	8	6	7
35	4	4	4	7	11	12	10	8	7
36	4	4	5	7	11	13	10	9	7
37	4	4	5	7	9	11	11	9	7
38	4	3	5	6	7	8	9	9	8
39	4	4	4	4	5	6	6	5	8
40	4	4	3	3	4	4	4	4	5
41	4	3	3	2	3	3	2	2	2
42	3	3	2	2	2	3	3	3	7
43	3	3	3	3	4	4	5	6	8
44	4	4	3	3	5	5	5	6	7
45	4	5	3	1	5	6	4	7	7

Table-a.2 the part the part that have been cut from the image in Fig-a.3.

	44	45	46	47	48	49	50	51	52
29	5	3	2	7	6	7	6	2	3
30	0	5	4	5	2	3	5	3	2
31	1	3	5	9	7	7	4	3	2
32	2	-1	6	8	12	9	3	6	2
33	1	5	6	5	12	11	5	3	3
34	3	1	2	6	11	12	11	8	9
35	5	3	5	9	11	15	12	4	6
36	5	4	10	6	13	12	12	9	3
37	3	8	4	4	16	16	13	12	10
38	2	6	4	6	10	10	4	9	8
39	4	2	4	6	8	5	6	8	8
40	4	6	6	3	11	4	4	5	5
41	7	2	6	3	2	3	0	7	5
42	3	3	5	5	4	4	4	3	-1
43	2	3	2	3	3	4	5	4	8
44	6	6	1	2	9	2	3	7	11
45	3	6	3	2	6	4	5	9	8

Table a.3 Noisy values.

APPENDIX B

The programming environment is provided by the MATLAB software which is suitable for matrix manipulations. The menu commands is depicted in the figure shown in Fig b.1.

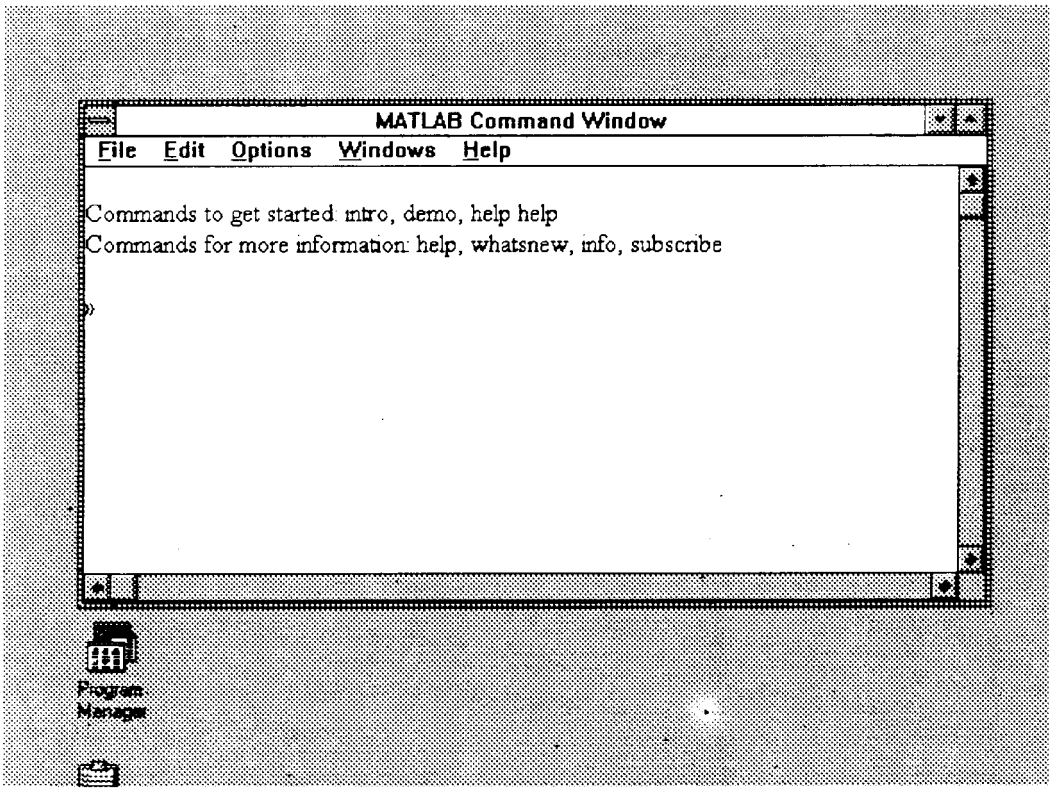


Fig b.1-MATLAB Ver. 4.0 for windows menu command

The original image which may be stored in an ASCII file can be read easily by the ***load*** command. The normal noise can be generated by the command ***randn***(x,y) where x and y are respectively the desired horizontal and vertical sizes of the noise matrix.

Recall from chapter 2 that the model of the noise degrading process is additive which can be written in matrix notations as:

$$G=F+n$$

In the matlab environment the above process can be simulated by a set of few instructions as:

```
» load F;
» n=randn(size(F));
» G=F+n;
» save G G;
```

The above commands provides at once the loading of the original matrix image F into a memory, generate a normal noise matrix n of zero mean and variance equal to 1, produce an observed image G from F with the additive noise n, and save it in a file called G..

Thus the MATLAB software implements numerically in a straightforward way what one may write mathematically as expressions. For instance, a vector of time can be generated quickly as follows:

»t=0:.01:1;

That is a raw vector of 100 elements starting from zero to 1 with step 0.01. If one is interested to generate a sine wave signal on the above time axis t with frequency f=5, he simply should type the following:

»f=5;

»y=sin(2*pi*f*t);

To plot the result, one just have to type the following:

»plot(t,y)

and the result gets out to look as Fig-b.2.

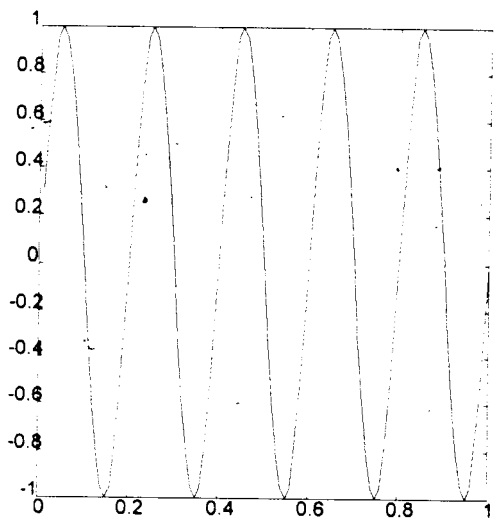


Fig-b.2-Plot command demo

The software matlab that we have used is version 4.0 of MATHWORKS INC which runs on windows a ıd take advantage of the available memory. This version is suitable for the digital image processing which comes ,vith an accompanying set of ready sampled images such as the famous digital image of baboon.

For example, if the covariance matrices of Qf and Qn, seen in chapter 2, of the original image and of the oise are available then the Weiner filtering approach can be implemented as few as the following instructions:

```
»FF=fft2(Qf);
»Pf=FF.*conj(FF);
»Fn=fft2(Qn);
»Pn=Fn.*conj(Fn);
»F_estimate=real(IFFT2(Pn/(Pn+PF).*FF));
```

The display of images can be performed by using the two basic commands ***colormap***(map) and ***image***(F_estimate) where map is nx3 matrix whose 3 columns correspond to the red, blue and green intensities for color images and n represents the numbers of equally spaced shade intensities. When the same vector is used for the 3 columns one may obtain gray level intensities images.

To show the straightforward implementation of the kernel regression estimator on the observed image G with λ=0.07, see chapter 4, we use the following

```
» % 17:53  19/12/94
» % By Fadli
»load G
»nsample=n+a;
»img07=kerest(G,.07);
»save img07 img07
```

where kerest is a an M-file whose content is:

```
% done at 4:52PM  12/9/94
% by fadli djamel-eddine.
% this is the fast quadratic kernel estimator
% the use is cestimated image =kerest(noisy image,
% smoothing parameter)c
function img=kerest(noimg,lambda)
size1=size(noimg);
weight=quadrat(size1(1),lambda);
size2=round(max(size(weight)/2));
size3=prod(size1);
cte=size3*lambda^2;
img=zeros(size1);
for k=size2:size1(1)-size2,
  for l=size2:size1(1)-size2,
        temp1=noimg(k-size2 + 1:k+size2-1 ,l-size2+1: l+ size2-1);
        if max(size(weight))==1
                temp=noimg(k,l);
        else

                temp=weight*temp1*weight'/(cte);

        end

        img(k,l)=temp;
  end
end
%ended for coding at 5:20PM  12/9/94
%debugged for algorithmic error at10:38AM  12/11/94
```

and where quadrat is the M-file implementation of the quadratic kernel as:

```
function weight1=quadrat(size1,lambda)
t=-1:1/size1:1;
t1=t/lambda;
clear t
weight=zeros(size(t1));
axis1=find((abs(t1)>=.5 & abs(t1)<=1));
weight(axis1)=(8/3)*(1-abs(t1(axis1))).^3;
clear axis1
axis2=find(abs(t1)<.5);
weight(axis2)=(4/3) - 8*t1(axis2).^2 + 8*abs(t1(axis2)).^3;
clear axis2
axis1=find(weight > 0);
weight1=weight(axis1);
```

ANNEXE

Liste et composition du jury en vue de la soutenance de la these de magister en ***Ingenierie des Systèmes Electroniques.***
Par ***Mr FADLI Djamel-Eddine Tayeb***

PRESIDENT / Dr MAAFI Abdelbaki, Docteur d'Etat (Maitre de Conferences a L'U.S.T.H.B)

RAPPORTEUR/ Dr DJEDDI Mohamed, Docteur d'Etat (Charge de recherche à L'I.N.E.L.E.C)

CO-RAPPORTEUR/ Mr BOUKLACHI Abbes, Master (Maitre Assistant a L'I.N.E.L.E.C)

EXAMINATEURS/ Dr BENTARZI Mohamed, Docteur d'Etat (Charge de cours à L'U.S.T.H.B)

Dr HARICHE Kamel, PhD (Charge de recherche a l'I.N.E.L.E.C)