Patrizio Campisi, Karen Egiazarian

Blind image deconvolution: theory and applications

CRC PRESS
Boca Raton London New York Washington, D.C.

Contributors

Dan R. Adam
Department of Biomedical Engineering, Technion, Haifa, Israel.

Bruno Amizic
Department of EECS, Northwestern University, Evanston, IL, USA.

Jaakko Astola
Institute of Signal Processing, Tampere University of Technology, Tampere, Finland.

S. Derin Babacan
Department of EECS, Northwestern University, Evanston, IL, USA.

Tom E. Bishop
School of Engineering and Electronics, The University of Edinburgh, Edinburgh, Scotland, UK.

Laure Blanc-Féraud
INRIA, Sophia-Antipolis, France.

Patrizio Campisi
Dipartimento di Elettronica Applicata, Università degli Studi "Roma TRE," Roma, Italy.

Tony Chan
UCLA Department of Mathematics, Los Angeles, CA, USA.

Stefania Colonnese
Dipartimento INFOCOM, Università degli Studi di Roma "La Sapienza," Roma, Italy.

Gabriel Cristóbal
Instituto de Óptica Consejo Superior de Investigaciones Científicas, Madrid, Spain.

Karen Egiazarian
Institute of Signal Processing, Tampere University of Technology, Tampere, Finland.

0-8493-0052-5/00/$0.00+$.50

Jan Flusser
Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic, Prague, Czech Republic.

Nikolas P. Galatsanos
Department of Computer Science, University of Ioannina, Ioannina, Greece.

Alexia Giannoula
Department of Electrical and Computer Engineering, University of Toronto, Toronto, Canada.

Jianxin Han
Department of Electrical and Computer Engineering, University of Toronto, Toronto, Canada.

Dimitrios Hatzinakos
Department of Electrical and Computer Engineering, University of Toronto, Toronto, Canada.

André Jalobeanu
PASEO research group, MIV team (LSIIT UMR 7005 CNRS-ULP), Illkirch, France.

Vladimir Katkovnik
Institute of Signal Processing, Tampere University of Technology, Tampere, Finland.

Aggelos K. Katsaggelos
Department of EECS, Northwestern University, Evanston, IL, USA.

Aristidis Likas
Department of Computer Science, University of Ioannina, Ioannina, Greece.

Oleg V. Michailovich
Department of Electrical and Computer Engineering, University of Alberta, Edmonton, Canada.

Rafael Molina
Departamento de Ciensas de la Computacion e I. A., Universidad de Granada, Spain.

Fionn Murtagh
School of Computer Science, Queen's University Belfast, Belfast, Northern Ireland, Observatoire de Strasbourg, Universit Louis Pasteur, Strasbourg,

France.

Alessandro Neri
Dipartimento di Elettronica Applicata, Università degli Studi "Roma TRE," Roma, Italy.

Michael K. Ng
Department of Mathematics, Hong Kong Baptist University, Hong Kong.

Gianpiero Panci
Dipartimento INFOCOM, Università degli Studi di Roma "La Sapienza", Roma, Italy.

Eric Pantin
Service d'Astrophysique, SAP/SEDI, CEA-Saclay, France.

Robert J. Plemmons
Department of Mathematics and Computer Science, Wake Forest University, Winston-Salem, NC, USA.

Gaetano Scarano
Dipartimento INFOCOM, Università degli Studi "La Sapienza," Roma, Italy.

Filip Šroubek
Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic, Prague, Czech Republic.

Jean-Luc Starck
Service d'Astrophysique, SAP/SEDI, CEA-Saclay, Gif-sur-Yvette Cedex, France.

Josiane Zerubia
Ariana research group (INRIA-I3S), Sophia Antipolis, France

Contents

0-8493-0052-5/00/$0.00+$.50

1

Blind Image Deconvolution: problem formulation and existing approaches

Tom E. Bishop
School of Engineering and Electronics, The University of Edinburgh, UK
e-mail: t.e.bishop@ed.ac.uk

S. Derin Babacan, Bruno Amizic, Aggelos K. Katsaggelos
Department of EECS, Northwestern University, Evanston, IL, USA
e-mail: (sdb, amizic)@northwestern.edu, aggk@eecs.northwestern.edu

Tony Chan
Deptartment of Mathematics, UCLA, Los Angeles, CA, USA
email: chan@math.ucla.edu

Rafael Molina
Departamento de Ciensas de la Computacion e I. A., Universidad de Granada
18071 Granada, Spain
e-mail: rms@decsai.ugr.es

1.1 Introduction

Images are ubiquitous and indispensable in science and everyday life. Mirroring the abilities of our own human visual system, it is natural to display observations of the world in graphical form. Images are obtained in areas

0-8493-0052-5/00/$0.00+$.50

ranging from everyday photography to astronomy, remote sensing, medical imaging, and microscopy. In each case, there is an underlying object or scene we wish to observe; the *original* or *true* image is the ideal representation of the observed scene.

Yet the observation process is never perfect: there is uncertainty in the measurements, occurring as blur, noise, and other degradations in the recorded images. Digital image restoration aims to recover an estimate of the original image from the degraded observations. The key to being able to solve this *ill-posed* inverse problem is proper incorporation of prior knowledge about the original image into the restoration process.

Classical image restoration seeks an estimate of the true image assuming the blur is known. In contrast, *blind* image restoration tackles the much more difficult, but realistic, problem where the degradation is unknown. In general, the degradation is nonlinear (including, for example, saturation and quantization) and spatially varying (non uniform motion, imperfect optics); however, for most of the work, it is assumed that the observed image is the output of a Linear Spatially Invariant (LSI) system to which noise is added. Therefore it becomes a Blind Deconvolution (BD) problem, with the unknown blur represented as a Point Spread Function (PSF).

Classical restoration has matured since its inception, in the context of space exploration in the 1960s, and numerous techniques can be found in the literature (for recent reviews see [1, 2]). These differ primarily in the prior information about the image they include to perform the restoration task. The earliest algorithms to tackle the BD problem appeared as long ago as the mid-1970s [3, 4], and attempted to identify known patterns in the blur; a small but dedicated effort followed through the late 1980s (see for instance [5, 6, 7, 8, 9]), and a resurgence was seen in the 1990s (see the earlier reviews in [10, 11]). Since then, the area has been extensively explored by the signal processing, astronomical, and optics communities. Many of the BD algorithms have their roots in estimation theory, linear algebra, and numerical analysis.

An important question one may ask is why is BD useful? Could we not simply use a better observation procedure in the first place? Perhaps, but there always exist physical limits, such as photonic noise, diffraction, or an observation channel outside of our control, and often images must be captured in suboptimal conditions. Also there are existing images of unique events that cannot be retaken that we would like to be able to recover (for instance with forensics or archive footage); furthermore in these cases it is often infeasible to measure properties of the imaging system directly. Another reason is that of cost. High-quality optics and sensing equipment are expensive. However, processing power is abundant today and opens the door to the application of increasingly sophisticated models. Thus BD represents a valuable tool that can be used for improving image quality without requiring complicated calibrations of the real-time image acquisition and processing system (i.e., in medical imaging, video conferencing, space exploration, x-ray imaging, bioimaging, and so on).

The BD problem is encountered in many different technical areas, such as astronomical imaging [12, 13], remote sensing [14], microscopy [15], medical imaging [16], optics [17, 18], photography [19, 20], superresolution applications [21], and motion tracking applications [22], among others.

For example, astronomical imaging is one of the primary applications of BD algorithms [12, 13]. Ground-based imaging systems are subject to blurring due to the rapidly changing index of refractions of the atmosphere. Extraterrestrial observations of the Earth and the planets are degraded by motion blur as a result of slow camera shutter speeds relative to the rapid spacecraft motion.

BD is used for improving the quality of the Poisson distributed film grain noise present in blurred x-rays, mammograms, and digital angiographic images. In such applications, most of the time the degradations are unavoidable because the medical imaging systems limit the intensity of the incident radiation in order to protect the patient's health [23].

In optics, BD is used to restore the original image from the degradation introduced by a microscope or any other optical instrument [17, 18]. The Hubble Space Telescope (HST) main mirror imperfections have provided an inordinate amount of images for the digital image processing community [12].

In photography, depth-of-field effects and misfocusing frequently result in blurred images. Furthermore, motion blur and camera shake are also problems during long exposures in low lighting. The parameters describing these processes are generally unknown; however, BD can enable restoration of these images [19, 20].

As a final example, in tracking applications the object being tracked might be blurred due to its own motion, or the motion of the camera. As a result, the track is lost with conventional tracking approaches and the application of BD approaches can improve tracking results [22].

In the rest of this chapter, we survey the field of blind image deconvolution. In conducting this review though, we develop and present most of the techniques within the Bayesian framework. Whilst many methods were originally derived by other means, this adds consistency to the presentation and facilitates comparison among the different methods. Even today, due to the difficulty of simultaneously estimating the unknown image and blur in the presence of noise, the problem still provides a very fertile ground for novel processing methods.

The chapter is organized as follows. In Section 1.2 we mathematically define the BD problem. In Section 1.3 we provide a classification of the existing approaches to BD. In Section 1.4 we formulate the BD problem within the Bayesian framework, and in Section 1.5 we survey the probabilistic models for the observation, the original image, the blur, and their associated unknown parameters. In Section 1.6 we discuss solutions to the BD problem as inference models under the Bayesian framework. Finally, in Section 1.7 we briefly review BD models which appeared in the literature and cannot be easily obtained from the Bayesian formulation. Conclusions are presented in Section 1.8.

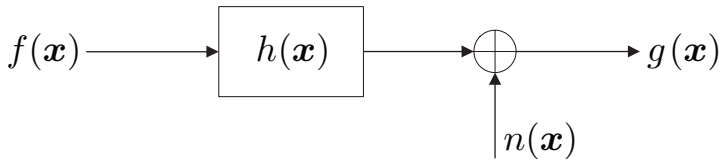

FIGURE 1.1: Linear space invariant image degradation model.

1.2 Mathematical Problem Formulation

In digital image processing, the general, discrete model for a linear degradation caused by blurring and additive noise is given by

$$g(\mathbf{x}) = \sum_{\mathbf{s}\in S_h} h(\mathbf{x},\mathbf{s}) f(\mathbf{s}) + n(\mathbf{x}), \qquad \mathbf{x} = (x_1, x_2) \in S_f \tag{1.1}$$

where $f(\mathbf{x})$, $g(\mathbf{x})$, $h(\mathbf{x},\mathbf{s})$, and $n(\mathbf{x})$ represent the original image, the observed image, the blur or PSF, and the observation noise, respectively, $S_f \subset \mathbb{R}^2$ is the support of the image, and $S_h \subset \mathbb{R}^2$ is the support of the PSF. The additive noise process $n(\mathbf{x})$ may originate during image acquisition, processing, or transmission. Common types of noise are electronic, photoelectric, film grain, and quantization noise. It is common to assume $n(\mathbf{x})$ is White Gaussian Noise (WGN), uncorrelated with the image (although certain types of noise may in practice be signal dependent).

The general objective of blind restoration is to estimate $\mathbf{f}$ and $\mathbf{h}$. The difficulty in solving this problem with a spatially varying blur motivates the use of a space-invariant model for the blur. This leads to the following expression for the degradation system:

$$g(\mathbf{x}) = (f * h)(\mathbf{x}) + n(\mathbf{x}) = \sum_{s\in S_h} h(\mathbf{x}-\mathbf{s}) f(\mathbf{x}) + n(\mathbf{x}), \tag{1.2}$$

where the operator $(*)$ denotes 2-D convolution. The block diagram of the LSI degradation model presented in Equation (1.2) is shown in Figure 1.1.

The image degradation model described by Equation (1.1) or Equation (1.2) is often represented in terms of a matrix-vector formulation, that is,

$$\mathbf{g} = \mathbf{H}\mathbf{f} + \mathbf{n}, \tag{1.3}$$

where the vectors $\mathbf{g}$, $\mathbf{f}$, and $\mathbf{n}$ represent the observed image, the original image, and the observation noise ordered lexicographically by stacking either the rows or the columns of each image into a vector. $\mathbf{H}$ is a Block Toeplitz with Toeplitz Blocks (BTTB) matrix when Equation (1.2) is used, which can

be approximated by a Block Circulant with Circulant Blocks (BCCB) matrix [24]. Since BCCB matrices are diagonalized using the 2-D Discrete Fourier Transform (DFT), this allows (1.2) to be expressed in the discrete frequency domain. Note that it will also be useful to write $\mathbf{Hf} = \mathbf{Fh}$, where $\mathbf{F}$ is a matrix formed from the image data and $\mathbf{h}$ is a vector parameterizing the blur.

The BD problem refers to finding estimates $\hat{f}(\mathbf{x})$ and $\hat{h}(\mathbf{x})$ for $f(\mathbf{x})$ and $h(\mathbf{x})$ based on $g(\mathbf{x})$ and prior knowledge about $f(\mathbf{x})$, $h(\mathbf{x})$, and $n(\mathbf{x})$. It should be noted that although the degradation model is LSI, the deconvolution algorithm may be nonlinear or spatially varying or both.

Blind deconvolution is an ill-posed problem: the solution of Equation (1.2) may not depend continuously on the data, may not be unique, or it may not even exist [1, 25]. With practical approaches, an approximate solution to the problem is estimated, so that the existence of the solution can be disregarded; however, the nonuniqueness and the sensitivity of the solution to the noise are still serious problems.

1.3 Classification of Blind Image Deconvolution Methodologies

We may classify BD approaches into two categories according to the stage at which we identify the blur: *a priori* or jointly with the image.

***A priori* blur identification methods** With this approach, the PSF is identified separately from the original image, and later used in combination with one of the classical image restoration algorithms in order to restore the original image. A parametric blur model may be used, for example, one of the general models to be described in Section 1.5.2; then the objective is to identify the most likely blur parameters $\mathbf{h}$ from the observation. This approach has been used in [3, 4], for example, and in a Bayesian context in [26].

Experimental approaches are also possible: images of one or more point sources are collected and used to obtain an estimate of the PSF (this, for example, was done with the HST). Furthermore if a good understanding of the imaging system is available for a specific application, we may make an *a priori* prediction of the blur; however unless we use this model as a prior in one of the other algorithms, this is not a blind procedure as such. This may be possible in microscopy, medical ultrasound, remote sensing, or optical telescope systems (e.g., Tiny Tim modeling of the HST).

Joint identification methods The majority of existing methods fall into this class, where the image and blur are identified simultaneously. How-

ever, in practice many methods in this category use an alternating approach to estimate $\mathbf{f}$ and $\mathbf{h}$ rather than truly finding the joint solution. Prior knowledge about the image and blur is typically incorporated in the form of models like those presented in Section 1.5.2 and Section 1.5.3. Parameters describing such models are also required to be estimated from the available data; often this is performed before image and blur identification, although simultaneous identification is possible, e.g., see [27].

In the following section, we use the Bayesian framework to present the different BD approaches proposed in the literature. As we will see, most of them can be seen as special cases of an application of Bayes' theorem: the main differences between them are the choices of the function to be optimized, and the prior distributions used to model the original image and the degradation process. Using the Bayesian framework allows us to describe the general BD problem in a systematic way, and to identify the similarities and differences between the proposed approaches in the above two categories. Each category may be seen as a particular inference model in the Bayesian paradigm.

1.4 Bayesian Framework for Blind Image Deconvolution

A fundamental principle of the Bayesian philosophy is to regard all parameters and observable variables as unknown stochastic quantities, assigning probability distributions based on subjective beliefs. Thus in BD, the original image $\mathbf{f}$, the blur $\mathbf{h}$, and the noise $\mathbf{n}$ in Equation (1.3) are all treated as samples of random fields, with corresponding *prior* Probability Density Functions (PDFs) that model our knowledge about the imaging process and the nature of images. These distributions depend on parameters which will be denoted by Ω. The parameters of the prior distributions are termed *hyperparameters*.

Often Ω is assumed known (or is first estimated separately from $\mathbf{f}$ and $\mathbf{h}$). Alternatively we may adopt the *hierarchical* Bayesian framework whereby Ω is also assumed unknown, in which case we also model our prior knowledge of its values. The PDFs for the hyperparameters are termed *hyperprior* distributions. This abstraction allows greater robustness to error when there is uncertainty, and is essential when we are less confident in the observed data (due to a lower Signal-to-Noise Ratio (SNR)). This *hierarchical* modeling allows us to write the joint global distribution

$$\mathrm{p}(\Omega, \mathbf{f}, \mathbf{h}, \mathbf{g}) = \mathrm{p}(\Omega)\mathrm{p}(\mathbf{f}, \mathbf{h}|\Omega)\mathrm{p}(\mathbf{g}|\Omega, \mathbf{f}, \mathbf{h}), \tag{1.4}$$

where $\mathrm{p}(\mathbf{g}|\Omega, \mathbf{f}, \mathbf{h})$ is termed the *likelihood* of the observations. Typically, we assume that $\mathbf{f}$ and $\mathbf{h}$ are *a priori* conditionally independent, given Ω, i.e., $\mathrm{p}(\mathbf{f}, \mathbf{h} \,|\, \Omega) = \mathrm{p}(\mathbf{f} \,|\, \Omega)\, \mathrm{p}(\mathbf{h} \,|\, \Omega)$. Then the task is to perform inference using

the posterior

$$\mathrm{p}\left(\mathbf{f},\mathbf{h},\Omega\,|\,\mathbf{g}\right)=\frac{\mathrm{p}\left(\mathbf{g}\,|\,\mathbf{f},\mathbf{h},\Omega\right)\mathrm{p}\left(\mathbf{f}\,|\,\Omega\right)\mathrm{p}\left(\mathbf{h}\,|\,\Omega\right)\mathrm{p}\left(\Omega\right)}{\mathrm{p}\left(\mathbf{g}\right)}. \tag{1.5}$$

Note that this corresponds to the joint estimation method described in Section 1.3. We can also marginalize either $\mathbf{f}$, to describe the *a priori* blur identification method as

$$p(\mathbf{h}|\mathbf{g})=\int\cdots\int\mathrm{p}\left(\mathbf{f},\mathbf{h},\Omega\,|\,\mathbf{g}\right)\mathrm{d}\mathbf{f}\cdot\mathrm{d}\Omega, \tag{1.6}$$

or marginalize $\mathbf{h}$ to obtain

$$p(\mathbf{f}|\mathbf{g})=\int\cdots\int\mathrm{p}\left(\mathbf{f},\mathbf{h},\Omega\,|\,\mathbf{g}\right)\mathrm{d}\mathbf{h}\cdot\mathrm{d}\Omega. \tag{1.7}$$

Note that by marginalizing $\mathbf{h}$ this approach seeks to bypass the blur identification stage and estimate the image $\mathbf{f}$ directly, by using appropriate constraints or prior knowledge, although this is less common in practice.

In the following sections we study first the various prior models for the image, blur, and hyperparameters that have appeared in the literature. We then analyze the estimation of these unknown quantities as inference models under the Bayesian framework.

1.5 Bayesian Modeling of Blind Image Deconvolution

1.5.1 Observation Model

The first stage of the Bayesian formulation is specifying the likelihood of the observed image, $\mathbf{g}$. Due to the model in Equation (1.3), the PDF of $\mathbf{g}$ is related to that of the observation noise, $\mathbf{n}$. A typically used model for $\mathbf{n}$ is zero mean independent WGN with distribution $\mathcal{N}\left(\mathbf{n}\,\middle|\,0,\beta^{-1}\mathbf{W}^{-1}\right)$, where β^{-1} denotes the variance, and $\mathbf{W}$ is a diagonal weights matrix included for generality. It allows us to represent spatially varying noise statistics, and it also introduces flexibility in the energy minimization formulation, as will be described later. For stationary noise, $\mathbf{W}=\mathbf{I}$. Thus we have

$$\mathrm{p}\left(\mathbf{n}\right)=\mathrm{p}\left(\mathbf{g}\,|\,\mathbf{f},\mathbf{h},\beta\right)=\det|\mathbf{W}|\left(\frac{\beta}{2\pi}\right)^{N/2}\exp\left[-\frac{1}{2}\beta\|\mathbf{g}-\mathbf{H}\mathbf{f}\|_{\mathbf{W}}^{2}\right], \tag{1.8}$$

where N is the size of the image vector, $\mathbf{f}$, and the weighted norm used in the PDF is defined as $\|\mathbf{x}\|_{\mathbf{W}}^{2}=\mathbf{x}^{T}\mathbf{W}^{T}\mathbf{W}\mathbf{x}$.

Alternative noise modeling, for instance Poisson noise arising in low-intensity imaging, is also assumed in certain BD problems. We will concentrate, however, on the Gaussian noise model presented above.

1.5.2 Parametric Prior Blur Models

The following analytical models are frequently used in Equation (1.2) to represent the LSI image degradation operator, i.e., the PSF. In this case the prior distribution for the blur is parameterized directly by the unknowns defining the parametric model.

Traditionally, when the parametric form of the PSF is assumed known, $\mathrm{p}(\mathbf{h})$ is usually a uniform distribution. Then the unknown parameters may be estimated using, for example, Maximum Likelihood (ML) methods (see [28, 29]). Alternatively the unknown quantities defining the parametric function may be estimated *a priori* if the real underlying image is known. For instance, in astronomy the parameters of the atmospheric turbulence blur are estimated using observed known point sources (stars); similarly, the parameters describing the HST PSF model could also be estimated.

Note in the case when such a parametric model is not used, the values of the PSF coefficients directly parameterize the prior distribution for the blur. Ideally this prior should embody the physical constraints arising from an imaging system: positivity of the coefficients and energy preservation (the PSF coefficients should sum to one). Other prior assumptions often used for the PSF coefficients are smoothness or piecewise smoothness, symmetry, and finite support size. Due to similarities to the priors used for images, these priors will be discussed in Section 1.5.3.

1.5.2.1 Linear Motion Blur

In general, relative motion of the camera and scene to be imaged results in a PSF representing temporal integration along this motion path. If the camera movement or object motion is fast relative to the exposure period, we may approximate this as a linear motion blur. This is represented as the 1-D local averaging of neighboring pixels. An example of a horizontal motion blur model is given by (L is an even integer):

$$h(\mathbf{x}) = \begin{cases} \dfrac{1}{L+1}, & -\dfrac{L}{2} \leq x_1 \leq \dfrac{L}{2}, \\ & x_2 = 0 \\ 0 \quad , & \text{otherwise.} \end{cases} \tag{1.9}$$

1.5.2.2 Atmospheric Turbulence Blur

This type of blur is common in remote sensing and aerial imaging applications. For long-term exposure through the atmosphere a Gaussian PSF model is used:

$$h(\mathbf{x}) = K\, e^{-\frac{|\mathbf{x}|^2}{2\sigma^2}}, \tag{1.10}$$

where K is a normalizing constant ensuring that the blur has a unit volume, and σ^2 is the variance that determines the severity of the blur. Alternative

atmospheric blur models have been suggested in [30, 31]. In these works the PSF is approximated by the function

$$h(\mathbf{x}) \propto (1 + \frac{|\mathbf{x}|^2}{R^2})^{-\delta}, \tag{1.11}$$

where δ and R are unknown parameters.

1.5.2.3 Out-of-Focus Blur

Photographical defocusing is another common type of blurring, primarily due to the finite size of the camera aperture. A complete model of the camera's image formation system depends on many parameters. These include the focal length, the camera aperture size and shape, the distance between object and camera, the wavelength of the incoming light, and effects due to diffraction (see [32] for further details). Furthermore, poor-quality optics introduce aberrations of their own. Accurate knowledge of all of these parameters is usually not available after the picture was taken. When the blur due to defocusing is large, the uniform circular PSF model is used as an approximation to these effects:

$$h\left(\mathbf{x}\right) = \begin{cases} \frac{1}{\pi r^2}\,, & |\mathbf{x}| \leq r \\ 0 \quad , & \text{otherwise.} \end{cases} \tag{1.12}$$

The uniform 2-D blur is sometimes used as a cruder approximation to an out-of-focus blur; it is also used as a model for sensor pixel integration in superresolution restoration. This model is defined (with L an even integer) as

$$h\left(\mathbf{x}\right) = \begin{cases} \frac{1}{(L+1)^2}, & -\frac{L}{2} \leq (x_1, x_2) \leq \frac{L}{2} \\ 0 \qquad , & \text{otherwise.} \end{cases} \tag{1.13}$$

1.5.3 Prior Image and Blur Models

The prior distributions $\mathrm{p}(\mathbf{f}|\Omega)$ and $\mathrm{p}(\mathbf{h}|\Omega)$ should reflect our beliefs about the nature of $\mathbf{f}$ and $\mathbf{h}$ and constrain the space of possible solutions for them to the most probable ones. This is necessary due to the ill-posed nature of the problem. Abstract descriptions of natural images have been made: smooth, piecewise-smooth, or textured, for instance (of course some applications may have other specific constraints). We can attempt to model these descriptions in a stochastic sense using the priors. Typically this is done by specifying probabilistic relations between neighboring image pixels or their derivatives. Similar procedures may be followed for the PSF.

We will consider a general exponential model of the form

$$\mathrm{p}\left(\mathbf{f}\,|\,\Omega\right) = \frac{1}{Z_f(\Omega)} \exp\left[-U_f(\mathbf{f},\Omega)\right] \tag{1.14a}$$

$$\mathrm{p}\left(\mathbf{h}\,|\,\Omega\right) = \frac{1}{Z_h(\Omega)} \exp\left[-U_h(\mathbf{h},\Omega)\right] \tag{1.14b}$$

to represent the image and blur priors. The normalizing terms Z_f and Z_h depend on the hyperparameters for each distribution. They may be treated as constants if we assume the hyperparameters to be known; otherwise they must be calculated as $\int \exp\left[-U_f(\mathbf{f},\Omega)\right]\,\mathrm{d}\mathbf{f}$ and $\int \exp\left[-U_h(\mathbf{h},\Omega)\right]\,\mathrm{d}\mathbf{h}$, respectively, which may cause difficulties in inference unless we assume a special form for $U(\cdot)$. Note that $U(\cdot)$ is sometimes termed the *energy function*.

Many different image and blur models in the literature can be put in the form of Equation (1.14); particular cases will now be considered.

1.5.3.1 Stationary Gaussian Models

The most common model is the class of Gaussian models provided by $U_f = \frac{1}{2}\alpha\|\mathbf{L}\mathbf{f}\|^2$. Then, if $\det|\mathbf{L}| \neq 0$, the term Z_f in Equation (1.14) becomes simply $(2\pi)^{\frac{N}{2}}\alpha^{-\frac{N}{2}}\det|\mathbf{L}|^{-1}$, which if we use a fixed stationary form for $\mathbf{L}$ is simple to calculate. These models are often termed Simultaneous Autoregression (SAR) or Conditional Autoregression (CAR) models [33].

In the most basic case, we can use $\mathbf{L} = \mathbf{I}$, the identity. This imposes constraints on the magnitude of the intensity distribution of $\mathbf{f}$. A more common usage is $\mathbf{L} = \mathbf{C}$, the discrete Laplacian operator, which instead constrains the derivative of the image. For instance, Molina et al. [27] used this model for both image and blur, giving

$$\mathrm{p}(\mathbf{f}|\alpha_{\mathrm{im}}) \propto \alpha_{\mathrm{im}}^{N/2} \exp\left[-\frac{1}{2}\alpha_{\mathrm{im}} \parallel \mathbf{C}\mathbf{f} \parallel^2\right] \tag{1.15a}$$

$$\mathrm{p}(\mathbf{h}|\alpha_{\mathrm{bl}}) \propto \alpha_{\mathrm{bl}}^{M/2} \exp\left[-\frac{1}{2}\alpha_{\mathrm{bl}} \parallel \mathbf{C}\mathbf{h} \parallel^2\right]. \tag{1.15b}$$

Note that in these two equations N and M should in practice be replaced by $N-1$ and $M-1$, respectively, because $\mathbf{C}^T\mathbf{C}$ is singular.

This SAR model is suitable for $\mathbf{f}$ and $\mathbf{h}$ if it is assumed that the luminosity distribution is smooth on the image domain, and that the PSF is a partially smooth function.

In [34] the SAR model was used for the image prior, and a Gaussian PDF for the PSF; that is,

$$\mathrm{p}(\mathbf{h}|\mu_{\mathrm{bl}},\alpha_{\mathrm{bl}}) = \mathcal{N}(\mu_{\mathrm{bl}},\alpha_{\mathrm{bl}}^{-1}\mathbf{I}). \tag{1.16}$$

The components of the blur are assumed statistically independent so that the covariance matrix is diagonal. Clearly, a Gaussian PDF with unknown

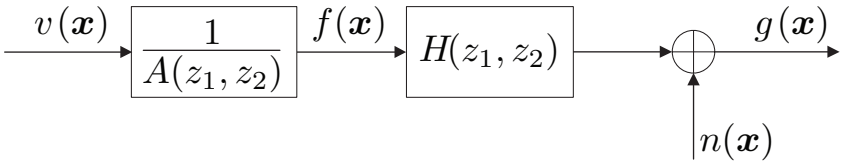

FIGURE 1.2: ARMA degradation model.

mean vector and fully populated covariance matrix might be used to model image and blur priors. However, this requires the simultaneous estimation of a very large number of hyperparameters, thus making the approach highly impractical unless we use additional hyperpriors (see Section 1.5.4) to impart *a priori* knowledge of the hyperparameters' values [27].

1.5.3.2 Autoregressive Models

A class of blind image deconvolution algorithms (see, e.g., [28, 29]) model the observed image $\mathbf{g}$ as an Autoregressive Moving Average (ARMA) process, as shown in Figure 1.2.

The observation Equation (1.2) forms the Moving Average (MA) part of the model. The original image is modeled as a 2-D Autoregressive (AR) process:

$$f(\mathbf{x}) = \sum_{\mathbf{s} \in S_a : \mathbf{s} \neq \mathbf{0}} a(\mathbf{s}) f(\mathbf{x} - \mathbf{s}) + v(\mathbf{x}), \tag{1.17}$$

or in matrix-vector form:

$$\mathbf{f} = \mathbf{A}\mathbf{f} + \mathbf{v}, \tag{1.18a}$$

or equivalently,

$$\mathbf{f} = \mathbf{F}\mathbf{a} + \mathbf{v}, \tag{1.18b}$$

where $S_a \subset \mathbb{R}^2$ is the support of the 2-D AR coefficients $a(\mathbf{x})$, and $\mathbf{A}$ has a BTTB form. The *excitation* noise signal, or modeling error, $v(\mathbf{x})$ is a zero-mean WGN process with diagonal covariance matrix Λ_v, that is independent of $f(\mathbf{x})$. Since $\mathbf{v} = (\mathbf{I} - \mathbf{A})\mathbf{f}$, the PDF of $\mathbf{f}$ is obtained via a probability transformation of $\mathbf{v}$, as a Gaussian: $\mathrm{p}(\mathbf{f} \,|\, \mathbf{a}, \Lambda_v) = \mathcal{N}(\mathbf{f} \,|\, 0, \Lambda_f)$. The image covariance matrix is defined as

$$\Lambda_f = (\mathbf{I} - \mathbf{A})^{-1} \Lambda_v (\mathbf{I} - \mathbf{A})^{-T}. \tag{1.19}$$

It should be clear that $\mathrm{p}(\mathbf{f})$ is in the form of Equation (1.14), with $U_f = \frac{1}{2}\|(\mathbf{I}-\mathbf{A})\mathbf{f}\|^2_{\Lambda_v}$ and $Z_f = (2\pi)^{\frac{N}{2}} \det|\Lambda_v|^{\frac{1}{2}} \det|\mathbf{I}-\mathbf{A}|^{-1}$. Unlike the SAR model above, however, where a deterministic form is used for the matrix $\mathbf{L}$, the AR coefficients defining $\mathbf{A}$ and the excitation noise covariance matrix Λ_v also have to be estimated (in [28] a flat prior distribution is assumed on them).

A related formulation to the stationary ARMA model is also considered by Katsaggelos and Lay in [29, 35, 36]. In these works, the AR model parameters are not estimated directly, but rather the defining sequence of the matrix Λ_f is found in the discrete frequency domain, along with the other parameters, under the assumption that the image model is stationary. Observe that this approach does not assume a known model support size S_a or blur support S_h.

The AR image model is good at representing textured images; however, like any of the other stationary Gaussian models in this section, it is not such a good model for an original image that has prominent edges as part of the scene. Therefore it is possible to consider nonstationary extensions to the AR model.

This was done, for example, in [26], where the image is partitioned into blocks each assumed to be homogeneous regions with their own AR coefficients and excitation variance. The equations above remain the same, although the matrix $\mathbf{A}$ is no longer BTTB and the sizes of $\mathbf{F}$ and $\mathbf{a}$ increase. A local mean may be assumed for each region to further better model a real image. In general, a segmentation of the image could be assumed such that the blocks in the model coincide with the natural regions in the image. The use of a nonstationary image model in conjunction with a stationary blur can also aid blur identification [26].

1.5.3.3 Markov Random Field Models

A class of models encountered extensively in image segmentation [37], classical image restoration [38], and also in superresolution restoration [39] and BD [40, 41] are the Markov Random Field (MRF) models [42]. They are usually derived using local spatial dependencies; however, we may see they are closely related to the other models in this section.

Defining $U = \sum_{c \in \mathcal{C}} V_c(\mathbf{f})$ in Equation (1.14), we have the definition of a *Gibbs distribution*. In this context, Z is termed the partition function. $V_c(\mathbf{f})$ is a *potential function* defined over *cliques*, c in the image [42]. Briefly speaking, this gives a simple way of specifying interactions over local neighborhoods in the image field. If we use quadratic potentials, $V_c(\mathbf{f}) = \left(\mathbf{d}_c^T \mathbf{f}\right)^2$, we may rewrite the Gibbs distribution as a Gaussian:

$$p(\mathbf{f}) = \frac{1}{Z} \exp\left[-\mathbf{f}^T \mathbf{B} \mathbf{f}\right] = \frac{1}{Z} \exp\left[-\sum_{c \in \mathcal{C}} \mathbf{f}^T \mathbf{B}_c \mathbf{f}\right] \tag{1.20}$$

where $\mathbf{B}_c$ is obtained from $\mathbf{d}_c$ and satisfies $[\mathbf{B}_c]_{\mathbf{x},\mathbf{s}}$ are only nonzero when pixels $\mathbf{x}$ and $\mathbf{s}$ are neighbors. Typically the vectors $\mathbf{d}_c$ represent finite difference operators. The partition function is now equal to $(2\pi)^{\frac{N}{2}} \det|\mathbf{B}|^{-\frac{1}{2}}$. This model is also then termed a Gaussian Markov Random Field (GMRF) [43] or CAR [33].

We may also use Generalized Gaussian MRFs (GGMRFs) with arbitrary nonquadratic potentials of a similar functional form: $V_c(\mathbf{f}) = \rho(\mathbf{d}_c^T \mathbf{f})$, where

ρ is some (usually convex) function, such as the *Huber function* [43] or p-norm (with $p \geq 1$) based function, $\rho(u) = |u|^p$. This is similar to the use of potential functions used in anisotropic diffusion methods, the motivation being edge preservation in the reconstructed image. Other extensions to the model consider hierarchical, or Compound GMRFs (CGMRFs), also with the goal of avoiding oversmoothing of edges [44, 38].

1.5.3.4 Anisotropic Diffusion and Total Variation Type Models

Non-quadratic image priors have been investigated using *variational integrals* in the *anisotropic diffusion* [45] or *Total Variation* (TV) [46] regularization frameworks, with the aim of preserving edges by not over-penalizing discontinuities, i.e. outliers in the image gradient distribution, see [47, 48] for a unifying view of the probabilistic and variational approaches. The main difference to the GGMRFs models mentioned above is that these usually begin with a formulation in the continuous image domain resulting in Partial Differential Equations (PDEs) that must be solved. However, eventual discretization is eventually necessary, and hence the constraints may be reformulated as non-Gaussian priors, or Gaussians with a non-stationary covariance matrix [20]. Alternatively, other methods propose formulating the TV norm directly in the discrete domain [49, 50].

The generalized regularization approach using anisotropic diffusion has been proposed by You and Kaveh [45]. In this formulation, convex functions $\kappa(\cdot)$ and $\upsilon(\cdot)$ of the image gradient $|\nabla f(x)|$ and the PSF gradient $|\nabla h(x)|$ respectively are used in defining regularization functionals:

$$\mathcal{E}(f) = \int_{S_f} \kappa\left(|\nabla f(\mathbf{x})|\right) \, \mathrm{d}\mathbf{x} \tag{1.21a}$$

$$\mathcal{E}(h) = \int_{S_h} \upsilon\left(|\nabla h(\mathbf{s})|\right) \, \mathrm{d}\mathbf{s}. \tag{1.21b}$$

This is in analogy with standard regularization procedures. However as the functionals are continuous, variational calculus is used to perform the differentiation needed for minimization [51]. This results in a PDE which must be solved for each variable. Consider for instance minimization of Equation (1.21a) for f; the solution must satisfy

$$\nabla_f \mathcal{E}(f) = \nabla \cdot \left(\frac{\kappa'(|\nabla f|)}{|\nabla f|} \nabla f \right) = 0, \tag{1.22}$$

with appropriate boundary conditions. One method of solution is imposing an artificial time evolution variable t, and using a steepest descent method, i.e., for f,

$$\frac{\partial \hat{f}}{\partial t} = -\nabla_f \mathcal{E}(\hat{f}). \tag{1.23}$$

This may be interpreted as representing a physical anisotropic diffusion process [45, 51, 52]. That is, as the time variable t progresses, directional smoothing occurs depending on the local image gradient. The strength and type of smoothing depends on the *diffusion coefficient* or *flux variable*, c, which is related to the potential function by

$$c(|\nabla f|) = \frac{\kappa'(|\nabla f|)}{|\nabla f|} \tag{1.24}$$

We may consider $c(|\nabla f|)$ to be the amount of smoothing perpendicular to the edges. Appropriate choice of c or equivalently κ can result in spatially-adaptive edge preserving restoration.

Consider two cases of the potential function κ and related diffusion coefficient c. In the first case, $\kappa(x) = \frac{1}{2}x^2$ and hence $c(|\nabla f|) = 1$, and $\nabla_f \mathcal{E}(f) = \nabla^2(f)$, i.e., a Laplacian operator [53]. This corresponds to standard spatially-invariant isotropic regularization, or a CAR model with the discrete Laplacian when discretized.

Another choice proposed for the BD problem by Chan and Wong [46] is given by the *Total Variation* (TV) norm. In this case, $\kappa(x) = x$ and hence $c(|\nabla f|) = \frac{1}{|\nabla f|}$. The result is that smoothing in the direction orthogonal to the edges is completely suppressed and is only applied parallel to the edge directions. This is demonstrated in [45] by decomposing the anisotropic diffusion equation, Equation (1.23) into components parallel and perpendicular to the edges. A very efficient way to solve the resulting optimization problem is shown in [46] in this particular case.

These two choices lead us to consider the following discrete prior image models:

$$\mathrm{p}(\mathbf{f}) \propto \exp\left[-\alpha_{\mathrm{im}} \sum_i ((\Delta_i^h \mathbf{f})_i^2 + (\Delta_i^v \mathbf{f})_i^2)\right] \tag{1.25}$$

for the Laplacian; and

$$\mathrm{p}(\mathbf{f}) \propto \exp\left[-\alpha_{\mathrm{im}} \sum_i \sqrt{(\Delta_i^h \mathbf{f})_i^2 + (\Delta_i^v \mathbf{f})_i^2)}\right] \tag{1.26}$$

for the TV norm, where Δ_i^h and Δ_i^v are linear operators corresponding to horizontal and vertical first order differences, at pixel i, respectively.

A combination of the two choices for c is considered in [45], resulting in a spatially-adaptive diffusion coefficient. The smoothing strength is increased using the Laplacian in areas with low gradient magnitude, and decreased using the TV norm in areas where large intensity transitions occur in order to preserve edges while still removing noise. This is analogous to the use of the Huber function in Section 1.5.3.3. Many other diffusion coefficients are proposed in the literature, including very complex structural operators (see [52] for a review).

Šroubek and Flusser [20] use a similar scheme to those already mentioned, but (using the half-quadratic approach [54, 55]) demonstrate how the anisotropic diffusion model may be written in the form of Equation (1.14) by discretization of the functional in Equation (1.21a):

$$\mathrm{p}\left(\mathbf{f}, c(\mathbf{f})\right) = \frac{1}{Z_f} \exp\left[-\frac{1}{2}\mathbf{f}^T \mathbf{B}\left(c\right)\mathbf{f}\right] \tag{1.27}$$

They equate the diffusion, or flux variable, to the hidden line process often used in CGMRFs, that is it represents the edge strength between two pixels in the image. Therefore it is possible to build a spatially-varying weights matrix $\mathbf{B}$ from the local image gradients. Note that as the flux variable is a function of $\mathbf{f}$, so is the covariance, so this is not strictly a Gaussian distribution unless $\mathbf{B}$ is assumed fixed. In practice, using an iterative scheme, $\mathbf{B}$ may be updated at each iteration.

A similar motivation was used in [19] to obtain a spatially-varying weights matrix based on the local image variance, as was previously suggested in [56, 57]. The difference here is that the regularization is isotropic; better performance can be expected with the anisotropic schemes.

1.5.4 Hyperprior Models

So far we have studied the distributions $\mathrm{p}(\mathbf{f}, \mathbf{h}|\Omega)$, $\mathrm{p}(\mathbf{g}|\Omega, \mathbf{f}, \mathbf{h})$ that appear in the Bayesian modeling of the BD problem in Equation (1.4). We complete this modeling by studying now the distribution $\mathrm{p}(\Omega)$.

An important problem is the estimation of the vector of parameters Ω when they are unknown. To deal with this estimation problem, the hierarchical Bayesian paradigm introduces a second stage (the first stage consisting again of the formulation of $\mathrm{p}(\mathbf{f}|\Omega)$, $\mathrm{p}(\mathbf{h}|\Omega)$, and $\mathrm{p}(\mathbf{g}|\mathbf{f}, \mathbf{h}, \Omega)$). In this stage the hyperprior $\mathrm{p}(\Omega)$ is also formulated.

A large part of the Bayesian literature is devoted to finding hyperprior distributions $\mathrm{p}(\Omega)$ for which $\mathrm{p}(\Omega, \mathbf{f}, \mathbf{h}|\mathbf{g})$ can be calculated in a straightforward way or be approximated. These are the so-called conjugate priors [58], which were developed extensively in Raiffa and Schlaifer [59].

Besides providing for easy calculation or approximations of $\mathrm{p}(\Omega, \mathbf{f}, \mathbf{h}|\mathbf{g})$, conjugate priors have, as we will see later, the intuitive feature of allowing one to begin with a certain functional form for the prior and end up with a posterior of the same functional form, but with the parameters updated by the sample information.

Taking the above considerations about conjugate priors into account, the literature in BD uses different *a priori* models for the parameters depending on the type of unknown parameters. For parameters, ω, corresponding to inverses of variances, the gamma distribution is used. This is defined by:

$$\mathrm{p}(\omega) = \Gamma(\omega|a_\omega, b_\omega) = \frac{(b_\omega)^{a_\omega}}{\Gamma(a_\omega)} \omega^{a_\omega - 1} \exp[-b_\omega\, \omega], \tag{1.28}$$

where $\omega > 0$ denotes a hyperparameter, $b_\omega > 0$ is the scale parameter, and $a_\omega > 0$ is the shape parameter. These parameters are assumed known. The gamma distribution has the following mean, variance, and mode:

$$E[\omega] = \frac{a_\omega}{b_\omega}\ , \qquad Var[\omega] = \frac{a_\omega}{(b_\omega)^2}\ , \qquad \text{Mode}[\omega] = \frac{a_\omega - 1}{b_\omega}. \tag{1.29}$$

Note that the mode does not exist when $a_\omega^o \leq 1$ and that mean and mode do not coincide.

For components of mean vectors the corresponding conjugate prior is a normal distribution. Additionally, for covariance matrices the hyperprior is given by an inverse Wishart distribution (see [60]).

We observe, however, that in general most of the methods proposed in the literature use the *uninformative* prior model

$$\mathrm{p}(\Omega) = constant. \tag{1.30}$$

1.6 Bayesian Inference Methods in Blind Image Deconvolution

There are a number of different ways that we may proceed to estimate the image and blur using Equation (1.5). Depending on the prior models chosen, finding analytic solutions may be difficult, so approximations are often needed. Many methods in the literature seek point estimates of the parameters $\mathbf{f}$ and $\mathbf{h}$. Typically, this reduces the problem to one of optimization. However, the Bayesian framework provides other methodologies for estimating the *distributions* of the parameters [60, 61, 62], which deal better with uncertainty; approximating or simulating the posterior distribution are two options. These different inference strategies and examples of their use will now be presented, proceeding from the simplest to the more complex.

1.6.1 Maximum *a Posteriori* and Maximum Likelihood

One possible point estimate is provided by the *Maximum A Posteriori* (MAP) solution, which are the values of $\mathbf{f}$, $\mathbf{h}$, and Ω that maximize the posterior probability density:

$$\{\hat{\mathbf{f}}, \hat{\mathbf{h}}, \hat{\Omega}\}_{\text{MAP}} = \underset{\mathbf{f},\mathbf{h},\Omega}{\text{argmax}}\ \mathrm{p}\left(\mathbf{g} \,|\, \mathbf{f}, \mathbf{h}, \Omega\right) \mathrm{p}\left(\mathbf{f} \,|\, \Omega\right) \mathrm{p}\left(\mathbf{h} \,|\, \Omega\right) \mathrm{p}\left(\Omega\right). \tag{1.31}$$

The *Maximum Likelihood* (ML) solution attempts instead to maximize the likelihood $\mathrm{p}\left(\mathbf{g} \,|\, \mathbf{f}, \mathbf{h}, \Omega\right)$ with respect to the parameters:

$$\{\hat{\mathbf{f}}, \hat{\mathbf{h}}, \hat{\Omega}\}_{\text{ML}} = \underset{\mathbf{f},\mathbf{h},\Omega}{\text{argmax}}\ \mathrm{p}\left(\mathbf{g} \,|\, \mathbf{f}, \mathbf{h}, \Omega\right). \tag{1.32}$$

Note, however, that in this case we can only estimate the parameters in Ω that are present in the conditional distribution $\mathrm{p}(\mathbf{g} \,|\, \mathbf{f}, \mathbf{h}, \Omega)$ but none of those present only in $\mathrm{p}(\mathbf{f}, \mathbf{h} \,|\, \Omega)$.

The above maximization of the likelihood is typically seen as a non-Bayesian method, although it is identical to the MAP solution with uninformative (flat) prior distributions. Some approaches may use flat priors for some parameters but not others. Assuming known values for the parameters is equivalent to using degenerate distributions (delta functions) for priors. For instance, a degenerate distribution on Ω is defined as

$$\mathrm{p}(\Omega) = \delta(\Omega, \Omega_0) = \begin{cases} 1, & \text{if } \Omega = \Omega_0 \\ 0, & \text{otherwise.} \end{cases} \tag{1.33}$$

Then, the MAP and ML solutions become, respectively,

$$\{\hat{\mathbf{f}}, \hat{\mathbf{h}}\}_{\mathrm{MAP}} = \underset{\mathbf{f},\mathbf{h}}{\operatorname{argmax}} \; \mathrm{p}(\mathbf{g} \,|\, \mathbf{f}, \mathbf{h}, \Omega_0)\, \mathrm{p}(\mathbf{f} \,|\, \Omega_0)\, \mathrm{p}(\mathbf{h} \,|\, \Omega_0) \tag{1.34}$$

$$\{\hat{\mathbf{f}}, \hat{\mathbf{h}}\}_{\mathrm{ML}} = \underset{\mathbf{f},\mathbf{h}}{\operatorname{argmax}} \; \mathrm{p}(\mathbf{g} \,|\, \mathbf{f}, \mathbf{h}, \Omega_0)\,. \tag{1.35}$$

Many deconvolution methods can fit into this Bayesian formulation. The main differences among these algorithms come from the form of the likelihood, the particular choice of priors on the image, blur, and the hyperparameters, and the optimization methods used to find the solutions.

Observe that the regularization-based approaches using the L_2 norm frequently found in the literature also fall into this category. In these approaches the blind deconvolution problem is stated as a constrained minimization problem, where a cost function is minimized with a number of regularization constraint terms.

In regularization approaches the cost function is chosen as the error function $\|\mathbf{g} - \mathbf{H}\mathbf{f}\|^2_W$, which ensures fidelity to the data. The regularization terms are used to impose additional constraints on the optimization problem. Generally, these constraints ensure smoothness of the image and the blur, that is, the high-frequency energy of the image and the blur is minimized. The effect of the regularization terms is controlled by the regularization parameters, which basically represent the trade-off between fidelity to the data and desirable properties (smoothness) of the solutions.

For example, in [19], the classical regularized image restoration formulation used in [56, 57, 63] was extended to the BD case by adding a constraint for the blur. The problem is stated, in a relaxed minimization form, as

$$\hat{\mathbf{f}}, \hat{\mathbf{h}} = \underset{\mathbf{f},\mathbf{h}}{\operatorname{argmin}} \; \left[\|\mathbf{g} - \mathbf{H}\mathbf{f}\|^2_W + \lambda_1 \|\mathbf{L}_f \mathbf{f}\|^2 + \lambda_2 \|\mathbf{L}_h \mathbf{h}\|^2 \right], \tag{1.36}$$

where λ_1 and λ_2 are the Lagrange multipliers for each constraint, and $\mathbf{L}_f$ and $\mathbf{L}_h$ are the regularization operators. In [19] each $\mathbf{L}$ is the Laplacian multiplied by a spatially varying weights term, calculated as in [63, 57, 64, 65] from the local image variance in order to provide some spatial adaptivity to avoid oversmoothing edges.

1.6.1.1 Iterated Conditional Modes

Let us consider again the solution of Equation (1.34). A major problem in the optimization is the simultaneous estimation of the variables $\mathbf{f}$ and $\mathbf{h}$. A widely used approach is that of *Alternating Minimization* (AM) of Equation (1.36) (or its continuous equivalent for PDE formulations), which follows the steepest descent with respect to one unknown while holding the other unknown constant. The advantage of this algorithm is its simplicity due to the linearization of the objective function. This optimization procedure corresponds to the Iterated Conditional Modes (ICM) proposed by Besag [66].

This estimation procedure has been applied to standard regularization approaches [19, 67], and to the anisotropic diffusion and TV type models described in Section 1.5.3.4, where the objective functional becomes

$$\int_{S_f}\left(g(\mathbf{x})-\int_{S_h}h(\mathbf{s}-\mathbf{x})*f(\mathbf{x})\mathrm{d}\mathbf{s}\right)^2\mathrm{d}\mathbf{x}+\lambda_1\int_{S_f}\kappa\left(|\nabla f(\mathbf{x})|\right)\mathrm{d}\mathbf{x}+\lambda_2\int_{S_h}\upsilon\left(|\nabla h(\mathbf{s})|\right)\mathrm{d}\mathbf{s}. \tag{1.37}$$

Partial derivatives with respect to f and h are taken to give the two PDEs for AM.

There are various numerical methods to solve the associated PDEs. These include the classical Euler, Newton, or Runge–Kutta methods; or recently developed approaches, such as time-marching [68], primal-dual methods [69], lagged diffusivity fixed point schemes [70], and half-quadratic regularization [54] (similar to the discrete schemes in [55, 71]). All of these methods employ techniques to discretize and linearize the PDEs to approximate the solution. The selection of a particular method depends on the computational limitations and speed requirements, since different techniques have different simplicity, stability, and convergence speed properties.

1.6.2 Minimum Mean Squared Error

The MAP estimate does not take into account the whole posterior PDF. If the posterior is sharply peaked about the maximum then this does not matter; however, in the case of high observation noise or a broad (heavy-tailed) posterior this estimate is likely to be unreliable. As mentioned in [72], for a Gaussian in high dimensions most of the probability *mass* is concentrated away from the probability *density* peak.

The Minimum Mean Squared Error (MMSE) estimate attempts to find the optimal parameter values that minimize the expected mean squared error between the estimates and the true values. In other words we aim at calculating the mean value of $\mathrm{p}(\mathbf{f},\mathbf{h},\Omega|\mathbf{g})$. In practice, finding MMSE estimates analytically is generally difficult, though it is possible with sampling-based methods (Section 1.6.5) and can be approximated using variational Bayesian methods (Section 1.6.4).

(a) (b) (c)

FIGURE 1.3: (a) Original Lena image; degraded images with Gaussian shaped PSF of variance 9, with: (b) BSNR=40 dB; (c) BSNR=20 dB.

1.6.3 Marginalizing Hidden Variables

In the discussion so far none of the three unknowns, $\mathbf{f}, \mathbf{h}$, and Ω have been marginalized out to perform inference on only a subset of $\mathbf{f}, \mathbf{h}$, and Ω.

We can, however, approach the BD inference problem by first calculating

$$\hat{\mathbf{h}}, \hat{\Omega} = \underset{\mathbf{h},\Omega}{\operatorname{argmax}} \int_{\mathbf{f}} \mathrm{p}(\Omega)\mathrm{p}(\mathbf{f}, \mathbf{h}|\Omega)\mathrm{p}(\mathbf{g}|\Omega, \mathbf{f}, \mathbf{h})\mathrm{d}\mathbf{f} \tag{1.38}$$

and then selecting as restoration the image

$$\hat{\mathbf{f}}\Big|_{\hat{\mathbf{h}},\hat{\Omega}} = \underset{\mathbf{f}}{\operatorname{argmax}}\, \mathrm{p}(\mathbf{f}|\hat{\Omega})\mathrm{p}(\mathbf{g}|\hat{\Omega}, \mathbf{f}, \hat{\mathbf{h}}). \tag{1.39}$$

We can also marginalize $\mathbf{h}$ and Ω to obtain

$$\hat{\mathbf{f}} = \underset{\mathbf{f}}{\operatorname{argmax}} \int_{\mathbf{h},\Omega} \mathrm{p}(\Omega)\mathrm{p}(\mathbf{f}, \mathbf{h}|\Omega)\mathrm{p}(\mathbf{g}|\Omega, \mathbf{f}, \mathbf{h})\mathrm{d}\mathbf{h} \cdot \mathrm{d}\Omega \tag{1.40}$$

The two above inference models are named Evidence- and Empirical-based analysis [73], respectively. The marginalized variables are called hidden variables.

The Expectation Maximization (EM) algorithm, first described in [74] is an incredibly popular technique in signal processing for iteratively solving ML and MAP problems that can be regarded as having *hidden data*. Its properties are well studied: convergence to a *local* maximum of the likelihood or the posterior distribution is guaranteed. It is particularly suited to inverse problems in image restoration and BD as it is obvious that the unobserved image, $\mathbf{f}$, represents a natural choice for the hidden data and in consequence for solving Equation (1.38).

The EM algorithm has been used in BD, for example, in [35, 36] in a general frequency domain formulation and in [28] using the ARMA model (Section

(a) (b)

FIGURE 1.4: Blind deconvolution with the method in [35]. (a) 40 dB BSNR case; (b) 20 dB BSNR case.

1.5.3.2). Some examples of the restorations possible using the method in [36] will now be demonstrated. Consider the original *Lena* image shown in Figure (1.3). This is synthetically degraded by blurring the original image by a Gaussian-shaped PSF with variance 9, and adding WGN with two different variances. This gives the images also in Figures (1.3b) and (1.3c), which have 40 dB and 20 dB Blurred-image SNR (BSNR) respectively. These test images will also be used in later sections. The restored images using the EM method [36] are shown in Figure (1.4).

Note that the Evidence-based analysis can also be used to marginalize the image $\mathbf{f}$ as well as the unknown parameters Ω to obtain $\mathrm{p}(\mathbf{h}|\mathbf{g})$, as in Equation (1.6), then the mode of this posterior distribution can be calculated. In order to estimate the original image $\mathbf{f}$ we can then use only the observation model (see [26] for details). An example of the results obtained by this method is shown in Figure (1.5). The original 256×256 pixel Cameraman image is blurred with a causal blur, and noise is added at 35 dB BSNR. The blind deconvolved image is shown in Figure (1.5).

It is rarely possible to calculate in closed form the integrals involved in the Evidence- and Empirical-based Bayesian inference. To solve this problem we can use approximations of the integrands. Let us consider the integral in Equation (1.38), then for each value of $\mathbf{h}$ and Ω we can calculate

$$\hat{\mathbf{f}}\Big|_{\mathbf{h},\Omega} = \underset{\mathbf{f}}{\operatorname{argmax}}\, \mathrm{p}(\mathbf{f}|\Omega)\mathrm{p}(\mathbf{g}|\mathbf{f},\mathbf{h},\Omega) \tag{1.41}$$

and perform the second-order Taylor's expansion of $\log \mathrm{p}(\mathbf{g}|\mathbf{f},\mathbf{h},\Omega)$ around $\hat{\mathbf{f}}$. As a consequence of the approximation, the integral in Equation (1.38) is performed over a distribution on $\mathbf{f}$ that is Gaussian and usually easy to calculate. This methodology is called Laplace distribution approximation [75, 76] and has been applied, for instance, by Galatsanos *et al.* [77, 78] to partially known blur deconvolution problems.

(a) (b) (c)

FIGURE 1.5: (a) Original Cameraman image; (b) degraded image; (c) reconstruction with the method in [26].

1.6.4 Variational Bayesian Approach

Variational Bayesian methods are generalizations of the EM algorithm to compute ML or MAP estimates. The EM algorithm has proven to be very useful in a wide range of applications; however, in many problems its application is not possible because the posterior distribution cannot be specified. The variational methods overcome this shortcoming by approximating $\mathrm{p}(\mathbf{f}, \mathbf{h}, \Omega|\mathbf{g})$ by a simpler distribution $\mathrm{q}(\mathbf{f}, \mathbf{h}, \Omega)$ obtained by minimizing the Kullback–Leibler (KL) divergence between the variational approximation and the exact distribution. Additionally to providing approximations to the estimates based on $\mathrm{p}(\mathbf{f}, \mathbf{h}, \Omega|\mathbf{g})$, the study of the distribution $\mathrm{q}(\mathbf{f}, \mathbf{h}, \Omega)$ allows us to examine the quality of these estimates.

Note that the Laplace approximation [77, 78] can be considered as an intermediate step between inference based on the true posterior and the one based on a variational distribution approximation.

The variational approximation applied to BD aims at approximating the intractable posterior distribution $\mathrm{p}(\mathbf{f}, \mathbf{h}, \Omega|\mathbf{g})$ by a tractable one denoted by $\mathrm{q}(\mathbf{f}, \mathbf{h}, \Omega)$. For an arbitrary PDF $\mathrm{q}(\mathbf{f}, \mathbf{h}, \Omega)$, the goal is to minimize the KL divergence, given by: [79]

$$\begin{aligned} KL(\mathrm{q}(\mathbf{f}, \mathbf{h}, \Omega) \parallel \mathrm{p}(\mathbf{f}, \mathbf{h}, \Omega|\mathbf{g})) &= \int \mathrm{q}(\mathbf{f}, \mathbf{h}, \Omega) \log \left(\frac{\mathrm{q}(\mathbf{f}, \mathbf{h}, \Omega)}{\mathrm{p}(\mathbf{f}, \mathbf{h}, \Omega|\mathbf{g})} \right) \mathrm{d}\mathbf{f} \cdot \mathrm{d}\mathbf{h} \cdot \mathrm{d}\Omega \\ &= \int \mathrm{q}(\mathbf{f}, \mathbf{h}, \Omega) \log \left(\frac{\mathrm{q}(\mathbf{f}, \mathbf{h}, \Omega)}{\mathrm{p}(\mathbf{f}, \mathbf{h}, \Omega, \mathbf{g})} \right) \mathrm{d}\mathbf{f} \cdot \mathrm{d}\mathbf{h} \cdot \mathrm{d}\Omega \\ &\quad + \text{const}, \end{aligned} \tag{1.42}$$

which is always nonnegative and equal to zero only when $\mathrm{q}(\mathbf{f}, \mathbf{h}, \Omega)$=$\mathrm{p}(\mathbf{f}, \mathbf{h}, \Omega|\mathbf{g})$, which corresponds to the EM result.

To reduce computational complexity and enable the approximate parameter distributions to be found in an analytic form, the PDF $\mathrm{q}(\mathbf{f}, \mathbf{h}, \Omega)$ is factorized

(a) (b)

FIGURE 1.6: Restorations with the variational Bayesian method in [34]: (a) 40 dB BSNR case; (b) 20 dB BSNR case

using the mean field approximation, such that

$$\mathrm{q}(\mathbf{f}, \mathbf{h}, \Omega) = \mathrm{q}(\mathbf{f})\mathrm{q}(\mathbf{h})\mathrm{q}(\Omega). \tag{1.43}$$

For a vector parameter $\theta \in \Theta = \{\mathbf{f}, \mathbf{h}, \Omega\}$, we denote by Θ_θ the subset of Θ with θ removed; for example, for $\theta = \mathbf{f}$, $\Theta_\mathbf{f} = \{\mathbf{h}, \Omega\}$ and $\mathrm{q}(\Theta_\mathbf{f}) = \mathrm{q}(\Omega)\mathrm{q}(\mathbf{h})$. An iterative procedure can be developed to estimate the distributions of the parameters $\{\mathbf{f}, \mathbf{h}, \Omega\}$. At each iteration, the distribution of the parameter θ is estimated using the current estimates of the distribution of Θ_θ:

$$\mathrm{q}^k(\theta) = \arg\min_{\mathrm{q}(\theta)} KL(\mathrm{q}^k(\Theta_\theta)\mathrm{q}(\theta) \parallel \mathrm{p}(\Theta \mid \mathbf{g})). \tag{1.44}$$

AM strategies can be employed. For example, in [34], a cascaded EM algorithm is proposed similar to the AM algorithm, where at each iteration the distributions of the parameters $\mathbf{f}$ and $\mathbf{h}$ are calculated in an alternating fashion while assuming one parameter to be constant. This approach can also be interpreted as an EM algorithm where at each stage the current estimate of one parameter is assumed "known" in estimating the other parameter, as in the classical image restoration problems.

Some example restorations using the variational method are now presented. Using the *VAR1* method presented by Likas and Galatsanos [34] produces the results shown in Figure (1.6). The *BR* method in Molina *et al.* [27] produces the restored images in Figure (1.7).

1.6.5 Sampling Methods

The most general approach to performing inference for the BD problem is to simulate the posterior distribution in Equation (1.5). This in theory allows us to perform inference on arbitrarily complex models in high-dimensional spaces, where no analytic solution is available. Markov Chain Monte Carlo

(a) (b)

FIGURE 1.7: Restorations with variational Bayesian method in [27]: (a) 40 dB BSNR case; (b) 20 dB BSNR case

(MCMC) methods (see, e.g., [61, 80, 81]) attempt to approximate the posterior distribution by the statistics of samples generated from a Markov Chain.

The most simple example of MCMC is the Gibbs sampler which has been used in classical image restoration in conjunction with MRF image models [38]. If we can write down analytic expressions for the *conditional* distributions of all the parameters we wish to estimate, given the others, we simply draw samples from each of the distributions in turn, conditioned on the most recently sampled values of the other parameters. For example if we want to simulate $\mathrm{p}\,(\mathbf{f}, \mathbf{h}, \Omega \,|\, \mathbf{g}\,)$, the iterations would proceed as follows:

$$
\begin{array}{rrcl}
\text{First iteration:} & \mathbf{f}^{(1)} & \leftarrow & \mathrm{p}(\mathbf{f} \mid \mathbf{h}^{(0)}, \Omega^{(0)}, \mathbf{g}) \\
& \mathbf{h}^{(1)} & \leftarrow & \mathrm{p}(\mathbf{h} \mid \mathbf{f}^{(1)}, \Omega^{(0)}, \mathbf{g}) \\
& \Omega^{(1)} & \leftarrow & \mathrm{p}(\Omega \mid \mathbf{f}^{(1)}, \mathbf{h}^{(1)}, \mathbf{g}) \\
\text{Second iteration:} & \mathbf{f}^{(2)} & \leftarrow & \mathrm{p}(\mathbf{f} \mid \mathbf{h}^{(1)}, \Omega^{(1)}, \mathbf{g}) \\
& \mathbf{h}^{(2)} & \leftarrow & \mathrm{p}(\mathbf{h} \mid \mathbf{f}^{(2)}, \Omega^{(1)}, \mathbf{g}) \\
& \Omega^{(2)} & \leftarrow & \mathrm{p}(\Omega \mid \mathbf{f}^{(2)}, \mathbf{h}^{(2)}, \mathbf{g}) \\
& \vdots & & \vdots \\
t^{\text{th}} \text{ iteration:} & \mathbf{f}^{(t)} & \leftarrow & \mathrm{p}(\mathbf{f} \mid \mathbf{h}^{(t-1)}, \Omega^{(t-1)}, \mathbf{g}) \\
& \vdots & & \vdots
\end{array}
$$

where the symbol $\leftarrow$ means the value is drawn from the distribution on the right. Notice the similarity to the iterative procedure for the Variational Bayesian approach in Equation (1.44), where instead of drawing samples we are taking expectations of the same distributions. Similarly, with the use of *Simulated Annealing* [38], the ICM formulation can be considered a deterministic approximation of the sampling, where the conditional distributions are

replaced by degenerate distributions at their modes (termed "instantaneous freezing" in [66]).

Once we have accumulated samples, point estimates and other statistics of the distribution may be found using Monte Carlo integration; for example, to find the MMSE estimate of the $\mathbf{f}$ we simply take the mean of the samples, $\frac{1}{n}\sum_{t=1}^{n}\mathbf{f}^{(t)}$.

Clearly, these methods can provide solutions closer to the optimal one than AM or any of the other methods. However, they are very computationally intensive by comparison, and although in theory convergence to the posterior is guaranteed, in practice it can be hard to tell when this has occurred; it may take a long time to explore the parameter space. Sampling methods could, for instance, be of use for the boundary condition model proposed in [82] where, because of the blur prior used, both direct inference as well as approximations by variational methods are difficult to perform.

1.7 Non–Bayesian Blind Image Deconvolution Models

In this section other blind deconvolution models, which have appeared in the literature and cannot be obtained from the Bayesian formulation described in the previous sections, are briefly reviewed.

1.7.1 Spectral and Cepstral Zero Methods

These algorithms fall into the *a priori* class of approaches described in Section 1.3, according to which the PSF of the blurring system is estimated separately from the image. They were the first examples of BD to be developed for images (see Stockham *et al.* [3] and Cannon [4]).

These blur identification algorithms are well suited to the problem when the frequency response of the blurring system has a known parametric form that is completely characterized by its frequency domain zeros: this is the case for linear motion blur and circular defocusing blur (Section 1.5.2). Let us rewrite Equation (1.2) in the frequency domain, while ignoring the noise term, i.e.,

$$G(\boldsymbol{\omega}) = F(\boldsymbol{\omega})\, H(\boldsymbol{\omega}). \tag{1.45}$$

Due to the multiplication of the spectra, the zeros of $G(\boldsymbol{\omega})$ are the zeros of $F(\boldsymbol{\omega})$ and $H(\boldsymbol{\omega})$ combined. Therefore the problem reduces to identifying which of the zeros of $G(\boldsymbol{\omega})$ belong to $H(\boldsymbol{\omega})$. The use of the parametric model makes this possible: the Fourier Transform (FT) of the linear and circular blurs are Sinc and Bessel functions that have periodic patterns of zeros in their spectra. The spacing between the zeros depends on the parameter L in Equation (1.9) or r in Equation (1.12). Hence if this pattern can be detected

in $G(\boldsymbol{\omega})$, the parameters can be identified.

In practice this type of identification often fails due to the presence of noise masking the periodic patterns. The *homomorphic* [3] or *Cepstral* [4] methods attempt to exploit the effect of the nonstationary image and stationary blur to alleviate this difficulty. The homomorphic procedure begins by partitioning the image into blocks f_i, of size larger than the PSF. Each blurred block g_i is equal to the blur convolved with the unblurred block:

$$g_i(\mathbf{x}) = f_i(\mathbf{x}) * h_i(\mathbf{x}) + n_i(\mathbf{x}). \tag{1.46}$$

This expression holds apart from at the block boundaries, due to contamination from neighbouring blocks (in practice the blocks are windowed to reduce these edge effects). The log operator may then be applied to the FT of the blurred image, which has the effect of converting the original convolution to an addition. The average of these blocks may then be calculated (assuming N blocks):

$$\frac{1}{N}\sum_{i=1}^{N}\log(G_i(\boldsymbol{\omega})) \approx \frac{1}{N}\sum_{i=1}^{N}\log(F_i(\boldsymbol{\omega})) + \log(H_i(\boldsymbol{\omega})). \tag{1.47}$$

This summation now consists of the average of the contributions from the blocks $H_i(\boldsymbol{\omega})$, which are assumed to be equal, and $F_i(\boldsymbol{\omega})$, which are not, since the image blocks have varying spectral content. Therefore the blur component should tend to dominate in the summation. It may be possible to remove this average image component, subtracting the average of a collection of representative unblurred images.

As an alternative to Equation (1.47), the *Power Cepstrum* may be used, which involves taking the FT of the log power spectra of the signals. Combined with the block-based method just described, the result is that a large spike will occur in the Cepstral domain wherever there was a periodic pattern of zeros in the original Fourier domain; the distance of this spike from the origin represents the spacing between the zeros and hence may be used to identify the parameters of the blurs. The large contribution from the blur term repeated in each block will dominate, and the high-frequency noise and spectrally averaged image content tend to be separated from the blur spike in the Cepstral domain. These methods have been extended to use the Power Bispectrum instead in [83] which shows improved performance in low SNR conditions. However, these methods are all limited to the parametric PSF models.

1.7.2 Zero Sheet Separation Algorithms

Lane and Bates [6] have shown that any signal $g(\mathbf{x})$, formed by multiple convolutions, in theory, is *automatically deconvolvable*, provided its dimension is greater than one. This argument rests on the analytical properties of the Z-Transform (ZT) of N-dimensional signals with finite support, which is

necessarily zero on $(2N-2)$-dimensional hypersurfaces in a $2N$-dimensional space.

The method assumes that there is no additive noise, i.e., the relation $G(z_1, z_2) = H(z_1, z_2)F(z_1, z_2)$ holds, where $F(z_1, z_2)$ and $H(z_1, z_2)$ are the 2-D ZTs of the original image and blur. Then the task is to separate the 2-D ZT of the blurred image, $G(z_1, z_2)$, into the two convolutive factors. As $G(z_1, z_2)$ is a polynomial in z_1, z_2, the solution is equivalent to factorizing the polynomial, i.e., identifying the zeros that belong to each component.

In order to do this, several assumptions are made: the convolutive factors (image and blur) should have compact (finite) support; the ZT of each factor should be zero on a single continuous surface, its *zero sheet*; that these zero sheets do not coalesce — they only intersect at discrete points (which it is suggested often holds in practice).

Zero sheets appear conceptually useful in analyzing the BD problem; however, in practice their applicability is limited. The main problem of this method is its high computational complexity and sensitivity to noise.

1.7.3 ARMA Parameter Estimation Algorithms

The estimation of the ARMA parameters in the model described in Section 1.5.3.2 can be done in many different ways; in addition to the Bayesian approaches already described, there also exist Generalized Cross-Validation (GCV) and neural network-based algorithms, which are based on second-order statistics. Also, Higher-Order Statistics (HOS) approaches can be used for estimating the ARMA parameters for non-Gaussian models (it should be noted that with the second-order statistics algorithms the phase of the PSF can not be recovered, unless it is assumed that the PSF is minimum phase). The GCV algorithm will now be described.

Cross-validation is a parameter estimation method that partitions the observed data into two sets. Given some parameter values, the estimation set is used to form a prediction, and then the validation set is used to test the validity of this prediction. If the parameter values used were close to the correct ones, the validation criterion will be small. To make full use of the data, the criterion may be averaged across all the choices of estimation sets, using only one data element for each validation set. Hence the procedure is also referred to as the "leave one out" algorithm. Prior to its use in BD, GCV was used in the estimation of the regularization parameter for classical restoration [84, 85].

When GCV is used to solve the BD problem [86], the original image $\hat{\mathbf{f}}$ is estimated using all but one of the pixels, $\mathbf{x}$, from the degraded image $\mathbf{g}$, for a particular value of the parameter set $\Theta = \{\mathbf{h}, \Omega\}$ under test. The validation criterion is then to test the difference between the pixel $\mathbf{x}$ of $\mathbf{g}$ that was left out, and the corresponding reblurred pixel in $\mathbf{H}\hat{\mathbf{f}}$. This is averaged over all choices of $\mathbf{x}$. With the GCV criterion defined, a numerical search technique

may be used to search for the parameter set that minimizes the criterion.

1.7.4 Nonparametric Deterministic Constraints Algorithms

This class of algorithms differs from the other joint blur and image identification methods in that they do not explicitly model the original image or the PSF with a stochastic or deterministic model. Instead, they typically use an iterative formulation to impose deterministic constraints on the unknowns at each step. These deterministic constraints may include nonnegativity, finite support, and energy bounds on the image, the blur, or both. These constraints are incorporated into an optimality criterion which is minimized with numerical iterative techniques.

Examples in the literature based on deterministic constraint principles include the Iterative Blind Deconvolution (IBD) algorithm [7, 87, 88], McCallum's simulated annealing algorithm [89], the Nonnegativity And Support constraints with Recursive Image Filtering (NAS-RIF) algorithm [90], and the blind superresolution algorithm [91], among others. Some of these methods will now be briefly described.

1.7.4.1 The Iterative Blind Deconvolution Algorithms

One of the early IBD algorithms is the one proposed by Ayers and Dainty [7]. In addition to nonnegativity and finite support, it uses Wiener-like constraints to estimate image and blur in the Fourier domain at each iteration. Beginning with an initial random PSF estimate $\widehat{h}_0(\mathbf{x})$ and image estimate $\widehat{f}_0(\mathbf{x})$, the following sequence defines the algorithm at iteration i:

1. Find $\widehat{F}_i(\mathbf{k})$, the DFT of $\widehat{f}_i(\mathbf{x})$.
2. Impose blur constraints in the Fourier domain (where $(.)^*$ denotes complex conjugation):

$$\widetilde{H}_i(\mathbf{k}) = \frac{G(\mathbf{k})\widehat{F}_i^*(\mathbf{k})}{\left|\widehat{F}_i(\mathbf{k})\right|^2 + \alpha / \left|\widehat{H}_{i-1}(\mathbf{k})\right|^2} \tag{1.48}$$

3. Find $\widetilde{h}_i(\mathbf{x})$, the IDFT of $\widetilde{H}_i(\mathbf{k})$.
4. Impose spatial domain positivity and finite support constraints on $\widetilde{h}_i(\mathbf{x})$ to give $\widehat{h}_i(\mathbf{x})$
5. Find $\widehat{H}_i(\mathbf{k})$, the DFT of $\widehat{h}_i(\mathbf{x})$.
6. Impose image constraints in the Fourier domain:

$$\widetilde{F}_i(\mathbf{k}) = \frac{G(\mathbf{k})\widehat{H}_i^*(\mathbf{k})}{\left|\widehat{H}_i(\mathbf{k})\right|^2 + \alpha / \left|\widehat{F}_i(\mathbf{k})\right|^2}, \tag{1.49}$$

7. Find $\widetilde{f}_i(\mathbf{x})$, the IDFT of $\widetilde{F}_i(\mathbf{k})$,
8. Impose spatial domain positivity and finite support constraints on $\widetilde{f}_i(\mathbf{x})$ to give $\widehat{f}_{i+1}(\mathbf{x})$.
9. Next iteration: set $i = i + 1$; go to step 1.

The real constant α represents the energy of the additive noise, and it has to be carefully chosen in order to obtain a reliable restoration. While this method is intuitively appealing, its convergence properties are undefined and tend to be highly sensitive to the initial guess [10]. Notice the similarity to the EM algorithm in the Fourier domain [29, 35]; however, the IBD algorithm is heuristically derived, and does not include estimation of the image and noise model parameters.

Further IBD-type algorithms using set-theoretic projection have been proposed by Yang *et al.* [92] and Lane [93], for the special case of astronomical speckle imaging.

1.7.4.2 The NAS-RIF Algorithms

The NAS-RIF method by Kundur and Hatzinakos [10, 90] is a similar method to IBD. It seeks to minimize a cost function at each step, by updating an FIR restoration filter, which is convolved with $\mathbf{g}$ to give an estimate $\hat{\mathbf{f}}$. The cost function is based on the same constraints as those in IBD, apart from that no assumptions are made on the PSF other than it having an inverse, both of which must be absolutely summable. The constraints are applied via the method of Projection Onto Convex Sets (POCS). NAS-RIF seems to have been fairly successful in its goals, with good convergence properties and reasonable quality restorations. However, it is only applicable to the class of images with finite support, i.e., it is entirely contained in the image frame on a uniform black background. This may include applications in medical imaging and astronomy, but prevents its widespread use with other natural images. Extensions to this method have appeared in [94, 95].

1.7.5 Nonparametric Algorithms based on Higher-Order Statistics

The principle of the HOS algorithms [96] – [99] is to use a nonlinear or non-Gaussian representation of the original image, allowing higher -order moments of the signal to be represented. These models have been typically applied when the image or its edges are modeled as sparse "spike-like" signals (for example, star fields).

In order to exploit HOS, an adaptive filter combined with a nonlinearity is used to restore the blurred image. The filter is updated to give a restored image that best fits the model. The adaptive filter structure for the HOS algorithms is shown in Figure 1.8.

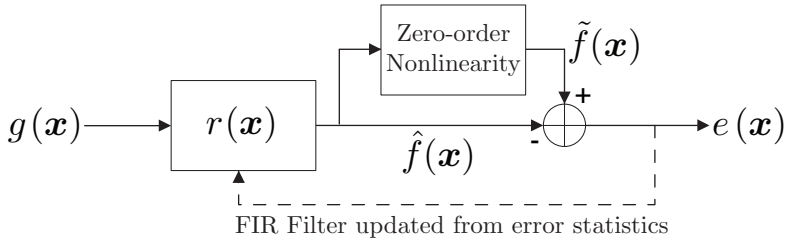

FIGURE 1.8: The blind image deconvolution based on the higher order statistics.

Note that $g(\mathbf{x})$ represents the degraded image, $\widehat{f}(\mathbf{x})$ represents the original image estimate, $\widetilde{f}(\mathbf{x})$ represents the output of the zero-order nonlinearity, $r(\mathbf{x})$ represents the restoration filter used to obtain the original image estimate, and $e(\mathbf{x})$ represents an error sample. The restoration filter $r(\mathbf{x})$ is optimized in order to minimize some cost function J that involves a sequence of error samples $e(\mathbf{x})$.

1.7.6 Total Least Squares (TLS)

Total Least Squares approaches are extensions of the standard least squares methods. The PSF is assumed to be the sum of a deterministic and a stochastic component, that is, $\mathbf{H} = \overline{\mathbf{H}} + \delta\mathbf{H}$. Using this, the degradation model in Equation (1.3) can be expressed as

$$\overline{\mathbf{g}} + \delta\mathbf{g} = (\overline{\mathbf{H}} + \delta\mathbf{H})\mathbf{f}, \tag{1.50}$$

where $\overline{\mathbf{g}}$ and $\delta\mathbf{g}$ are the deterministic and stochastic components of the observation $\mathbf{g}$, respectively. The problem is formulated as a minimization of $\delta\mathbf{g}$ and $\delta\mathbf{H}$ subject to Equation (1.50). Generally, Regularized Constrained Total Least Squares (RCTLS) filters [100] are applied to find the minimum values, where the matrices are assumed to have a special form, such as BTTB. In addition to the TLS solutions already presented that use the hierarchical Bayesian framework [78], linear algebra-based solutions are applied in [101].

1.7.7 Learning-Based Algorithms

Learning-based algorithms for image restoration and blind image restoration have been recently proposed [102, 103]. The basic idea with such an approach is that the prior knowledge required for solving various (inverse) problems can be learned from training data, i.e., a set of prototype images belonging to the same (statistical) class of images to the ones processed.

Original images and their degraded versions by the known degradation operator (restoration problem) are used for designing Vector Quantizer (VQ) codebooks. The codevectors are designed using the blurred images. For each

such vector, the high frequency information obtained from the original images is also available. During restoration, the high-frequency information of a given degraded image is estimated from its low-frequency information based on the codebooks. For the BD problem, a number of codebooks are designed corresponding to various versions of the blurring function. Given a noisy and blurred image, one of the codebooks is chosen based on a similarity measure, therefore providing the identification of the blur. To make the restoration process computationally efficient, Principal Component Analysis (PCA) and VQ-Nearest Neighborhood approaches are utilized in [103].

1.7.8 Methods for Spatially Varying Degradation

Blind deconvolution in case of Spatially Varying (SV) degradation is a more difficult problem than the spatially invariant case. The blur is generally assumed to be varying smoothly or piecewise-smoothly, and the variation to be slow in the spatial domain. The standard EM procedure has been extended to use sectioned methods where the image is divided into blocks [104, 105]. A hierarchical sliding window approach with the local Fourier transform is employed in [106]. In [107], SV PSF identification for a known image is considered using an MRF model for the parameterization of the SV blur. Some of the spatially invariant methods described in previous sections are also extended to the SV blur case, for example, a parameterized piecewise-smooth degradation model is used to extend the anisotropic regularization-based restoration method in [45].

1.7.9 Multichannel Methods

Multichannel images are typically acquired using an imaging system with multiple sensors, multiple time instants, or multiple frequency bands. Examples of multichannel images include multispectral images, where different channels present different frequency bands, wave radiometric images, and image sequences, such as video. Reviews of classical and blind multichannel restoration methods are presented in [108, 109]. Multichannel methods can be classified into two approaches, Single-Input-Multiple Output (SIMO) and Multiple-Input-Multiple-Output (MIMO).

The EM approach in [35] is combined with the MIMO restoration method in [110] to obtain a blind multichannel method in [111], where cross-channel blurs are also taken into account. The SIMO multichannel restoration problem is addressed in [112] by an ARMA model, where the likelihood function maximization is performed using a steepest descent algorithm.

Another class of algorithms have been developed using Greatest Common Divisor (GCD) methods. Under the *relative coprimeness* condition — the channel PSFs share no common factor other than a scalar — the GCD of the outputs of the channels will be the original image. By exploiting the commutativity of the convolution operator, a matrix equation is formed involving

channel outputs and PSFs. Then, as an extension of 1-D blind equalization methods, eigenstructure properties of the matrices may be used to directly estimate the PSFs and the original image. Perfect restoration algorithms for noise-free degradation with FIR PSFs have been proposed in [113] and [114]. A similar approach with a direct estimation method has been proposed in [115]. The noise is a major problem for these approaches. To deal with noise amplification, direct vector-space methods have been proposed in [116] and [117], and sufficiency conditions are derived for exact restoration in [118]. Rav-Acha and Peleg [119] have proposed a multichannel method with an *a priori* blur identification method via an exhaustive search, and a coprimeness condition is imposed on the channel model.

Recently an extension of [113] and [114] has been proposed in [120], that exploits anisotropic regularization priors mentioned in Section 1.5.3.4. As well as the standard TV form, a more advanced Mumford-Shah regularization term is also used. Very good restoration results are achieved even in low SNR conditions. This method is extended to deal with unknown PSF support and global translational motion in [20], where a MAP formulation is utilized.

Observe that there is an inherent advantage with multichannel methods in the amount of information available to aid both blur identification and image restoration. The coprimeness condition ensures that the problem becomes less ill-posed; therefore we should expect better results than in single-channel methods.

1.8 Conclusions

In this chapter we have analyzed the methods proposed in the literature to tackle BD problems from the point of view of Bayesian modeling and inference. We have shown that most of the proposed methods can be considered as particular selections of probability distributions and inference models within the Bayesian framework. The study of inference models that go from single-point estimates to distribution simulations makes possible the introduction of image and blur models encapsulating information that goes beyond simple prior constraints.

References

[1] A. K. Katsaggelos, ed., *Digital Image Restoration.* Springer–Verlag, 1991.

[2] M. R. Banham and A. K. Katsaggelos, "Digital image restoration," *IEEE Signal Processing Magagine*, vol. 14, no. 2, pp. 24–41, 1997.

[3] T. G. Stockham, Jr., T. M. Cannon, and R. B. Ingebretsen, "Blind deconvolution through digital signal processing," *Proceedings IEEE*, vol. 63, no. 4, pp. 678–692, 1975.

[4] M. Cannon, "Blind deconvolution of spatially invariant image blurs with phase," *IEEE Transactions on Acoustic, Speech, and Signal Processing*, vol. 24, no. 1, pp. 58–63, 1976.

[5] A. M. Tekalp, H. Kaufman, and J. W. Woods, "Identification of image and blur parameters for the restoration of noncausal blurs," *IEEE Transactions on Acoustic, Speech, and Signal Processing*, vol. 34, pp. 963–972, August 1986.

[6] R. G. Lane and R. H. T. Bates, "Automatic multidimensional deconvolution," *Journal of the Optical Society of America-A*, vol. 4, no. 1, pp. 180–188, 1987.

[7] G. R. Ayers and J. C. Dainty, "Iterative blind deconvolution method and its applications," *Optics Letters*, vol. 13, no. 7, pp. 547–549, 1988.

[8] K. T. Lay and A. K. Katsaggelos, "Simultaneous identification and restoration of images using maximum likelihood estimation and the EM algorithm," in *Proc. 26th Annual Allerton Conf. on Commun., Control and Computing* (Monticello, IL), pp. 661–662, September 1988.

[9] R. L. Lagendijk, J. Biemond, and D. E. Boekee, "Blur identification using the expectation-maximization algorithm," *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing — Proceedings*, vol. 3, pp. 1397–1400, 1989.

[10] D. Kundur and D. Hatzinakos, "Blind image deconvolution," *IEEE Signal Processing Magazine*, vol. 13, no. 3, pp. 43–64, 1996.

[11] D. Kundur and D. Hatzinakos, "Blind image deconvolution revisited," *IEEE Signal Processing Magazine*, vol. 13, no. 6, pp. 61–63, 1996.

[12] J. Krist, "Simulation of HST PSFs using Tiny Tim," in *Astronomical Data Analysis Software and Systems IV* (R. A. Shaw, H. E. Payne, and J. J. E. Hayes, eds.), (San Francisco, USA), pp. 349–353, Astronomical Society of the Pacific, 1995.

[13] T. J. Schultz, "Multiframe blind deconvolution of astronomical images," *Journal of the Optical Society of America-A*, vol. 10, pp. 1064–1073, 1993.

[14] T. Bretschneider, P. Bones, S. McNeill, and D. Pairman, "Image-based quality assessment of SPOT data," in *Proceedings of the American Society for Photogrammetry & Remote Sensing*, 2001. Unpaginated CD-ROM.

[15] F. S. Gibson and F. Lanni, "Experimental test of an analytical model of aberration in an oil-immersion objective lens used in three-dimensional light microscopy," *Journal of the Optical Society of America-A*, vol. 8, pp. 1601–1613, 1991.

[16] O. Michailovich and D. Adam, "A novel approach to the 2-D blind deconvolution problem in medical ultrasound," *IEEE Transactions on Medical Imaging*, vol. 24, no. 1, pp. 86–104, 2005.

[17] M. Roggemann, "Limited degree-of-freedom adaptive optics and image reconstruction," *Applied Optics*, vol. 30, pp. 4227–4233, 1991.

[18] P. Nisenson and R. Barakat, "Partial atmospheric correction with adaptive optics," *Journal of the Optical Society of America-A*, vol. 4, pp. 2249–2253, 1991.

[19] Y. L. You and M. Kaveh, "A regularization approach to joint blur and image restoration," *IEEE Transactions on Image Processing*, vol. 5, no. 3, pp. 416–428, 1996.

[20] F. Šroubek and J. Flusser, "Multichannel blind deconvolution of spatially misaligned images," *IEEE Transactions on Image Processing*, vol. 7, pp. 45–53, July 2005.

[21] C. A. Segall, R. Molina, and A. K. Katsaggelos, "High-resolution images from low-resolution compressed video," *IEEE Signal Processing Magazine*, vol. 20, no. 3, pp. 37–48, 2003.

[22] S. Dai, M. Yang, Y. Wu, and A. K. Katsaggelos, "Tracking motion-blurred targets in video," in *International Conference on Image Processing (ICIP'06)*, Atlanta, GA, October 2006.

[23] K. Faulkner, C. J. Kotre, and M. Louka, "Veiling glare deconvolution of images produced by X-ray image intensifiers," *Third Int. Conf. on Image Proc. and Its Applications*, pp. 669–673, 1989.

[24] A. K. Jain, *Fundamentals of Digital Image Processing*. New Jersey: Prentice Hall, 1 ed., 1989.

[25] M. Bertero and P. Boccacci, *Introduction to Inverse Problems in Imaging*. Institute of Physics Publishing, 1 ed., 1998.

[26] T. E. Bishop and J. R. Hopgood, "Blind image restoration using a block-stationary signal model," in *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing — Proceedings*, May 2006.

[27] R. Molina, J. Mateos, and A. Katsaggelos, "Blind deconvolution using a variational approach to parameter, image, and blur estimation," *IEEE Transactions on Image Processing*, vol. 15, no. 12, pp. 3715–3727, December 2006. .

[28] R. L. Lagendijk, J. Biemond, and D. E. Boekee, "Identification and restoration of noisy blurred images using the expectation-maximization algorithm," *IEEE Transactions on Acoustic, Speech, and Signal Processing*, vol. 38, pp. 1180-1191, July 1990.

[29] A. K. Katsaggelos and K. T. Lay, "Maximum likelihood identification and restoration of images using the expectation-maximization algorithm," in *Digital Image Restoration* (A. K. Katsaggelos, ed.), Springer-Verlag, 1991.

[30] A. F. J. Moffat, "A theoretical investigation of focal stellar images in the photographic emulsion and application to photographic photometry," *Astronomy and Astrophysics*, vol. 3, pp. 455–461, 1969.

[31] R. Molina and B. D. Ripley, "Using spatial models as priors in astronomical image analysis," *Journal of Applied Statistics*, vol. 16, pp. 193–206, 1989.

[32] H.-C. Lee, "Review of image-blur models in a photographic system using the principles of optics," *Optical Engineering*, vol. 29, pp. 405–421, May 1990.

[33] B. D. Ripley, *Spatial Statistics*, pp. 88–90, JohnWiley, 1981.

[34] A. C. Likas and N. P. Galatsanos, "A variational approach for Bayesian blind image deconvolution," *IEEE Transactions on Signal Processing*, vol. 52, no. 8, pp. 2222–2233, 2004.

[35] K. T. Lay and A. K. Katsaggelos, "Image identification and image restoration based on the expectation-maximization algorithm," *Optical Engineering*, vol. 29, pp. 436–445, May 1990.

[36] A. K. Katsaggelos and K. T. Lay, "Maximum likelihood blur identification and image restoration using the EM algorithm," *IEEE Transactions on Signal Processing*, vol. 39, no. 3, pp. 729–733, 1991.

[37] H. Derin and H. Elliott, "Modelling and segmentation of noisy and textured images using Gibbs random fields," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-9, pp. 39–55, January 1987.

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

[39] R. R. Schultz and R. L. Stevenson, "Extraction of high-resolution frames from video sequences," *IEEE Transactions on Image Processing*, vol. 5, no. 6, pp. 996–1011, 1996.

[40] J. Zhang, "The mean field theory in EM procedures for blind Markov random field image restoration," *IEEE Transactions on Image Processing*, vol. 2, no. 1, pp. 27–40, 1993.

[41] B. A. Chipman and B. D. Jeffs, "Blind multiframe point source image restoration using MAP estimation," *Conference Record of the Asilomar Conference on Signals, Systems and Computers*, vol. 2, pp. 1267–1271, 1999.

[42] C. S. Won and R. M. Gray, *Stochastic Image Processing*. Information Technology: Transmission, Processing, and Storage, Kluwer Academic / Plenum Publishers, 2004.

[43] C. A. Bouman and K. Sauer, "Generalized Gaussian image model for edge-preserving MAP estimation," *IEEE Transactions on Image Processing*, vol. 2, no. 3, pp. 296–310, 1993.

[44] F.-C. Jeng and J. W. Woods, "Compound Gauss-Markov random fields for image estimation," *IEEE Transactions on Signal Processing*, vol. 39, no. 3, pp. 683–697, 1991.

[45] Y. L. You and M. Kaveh, "Blind image restoration by anisotropic regularization," *IEEE Transactions on Image Processing*, vol. 8, no. 3, pp. 396–407, 1999.

[46] T. F. Chan and C.-K. Wong, "Total variation blind deconvolution," *IEEE Transactions on Image Processing*, vol. 7, no. 3, pp. 370–375, 1998.

[47] A. Hamza, H. Krim, and G. Unal, "Unifying probabilistic and variational estimation," *IEEE Signal Processing Magazine*, vol. 19, no. 5, pp. 37–47, 2002.

[48] T. F. Chan and J. Shen, *Image Processing and Analysis: Variational, Pde, Wavelet, and Stochastic Methods*. SIAM, 2005.

[49] T. F. Chan, S. Osher, and J. Shen, "The digital TV filter and nonlinear denoising," *IEEE Transactions on Image Processing*, vol. 10, no. 2, pp. 231–241, 2001.

[50] J. M. Bioucas-Dias, M. A. T. Figueiredo, and J. P. Oliveira, "Total variation-based image deconvolution: a majorization-minimization approach," in *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing — Proceedings*, May 2006.

[51] Y. You, W. Xu, A. Tannenbaum, and M. Kaveh, "Behavioral analysis of anisotropic diffusion in image processing," *IEEE Transactions on Image Processing*, vol. 5, no. 11, pp. 1539–1553, 1996.

[52] J. Weickert, "A review of nonlinear diffusion filtering," in *SCALE-SPACE '97: Proceedings of the First International Conference on Scale-*

Space Theory in Computer Vision, London, UK, pp. 3–28, Springer-Verlag, 1997.

[53] Y.-L. You and M. Kaveh, "Ringing reduction in image restoration by orientation-selective regularization," *IEEE Signal Processing Letters*, vol. 3, no. 2, pp. 29–31, 1996.

[54] A. Chambolle and P.-L. Lions, "Image recovery via total variation minimization and related problems," *Numerische Mathematik*, vol. 76, no. 2, pp. 167–188, 1997.

[55] D. Geman and G. Reynolds, "Constrained restoration and the recovery of discontinuities," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 3, pp. 367–383, 1992.

[56] R. L. Lagendijk, J. Biemond, and D. E. Boekee, "Regularized iterative image restoration with ringing reduction," *IEEE Transactions on Acoustic, Speech, and Signal Processing*, vol. 36, pp. 1874–1888, December 1988.

[57] A. K. Katsaggelos, J. Biemond, R. W. Schafer, and R. M. Mersereau, "A regularized iterative image restoration algorithm," *IEEE Transactions on Signal Processing*, vol. 39, pp. 914–929, April 1991.

[58] J. O. Berger, *Statistical Decision Theory and Bayesian Analysis*, ch. 3 and 4. New York, Springer-Verlag, 1985.

[59] H. Raiffa and R. Schlaifer, *Applied Statistical Decision Theory.* Division of Research, Graduate School of Business, Administration, Harvard University, Boston, 1961.

[60] A. Gelman, J. B. Carlin, H. S. Stern, and D. R. Rubin, *Bayesian Data Analysis*, Chapman & Hall, 2003.

[61] R. M. Neal, "Probabilistic inference using Markov chain Monte Carlo methods," Tech. Rep. CRG-TR-93-1, Dept. of Computer Science, University of Toronto, 1993. available online at http://www.cs.toronto.edu/∼radford/res-mcmc.html.

[62] M. I. Jordan, Z. Ghahramani, T. S. Jaakola, and L. K. Saul, "An introduction to variational methods for graphical models," in *Learning in Graphical Models*, pp. 105–162, MIT Press, 1998.

[63] A. K. Katsaggelos, *Iterative Image Restoration Algorithms*, PhD thesis, Georgia Institute of Technology, School of Electrical Engineering, August 1985.

[64] S. N. Efstratiadis and A. K. Katsaggelos, "Adaptive iterative image restoration with reduced computational load," *Optical Engineering*, vol. 29, pp. 1458–1468, 1990.

[65] M. G. Kang and A. K. Katsaggelos, "General choice of the regularization functional in regularized image restoration," *IEEE Transactions on Image Processing*, vol. 4, no. 5, pp. 594–602, 1995.

[66] J. Besag, "On the statistical analysis of dirty pictures," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 48, no. 3, pp. 259–302, 1986.

[67] L. Chen and K.-H. Yap, "A soft double regularization approach to parametric blind image deconvolution," *IEEE Transactions on Image Processing*, vol. 14, no. 5, pp. 624–633, 2005.

[68] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," in *Proceedings of the eleventh annual international conference of the Center for Nonlinear Studies on Experimental mathematics: computational issues in nonlinear science*, Amsterdam, The Netherlands, pp. 259–268, Elsevier North-Holland, Inc., 1992.

[69] T. F. Chan, G. H. Golub, and P. Mulet, "A nonlinear primal-dual method for total variation-based image restoration," *SIAM Journal on Scientific Computing*, vol. 20, pp. 1964–1977, November 1999.

[70] C. R. Vogel and M. E. Oman, "Iterative methods for total variation denoising," *SIAM Journal on Scientific Computing*, vol. 17, no. 1, pp. 227–238, 1996.

[71] D. Geman and C. Yang, "Nonlinear image recovery with half-quadratic regularization," *IEEE Transactions on Image Processing*, vol. 4, no. 7, pp. 932–946, 1995.

[72] R. Molina, A. K. Katsaggelos, and J. Mateos, "Bayesian and regularization methods for hyperparameter estimation in image restoration," *IEEE Transactions on Image Processing*, vol. 8, no. 2, pp. 231–246, 1999.

[73] R. Molina, "On the hierarchical Bayesian approach to image restoration. Applications to Astronomical images," *IEEE Transactions on Pattern Analysis and Machine Intell.*, vol. 16, no. 11, pp. 1122–1128, 1994.

[74] A. D. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the E-M algorithm," *Journal of the Royal Statistical Society: Series B*, vol. 39, pp. 1–37, 1977.

[75] R. Kass and A. E. Raftery, "Bayes factors," *Journal of the American Statistical Association*, vol. 90, pp. 773–795, 1995.

[76] D. J. C. MacKay, "Probable networks and plausible predictions: a review of practical Bayesian methods for supervised neural networks," *Network: Computation in Neural Systems*, no. 6, pp. 469-505, 1995.

[77] N. P. Galatsanos, V. Z. Mesarovic, R. Molina, A. K. Katsaggelos, and J. Mateos, "Hyperparameter estimation in image restoration problems with partially-known blurs," *Optical Engineering*, vol. 41, no. 8, pp. 1845–1854, 2002.

[78] N. P. Galatsanos, V. Z. Mesarovic, R. Molina, and A. K. Katsaggelos, "Hierarchical Bayesian image restoration for partially-known blur," *IEEE Transactions on Image Processing*, vol. 9, no. 10, pp. 1784–1797, 2000.

[79] S. Kullback, *Information Theory and Statistics*, New York, Dover Publications, 1959.

[80] C. Andrieu, N. de Freitras, A. Doucet, and M. Jordan, "An introduction to MCMC for machine learning," *Machine Learning*, vol. 50, pp. 5–43, 2003.

[81] J. J. K. Ó Ruanaidh and W. Fitzgerald, *Numerical Bayesian Methods Applied to Signal Processing*. Springer Series in Statistics and Computing, New York: Springer, 1 ed., 1996.

[82] R. Molina, A. K. Katsaggelos, J. Abad, and J. Mateos, "A Bayesian approach to blind deconvolution based on Dirichlet distributions," in *1997 International Conference on Acoustics, Speech and Signal Processing (ICASSP'97)*, vol. IV, Munich, Germany, pp. 2809–2812, 1997.

[83] M. M. Chang, A. M. Tekalp, and A. T. Erdem, "Blur identification using the bispectrum," *IEEE Transactions on Signal Processing*, vol. 39, no. 10, pp. 2323–2325, 1991.

[84] N. P. Galatsanos and A. K. Katsaggelos, "Methods for choosing the regularization parameter and estimating the noise variance in image restoration and their relation," *IEEE Transactions on Image Processing*, vol. 1, pp. 322–336, 1992.

[85] S. J. Reeves and R. M. Mersereau, "Optimal estimation of the regularization parameters and stabilizing functional for regularized image restoration," *Optical Optical Engineering*, vol. 29, no. 5, pp. 446–454, 1990.

[86] S. Reeves and R. Mersereau, "Blur identification by the method of generalized cross-validation," *IEEE Transactions on Image Processing*, vol. 1, pp. 301–311, July 1992.

[87] N. Miura and N. Baba, "Extended-object reconstruction with sequential use of the iterative blind deconvolution method," *Optics Communications*, vol. 89, pp. 375–379, 1992.

[88] N. M. F. Tsumuraya and N. Baba, "Iterative blind deconvolution method using Lucy's algorithm," *Astronomy and Astrophysics*, vol. 282, no. 2, pp. 699–708, 1994.

[89] B. C. McCallum, "Blind deconvolution by simulated annealing," *Optics Communications*, vol. 75, no. 2, pp. 101–105, 1990.

[90] D. Kundur and D. Hatzinakos, "A novel blind deconvolution scheme for image restoration using recursive filtering," *IEEE Transactions on Signal Processing*, vol. 46, no. 2, pp. 375–390, 1998.

[91] K. Nishi and S. Ando, "Blind superresolving image recovery from blur-invariant edges," *IEEE Transactions on Acoustic Speech, and Signal Processing*, vol. 5, pp. 85–88, 1994.

[92] Y. Yang, N. P. Galatsanos, and H. Stark, "Projection based blind deconvolution," *Journal of the Optical Society of America-A*, vol. 11, no. 9, pp. 2401–2409, 1994.

[93] R. G. Lane, "Blind deconvolution of speckle images," *Journal of the Optical Society of America-A*, vol. 9, pp. 1508–1514, September 1992.

[94] M. Ng, R. Plemmons, and S. Qiao, "Regularization of RIF blind image deconvolution," *IEEE Transactions on Image Processing*, vol. 9, no. 6, pp. 1130–1134, 2000.

[95] C. A. Ong and J. Chambers, "An enhanced NAS-RIF algorithm for blind image deconvolution," *IEEE Transactions on Image Processing*, vol. 8, no. 7, pp. 988–992, 1999.

[96] P. Campisi and G. Scarano, "A multiresolution approach for texture synthesis using the circular harmonic functions," *IEEE Transactions on Image Processing*, vol. 11, pp. 37–51, January 2002.

[97] G. Panci, P. Campisi, C. Colonnese, and G. Scarano, "Multichannel blind image deconvolution using the Bussgang algorithm: spatial and multiresolution approaches," *IEEE Transactions on Image Processing*, vol. 12, pp. 1324–1337, November 2003.

[98] H. S. Wu, "Minimum entropy deconvolution for restoration of blurred two-tone images," *Electronics Letters*, vol. 26, no. 15, pp. 1183–1184, 1990.

[99] R. A. Wiggins, "Minimum entropy deconvolution," *Geoexploration*, vol. 16, pp. 21–35, 1978.

[100] V. Z. Mesarovic, N. P. Galatsanos, and A. K. Katsaggelos, "Regularized constrained total least–squares image restoration," *IEEE Transactions on Image Processing*, vol. 4, pp. 1096–1108, August 1995.

[101] N. Mastronardi, P. Lemmerling, S. V. Huffel, A. Kalsi, and D. O'Leary, "Implementation of regularized structured total least squares algorithms for blind image blurring," *Linear Algebra and Its Applications*, vol. 391, no. 1–3, pp. 203–221, 2004.

[102] K. Panchapakesan, D. G. Sheppard, M. W. Marcellin, and B. R. Hunt, "Blur identification from vector quantizer encoder distortion," *IEEE Transactions on Image Processing*, vol. 10, pp. 465–470, March 2001.

[103] R. Nakagaki and A. K. Katsaggelos, "A VQ-based blind image restoration algorithm," *IEEE Transactions on Image Processing*, vol. 12, pp. 1044–1053, September 2003.

[104] R. L. Lagendijk and J. Biemond, "Block–adaptive image identification and restoration," *Proceedings ICASSP, IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 4, pp. 2497–2500, 1991.

[105] Y. P. Guo, H. P. Lee, and C. L. Teo, "Blind restoration of images degraded by space-variant blurs using iterative algorithms for both blur identification and image restoration," *Image and Vision Computing*, vol. 15, pp. 399–410, May 1997.

[106] M. K. Ozkan, A. Tekalp, and M. Sezan, "Identification of a class of space-variant image blurs," *Proceedings of SPIE – The International Society for Optical Engineering*, vol. 1452, pp. 146–156, 1991.

[107] A. Rajagopalan and S. Chaudhuri, "MRF model-based identification of shift-variant point spread function for a class of imaging systems," *Signal Processing*, vol. 76, no. 3, pp. 285–299, 1999.

[108] F. Šroubek and J. Flusser, "An overview of multichannel image restoration techniques," in *Week of Doctoral Students* (J. Safrnkov, ed.), Prague, pp. 580–585, Matfyzpress, 1999.

[109] N. P. Galatsanos, M. Wernick, and A. K. Katsaggelos, "Multi-channel image recovery," in *Handbook of Image and Video Processing* (A. Bovik, ed.), ch. 3.7, pp. 161–174, Academic Press, 2000.

[110] A. K. Katsaggelos, K. T. Lay, and N. Galatsanos, "A general framework for frequency domain multichannel signal processing," *IEEE Transactions on Image Processing*, vol. 2, no. 3, pp. 417–420, 1993.

[111] B. C. Tom, K. Lay, and A. K. Katsaggelos, "Multichannel image identification and restoration using the expectation-maximization algorithm," *Optical Engineering*, vol. 35, no. 1, pp. 241–254, 1996.

[112] A. Rajagopalan and S. Chaudhuri, "A recursive algorithm for maximum likelihood-based identification of blur from multiple observations," *IEEE Transactions on Image Processing*, vol. 7, no. 7, pp. 1075–1079, 1998.

[113] G. Harikumar and Y. Bresler, "Perfect blind restoration of images blurred by multiple filters: theory and efficient algorithms," *IEEE Transactions on Image Processing*, vol. 8, pp. 202–219, February 1999.

[114] G. Harikumar and Y. Bresler, "Exact image deconvolution from multiple FIR blurs," *IEEE Transactions on Image Processing*, vol. 8, no. 6, pp. 846–862, 1999.

[115] G. Giannakis and R. J. Heath, "Blind identification of multichannel FIR blurs and perfect image restoration," *IEEE Transactions on Image Processing*, vol. 9, no. 11, pp. 1877–1896, 2000.

[116] H. Pai and A. C. Bovik, "Exact multichannel blind image restoration," *IEEE Signal Processing Letters*, vol. 4, no. 8, pp. 217–220, 1997.

[117] H. Pai and A. C. Bovik, "On eigenstructure-based direct multichannel blind image restoration," *IEEE Transactions on Image Processing*, vol. 10, no. 10, pp. 1434–1446, 2001.

[118] H. Pai, J. Havlicek, and A. C. Bovik, "Generically sufficient conditions for exact multichannel blind image restoration," *Proceedings of the 1998 IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 5, pp. 2861–2864, 1998.

[119] A. Rav-Acha and S. Peleg, "Two motion-blurred images are better than one," *Pattern Recognition Letters*, vol. 26, pp. 311–317, 2005.

[120] F. Šroubek and J. Flusser, "Multichannel blind iterative image restoration," *IEEE Transactions on Image Processing*, vol. 12, no. 9, pp. 1094–1106, 2003.