BAYESIAN PARADIGMS IN IMAGE PROCESSING

ZHI-QIANG LIU

Computer Vision and Machine Intelligence Lab. (CVMIL)
Department of Computer Science, The University of Melbourne
Parkville, Victoria 3052, Australia
E-mail: zliu@cs.mu.OZ.AU

A large number of image and spatial information processing problems involves the estimation of the intrinsic image information from observed images, for instance, image restoration, image registration, image partition, depth estimation, shape reconstruction and motion estimation. These are *inverse* problems and generally ill-posed. Such estimation problems can be readily formulated by Bayesian models which infer the desired image information from the measured data. Bayesian paradigms have played a very important role in spatial data analysis for over three decades and have found many successful applications. In this paper, we discuss several aspects of Bayesian paradigms: uncertainty present in the observed image, prior distribution modeling, Bayesian-based estimation techniques in image processing, particularly, the maximum *a posteriori* estimator and the Kalman filtering theory, robustness, and Markov random fields and applications.

Keywords: Bayesian models, Markov random fields, image processing, Gaussian distribution, robustness, Kalman filter.

1. INTRODUCTION

In low-level vision tasks there are basically two types of processes: *image processing* such as image restoration, partition, and registration; *intrinsic image formation*[2] which includes estimation of depth, surface, motion, and stereo. All these processes require to extract or to recover certain properties given the observed image data. For instance, in image restoration, we are required to remove degradations due to noise and blurring introduced by the imaging sensor; whereas in image partition our task is to partition the image field into patches based on some homogeneity criteria (e.g. textures, surfaces in the image). In motion estimation, we need to infer qualitative and/or quantitative object motion information from observed image sequences. As low-level vision is the basic and integral part of any vision system, it has great impact on the usefulness of the entire system and deserves considerable attention and efforts in designing vision systems.

Due mainly to the physical limitations present in the imaging device and imaging conditions, the observed image is usually the result of mixed and distorted sources of information such as illumination, reflectivity, surface shape, and projective geometry. The low-level vision process produces an improved version of the observed image in terms of reduced noise, good contrast, or less distortion; and useful intrinsic image information in terms of depth, surface shape, or object motion. These in turn make higher-level vision tasks possible.

The problems associated with the low-level vision are to a large extent of an *inverse* nature.[42] The inverse problem arising in low-level vision is generally ill-posed[81] and difficult to solve due to lack of constraint and incomplete information — sometimes very little knowledge is available about what is being viewed. There

Spatial Computing: Issues in Vision, Multimedia and Visualization Technologies (1997) 3–33

are basically two approaches to this problem: regularization[93] which imposes additional weak smoothness constraints in the form of stabilizers. In order to solve such problems certain assumptions are necessary based on experience or previous data. For example, in 3D surface reconstruction from depth maps, surface smoothness is usually assumed. Another approach is based on the Bayesian theorem, which assumes a prior statistical distribution for the data being estimated, and models the image formation and sensing phenomena as stochastic or noisy processes. For example, when solving the image restoration problem it is assumed that the probability distribution of the contaminating noise is known and that the blurring effects can be modeled by some transfer functions. It is interesting to also notice that regularization is in fact a special case of the more general Bayesian approach to the formulation of inverse problems.

As mentioned before, the physical limitations in the imaging process often result in uncertainties in the image data. Imaging sensory induced uncertainties are (a) noise, (b) blurring, (c) incomplete data to sufficiently obtain a unique solution or it may require additional measurements (data-fusion). From the modeling point of view, we may have the following uncertainties:

(a) inaccurate prior models and unknown parameters due to insufficient knowledge about the nature of the process,
(b) space and time variables in many image processing problems.

In order to develop highly robust vision systems it is imperative that these uncertainties be handled and quantified at the early stages. Estimation techniques should be able to recover as much as possible the intrinsic nature or features in the image. Bayesian paradigms allow us to handle these uncertainties and provide a natural framework for developing powerful statistical estimation methods. As the Bayesian approach uses explicit probability models to incorporate our prior knowledge about the world and sensor in image processing, it combines effects of both prior probabilities and likelihood. This enables us to use the Bayesian formula to integrate our belief in (or intuition or knowledge about) the world. Such a framework allows us to consider image processing tasks as statistical inference problems. One important aspect of the Bayesian framework is its ability to assign confidence interval estimates through probability measures. This provides estimation algorithms with additional flexibility and control.[7] Further, the Bayesian theory provides us with a unified framework for developing a variety of algorithms for image analysis. Many recursive estimation algorithms have been developed in the Bayesian framework. These methods have been successfully applied to image processing problems such as image enhancement, feature extraction, image sequence processing, depth map generation, and real-time tracking of moving objects.

In this paper we will discuss some general aspects of Bayesian paradigms in low-level vision, or more specifically, in image processing tasks. These include image restoration, texture analysis, edge and boundary description, depth estimation, surface reconstruction, stereo, object motion, and shape, etc. We will describe the modeling process, some major estimation algorithms, the problems and dilemma, and recent developments and applications.

2. MODELS IN IMAGE PROCESSING

As shown in Fig. 1, in the imaging process we generally have two planes, namely, the object plane where the object being viewed is present and the image plane which is the image sensing media such as film. The object is perceived by an intensity distribution of some radiated energy, for instance, light illumination in normal photography, laser light in range finders, and positron emission energy in PET. The energy is then collected through the imaging system (e.g. lenses) to form the image of the object.

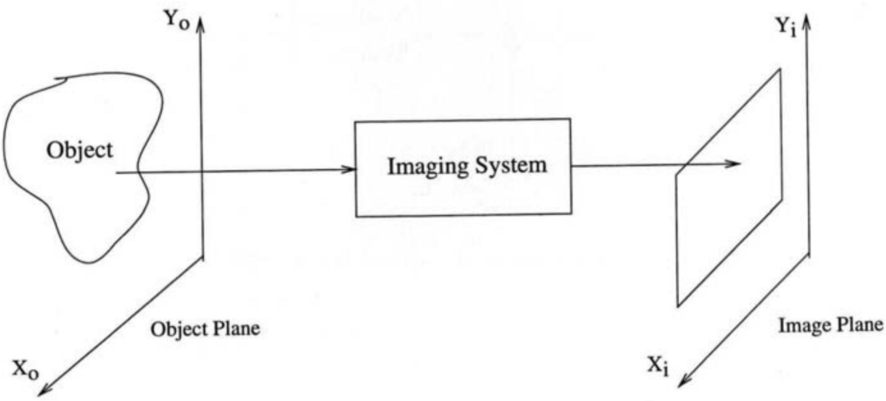

Fig. 1. Image formation.

Such a system can be depicted in terms of functional block diagrams shown in Fig. 2. The original image in fact represents the *ideal* distribution of the radiated energy (signal) f from the object. The imaging system (which may include the imaging environment, e.g. under water, atmosphere) is considered as a filter with a particular transfer (point-spread) function H which passes f. This results in blurring and geometrical distortions. For example, long term exposure of atmospheric turbulence can be modeled as

$$H_a(w_x, w_y) = \exp\left\{-\frac{w_x^2 + w_y^2}{\sigma^2}\right\}, \tag{1}$$

which is also called the Gaussian blur, whereas, horizontal camera motion blur of length $2d$ is modeled as follows,

$$h_b(i,j) = \begin{cases} 0 & j \neq 0 \ -\infty \leq i \leq \infty, \\ \dfrac{1}{2d} & j = 0 \ -d \leq i \leq d. \end{cases} \tag{2}$$

For defocused lens system with circular aperture, the blur can be approximated by a cylinder where the radius r depends on the focus defect extent:

$$h_b(i,j) = \begin{cases} 0 & \sqrt{i^2 + j^2} > r, \\ \dfrac{1}{\pi r^2} & \sqrt{i^2 + j^2} \leq r. \end{cases} \tag{3}$$

It is very difficult to identify blurring models and parameters in practice. In addition, as the image sensors (e.g. film, CCD cameras) are not ideal devices, they will introduce noise n into the final observed image g. It is often assumed that the noise is white Gaussian. However, in many imaging systems the image is formed based on counts of photon emissions (e.g. PET, SPECT). In such systems, the noise can be assumed to follow a Poisson distribution. Finally, the entire system is in general nonlinear ϕ.

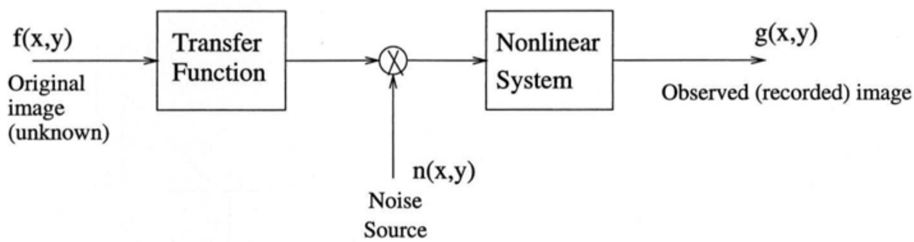

Fig. 2. Block diagram of a general imaging system.

In summary, an observed image can be generally represented by (see Fig. 2)

$$g(x,y) = \phi(Hf(x,y), n(x,y)), \tag{4}$$

where H represents the point-spread function of the imaging environment, f the original (ideal but unknown) image, and n is the noise with some assumed properties.

In many applications, it is both natural and simple to assume additive noise. Under such an assumption, (4) can be written as follows

$$g(x,y) = \phi(Hf(x,y)) + n(x,y). \tag{5}$$

The additive noise assumption in (5) is justified on the basis of the nature of image sensors. In photoelectronic systems, the output current of the sensor is usually subject to the usual sources of thermal noise, amplifier electronic noise, etc. Such noise is to a large extent additive.[86] Although film-grain noise is signal dependent and very complex, studies have shown that assuming the signal-dependent nature of the noise does not gain much in terms of processed results.[97]

Further, if we assume that the system is linear and stationary with an additive noise as shown in Fig. 3, the general model (4) takes a much simpler form

$$g(v) = H(\theta)f(v) + n(v). \tag{6}$$

where θ is the parameter vector to be estimated, and v is the variable vector: for 2D image processing problems $v = (x,y)$; for spatial-temporal image sequences, $v = (x,y,t)$, where t is time; for 3D surfaces, $v = (x,y,u(x,y))$, and (x,y) represents the coordinate of the pixel being computed.

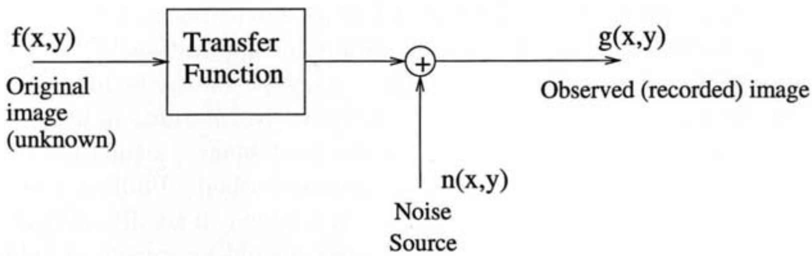

Fig. 3. Block diagram of a simplified imaging system.

3. BAYESIAN ESTIMATION

In image processing, there are different approaches to the model in (4). We will in this section first describe the classic least squares estimation method and discuss its problems and modifications. This leads naturally to the Bayesian approach which take both the prior model and the sensor model into account.

3.1. Least Squares Estimation

If (6) is to be considered as a deterministic problem, it is only a matter of finding the solution for the system. For instance, we may consider solving the following equation

$$e = g - Hf\,, \tag{7}$$

where we have dropped the variables for convenience of discussion. Letting the derivative of $||e||^2$ equal zero gives

$$H^T H f = H^T g\,, \tag{8}$$

which leads to the well-known result of least-squares estimator (LSE):

$$\hat{f} = (H^T H)^{-1} H^T g\,, \tag{9}$$

where T denotes transpose of a vector or a matrix, and $\hat{f}$ indicates that the solution is indeed an estimate of f.

It is known from the Gauss-Markov theorem that LSE is optimal in all linear estimators.[36] In addition, it does not require to know or to assume the probability models for the parameter and data. However, the least-squares solution has two critical problems:

- It requires that H must be of full rank. Otherwise, (9) cannot be solved without additional constraints, which happens very often in many applications. Furthermore, even if H is of full rank, $H^T H$ will often be ill-conditioned resulting in numerical instability.
- It does not take into account of the prior knowledge about the process and the error distribution, which is frequently available in our practice. The use of prior knowledge will in many cases simplify the computation and obtain better estimation results.

Nevertheless, LSE has been widely used in many statistical analysis ever since its introduction by Legendre in 1805. It also found some applications[46,61] in the early days of image processing. However, Sondhi[88] reported that it is difficult to solve the image restoration problem due to the ill-conditioned nature. In order to solve the numerical instability problem present in the least squares estimator, Phillips[79] proposed the *constrained* least squares estimation method. Phillips argued that instability in the numerical solutions due to the inherent ill-conditioning is in fact contrary to our prior experience that the solution should be *smooth*. Phillips proposed to use a smoothness constraint in the original criterion for LSE solutions. This smoothness constraint is sensitive to the relative smoothness of the solution vector, and as a result, improves the numerical stability of LSE.

Hunt[46] extended the constrained LSE (CLSE) to 2D systems and applied it to image restoration with some success. In the following we briefly discuss this method.

Given the estimate $\hat{f}$ *from* (9), *we reformulate the original problem to solve the* following problem. Find $\hat{f}$, such that the following cost function is minimized

$$J(\hat{f}, C) = \hat{f}^T C^T C \hat{f}\,, \tag{10}$$

subject to

$$(g - H\hat{f})^T (g - H\hat{f}) = e^T e\,. \tag{11}$$

A simple Lagrangian minimization gives the following solution

$$\hat{f} = (H^T H + \lambda C^T C)^{-1} H^T g\,, \tag{12}$$

where $\lambda = \frac{1}{\gamma}$ and γ is a Lagrange multiplier. As the constraint matrix C was defined for 1D problems[79] which cannot be used directly for 2D estimation problems, Hunt used two different approaches, namely, the Laplacian operator and radial symmetry, to generate C.[46] Their experiments show that the proposed CLSE is effective for Gaussian and Dirac impulse point-spread functions and uniformly distributed white noise.

3.2. Bayesian Approach

If we consider that the image processing problem is of a stochastic nature and nonlinear (as shown in Eq. (5)) which requires us to estimate f given observation g and based on the prior knowledge about the noise. This leads naturally to the Bayesian approach:

$$p(f|g) = \frac{p(g|f)p(f)}{p(g)}\,, \tag{13}$$

where $p(g) = \sum_f p(g|f)$.

The Bayesian model (13) describes the problem of estimation of f given the observation g in terms of conditional probabilities with two major components:

- the prior model $p(f)$,
- the sensor model $p(g|f)$.

In other words, the *a posteriori* probability $p(f|g)$ is an expression relating to what we have learned (the prior knowledge) and what we have measured (observation) as shown in Fig. 4.

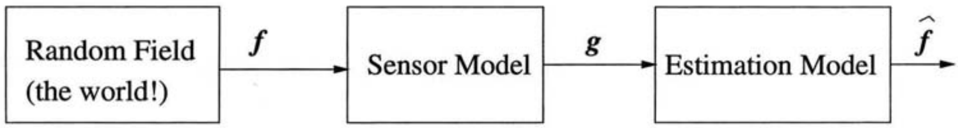

Fig. 4. The Bayesian estimation structure.

The use of Bayesian modeling for estimation in both signal processing and system identification has been well-known. Given such a model, a minimum error estimator indicates that an instance $\hat{f}$ of f is present in g if $p(\hat{f}|g)$ is maximum over all possible models. Such an estimator is called a *maximum a posteriori* (MAP) estimator. Given that the probability density of g is independent of the hypothesis about f, our task is then to find $\hat{f}$ based on the minimum error criterion such that $p(g|\hat{f})p(\hat{f})$ is maximized. It is obvious that the *a posteriori* probability density function plays a central role in the general nonlinear estimation problem in the Bayesian framework.

If $p(f|g)$ is *available*, an estimate of f can be determined from the posterior distribution in several ways. For instance, a common choice is the mean of the posterior distribution, i.e. the conditional expectation

$$\hat{f}_k = E(f_k|g_0, g_1, \ldots, g_{k-1}), \tag{14}$$

where the subscript denotes discrete time sequence. Particularly, for symmetric posterior distributions, such as Gaussian, the estimate coincides with that given in (14). Interestingly, $\hat{f}_k$ is also the value that minimizes the parameter error variance $E|f - \hat{f}_k|^2$ under very general conditions.

In practice, if the model for the prior model $p(f)$ is not known precisely, we may assume that all models are *equally* likely. This assumption results in a uniform prior distribution (which is also called *non-informative* prior) for the *a priori* probability. Consequently, we will obtain the maximum likelihood (ML) estimation which in this case is equivalent to MAP. Furthermore, it can be easily shown that for the special case of additive white Gaussian noise, MLE is equivalent to LSE.[75,90]

To summarize, the Bayesian estimation approach requires us to specify the following information:

- the world model (or the random field) parameter vector:
  - image fields for grey-level, color, or multi-spectrum images,
  - spatial-temporal random fields for dynamic image sequences,
  - different modality models.
- the sensory system parameter vector: camera, laser range finder, tomographic imaging devices (CT, MRI, PET, SPECT, etc.), remote sensing devices, etc.

Proper assumptions about the prior model are essential for the Bayesian approach, which will help the estimator to perform efficiently with good accuracy. Under the Bayesian estimation structure, the estimation paradigm is to

- construct a cost function $J(\hat{f}, f)$,
- find the estimate value $\hat{f}$ of f,
- the objective is to minimize (or maximize) the averaged or expected value $E[J(\hat{f}, f)]$.

If the posterior distribution function $p(f|g)$ is available, if at all, the task in the Bayesian framework is then to find the value (estimate) $\hat{f}$ of f, such that the *a posteriori* $p(f|g)$ is maximized (MAP):

$$max_f p(f|g) = max_f\{\log p(g|f) + \log p(f)\}\,. \tag{15}$$

It is very important to realize, however, that in most applications, the posterior $p(f|g)$ is very difficult to obtain without considering specific random fields. As mentioned above, if our assumption is too simple and unrealistic it is likely that the MAP algorithm we developed for a particular image processing problem will be equivalent to LSE which has been shown[36,44,90] to be non-robust — a small aberration in data will tip over the scale resulting in the wrong estimate.

4. IMAGE PROCESSING USING MAP

As discussed in the previous section, given observation g in (5) and the *a posteriori* model $p(f|g)$ we can develop estimation algorithms based on the Bayes theory, namely,

$$p(f|g) = \frac{p(g|f)p(f)}{p(g)}. \tag{16}$$

Obviously, in such a framework, the estimate depends primarily on two models, namely, the prior probability density function $p(f)$ and the likelihood function $p(g|f)$ which according to (5) is governed by $p(n)$ the density function for the additive noise n. In choosing particular forms of density functions we often have to base our judgments on the following criteria,

(a) the underlying physics that govern the imaging process,
(b) tractable mathematical representation,
(c) computational feasibility,
(d) our knowledge and experience about the nature of the object being viewed.

For instance, it is possible to model the noise as a white Gaussian process for both film and photoelectronic imaging media,[8,72] whereas in many imaging systems based on counts of photon emission, for example, X-ray, PET (Positron Emission Tomography) and SPECT (Single-Photon Emission Computerized Tomography), the process can be modeled by a Poisson distribution.[34,52] However, the Gaussian assumption applies to MRI (Magnetic Resonance Imaging) where the major source of noise is the thermal noise associated with the resistive loading of the collecting coil.[26,43] Most statistical image processing, or for that matter, signal processing methods have used the Gaussian model for noise. Although Hunt and Cannon[48] have pointed out that Gaussian assumption for modeling images is not evident, the Gaussian modeling is nevertheless justified by the need for a *workable* mathematical formulation.

The Gaussian or normal model occupies a unique position in both mathematics and engineering. In image processing it is not an exception. It is due to the following reasons that the Gaussian distribution is considered as a suitable model: the central-limit theorem states that the density function $p(s)$ for the sum of the independent random variable s_i

$$s = \sum_{i=0}^{M} s_i \,, \tag{17}$$

approaches a bell-shaped curve referred to as the Gaussian distribution as $M \to \infty$. However, as this condition cannot be reached in reality, we may have to settle for a *sufficiently* large M in practice. It can be shown that passing a Gaussian random signal through a linear filter results in a new signal that is also a Gaussian process. It is possible to generate non-Gaussian signals by passing a Gaussian random signal through a nonlinear system. Computationally, the Gaussian model is most efficient as it is completely determined by the statistic mean μ and covariance σ^2. Furthermore, it is known that in information theory terms, Gaussian noise is the *noisest* waveform, as this is the only noise that reaches the theoretical limit where its entropy power equals its variance.[9] We have to stress at this point that Gaussian modeling is an approximation based in many ways on our belief about how the signal *ought* to behave in the natural world.

Hunt in his pioneering work[47] proposed to model the noise process n by a multivariate normal probability density: $n \sim G(0, R_n)$:

$$p(n) = \frac{1}{(2\pi)^{M/2}|R_n|^{1/2}} \exp\left(-\frac{1}{2} n^T R_n^{-1} n \right) , \tag{18}$$

and to model the prior distribution for the image as $f \sim G(\mu_f, R_f)$:

$$p(f) = \frac{1}{(2\pi)^{M/2}|R_f|^{1/2}} \exp\left\{ -\frac{1}{2}(f - \mu_f)^T R_f^{-1}(f - \mu_f) \right\} , \tag{19}$$

where μ_f is the mean vector for the original image f. More recently, MAP has been used in 3D surface estimation from image sequences.[45] In this case they assumed joint distribution $p(g_i|f_i)$ similar to (18) (or equivalently, $p(n_i)$) for the observed image sequence g_i:

$$p(g_i(u)|f_i(u)) = \frac{1}{(2\pi\sigma^2)^{d_i/2}} \exp \sum_{u \in D_i} \left\{ -\frac{1}{2\sigma^2}[g_i(u) - f_i(u)]^2 \right\} , \tag{20}$$

where d_i represents the number of pixels in the selected region D_i of the ith image, and $u \in D_i$.

Given that for a particular image f, the variation in g is in fact caused solely by the noise n (refer to (5)), according to (18), the *a priori* model for $p(g|f)$ can be represented as follows

$$p(g|f) = \frac{1}{(2\pi)^{M/2}|R_n|^{1/2}} \exp\left\{-\frac{1}{2}(g - \phi(Hf))^T R_n^{-1}(g - \phi(Hf))\right\}. \tag{21}$$

The necessary condition for MAP estimate is

$$\frac{\partial \ln p(f|g)}{\partial f} = \frac{\partial \ln p(g|f)}{\partial f} + \frac{\partial \ln p(f)}{\partial f} = 0. \tag{22}$$

From (21) and (19), we may solve (22) for the MAP estimate:

$$\hat{f} = \mu_f + R_f H^T \Phi R_n^{-1}[g - \phi(H\hat{f})], \tag{23}$$

where Φ is a diagonal matrix consisting of derivatives of the form $\frac{\partial \phi(x)}{\partial x}$.

Hunt *et al.* has studied the effectiveness of MAP algorithms for image restoration with Gaussian noise models and various prior means μ_f under different blurring degradations.[48,94,95] Bresler *et al.*[13] applied the MAP algorithm to 3D reconstruction problems from noisy and incomplete projections. They were able to obtain satisfactory results from only four views in a 135° sector with an SNR as low as 1.5.

Hunt *et al.* argued that in the Bayesian framework it is possible to use the image prior model (19) and to produce reasonable results with a stationary variance about the mean μ_f which is not necessarily constant. The non-uniform image mean assumption is interesting in that it maintains the *structural* information about the image ensemble, which is related to visible characteristics of the image. For instance, instead of giving an averaged (constant) value of a scene, the nonstationary mean provides *general* shapes and shadings to represent objects present in the scene. Incidentally, the solution in (23) is in fact a discrete version of the Wiener filter with the *a priori* mean μ_f, which can be easily carried out by FFT for computational efficiency. This coincidence is the result of the Gaussian distribution assumption in developing the MAP estimator.[96] It is also of interest to notice that MAP estimates in many situations are the same as minimum mean squares error (MMSE) estimates which are the mean of the *a posteriori* density (the conditional mean), and particularly, for linear image models (6) under the Gaussian assumption, MAP and MMSE are equivalent.

5. KALMAN FILTER

The Kalman filter[54] is an estimation technique developed in the Bayesian framework, which is able to solve the estimation problem for nonstationary random processes. It uses conditional expectation and state transfer to solve estimation problems. Due to its high computational feasibility and efficiency, the Kalman filter theory has long been applied to control systems and signal processing. Extending Kalman filter to multi-dimensional signals (e.g. images, spatial-temporal signals) gained considerable popularity in the early 70s. Nahi[76] and Habibi[35] probably were among the first to use the Kalman filter for image estimation problems. Further work on the development of two-dimensional Kalman filters has been carried out by many others, such as Powell and Silverman,[82] Murphy and Silverman,[74] and Liu *et al.*[63,64] Willsky[98] has done a thorough survey of the related research. Woods and his colleagues[57,100–103] have done probably the most extensive study of two-dimensional Kalman-type filter theory and techniques for image estimation.

In dynamic image sequence processing, motion compensated spatial-temporal filters have been developed based on the Kalman filtering theory,[12] and more recently in dynamic object tracking.[18,104] The Kalman filter also found applications in edge tracking.[3,4,69] Matthies *et al.*[70] used the Kalman for computing depth-from-motion. In 3D reconstruction using image sequences, many systems have been developed based on Kalman filtering theory.[1,14,29,30]

In this section we present a Kalman-type filter for image estimation. The image is modeled by an autoregressive model.

5.1. Image Modeling

We model the image based on a finite nonsymmetric half plane (NSHP)[27] which is illustrated in Fig. 5. For any spatial position of the scanner, all the previous processed pixels are considered "past" states, while the pixel which is presently being processed is called the "present" state, and all the pixels to be processed are referred to as the "future" states. Obviously, this definition is rather artificial, as the scanning operation can take any order.

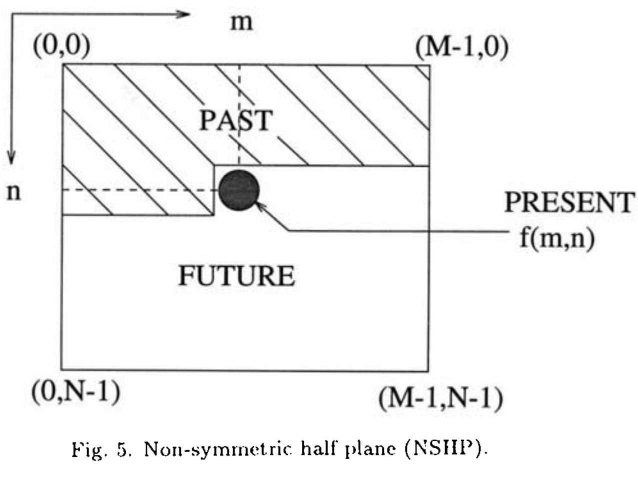

Fig. 5. Non-symmetric half plane (NSHP).

Let (m, n) be the present position of the scanner, we define $P(m, n)$ as past region with respect to (m, n),

$$P(m,n) = [(i,j) : \{1 < i < M, 1 < j < n-1\} \bigcup \{1 < i < m-1, j = n\}], \quad (24)$$

where $\bigcup$ is union. The image signal at (m,n) can be modeled by an autoregressive model, namely,

$$f(m,n) = \sum_{(i,j)\in P(m,n)} a(i,j)f(m-i,n-j) + w_s(m,n)\,, \tag{25}$$

where w_s is the modeling error and is assumed to be Gaussian and white,[a] and $a(i,j)$ are the modeling coefficients.

Let $g(m,n)$ represent the observed image which is corrupted by an additive noise $v(m,n)$. We have

$$g(m,n) = f(m,n) + v(m,n)\,, \tag{26}$$

where $v(m,n)$ is assumed to be a Gaussian, white noise array with zero mean and variance $\sigma_v^2(m,n)$, and is independent of $f(m,n)$ and $w_s(m,n)$.

Usually the dimension of the image is very large, the simple use of (25) as the image model in (26) will lead to high dimensionality problem, and result in excessive computation. However, it is assumed[83,100] that only those pixels in a small neighborhood of the present pixel contribute significantly to the modeling of $f(m,n)$, we may reduce (25) to

$$f(m,n) = \sum_{(i,j)\in\Re} a(i,j)f(m-i,n-j) + w_s(m,n)\,, \tag{27}$$

where $\Re \triangleq (k,l)$ is the sub-region of the NSHP shown in Fig. 6, and w_s is the modeling error with zero mean and variance σ_{ws}^2. Without loss of generality, we assume $E[f(m,n)] = 0$.

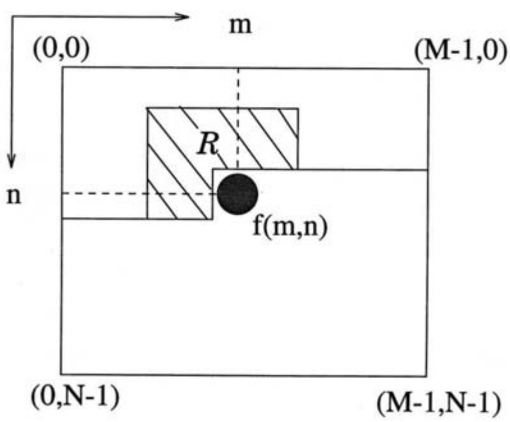

Fig. 6. Subregion of the NSHP for modeling the image.

We have used a spiral scanning operation which runs continuously along a square-shaped spiral line (see Fig. 7) and ends at the geometrical center of the image. Referring to Fig. 8, for simplicity of notation we denote P as the "past",

[a]Valid for the autoregressive model only, otherwise $w_s(m,n)$ will not be white.[76]

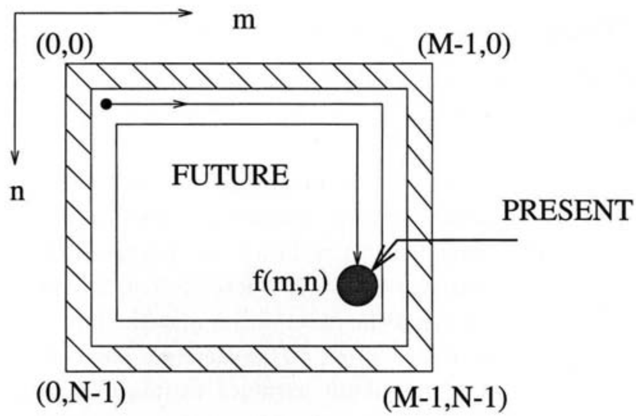

Fig. 7. Scanning process.

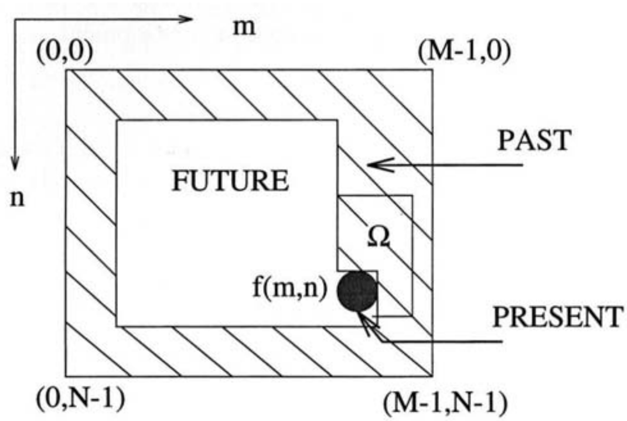

Fig. 8. Causal image field.

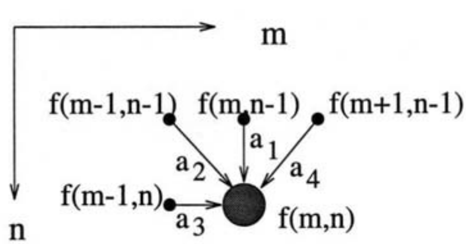

Fig. 9. Neighborhood of $f(m,n)$.

and F as the "future". For the particular spiral scanning configuration, the entire image can be considered as a long string of pixels. The image model (27) can be rewritten for the subregion Ω of P (see Fig. 8) as follows,

$$f(m,n) = \sum_{(i,j)\in\Omega} a(i,j)f(m-i,n-j) + w_s(m,n)\,. \tag{28}$$

Figure 9 shows a possible neighborhood of pixels that can be used in (28).

5.2. The Two-Dimensional Adaptive Kalman Filter

The Kalman filtering technique is based on the linear minimum variance (or, sometimes, minimum mean squares error) estimate of $f(m,n)$, given an observation $g(m,n)$.

This estimate is obtained by using the orthogonal projection principle[54,78], which states that the linear minimum variance estimate $\hat{y} = Ax + B$ of y is such that the estimate error $(y - \hat{y})$ is orthogonal to the data x, or $\hat{y}$ is the orthogonal projection of y onto X, in the sense that $E[y - \hat{y}] = 0$, where $X = x$. In the present context, we require an estimate $\hat{f}$ of f given the observation g, such that $\hat{f}$ is the orthogonal projection of f onto P. That is, $\hat{f} = \hat{E}\{f|P\}$. Here we have used $\hat{E}$ rather than E to indicate that the linear minimum variance estimate is in general not the true conditional mean in the sense that the linear minimum variance estimate is a weighted linear combination of the observations.

The image models defined by Eqs. (26) and (28) over the causally partitioned image field provide the basis for all our succeeding development.

Problem Statement

Given the models (26) and (28), determine an estimate $\hat{f}$ (of f) which is a linear combination of the observation g and estimates $\hat{f}$. The estimate is optimal subject to the criterion that the expected value of the sum of the variance between $\hat{f}$ and f is a minimum. That is, $\hat{f}$ is to be chosen so that

$$\hat{E}[f - \hat{f}]^2 = minimum\,. \tag{29}$$

Signal Prediction and Update

We may write the estimate $\hat{f}(m,n)$ as

$$\hat{f}(m,n) = \hat{f}^-(m,n) + K(m,n)e(m,n)\,, \tag{30}$$

where $K(m,n)$ is an unknown weighting function (or the gain in the Kalman filter theory), the term $\hat{f}^-(m,n)$ defines the predicted estimate at (m,n), and is given by

$$\hat{f}^-(m,n) = \sum_{(i,j)\in\Omega} a(i,j)\hat{f}(m-i,n-j)\,, \tag{31}$$

and

$$e(m,n) = g(m,n) - \hat{f}^-(m,n) \tag{32}$$

which is called the Kalman innovation process.[65]

It is easy to show that, if the minimum variance estimator is *unbiased* (meaning that $E\{\hat{f}\} = f$), the innovation process is also white, with zero mean provided that the noise v is white with zero mean. We may write (32) as follows

$$e(m,n) = \tilde{f}^-(m,n) + v(m,n)\,, \tag{33}$$

where

$$\tilde{f}(m,n) = f(m,n) - \hat{f}(m,n)\,.$$

By the orthogonality principle and from equations (30) and (33), we have

$$E\{e(m,n)e(k,l)\} = E\{v(m,n)v(k,l)\} = 0, \forall m \neq k, n \neq l \tag{34}$$

and obviously, since $E\{w_s\} = 0$,

$$E\{e(m,n)\} = 0\,. \tag{35}$$

Equation (34) implies that e is a white process, which may be written, in general, as

$$\begin{aligned} E\{e(m,n)e(k,l)\} &= [V_{\tilde{f}}^{-}(m,n;m,n) + \sigma_v^2(m,n)]\delta_{mk}\delta_{nl} \\ &= R_e(m,n) + \delta_{mk}\delta_{nl}\,, \end{aligned} \tag{36}$$

where $R_e(m,n) \triangleq [V_{\tilde{f}}^{-}(m,n;m,n) + \delta_v^2(m,n)]$ and

$$V_{\tilde{f}}^{-}(m,n;k,l) \triangleq E\{\tilde{f}^{-}(m,n)\tilde{f}^{-}(k,l)\} \tag{37}$$

which is the *a priori* error variance, and δ_{rs} is the Kronecker delta function,

$$\delta_{rs} = \begin{cases} 1, & \text{if } r = s, \\ 0, & \text{otherwise.} \end{cases}$$

Gain Equation

From Eq. (30) we have

$$E\{[f(m,n) - \hat{f}(m,n)]e(k,l)\} = E\{[\tilde{f}^{-}(m,n) - K(m,n)e(m,n)]e(k,l)\} = 0\,, \tag{38}$$

thus

$$E\{[\tilde{f}^{-}(m,n)e(k,l)\} = E\{K(m,n)e(m,n)e(k,l)\} = K(m,n)R_e(m,n)\delta_{mk}\delta_{nl}\,. \tag{39}$$

If $m = k, n = l$, we have

$$K(m,n) = E\{\tilde{f}^{-}(m,n)e(m,n)\}R_e^{-1}(m,n)\,, \tag{40}$$

where

$$E\{\tilde{f}^{-}(m,n)v(m,n)\} = 0\,, \tag{41}$$

and

$$e(m,n) = \tilde{f}^{-}(m,n) + v(m,n)\,.$$

The Kalman gain can now be written as

$$K(m,n) = V_{\tilde{f}}^{-}(m,n;m,n)R_e^{-1}(m,n)\,. \tag{42}$$

A Priori Variance Algorithm

In Eq. (37) we have defined the *a priori* variance, in the following we express the *a priori* variance in terms of estimate error $\tilde{f}$.

Recall that $\tilde{f}^- = f - \hat{f}^-$, and from the image model (28) we have

$$f(m,n) = \sum_{(i,j)\in\Omega} a(i,j)f(m-i,n-j) + w_s(m,n)\,. \tag{43}$$

Remember that w_s is a white random process with the variance σ^2_{ws}, thus

$$\begin{aligned} V^-_{\tilde{f}}(m,n;m,n) &= \sum_{(k,l)\in\Omega}\sum_{(i,j)\in\Omega} a(i,j)E[\tilde{f}(m-i,n-j)\tilde{f}(m-k,n-l)]a(k,l) + \sigma^2_{ws} \\ &= \sum_{(k,l)\in\Omega}\sum_{(i,j)\in\Omega} a(i,j)V_{\tilde{f}}(r,s,t,u)a(k,l) + \sigma^2_{ws}\,, \end{aligned} \tag{44}$$

where $r = m - i$, $s = n - j$, $t = m - k$, $u = n - l$; and V is called the *a posteriori variance* which is readily available.

We define

$$\tilde{f}(m,n) = f(m,n) - \hat{f}(m,n)\,. \tag{45}$$

Substituting (33) into (28) yields

$$\tilde{f}(m,n) = [f(m,n) - \hat{f}^-(m,n)] - K(m,n)e(m,n)\,,$$

or

$$\tilde{f}(m,n) = \tilde{f}^-(m,n) - K(m,n)e(m,n)\,. \tag{46}$$

Thus

$$E\{\tilde{f}^2(m,n)\} = E\{[\tilde{f}^-(m,n) - K(m,n)e(m,n)]^2\}\,. \tag{47}$$

It is known that v is white noise, and from Eqs. (37), (41) and (44), we obtain

$$V_{\tilde{f}}(m,n;m,n) = E\{\tilde{f}^2(m,n)\} = [1 - K(m,n)]V^-_{\tilde{f}}(m,n;m,n)\,. \tag{48}$$

The equations (30), (42), (44) and (48) are the sequential Kalman filter for the image model (28) and the observation (26).

5.3. Identification of the Modeling Coefficients

In the image model (28), the modeling coefficients are normally not known, and must be estimated. We may write Eq. (28) in vector form, as,

$$f(m,n) = \mathbf{A}^\top(m,n)\mathbf{F}(m,n) + w_s(m,n)\,, \tag{49}$$

where $\mathbf{A}$ is the vector composed of the modeling coefficients, and $\mathbf{F}$ the signal vector. The structures of $\mathbf{A}$ and $\mathbf{F}$ depends on the specific scanning operation and the neighborhood configuration. The Eq. (28) is causal, and the original image is also assumed to be an ergodic random field. We may use the identification algorithm[65] to estimate $\{a(i,j)\}$. In order to make the notations simple, we define $k \triangleq (m,n)$ and use $k+j$ to represent moving j steps further from (m,n) or k. With this definition, equation (49) can be written as

$$f(k) = \mathbf{A}(k)\mathbf{F}(k) + w_s(k)\,. \tag{50}$$

From (50), a general parameter identification algorithm is readily available as follows, assuming that the modeling coefficients $\{a(i,j)\}$ are constants,

$$\left.\begin{aligned}
&\hat{\mathbf{A}}(k+1 \mid k+1) = \hat{\mathbf{A}}(k \mid k) + \Phi(k+1)[f(k) - \mathbf{F}(k)\hat{\mathbf{A}}(k)]\,,\\
&\Phi(k+1) = \mathbf{R}(k+1 \mid k)\mathbf{F}^{\top}(k)[\mathbf{F}(k)\mathbf{R}(k+1 \mid k)\mathbf{F}^{\top}(k) + \sigma_{ws}^2(k+1)\mathbf{I}]^{-1}\,,\\
&\mathbf{R}(k+1 \mid k) = \mathbf{R}(k \mid k) + \sigma_{ws}^2(k)\mathbf{I}\,,\\
&\mathbf{R}(k+1 \mid k+1) = [\mathbf{I} - \Phi(k+1)\mathbf{F}(k)]\mathbf{R}(k+1 \mid k)\,,
\end{aligned}\right\} \tag{51}$$

where

$$\mathbf{R}(k \mid k) = E\{[\mathbf{A}(k) - \hat{\mathbf{A}}(k \mid k)][\mathbf{A}(k) - \hat{\mathbf{A}}(k \mid k)]^{\top}\}\,, \tag{52}$$

and σ_{ws}^2 is constant for stationary white process.

From Eq. (51) we can see that this algorithm actually uses the original image to estimate $\mathbf{A}(k)$. Unfortunately, in restoration problems, the original image is not available at all, the direct use of (51) is seemingly meaningless. However, in many practical applications the gross structure of the original image may be observed, and we may apply (51) over images with similar structure to obtain $\{a(i,j)\}$.

In most cases we may have or assume some prior knowledge about the original image, e.g. the statistical characteristics of the image. Experiments have shown that many images can be regarded as wide-sense stationary random fields with the following autocorrelation function,[31]

$$R(i,j) = \sigma_s^2 exp[-\alpha|i| - \beta|j|]\,, \tag{53}$$

where σ_s^2 is the global image variance, i and j are the horizontal and vertical displacements, respectively, and α and β are constants.

With (53) given we are now in a position to introduce another parameter estimation method which is more realistic than (51). From (49) we have

$$w_s(m,n) = f(m,n) - \mathbf{A}(m,n)\mathbf{F}(m,n)\,, \tag{54}$$

we seek $\mathbf{A}$ such that

$$J(w_s(m,n)) = E\{w_s^2(m,n)\}\,, \tag{55}$$

is minimized. This gives

$$\mathbf{A}(m,n) = [\sum_{\Omega} \mathbf{F}(m,n)\mathbf{F}^{\top}(m,n)]^{-1}[\sum_{\Omega} f(m,n)\mathbf{F}(m,n)]. \tag{56}$$

The right side of Eq. (56) is, in fact, composed of the autocorrelation $\{R(i,j); (i,j) \in \Omega(m,n)\}$. Since (53) is known (56) can be solved.

However, the Kalman filter in general has high computational complexity and — being a type of recursive least squares (RLS) algorithm — is very sensitive to round off errors which means that it can become unstable in fixed-point implementations.

6. MARKOV RANDOM FIELDS

In the previous section we discussed that under the Bayesian framework, the estimation problem is essentially to find $\hat{f}$ given observation such that $p(f|g)$ is maximized. This results in a very simple and straightforward maximum *a posteriori* (MAP) estimation method. However, in order to solve the MAP estimation problem we have to know the distribution functions involved in the model. This is a very difficult problem in most applications. Although the Bayesian framework is natural and philosophically correct for dealing with image estimation problems, choosing correct probability functions is rather *ad hoc* and of a heuristic nature as it primarily relies on our intuition. As a result, the distribution functions we assume may not reflect the true reality and represent our *ignorance*. It is therefore more desirable to find a sound modeling structure such that the problem can be modeled properly. Markov random fields (MRFs) have been shown to be suitable for the problem of spatial statistical modeling. MRF has the following properties that make it most suitable for probabilistic modeling of spatial information[32,68]:

1. The interaction in Markov random fields between pixel elements is local which satisfies the piecewise smoothness requirement.
2. The local interaction characteristic makes it possible to be implemented parallely for computational efficiency.[17,32]. More importantly though, this also ensures the existence of priors rich enough for modeling different spatial behaviors.
3. In MRF the prior distribution function is proven to be Gibbsian[5] under a very mild condition. This makes the modeling process in the Bayes theorem straightforward.
4. With a sound distribution model it is possible to carry out optimization procedures (e.g. stochastic relaxation, simulated annealing) and to systematically verify the consistency between the model and our prior knowledge.

In recent years, Markov random field has gained considerable popularity as a method for modeling images in a variety of applications, for instance, image restoration,[32,40,50,51,85,106] texture segmentation and modeling,[16,20,28,55,56,67,77] edge detection,[21,33,49,91,105] medical image analysis,[34,37,39,62,87] motion estimation,[41,59,66,84,107] document processing,[92] and even in image interpretation,[73] just to mention a few.

6.1. Neighborhood System

The theory of Markov random fields has been around for about thirty years due to Dorushin.[24,58] MRF is defined on a neighborhood system.[6,32] In time sequence we have a causal ordering of states: $\Omega = \{\omega_1, \omega_2, ..., \omega_L\}$, the neighborhood for this sequence in terms of first order homogeneous Markov chain is given by $\aleph_i =$

$\{i-1, i+1, 2 \leq i \leq L-1\}$ with two boundary neighborhoods: $\aleph_1 = \{2\}$ and $\aleph_L = \{L-1\}$,

$$p(\omega_i|\omega_j, j < L) = p(\omega_i|\omega_{i-1}) .$$

It can be shown that

$$p(\omega) = p(\omega_1)\prod_{i=2}^{L} p(\omega_i|\omega_{i-1}).$$

However, in a two-sided neighborhood system $\aleph = \{\aleph_1, \aleph_i, \aleph_L\}$ (see Fig. 10), the Markov chain in general is also dependent on ω_{i+1}[15]:

$$p(\omega) = p(\omega_1)\prod_{i=2}^{L} p(\omega_i|\omega_{i-1}, \omega_{i+1}) .$$

This can be applied to k-nearest neighbors[32].

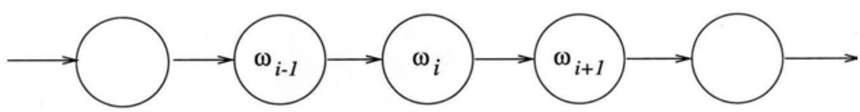

Fig. 10. Two-sided neighborhood of ω_i.

In most spatial computing tasks we are dealing with two-dimensional information, it is necessary to extend the usual Markov chains to 2D cases. In fact in the previous section, Fig. 9 shows a typical nearest neighborhood system for the 2D Markov chain. However, in such a *nonsymmetric* neighborhood system[99] it is difficult to verify the Markovian property and does not provide a natural description of the image (spatial) field.

Let $S = \{s_1, s_2, \ldots, s_L\}$ be a set of lattice points (e.g. pixel locations (i, j) in images) and $f_{s_i} \in \Re$ is the value at point s_i, for instance, in grey-level images $f_{s_i} \in [0, 255]$. Define a neighborhood system $\aleph = \{\aleph_s, s \in S\}$. We require that the neighborhood system $\aleph$ be symmetric:

$$\forall t \in \aleph_s \Rightarrow s \in \aleph_t \text{ and } s \notin \aleph_s . \tag{57}$$

In addition we define clique $C \subseteq S$ as a set of points, $c \in C$ with the following property,

$$\forall s, c \in C, c \in \aleph_s. \tag{58}$$

That is, from (57), they are neighbors to each other.

Figure 11 shows some typical nearest-neighborhood configurations used in many image processing applications. Besag[7] suggested a neighborhood system defined on a hexagonal lattice, which offers one more orientation for line (edge) sites and removes the problem of four or more lines converging at a point.

Figure 12 shows some clique types. It is apparent that for the configuration in Fig. 11(a) the possible cliques are {(a), (b), (c)} in Fig. 12. The 8-connected configuration in Fig. 11(b) has the clique set C_8 = {(a), (b),(c),(d),(e),(f),(g),(h),(i), (j)} in Fig. 12.

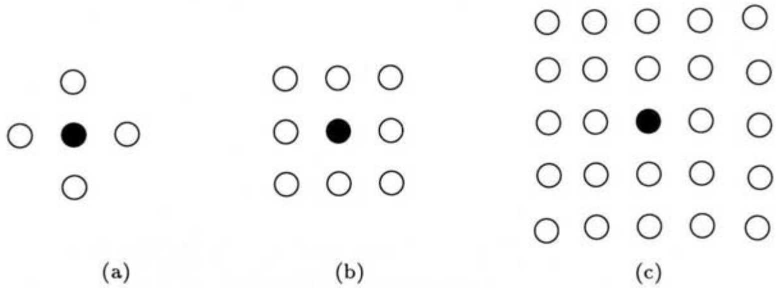

Fig. 11. Some typical nearest-neighborhood configurations.

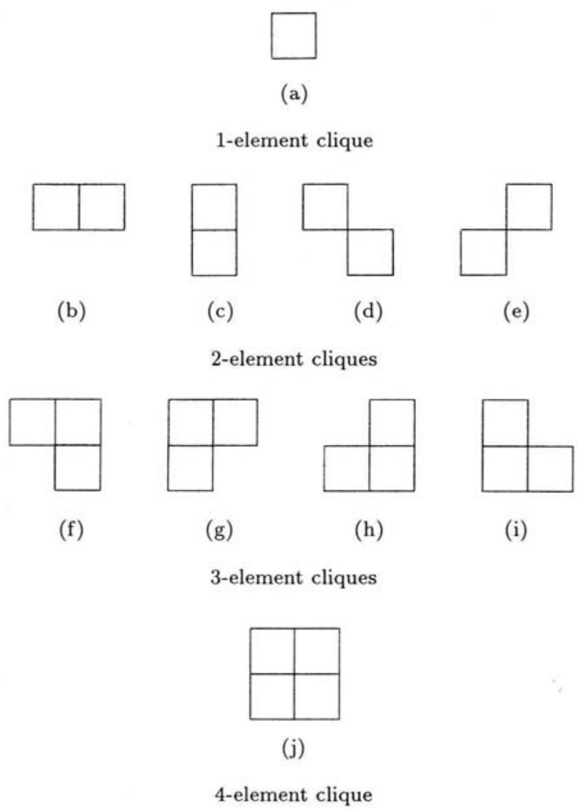

Fig. 12. Examples of cliques.

In such a neighborhood system, it is possible to model different image processes in addition to intensity at a pixel site. For instance, between each pair of pixel sites we can assign line (edge) sites (Fig. 13), orientation discontinuities, and depth discontinuities in 3D surface reconstruction. Combining these fields we obtain a Markov random field.

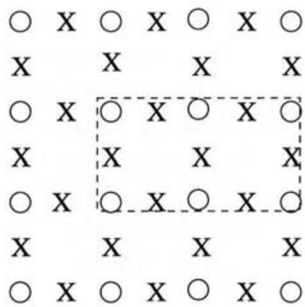

Fig. 13. Line (edge) sites are indicated by Xs. Each line site has six line sites as its neighbors.

6.2. Gibbs Distribution

The Markov random field is defined under the neighborhood system described above if $p(\omega) > 0$ and $p(\omega_s|\omega_r, r \neq s) = p(\omega_s|\omega_r, r \in \aleph_s)$, where $\omega = \{\omega_i \in \Re, i \in [1, L]\}$ is a realization of possible configurations (e.g. an image). The Hammersley–Clifford theorem[5] states that in the neighborhood system $\aleph = \{\aleph_s, s \in S\}$, MRFs can be modeled as by a Gibbs (Boltzmann) distribution[b]:

$$p(f) = \frac{1}{Z}\exp\{-\sum_{c\in C} V_c(f_c)\} = \frac{1}{Z}\exp\{-U(f)\}\,, \tag{59}$$

where C is the set of all cliques, Z is the normalizing constant for the density, also known as the partition function, and $U(f)$ the energy function:

$$U(f) = \sum_{c\in C} V_c(f_c)\,,$$

where potentials V_c are assigned certain values depending on the nature of the application. Figure 14 shows the potentials used in Geman and Geman's pioneering work[32] on the use of MRF in image segmentation and restoration. Obviously, smoothness of image regions and continuity of edges are the basis for assigning these potentials. However, the presence of an edge at an edge site has to be derived based on the pixel values *observed* (measured) in the neighborhood of the edge site. For instance, Jain and Nadabar[49] have discussed this problem in detail and defined two different statistics to measure the strength of jump edges and crease edges for their work on range image segmentation.

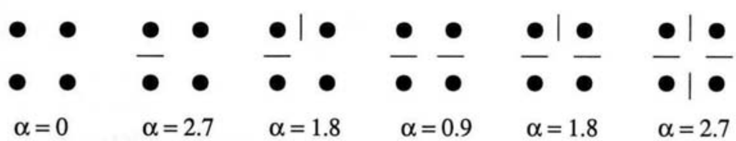

Fig. 14. Potential values assigned to different cliques.

[b] Spitzer[89] has shown that Gibbs random fields and Markov random fields are equivalent.

In Ref. 49 Jain and Nadabar used a second order neighborhood system (Fig. 15) to model edge sites for detecting jump and crease edges in range images. In this case, MRF is used to represent the prior knowledge about the nature of edges in range images and the specific consideration is given to continuity and smoothness of edges. For the second order neighborhood system they proposed a set of 14 cliques of edge and non-edge type as shown in Fig. 16 and assigned a specific value for each clique. The rationale here is to assign lower values to continuous line segments and higher values for *broken* line sites. This scheme reinforces our belief in continuity and smoothness of certain intrinsic features in the spatial data.

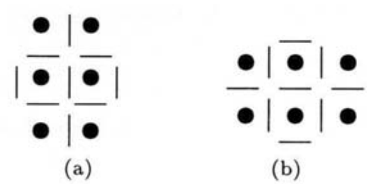

Fig. 15. Edge sites in the second-order neighborhood system (a) vertical edge site, (b) horizontal edge site.

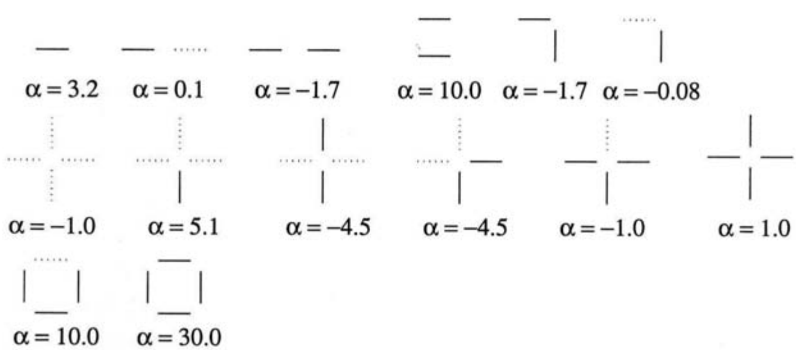

Fig. 16. Potential values assigned to different cliques of edge (solid lines) and non-edge (dotted lines) type.

6.3. Conditional Probabilities

The conditional probability of a pixel given neighbors is defined as:

$$p(f_s|f_t, s \neq r) = \frac{p(f)}{\sum_{f_s \in \Re} p(f)}, \tag{60}$$

where $p(f)$ is the prior Gibbs distribution (59), and $\Re$ may represent, for instance, the set of grey levels or edge orientations. For example, considering an 8-connected nearest neighborhood (as that shown in Fig. 11(b)) with the cliques formed by two pixels, there will be eight cliques that contain f_s, since by definition any pixel f_t at site t in a clique containing f_s at site s must be a neighbor of f_s. We can use (60) to calculate the conditional probability of the pixel f_s as follows

$$p(f_s|f_t, s \neq t) = \frac{\exp\{-\sum_{t=1}^{8} \alpha_t \delta(f_s \neq f_t)\}}{\sum_{f_s=0}^{255} \exp\{-\sum_{t=1}^{8} \alpha_t \delta(f_s \neq f_t)\}}, \tag{61}$$

where α_t are the potential value assignments for the potential function $V(f_t)$, and $\delta(x)$ is defined as

$$\delta(x) = \begin{cases} 1 & \text{if } x \neq 0, \\ 0 & \text{otherwise.} \end{cases}$$

Potential assignment is difficult in most applications and requires careful analysis[32,49] of the problem at hand.

In the Bayesian estimation paradigm, most researchers use MAP as the optimality criterion. The *a posteriori* distribution function $p(f|g)$ is conditioned on observation. In Markov random fields, the prior $p(f)$ can be modeled as a Gibbs distribution. Furthermore, using Markov random fields it can be shown[32] that for the image model (5), the *a posteriori* distribution $p(f|g)$ is also Gibbsian:

$$p(f|g) = \frac{1}{Z_g} e^{-U(f|g)}, \tag{62}$$

with an energy function given as follows

$$U(f|g) = U(f) + \|\mu - \Phi(g, \phi(Hf))\|^2 / 2\sigma^2, \tag{63}$$

where μ is mean vector for the white Gaussian noise n, and $\Phi(g, \phi(Hf)) = g - \phi(Hf)$ for the model in (5) with an additional noise, and Φ represents an inverse operation. Although (63) is derived based on white Gaussian noise assumption, Geman and Geman[32] stressed that extension to other forms of noise is only notational.

This result enables us to formulate the MAP estimation problem in Markov random fields with a sound knowledge about specific distribution functions that can be used in the modeling process. The problem of maximizing the posterior distribution for a given g is to minimize (63). However, this is a computationally very demanding operation even for images of moderate sizes. There are many algorithms[25] available for this minimization problem, for instance, simulated annealing, stochastic relaxation,[32] Monte Carlo approach to maximizer of *a posteriori* marginals (MPM),[68] iterated conditional modes (ICM),[6] and high-confidence first (HCF).[19]

7. CONCLUSIONS

As with most other signal processing and control problems, image processing is in general of an inverse nature. This type of problems is ill-posed due to lack of constraints, incomplete observed data, and limited knowledge about the process. To solve such problems we have to use our experience and prior knowledge in addition to measured signals. The Bayesian framework is a natural and unified environment for modeling uncertainties in image and spatial data processing tasks. It provides us with a structure for inference given observation.

In this paper we have discussed several important aspects of Bayesian methods in image and spatial information processing. The Bayesian paradigm is based on two probabilistic models, namely, the prior model $p(f)$ which describes the world where the original data f is present and the sensor model $p(g|f)$ which represents the sensor behavior under the influence of f. According to the Bayes theorem, these two models determine the *a posteriori* probability $p(f|g)$ which expresses the relationship between the observed (measured) data and the original data f. As such, this framework allows us to consider the problem of estimation as a statistical inference that requires the incorporation of our intuition and knowledge for the solution. Although our knowledge in some cases is biased due to limited experience, judicious use of the prior knowledge will in many practical applications help us to constrain the ill-posed problem and to obtain satisfactory results efficiently.

In the Bayesian milieu, many estimation algorithms have been developed based on different optimality criteria. One of the most frequently used algorithm is the maximum *a posteriori* estimator (MAP), which attempts to find $\hat{f}$ such that

$$maxp(\hat{f}|g) = max\{\log p(g|\hat{f}) + \log p(\hat{f})\}.$$

However, in order to solve this maximization problem it is necessary to specify the probability distribution functions involved. This has been the most difficult problem in the Bayesian approach, and as stated above, requires considerable knowledge and experience. In the estimation theory it is customary to assume some simple distribution functions based primarily on the following factors: (i) our expectation of the underlying physics involved in the data collection process, (ii) tractable or workable mathematics, (iii) computational feasibility. For instance, a very popular assumption for $p(g|f)$ is the ubiquitous Gaussian distribution, and occasionally $p(f)$ is assumed to be uniformly distributed. Hunt *et al.*[47,48,94,95] were among the pioneers in using the Bayesian approach to image estimation problems. They developed MAP for 2D signals based on Gaussian priors. Apart from some other nice properties, the fundamental reason for using the Gaussian model is due to the central limit theorem.[78] While this is true (at least theoretically) for independent random variables over infinite observations, such an assumption often leads to least-squares (LS) type estimators which have been shown to be non-robust due to their sensitivity to minor deviations from the Gaussian model. As a consequence, many image processing algorithms, albeit theoretically elegant, are unfortunately very unreliable in real applications. This has been recognized as a major problem facing the computer vision community and has been addressed recently (inadequately though) by several researchers.[38,71,108]

Modeling the image field as a nonsymmetric half plane, Woods *et al.*[57,100–103] proposed a 2D Markov chain which imposes causality in the image field that is otherwise non-causal. Based on this causal Markov random field a number of 2D Kalman-type filters has been developed and applied to image estimation problems. In this paper we have derived such a filter in detail. Kalman-type filters have subsequently been used in many image processing problems that involve dynamics, for instance, dynamic feature tracking, image sequence filtering, edge and object tracking, and depth-from-motion applications. However, despite its computational flexibility and modeling capability, the Kalman filter is very difficult to use in real applications due mainly to its numerical instability and computational complexity. This is a pathological characteristic of recursive least squares (RLS) type filters.

Geman and Geman's work on using Markov random fields[32] in image processing and computer vision has provided a solid foundation for image modeling and inspired many new developments in recent years. We have in this paper discussed some major aspects of the Markov random field. Motivated by the intuition that pixels interact in a small local region, we define a neighborhood system in which cliques are formed according to specific requirements. The Markov random field is defined in such a neighborhood system. More significantly, the Hammersley–Clifford theorem, under a minor technical condition, *guarantees* that MRF can be modeled by a Gibbs distribution. Under the Bayesian framework, we are able to use Gibbs distribution to model both the prior $P(f)$ and the *a posteriori* $p(f|g)$. As Markov random fields combine both pixel sites and line sites, it is possible for us to integrate contextual information and our knowledge about the underlying process through specifying clique types and assigning potential values. The estimation problem in MRFs can be solved by a number of approaches, including simulated annealing, stochastic relaxation, maximizer of posterior marginals (MPM), iterative conditional modes (ICM), and high confidence first (HCF). The use of MRF in the Bayesian framework provides us with a better control of the modeling process and greater flexibility in algorithm design. As a result, Bayesian methods have gained considerable attention recently and have found applications in a variety of image processing problems.

An important consequence of applying the Bayesian paradigm to Markov random fields is the opportunity to gain better insights into the regularization theory and to handle the ill-posed problem with greater certainty. As we discussed before, image processing is ill-posed, solutions for such problems require that additional constraints be imposed by the system designer. The ill-posed problem normally solved by regularization methods which for the most part impose the smoothness constraint.[10,81,93] In the Markov random field, through the Bayesian modeling and the Gibbs distribution it can be shown that this type of regularization is in fact a special case of the Bayesian paradigm.[68] This is to be expected as in MRF we assign the potentials in the Gibbs distribution according to continuity and smoothness assumptions.[49]

Indeed, the Bayesian approach is a powerful tool for modeling and inference in many image processing problems, particularly, in Markov random fields, where we have a sound knowledge about the distribution function, namely, the Gibbs distribution. This significantly reduces the ignorance present in the Bayesian paradigms. However, we are still left with the task of determining *proper* cliques and assigning *reasonable* potentials for the energy function in the Gibbs distribution. This on one hand gives us the flexibility and control in the design process, on the other hand we are still left with uncertainties as to what constitutes the proper or reasonable choice. It would be better if we could develop some learning algorithms that can learn and decide on types of cliques and potentials for some *domain specific* applications. Recently, Devijer and Dekesel[22] have proposed a learning algorithm for hidden Markov random fields. Their approach is applicable to Pickard random fields that are both causal and non-causal. The learning process is similar to those in speech recognition. It is also desirable to investigate techniques for learning cliques and potentials based on known spatial information and our heuristics.

Another important issue which deserves urgent considerations is the robustness issue in MRF-based Bayesian approaches. As indicated by Geman and Geman,[32] their derivation of the Gibbs posterior is based on the more traditional white Gaussian noise assumption. Indeed, in many applications, most researchers followed the same assumption.[68] It is interesting to investigate the effects on the system's performance if the Gaussian model is modified as follows[44]:

$$G_f(\mu, \sigma^2, \epsilon) = (1 - \epsilon)G(\mu, \sigma^2) + \epsilon D, D \in P\,,$$

where $G(\mu, \sigma^2)$ is the Gaussian distribution and D represents the set of unknown distributions. With this modeling function it is possible to study effects of Gaussian noise assumption on the Gibbs posterior and to develop techniques to handle modeling uncertainties.[108] Investigation of robustness in this context represents a significant challenge to researchers in the area of image processing, or more generally, in spatial computing.

REFERENCES

1. N. Ayache and O. D. Faugeras, "Maintaining representations of the environment of mobile robot", *Int. Symp. Robotics Res.*, 1987, MIT Press, also INRIA Tech. Rep., France, 1987.
2. H. G. Barrow and J. M. Tenenbaum, "Recovering intrinsic scene characteristics from images", in *Computer Vision Systems*, eds. A. R. Hanson and E. M. Riseman, New York: Academic Press, 1978, pp. 3–26.
3. H. H. Baker, "Building surfaces of evolution: the weaving wall", *Int. J. of Computer Vision* **3**, 1 (1989).
4. H. H. Baker and R. C. Bolles, "Generalizing epipolar-plane image analysis on the spatiotemporal surface", *Int. J. of Computer Vision* **3**, 1 (1989).
5. J. Besag, "Spatial interaction and the statistical analysis of lattice systems", *J. Royal Statistical Society B* **36**, 2 (1974) 192–236.
6. J. Besag, "On the statistical analysis of dirty pictures", *J. Royal Statistical Society B* **48**, 3(1986) 259–302.
7. J. Besag, "Towards Bayesian image analysis", *J. Applied Statistics* **16**, 3 (1989) 395–407.
8. L. M. Biberman and S. Nudelman, *Photoelectronic Imaging Devices*, Vols. 1 and 2, Plenum, 1971.
9. R. Blahut, *Principles and Practice of Information Theory*, Addison-Wesley Publishing Company, 1988.
10. A. Blake and A. Zisserman, *Visual Reconstruction*, MIT Press, 1987.
11. G. E. P. Box, T. Leonard, and C.-F. Wu, eds., *Scientific Inference, Data Analysis, and Robustness*, Academic Press, 1983.

12. J. C. Brailean, R. P. Kleihorst, S. Efstratiadis, A. K. Katsaggelos, and R. L. Lagendijk, "Noise reduction filters for dynamic image sequences: A review", *Proc. IEEE* **83**, 9 (1995) 1272–1292.
13. Y. Bresler, J. A. Fessler, and A. Macovski, "A Bayesian approach to reconstruction from incomplete projections of a multiple object 3D domain", *IEEE Trans. Pattern Analysis and Machine Intelligence* **PAMI-11**, 8 (1989) 840–858.
14. T. J. Broida and R. Challappa, "Kinematics and structure of a rigid object from a sequence of noisy images", *Proc. IEEE Workshop on Motion: Representation and Analysis*, Charleston, May 1986, pp. 95–100.
15. D. Brook, "On the distinction between the conditional probability and the joint probability approaches in the specification of nearest-neighbour systems", *Biometrika* **51**, 3&4 (1964), pp. 481–483.
16. R. Chellappa and S. Chatterjee, "Classification of textures using Gaussian Markov random fields", *IEEE Trans. on Acoust. Speech and Signal Proc.* **ASSP-33**, 4 (1985) 959–963.
17. R. Chellappa and A. Jain, eds., *Markov Random Fields: Theory and Application*, Academic Press, 1993.
18. T. K. Cheng, L. Kitchen, and Z. Q. Liu, "Kalman filtering and clustering for 3D motion tracking using stereo", *Proc. Asian Conf. on Computer Vision*, **ACCV'95**, vol. I, Singapore, Dec. 1995, pp. 34–38.
19. P. Chou, C. Brown, and R. Raman, "A confidence-based approach to the labeling problem", *Proc. IEEE Workshop on Computer Vision*, Miami Beach, Florida, 1987, pp. 51–56.
20. F. S. Cohen, Z. Fan, and M.A. Patel, "Classification of rotated and scaled textured images using Gaussian Markov random field models", *IEEE Trans. Pattern Analysis and Machine intelligence* **PAMI-13**, 2 (1991) 192–202.
21. D. B. Cooper, H. Elliott, F. Cohen, L. Reiss, and P. Symosek, "Stochastic boundary estimation and object recognition", *Computer Vision Graphics and Image Process* **12** (1980) 326–356.
22. P. A. Devijer and M. Dekesel, "Learning the parameters of a hidden Markov random field image model: a simple example", *Pattern Recognition Theory and Applications*, Springer-Verlag, 1987, pp. 141–163.
23. P. A. Devijer and J. Kittler, eds., *Pattern Recognition Theory and Applications*, Springer-Verlag, 1987.
24. R. L. Dorushin, "Description of a random field by means of conditional probabilities and the conditions govern its regularity", *Theory of Probability and its Applications*, (English translation of *Teoriia Veriatn*) **13** (1968) 197–224.
25. R. Dubes, A. Jain, S. Nadabar, and C. Chen, "MRF model-based algorithms for image segmentation", *Proc. 10th Int. Conf. on Pattern Recognition*, Atlantic City, NJ, June 1990, pp. 808–814.
26. W. A. Edelstein *et al.*, "Signal, noise, and contrast in NMR imaging", *J. Comput. Assisted Tomography* **7** (1983) 391–401.
27. M. P. Ekstrom and J. W. Woods, "Two-dimensional spectral factorization with applications in recursive digital filtering", *IEEE Trans. Acoustics, Speech, and Signal Processing* **24**, 2 (1976) 115–128.
28. I. M. Elfadel and R. W. Picard, "Gibbs random fields, co-occurrences and texture modeling", *IEEE Trans. Pattern Analysis and Machine Intelligence* **PAMI-16**, 1 (1994) 24–37.
29. O. D. Faugeras, N. Ayache, and B. Faverjon, "Building visual maps by combining noisy stereo measurements", *Proc. IEEE Int. Conf. Robotics Automation*, San Francisco, April 1986, pp. 1433–1438.

30. O. D. Faugeras, F. Lustman, and G. Toscani, "Motion and structure from motion from point and line matches", *Proc. 1st Int. Conf. Computer Vision*, London, April 1987, pp. 1433–1438.
31. L. E. Franks, "A model for the random video process", *Bell Syst. Tech. J.* **45** (1966) 609–630.
32. S. Geman and D. Geman, "Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images", *IEEE Trans. Pattern Analysis and Machine Intelligence* **PAMI-6**, 6 (1984) 721–741.
33. D. Geman, S. Geman, C. Graffigne, and P. Dong, "Boundary detection by constrained optimization", *IEEE Trans. Pattern Analysis and Machine Intelligence* **PAMI-12**, 7 (1990) 609–628.
34. P. J. Green, "Bayesian reconstructions from emission tomography data using a modified EM algorithm", *IEEE Trans. Medical Imaging* **9**, 1 (1990).
35. A. Habibi, "Two-dimensional Bayesian estimation of images", *Proc. IEEE* **60** (1972) 878–883.
36. F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel, *Robust Statistics: The Approach Based on Influence Functions*, John Wiley & Sons, Inc., 1986.
37. K. Hanson and G. Wecksung, "Bayesian approach to limited-angle reconstruction computed tomography", *J. Opt. Soc. Am.* **73**, 11 (1983) 1501–1509.
38. R. M. Haralick, "Computer vision theory: the lack of it", *Computer Vision, Graph., and Image Processing* **36** (1986) 372–386.
39. T. Hebert and R. Leahy, "A generalized EM algorithm for 3-D Bayesian reconstruction from Poisson data using Gibbs priors", *IEEE Trans. Medical Imaging* **8**, 2 (1989) 194–202.
40. T. J. Hebert and K. Lu, "Expectation-maximization algorithms, null spaces, and MAP image restoration", *IEEE Trans. Image Processing* **4**, 8 (1995) 1084–1095.
41. F. Heitz and P. Bouthemy, "Multimodal estimation of discontinuous optical flow using Markov random fields", *IEEE Trans. Pattern Analysis and Machine Intelligence* **15**, 12 (1993) 1217–1232.
42. B. K. P. Horn, "Understanding image intensities", *Artificial Intelligence* **8** (1977) 201–231.
43. D. I. Hoult and R. E. Richards, "Signal-to-noise ratio of nuclear magnetic resonance experiment", *J. Magn. Reson.* **24** (1976) 71–85.
44. P. Huber, *Robust Statistics*, John Wiley & Sons, Inc., 1981.
45. Y. P. Hung, D. B. Cooper, and B. Cernuschi-frias, "Asymptotic Bayesian surface estimation using an image sequence", *Int. J. of Computer Vision* **6**, 2 (1991) 105–132.
46. B. R. Hunt, "The application of constrained least squares estimation to image restoration by digital computer", *IEEE Trans. Computers* **C-22**, 9 (1973) 805–812.
47. B. R. Hunt, "Bayesian methods in nonlinear digital image restoration", *IEEE Trans. Computers* **C-26**, 3 (1976) 219–229.
48. B. R. Hunt and T. M. Cannon, "Nonstationary assumptions for Gaussian models", *IEEE Trans. Systems, Man, and Cybernetics* **6**, 12 (1976) 876–882.
49. A. Jain and S. G. Nadabar, "Range image segmentation using MRF models", in *Markov Random Fields: Theory and Application*, eds. R. Chellappa and A. Jain, Academic Press, 1993, pp. 543–572.
50. B. D. Jeffs and M. Gunsay, "Restoration of blurred star field images by maximally sparse optimization", *IEEE Trans. Image Process.* **2**, 2 (1993) 202–211.
51. F.-C. Jeng and W. Woods, "Compound Gauss-Markov random fields for image estimation", *IEEE Trans. Signal Processing* **39**, 3 (1991) 683–697.

52. V. E. Johnson, W. H. Wong, X. Hu, and C.-T. Chen, "Image restoration using Gibbs priors: boundary modeling, treatment of blurring, and selection of hyperparameter", *IEEE Trans. Pattern Analysis and Machine Intelligence* **13**, 5 (1991) 413-425.
53. J. B. Kadane, ed., *Robustness of Bayesian Analysis*, Elsevier North-Holland, 1984.
54. R. E. Kalman, "A new approach to linear filtering and prediction problems", *J. Basic Engineering, (ASME Trans.)* (1960) 35 45.
55. R. Kashyap, R. Chellappa, and A. Khotanzad, "Texture classification using features derived from random field models", *Pattern Recogn. Lett.* **1**, 1 (1982) 43–50.
56. R. Kashyap and R. Chellappa, "Estimation and choice of neighbors in spatial-interaction models of images", *IEEE Trans. Information Theory* **IT-29**, 1 (1983) 60-72.
57. H. Kaufman, J. W. Woods, S. Dravida, and A. M. Tekalp, "Estimation and identification of two-dimensional images", *IEEE Trans. Automatic Control* **AC-28**, 7 (1983) 745-756.
58. R. Kindermann and J. Snell, *Markov Random Fields and Their Applications*, American Mathematical Society, 1980.
59. J. Konrad and E. Dubois, "Bayesian estimation of motion vector fields", *IEEE Trans. Pattern Analysis and Machine Intelligence* **PAMI-14**, 9 (1992) 910–927.
60. D. T. Kuan, A. A. Sawchuk, T. C. Strand, and P. Chavel, "Nonstationary 2-D recursive restoration of images with signal-dependent noise", *Int. Conf. on Acoustics, Speech, and Signal Processing*, San Diego, CA, April 1984, pp. 37.4.1 37.4.4.
61. M. J. Lahart, "Local image restoration by a least-squares method", *J. Opt. Soc. Am.* **69**, 10 (1979) 1333–1339.
62. K. Large, "Convergence of EM image reconstruction algorithms with Gibbs smoothing", *IEEE Trans. Medical Imaging* **9**, 4 (1990) 439–446.
63. Z. Q. Liu and T. M. Caelli, "A sequential recursive filter for image restoration", *Computer Vision, Graph., and Image Processing* **44**, 3 (1988) 332–349.
64. Z. Q. Liu, V. G. Gourishankar, and T. M. Caelli, "A new adaptive iterative algorithm for image restoration", *Image and Vision Computing* **6**, 3 (1988) 151 161.
65. L. Ljung and T. Söderström, *Theory and Practice of Recursive Identification*, MIT Press, 1983.
66. M. R. Luettgen, W. C. Karl, and A. S. Willsky, "Efficient multiscale regularization with applications to the computation of optical flow", *IEEE Trans. Image Processing* **3**, 1 (1994).
67. J. Mao and A. K. Jain, "Texture classification and segmentation using multiresolution simultaneous autoregressive models", *Pattern Recognition* **25**, 2 (1992) 173 188.
68. J. Marroquin, J. Mitter, and T. Poggio, "Probabilistic solution of ill-posed problems in computational vision", *J. American Statistical Association* **82**, 397 (1987) 76-89.
69. L. H. Matthies and S. A. Sharer, "Error modeling in stereo navigation", *IEEE J. Robotics and Automation* **RA-3**, 3 (1987) 239-248.
70. L. H. Matthies, R. Szeliski, and T. Kanade, "Kalman filter-based algorithms for estimating depth from image sequences", *Int. J. of Computer Vision* **3** (1989) 209–236.
71. P. Meer, D. Mintz, and A. Rosenfeld, "Robust-regression methods for computer vision: a review", *Int. J. of Computer Vision* **6**, 1 (1991) 59–70.
72. C. E. K. Mees, *The Theory of Photographic Process*, MacMillan, 1954.
73. J. W. Modestino and J. Zhang, "A Markov random field model-based approach to image interpretation", in *Markov Random Fields: Theory and Applications*, eds. R. Chellappa and A. Jain, Academic Press, Inc., Boston, 1993, pp. 369–408.
74. M. S. Murphy and L. M. Silverman, "Image model representation and line-by-line recursive restoration", *IEEE Trans. Automatic Control* **AC-23** (1978) 809–816.
75. N. E. Nahi, *Estimation Theory and Applications*, John Wiley & Sons, Inc., 1969.

76. N. E. Nahi, "Role of the recursive estimation in statistical image enhancement", *Proc. IEEE* **60** (1972) 872–877.
77. D. Panjwani and G. Healey, "Markov random field models for unsupervised segmentation of texured color images", *IEEE Trans. Pattern Analysis and Machine Intelligence* **PAMI-17**, 10 (1992) 939–954.
78. A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, 2nd edn., McGraw-Hill Book Company, 1984.
79. D. L. Phillips, "A technique for numerical solution of certain integral equations of the first kind", *J. Assoc. Comput. Mach.* **9** (1962) 97–101.
80. J. Pliz, *Bayesian Estimation and Experimental Design in Linear Regression Models*, John Wiley & Sons, Inc., 1991.
81. T. Poggio and V. Torre, "Ill-posed problems and regularization analysis in early vision", *Proc. Image Understanding Workshop*, Science Applications International Corporation, New Orleans, Louisiana, 1984, pp. 257 263.
82. S. R. Powell and L. M. Silverman, "Modeling of two-dimensional covariance functions with application to image restoration", *IEEE Trans. Automatic Control*, **AC-19**, 1 (1974) 8–13.
83. A. Rosenfeld and A. C. Kak, *Digital Picture Processing*, Academic Press, 1982.
84. B. Rouchouze, P. Mathieu, T. Gaidon, and M. Bariand, "Motion estimation based on Markov random fields", *Proc. IEEE Int. Conf. on Image Proc.*, vol. 3, Austin, TX, Nov. 1994, pp. 270–274.
85. R. R. Schultz and R. L. Stevenson, "Stochastic modeling and estimation of multispectral image data", *IEEE Trans. Image Processing* **4**, 8 (1995) 1109–1119.
86. M. Schwartz, *Information Transmission, Modulation, and Noise*, McGraw-Hill, 1962.
87. K. Sauer and C. A. Bouman, "A local update strategy for iterative reconstruction from projections", *IEEE Trans. Signal Processing* **41**, 2 (1993).
88. M. M. Sondhi, "Image restoration: the removal of spatially invariant degradations", *Proc. IEEE* **60** (1972) 842 853.
89. F. Spitzer, "Markov random fields and Gibbs ensembles", *Amer. Math. Mon.* **78** (1971) 142–154.
90. R. G. Staudte and S. J. Sheather, *Robust Estimation and Testing*, John Wiley & Sons, Inc., 1990.
91. H. Tan, S. Gelfand, and E. Delp, "A comparative cost function approach to edge detection", *IEEE Trans. Systems, Man and Cybernetics* **SMC-19**, 6 (1989) 1337–1349.
92. T. Taxt, P. Flynn, and A. Jain, "Segmentation of document images", *IEEE Trans. Pattern Analysis and Machine Intelligence* **11**, 12 (1989) 1322–1329.
93. A. N. Tikhonov and V. Y. Arsenin, *Solutions of Ill-Posed Problems*, V. H. Winston, 1977.
94. H. J. Trussell and B. R. Hunt "Sectioned methods for image restoration", *IEEE Trans. Acoustics, Speech, and Signal Processing* **ASSP-26**, 2 (1978) 157–164.
95. H. J. Trussell, "Notes on linear image restoration by maximizing the *a posteriori* probability", *IEEE Trans. Acoustics, Speech, and Signal Processing* **ASSP-26**, 2 (1978) 174–176.
96. H. L. Van Trees, *Detection, Estimation, and Modulation Theory*, John Wiley & Sons, Inc., 1968.
97. J. F. Walkup and R. C. Choens, "Image processing in signal-dependent noise", *Opt. Eng.* **13** (1974) 258–266.
98. A. S. Willsky, "Relationships between digital signal processing and control and estimation theory", *Proc. IEEE* **66**, 9 (1979) 996–1917.

99. J. W. Woods, "Two-dimensional discrete Markovian fields", *IEEE Trans. Information Theory* **18** (1972) 232–240.
100. J. W. Woods and C. H. Radewan, "Kalman filtering in two dimensions", *IEEE Trans. Information Theory* **23**, 4 (1977) 473–482.
101. J. W. Woods, "Correction to 'Kalman filtering in two dimensions'", *IEEE Trans. Information Theory* **25**, 5 (1979) 628–629.
102. J. W. Woods, "Two-dimensional Kalman filtering", *Two-Dimensional Digital Signal Processing: I*, ed. S. T. Huang, Springer-Verlag, 1981, pp. 155–205.
103. J. W. Woods and V. K. Ingle, "Kalman filtering in two dimensions: further results", *IEEE Trans. Acoustics, Speech, and Signal Processing* **29**, 2 (1981) 188–197.
104. J. J. Wu, R. E. Rink, T. C. Caelli, and V. G. Gourishakar, "Recovery of the 3D location and motion of rigid object through camera image (an extended Kalman filter approach)", *Int. J. of Computer Vision* **3** (1988) 373–394.
105. J. Zerubia and R. Chellappa, "Mean field annealing using compound Gauss Markov random fields for edge detection and image estimation", *IEEE Trans. Neural Networks* **4**, 4 (1993) 703–709.
106. J. Zhang, "The mean field theory in EM procedures for blind Markov random field image restoration", *IEEE Trans. Image Processing* **2**, 1 (1993) 27–40.
107. J. Zhang and G. G. Hanauer, "The application of mean field theory to image motion estimation", *IEEE Trans. Image Processing* **4**, 1 (1995) 19–33.
108. X. H. Zhuang, T. Wang, and P. Zhang, "A highly robust estimator through partially likelihood function modeling and its application in computer vision", *IEEE Trans. Pattern Analysis and Machine Intelligence* **PAMI-14**, 1 (1992) 19–35.

Zhi-Qiang Liu is an Associate Professor with the Department of Computer Science, The University of Melbourne. He received a M.A.Sc. degree in aerospace engineering from the Institute for Aerospace Studies, The University of Toronto, and a Ph.D. degree in electrical engineering from The University of Alberta, Canada.

Dr. Liu is the leader and the Principal Investigator of the Computer Vision and Machine Intelligence Laboratory (CVMIL) at the Department of Computer Science, The University of Melbourne.

His research interests include image processing, computer vision, neural networks, image information management, and intelligent and knowledge-based systems. Dr. Liu has over 100 refereed publications and one book in these areas.