

Bayesian Approaches for Blind Restoration of Nonstationary Images

By Tom E. Bishop and James R. Hopgood

Institute for Digital Communications, Joint Research Institute for Signal & Image Processing,
School of Engineering & Electronics, King's Buildings, The University of Edinburgh,
Edinburgh EH9 3JL t.e.bishop@ed.ac.uk, James.Hopgood@ed.ac.uk

Abstract

Blind image restoration is an ill-posed problem where the degree of uncertainty must be countered by explicit use of prior knowledge. The hierarchical Bayesian paradigm allows for a robust way to estimate both image, blur, and model parameters in the presence of uncertainty. We consider a nonstationary image model that aids restorations, and describe methods available in this framework to perform inference.

1. Introduction

Many images taken with conventional and digital cameras undergo blurring and other degradations. The rapidly expanding digital photography and camera phone industries are placing tough demands on improvements in image quality at lower costs; digital image restoration can help give sharper pictures, and reconstruct blurry images of unique events.

In practice however, the exact nature of the blur in an image is unknown and must be estimated from the image itself: the blind image restoration (BIR) problem. By mathematically modelling both the image and blur we can obtain probabilistic estimates for both. Use of a fully stochastic Bayesian approach allows for robust parameter estimates in the presence of uncertainty. When the degree of uncertainty is great, as in the BIR problem, it is necessary to impose as much structure as possible representing our prior beliefs in order to recover the original image and blur. This may be done effectively with a hierarchical model. Most existing BIR techniques have assumed a stationary image model. However investigations suggest that using a nonstationary model both aids accurate blur identification, and improves the restored image.

Under this framework several restoration options may be considered; a full review of BIR methods may be found in Bishop *et al.* (2007). In the following sections, we outline the maximum marginalised *a posteriori* (MMAP) blur identification method, the Markov chain Monte Carlo (MCMC) approaches, and the Variational Bayesian (VB) approach.

2. Problem Formulation

The BIR problem consists of estimating an unobserved true image, $f(i, j)$, and blur or point spread function (PSF), $h(i, j)$ from an observed degraded image $g(i, j)$, which is modelled as a discrete convolution $f(i, j) * h(i, j)$ plus additive white Gaussian noise (WGN), $w(i, j)$. This may be expressed in matrix-vector form with lexicographically ordered image vectors and a Block Toeplitz with Toeplitz Blocks (BTTB) PSF matrix as $\mathbf{g} = \mathbf{H}\mathbf{f} + \mathbf{w}$. Since $\mathbf{w} \sim \mathcal{N}(0, \mathbf{Q}_w)$, the likelihood of the observations conditioned on the true image may be written $p(\mathbf{g} | \mathbf{f}, \mathbf{H}, \mathbf{Q}_w) = \mathcal{N}(\mathbf{g} | \mathbf{H}\mathbf{f}, \mathbf{Q}_w)$.

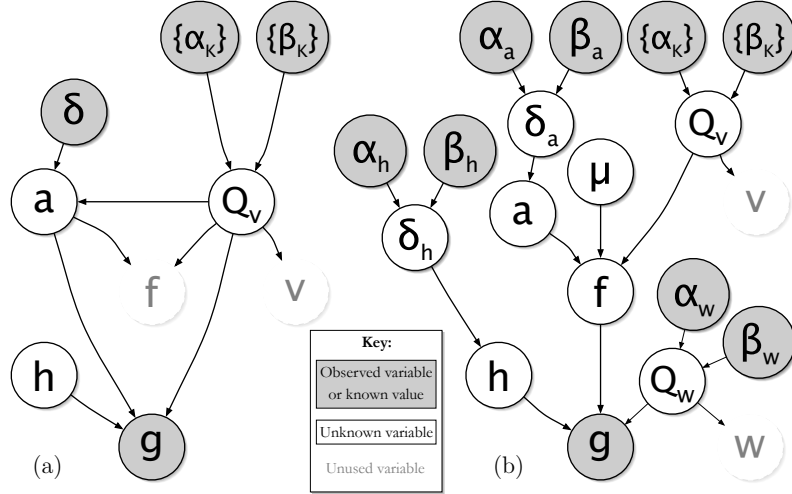


FIGURE 1. (a) Model 1: Simplified hierarchical Bayesian network for MMAP approach; (b) Model 2: Full model for Variational Bayes and Gibbs sampling approaches

In the Bayesian framework we introduce prior models for the unknown parameters, which in turn depend on *hyperparameters*. A further stage in the hierarchical Bayesian framework is to regard the hyperparameters as unknown variables and model our prior knowledge of their distributions with *hyperpriors*. This abstraction allows for greater robustness of estimates when we are less confident in the observations and the hyperparameters themselves.

Two models will be considered, as shown graphically in Fig. 1. The first (Model 1) is a simplification of the second (Model 2) in which the true image is marginalised from the model. The meanings of the various variables and parameters will now be explained.

2.1. Nonstationary Image model

A block-stationary autoregressive (BSAR) model, belonging to the class of Markov random field (MRF) models, is used for the true image \mathbf{f} . The model segments f into R blocks or regions. Each region is modelled by a 2D AR process plus a mean. Defining the vector of lexicographically ordered AR coefficients in each region as $\mathbf{a}_K = \text{vec}\{a_K(k, l)\}$, the stacked vector of \mathbf{a}_K 's as \mathbf{a} , the image model may then be written as

$$\mathbf{f} = \boldsymbol{\mu} + \mathbf{A}(\mathbf{f} - \boldsymbol{\mu}) + \mathbf{v} = \boldsymbol{\mu} + (\mathbf{F} - \mathbf{M})\mathbf{a} + \mathbf{v}, \quad (2.1)$$

where \mathbf{v} is a WGN excitation signal $\mathbf{v} \sim \mathcal{N}(0, \mathbf{Q}_v)$, $\boldsymbol{\mu}$ is a vector of block means, \mathbf{M} is a block-diagonal mean matrix, \mathbf{F} is a data matrix and \mathbf{A} is similar in form to \mathbf{H} but non-stationary. Thus the probability density function (PDF) of \mathbf{f} may be written as $p(\mathbf{f} | \mathbf{a}, \mathbf{Q}_v, \boldsymbol{\mu}) = \mathcal{N}(\mathbf{f} | \boldsymbol{\mu}, \boldsymbol{\Sigma}_f)$ where $\boldsymbol{\Sigma}_f = \mathbb{E}[\mathbf{f}\mathbf{f}^T] = (\mathbf{I} - \mathbf{A})^{-1}\mathbf{Q}_v(\mathbf{I} - \mathbf{A})^{-T}$. A causal image model (lower triangular \mathbf{A}) is used such that the Jacobian $\det|\mathbf{I} - \mathbf{A}|$ is unity.

2.2. Prior blur models

In Model 2, the prior used for the PSF is a Gaussian smoothness prior, of a similar type to that for the image, but with a stationary covariance described by the discrete Laplacian operator \mathbf{C} , i.e. $p(\mathbf{h} | \delta_h) = \mathcal{N}(\mathbf{h} | 0, (\delta_h \mathbf{C}^T \mathbf{C})^{-1})$, where δ_h controls the strength of the prior. The choice of prior for the blur in Model 1 depends upon its parameterisation, however for the MMAP method described, a flat prior has been used for simplicity.

2.3. Hyperprior models

To remove dependence on exact specification of the prior models, whose hyperparameters are in practice unknown, the hyperpriors may be specified. A standard way to do this, resulting in more tractable solutions, is to use conjugate priors.

For the AR parameter vector, a Gaussian $p(\mathbf{a} | \delta_a) = \prod_{K=1}^R \mathcal{N}(\mathbf{a}_K | 0, \delta_a^{-1} \mathbf{I})$ is used. A flat prior is used for the regional means, $\boldsymbol{\mu}$, as they are readily obtained from \mathbf{f} . For the other hyperparameters, which are all (inverse) variances of Gaussian distributions, the standard conjugate priors are (inverse) Gamma distributions; thus we have:

$$\begin{aligned} p(\delta_a) &= \mathcal{G}(\delta_a | \alpha_a, \beta_a) & p(\delta_h) &= \mathcal{G}(\delta_h | \alpha_h, \beta_h) \\ p(\mathbf{Q}_v) &= \prod_{K=1}^R \mathcal{IG}(\sigma_{v_K}^2 | \alpha_{v_K}, \beta_{v_K}) & p(\mathbf{Q}_w) &= \mathcal{IG}(\sigma_w^2 | \alpha_w, \beta_w) \end{aligned}$$

where $\sigma_{v_K}^2$ and σ_w^2 are the variances of each block of \mathbf{v} and of \mathbf{w} respectively.

3. MMAP solution

The MMAP solution found using Model 1 has been investigated in Bishop & Hopgood (2006), and the approach is summarised here. In Model 1, we consider firstly marginalising \mathbf{f} from the full likelihood, giving $p(\mathbf{g} | \mathbf{a}, \mathbf{h}, \mathbf{Q}_v, \mathbf{Q}_w, \boldsymbol{\mu}) = \mathcal{N}(\mathbf{g} | \mathbf{H}\boldsymbol{\mu}, \boldsymbol{\Sigma}_g)$, where $\boldsymbol{\Sigma}_g = \mathbf{H}(\mathbf{I} - \mathbf{A})^{-1} \mathbf{Q}_v (\mathbf{I} - \mathbf{A})^{-T} \mathbf{H}^T + \mathbf{Q}_w$. This result may also be found by combining $p(\mathbf{f} | \mathbf{a}, \mathbf{Q}_v, \boldsymbol{\mu})$ with $p(\mathbf{w})$ under the original observation equation $\mathbf{g} = \mathbf{H}\mathbf{f} + \mathbf{w}$.

However in order to find the MMAP estimate, it is also necessary to approximate $\boldsymbol{\Sigma}_g$ by omitting \mathbf{Q}_w . This amounts to using an inverse filter estimate $\hat{\mathbf{f}} = \mathbf{H}^{-1}\mathbf{g}$ in the MMAP formulae. Note that in practice at this stage we also remove the block-based sample mean from $\hat{\mathbf{f}}$ and $\hat{\mathbf{F}}$, thus removing the dependence on $\boldsymbol{\mu}$. The posterior is then found as $p(\mathbf{h}, \mathbf{a}, \mathbf{Q}_v | \mathbf{g}) \propto p(\mathbf{g} | \mathbf{h}, \mathbf{a}, \mathbf{Q}_v) p(\mathbf{h}) p(\mathbf{a} | \mathbf{Q}_v) p(\mathbf{Q}_v)$. The *nuisance parameters* \mathbf{a} and \mathbf{Q}_v may now be marginalised out to give the marginalised posterior expression:

$$p(\mathbf{h} | \mathbf{g}) = \int \cdots \int p(\mathbf{h}, \mathbf{a}, \mathbf{Q}_v | \mathbf{g}) d\mathbf{a} \cdot d\mathbf{Q}_v. \quad (3.1)$$

The resulting analytic expression (found with standard Gamma and Gaussian integrals) may then be maximised numerically to obtain the MMAP estimate. The key advantage of the MMAP approach is in reduction of the parameter search space to a manageable size; however there are clearly some approximations that have been made from the full model in order to obtain an analytic solution to (3.1).

It is the presence of the $\det |\mathbf{H}|$ term in the marginalised posterior though that causes the most difficulties when trying to use a non-causal blur model, which will more accurately model a real blur. The problem arises because \mathbf{H} is in fact often singular, or at best badly conditioned — the omission of \mathbf{Q}_w has removed the regularising effect in the determinant. Furthermore, the determinant is highly sensitive to the boundary conditions used in the model, which are the subject of further investigation.

4. MCMC approaches

MCMC procedures are another method available under the Bayesian framework to estimate the unknown parameters, by drawing samples from the posterior distribution. They can in theory provide solutions to arbitrarily difficult problems. In the causal blur case, a Gibbs sampler may be applied to the posterior of Model 1, $p(\mathbf{h}, \mathbf{a}, \mathbf{Q}_v | \mathbf{g})$, to simultaneously estimate the blur and other parameters; this approach is analogous to that used in the 1D case in Hopgood & Rayner (2003). However this does not get around the difficulty of the Jacobian $\det |\mathbf{H}|$ that is present under a non-causal blur model.

To sample the posterior distribution including the Jacobian would require using a Metropolis-Hastings step to sample from the conditional $p(\mathbf{h} | \mathbf{a}, \mathbf{Q}_v, \mathbf{g})$, since it is a non-standard distribution in \mathbf{h} . However the proposal distribution would have to model the effect of the Jacobian, and finding such a distribution that doesn't lead to samples being continuously rejected is no simple task. This Jacobian term actually results due to the marginalisation of \mathbf{f} from the model; as such it may be simpler to consider the full model in Model 2.

Under Model 2, the Gibbs sampler may be run even with a non-causal blur. Sampling for \mathbf{f} is now included, resulting in a stochastic image restoration step, and an equivalent blur estimate conditional on \mathbf{f} , now independent of $\det |\mathbf{H}|$. All the required conditionals, namely $p(\mathbf{h} | \mathbf{f}, \mathbf{Q}_w, \delta_h, \mathbf{g})$, $p(\mathbf{f} | \mathbf{h}, \mathbf{a}, \mathbf{Q}_v, \mathbf{Q}_w, \boldsymbol{\mu}, \mathbf{g})$, $p(\mathbf{Q}_w | \mathbf{f}, \mathbf{h}, \mathbf{g})$, $p(\mathbf{a} | \mathbf{f}, \boldsymbol{\mu}, \mathbf{Q}_v, \delta_a)$, $p(\boldsymbol{\mu} | \mathbf{f}, \mathbf{a}, \mathbf{Q}_v)$, $p(\mathbf{Q}_v | \mathbf{f}, \mathbf{a}, \boldsymbol{\mu}, \{\alpha_K\}, \{\beta_K\})$, $p(\delta_h | \mathbf{h}, \alpha_h, \beta_h)$, and $p(\delta_a | \mathbf{a}, \alpha_a, \beta_a)$ may be found in terms of standard distributions and hence sampled. The only remaining problem is the computational complexity of sampling for \mathbf{f} .

5. Variational Bayesian approach

The VB approach is a fairly new alternative to using MCMC; rather than sampling, we approximate the intractable posterior distribution by a separable function, hoping to achieve a good compromise between accuracy and speed. It may also be seen as a fully stochastic version of the expectation maximisation (EM) algorithm, where the distributions of all the parameters and hidden variables are estimated rather than just using point estimates for the parameters.

Molina *et al.* (2006) recently applied the VB algorithm to the BIR problem. The model used therein is the same as that of Model 2, except a simpler stationary Gaussian prior (of the same type described in §2.2) is used for the image model. The stationary models allow for rapid calculations in the discrete frequency domain. We are currently working on extending this to the nonstationary image prior shown in Model 2, to improve the restorations and estimates. This requires finding efficient and accurate solutions in the spatial domain, taking account of boundary effects. The connection to the Gibbs sampling approach is interesting; at each iteration rather than sampling, expectations and moments are found of the same conditional distributions as those mentioned above.

6. Conclusions

Two different hierarchical models, and three Bayesian approaches for estimating their parameters for BIR have been discussed. Each has its relative merits in terms of complexity and accuracy. The nonstationary model has been shown to aid blur identification in the MMAP case, and is currently being applied to the other two cases.

REFERENCES

- BISHOP, T. E., BABACAN, S. D., AMIZIC, B., CHAN, T., MOLINA, R., & KATSAGGELOS, A. K. 2007 Blind image deconvolution: problem formulation and existing approaches. In *Blind image deconvolution: Theory and Applications* (ed. P. Campisi & K. Egiazarian). CRC.
- BISHOP, T. E., & HOPGOOD, J. R. 2006 Blind image restoration using a block-stationary signal model. In *Proc. ICASSP 2006, IEEE International Conference on Acoustics, Speech and Signal Processing*. vol. 2, pp. II-853-II-856.
- HOPGOOD, J. R., & RAYNER, P. J. W. 2003 Blind single channel deconvolution using nonstationary signal processing. *IEEE Transactions on Speech and Audio Processing* **11**, pp. 476-488.
- MOLINA, R., MATEOS, J., & KATSAGGELOS, A. K. 2006 Blind deconvolution using a variational approach to parameter, image, and blur estimation. *IEEE Trans Image Process.* **15**, pp. 3715-3727.