

BLIND RESTORATION OF BLURRED PHOTOGRAPHS VIA AR MODELLING AND MCMC

Tom E. Bishop^a, Rafael Molina^b, James R. Hopgood^a

a) IDCOM, Joint Research Institute for Signal & Image Processing,
School of Engineering & Electronics, The University of Edinburgh, Edinburgh, EH9 3JL, UK

b) Dept. Ciencias de la Computación e I. A., Univ. de Granada, 18071 Granada, Spain.

t.e.bishop@ed.ac.uk, rms@decsai.ugr.es, james.hopgood@ed.ac.uk

ABSTRACT

We propose a new image and blur prior model, based on non-stationary autoregressive (AR) models, and use these to blindly deconvolve blurred photographic images, using the Gibbs sampler. As far as we are aware, this is the first attempt to tackle a real-world blind image deconvolution (BID) problem using Markov chain Monte Carlo (MCMC) methods. We give examples with simulated and real out-of-focus images, which show the state-of-the-art results that the proposed approach provides.

Index Terms— Blind Deconvolution, Learned Image Prior, Nonstationary Image Models, Bayesian methods, Gibbs Sampler

1. INTRODUCTION

Photographic images are commonly degraded by motion or out-of-focus blurs. In practice, the point-spread functions (PSFs) that cause these degradations are unknown, and the need to estimate them simultaneously with a restoration of the image results in the blind image deconvolution (BID) problem [1]. This is an extremely ill-posed problem, and as such, proper incorporation of as much prior knowledge as possible regarding the types of image and blur that are likely to be encountered is essential to properly constrain the solution.

In this paper, we develop nonstationary models incorporating AR parameter estimation, which are able to adapt very locally to the variety of smooth, textured, and edge regions encountered in typical images and PSFs. The new model is an extension of the models used in previous work [2, 3].

An overview and collection of the many recent methods proposed for tackling BID is given in [1]. Many of these, e.g. [4, 5, 2], have used the variational Bayesian (VB) approach as a means to mitigate the intractability of a direct solution of these models.

One major problem with deterministic alternating minimization (AM) or expectation maximization (EM) type algorithms is that they are too easily trapped in local optima, especially when the state space is very large and the likelihood is non-linear in the parameters, as is the case for BID problems. Fully stochastic methods allow searching beyond the local optimum. Although MCMC methods are often notoriously computationally intense, we also use simulated annealing (SA) to improve convergence rates to a level that is now becoming feasible on modern hardware. See [6] for an attempt to use SA to find maximum likelihood (ML) estimates in the BID problem.

The main differences of the model in the proposed approach from the model in [2] are as follows:

This work has been supported by a UK-EPSC Studentship, by grant EPSRC-EP/F023073/1(P), and by the Spanish research programme Consolider Ingenio 2010: MIPRCV (CSD2007-00018) and the Ministerio de Educacion y Ciencia under contract TIN2007-65533.

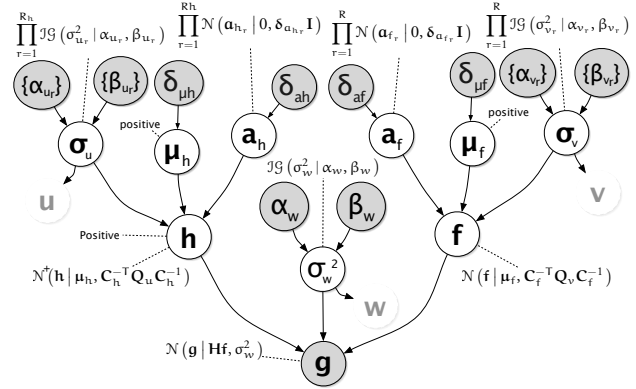


Fig. 1. Graphical model showing relationships between variables

- AR parameter estimation is included, in order to model textures in the image and better represent edge details;
- Local region means are also estimated to enhance model fit.
- The grid used to define regions in the image is not fixed, but chosen randomly at each iteration, avoiding block artefacts.
- PSF positivity is imposed via rectified Gaussian distributions.

2. PROBLEM FORMULATION

The BID 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), $w(i, j)$. In matrix-vector form this may be written as

$$\mathbf{g} = \mathbf{H}\mathbf{f} + \mathbf{w} = \mathbf{F}_h\mathbf{h} + \mathbf{w} \quad (1)$$

where the vectors \mathbf{g} (of length L) and \mathbf{f} (of length L_+) are lexicographically ordered, and \mathbf{H} is a block Toeplitz with Toeplitz blocks (BTTB) matrix, with defining sequence \mathbf{h} of length L_h . An extended observation model is used [2] to take into account the influence of the parts of the original image outside the observation boundary that blur into the field of view. This implies that \mathbf{H} is non-square, of size $L \times L_+$, and the support \mathcal{S}_g of the observation \mathbf{g} is smaller than the support \mathcal{S}_f of \mathbf{f} , by the size of the blur support, \mathcal{S}_h . Based on the WGN assumption, $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \sigma_w^2 \mathbf{I})$, the likelihood of the observations conditioned on the true image may then be written

$$p(\mathbf{g} | \mathbf{f}, \mathbf{h}, \sigma_w^2) = \mathcal{N}(\mathbf{g} | \mathbf{H}\mathbf{f}, \sigma_w^2 \mathbf{I}). \quad (2)$$

2.1. Nonstationary Image Model and Blur Model

The image and blur priors in the proposed model are based on a local (nonstationary) modelling. The model is symmetric in the image and blur, in that the structure of the hierarchical model for each is the same (see Fig. 1). In previous work, a causal block-stationary AR (BSAR) model was used to represent the image in [3]. We will use a similar, non-causal version, but now including a local mean, making the model more locally adaptive.

The image model will be described first in detail. Assume we have a segmentation of the canvas \mathcal{S}_f , with R blocks or regions. The ordered vector of pixels inside each region $\mathcal{R}_r \subset \mathcal{S}_f$ is denoted \mathbf{f}_r , and is of length L_r , for $r \in \{1 \cdots R\}$. The principle of the BSAR model is to find a set of coefficients \mathbf{a}_f , that model the image in each local region \mathcal{R}_r , such that each pixel is the sum of an excitation signal $v_r(i, j)$ and the weighted combination of the pixel's neighbours within a support region of shape \mathcal{S}_{a_f} , whose size $L_{a_f} = (2P + 1)^2 - 1$. If the model is a good fit, the excitation signal, or modelling error, will be small and uncorrelated with the image, and resemble WGN.

Block boundaries Because the signal at the boundary of \mathcal{R}_r depends on its neighbours outside of the region, according to the extent of \mathcal{S}_{a_f} , we define the extended region \mathcal{R}_{r+} as a dilation of the set \mathcal{R}_r by \mathcal{S}_{a_f} . The intersection of \mathcal{R}_{r+} and \mathcal{R}_r is the set of boundary pixels \mathcal{R}_{rb} exterior to \mathcal{R}_r (see [2]). Image pixels within \mathcal{R}_{rb} are vectorised as \mathbf{f}_{rb} , of length L_{rb} , then we let $\mathbf{f}_{r+} = [\mathbf{f}_r^T \mathbf{f}_{rb}^T]^T$, which is a vector of length L_{r+} containing all the pixels in \mathcal{R}_{r+} .

2D AR model in each block For a zero-mean signal in \mathcal{R}_r , $\mathbf{f}_r^*(i, j)$, the 2D AR model of order P is defined according to:

$$\mathbf{f}_r^*(i, j) = \sum_{\forall(k, l) \in \mathcal{S}_{a_f}} \mathbf{f}_r^*(i - k, j - l) \mathbf{a}_{f_r}(k, l) + v_r(i, j) \quad (3)$$

Since natural images are inherently locally non-zero-mean, a mean μ_{f_r} is introduced to the model, such that $\mathbf{f}_r^*(i, j) = \mathbf{f}_r(i, j) - \mu_{f_r}$. It is critical, in order to avoid introducing artificial discontinuities in the model, that pixel values $\mathbf{f}_r^*(i - k, j - l)$ on the RHS of (3) have a common mean subtracted, whether they are in \mathcal{R}_r or \mathcal{R}_{rb} . Observe that another neighbouring region, $\mathcal{R}_{r'}$, will also contain some of these pixels, but the mean will be different in this case.

In vector form, with suitable ordering, (3) is written as

$$\mathbf{f}_r - \mu_{f_r} = \mathbf{A}_{f_{r+}} (\mathbf{f}_{r+} - \mu_{f_{r+}}) + \mathbf{v}_r \quad (4)$$

where $\mu_{f_r} = \mu_{f_r} \mathbf{1}_{L_r}$, $\mu_{f_{r+}} = \mu_{f_r} \mathbf{1}_{L_{r+}}$, and $\mathbf{A}_{f_{r+}}$ is a $L_r \times L_{r+}$ BTTB matrix with defining sequence \mathbf{a}_{f_r} .¹ Defining $\mathbf{I}_{r+} = [\mathbf{I}_{L_r} \mathbf{0}_{L_r \times L_{rb}}]$, the excitation signal in \mathcal{R}_r may then be written

$$\mathbf{v}_r = (\mathbf{I}_{r+} - \mathbf{A}_{f_{r+}}) (\mathbf{f}_{r+} - \mu_{f_{r+}}) = \mathbf{C}_{f_{r+}} \mathbf{f}_{r+} - \boldsymbol{\eta}_{f_r} \quad (5)$$

where $\mathbf{C}_{f_{r+}} = (\mathbf{I}_{r+} - \mathbf{A}_{f_{r+}})$ and $\boldsymbol{\eta}_{f_r} = \mathbf{C}_{f_{r+}} \mu_{f_{r+}}$. We will also write $\mathbf{A}_{f_{r+}} (\mathbf{f}_{r+} - \mu_{f_{r+}}) = (\mathbf{F}_{a_r} - \mu_{f_r} \mathbf{1}_{L_r \times L_a}) \mathbf{a}_{f_r} = \mathbf{F}_{a_r}^* \mathbf{a}_{f_r}$.

Whole image model Now define $\mathbf{v}' = [v_1^T, \dots, v_R^T]^T$, the concatenation of the excitation across all the regions, and re-order this vector as $\mathbf{v} = \mathbf{P} \mathbf{v}'$. \mathbf{P} is a permutation matrix that converts the block-wise scanning to lexicographic ordering across \mathcal{S}_f . Then we have

$$\mathbf{v} = \mathbf{C}_f \mathbf{f} - \boldsymbol{\eta}_f \quad (6)$$

¹We use the notation $\mathbf{1}_M$ or $\mathbf{0}_M$ to represent a vector of M ones or zeros, and $\mathbf{1}_{M \times N}$ or $\mathbf{0}_{M \times N}$ an $M \times N$ matrix of ones or zeros. \mathbf{I}_M is the $M \times M$ identity matrix.

where $\boldsymbol{\eta}_f = \mathbf{P} [\boldsymbol{\eta}_{f_1}^T, \dots, \boldsymbol{\eta}_{f_R}^T]^T$, and \mathbf{C}_f is a non-stationary BTTB matrix with entries taken from the matrices $\mathbf{C}_{f_{r+}}$ for $r = 1 \cdots R$, relating pixels in \mathbf{v} and \mathbf{f} according to (5).

Now, using the probability density function (PDF) of the excitation signal in each block $\mathbf{v}_r \sim \mathcal{N}(\mathbf{0}, \sigma_{v_r}^2 \mathbf{I}_{L_r})$, and the fact that these are independent, we may write

$$p(\mathbf{v} | \boldsymbol{\sigma}_v) = \prod_{r \in \{1 \cdots R\}} (2\pi)^{-\frac{L_r}{2}} \sigma_{v_r}^{-L_r} \exp\left[-\frac{1}{2} \boldsymbol{\sigma}_{v_r}^{-2} \mathbf{v}_r^T \mathbf{v}_r\right] \quad (7)$$

$$= (2\pi)^{-L+} \det|\mathbf{Q}_v|^{-1} \exp\left[-\frac{1}{2} \mathbf{v}^T \mathbf{Q}_v^{-1} \mathbf{v}\right] \quad (8)$$

where $\boldsymbol{\sigma}_v = [\sigma_{v_1}^2, \dots, \sigma_{v_R}^2]^T$, and the diagonal matrix \mathbf{Q}_v is formed from the excitation variances in each block, $\sigma_{v_r}^2$. Then by applying a probability transformation to (8), using (6), the extended image prior PDF may be written as

$$p(\mathbf{f} | \mathbf{a}_f, \boldsymbol{\sigma}_v, \boldsymbol{\mu}_f) = (2\pi)^{-L+} \det|\boldsymbol{\Sigma}_f^{-1}| \cdot \exp\left[-\frac{1}{2} (\mathbf{C}_f \mathbf{f} - \boldsymbol{\eta}_f)^T \mathbf{Q}_v^{-1} (\mathbf{C}_f \mathbf{f} - \boldsymbol{\eta}_f)\right] \quad (9)$$

$$= \mathcal{N}(\mathbf{f} | \bar{\mathbf{f}}, \boldsymbol{\Sigma}_f), \quad (10)$$

where $\boldsymbol{\Sigma}_f = \mathbb{E}[(\mathbf{f} - \bar{\mathbf{f}})(\mathbf{f} - \bar{\mathbf{f}})^T] = \mathbf{C}_f^{-1} \mathbf{Q}_v \mathbf{C}_f^{-T}$ (11)

$$\text{and } \bar{\mathbf{f}} = \mathbf{C}_f^{-1} \boldsymbol{\eta}_f. \quad (12)$$

Here we denote $\boldsymbol{\mu}_f = [\mu_{f_1}, \dots, \mu_{f_R}]^T$, i.e. the vector of coefficients parameterising the mean image, rather than the mean image within \mathcal{R} ; also we define $\mathbf{a}_f = [\mathbf{a}_{f_1}^T, \dots, \mathbf{a}_{f_R}^T]^T$. Note that the while the excitation signal \mathbf{v} is independent in each region, the image in each region depends on its neighbours.

2.2. Hyperprior models

The image prior model depends upon three hyperparameters: \mathbf{a}_f , $\boldsymbol{\sigma}_v$, $\boldsymbol{\mu}_f$. Because in the hierarchical Bayesian approach it is considered that these parameters are also unknowns, the next stage is specification of hyperpriors defining the distributions of their likely values. Conjugate priors are used to ensure that the resulting conditional distributions which will be sampled from will be of the same known forms.

The AR parameters are assumed to be Gaussian distributed. While an uninformative uniform distribution may be used reliably for the means (there are a large number of observed data points available for their estimation), we choose to constrain them to be positive by using a vague (large variance) rectified Gaussian, denoted by \mathcal{N}^+ . The distributions are then:

$$p(\mathbf{a}_f | \boldsymbol{\delta}_{a_f}) = \prod_{r \in \{1 \cdots R\}} \mathcal{N}(\mathbf{a}_{f_r} | 0, \boldsymbol{\delta}_{a_{f_r}} \mathbf{I}_{L_{a_f}}) \quad (13)$$

$$p(\boldsymbol{\mu}_f | \boldsymbol{\delta}_{\mu_f}) = \prod_{r \in \{1 \cdots R\}} \mathcal{N}^+(\mu_{f_r} | 0, \boldsymbol{\delta}_{\mu_f}) \quad (14)$$

For the other parameters, which are variances of Gaussian distributions, the standard conjugate priors are inverse-Gamma (IG):

$$p(\boldsymbol{\sigma}_v | \boldsymbol{\alpha}_v, \boldsymbol{\beta}_v) = \prod_{r \in \{1 \cdots R\}} \mathcal{IG}(\sigma_{v_r}^2 | \alpha_{v_r}, \beta_{v_r}) \quad (15)$$

$$p(\sigma_w^2 | \alpha_w, \beta_w) = \mathcal{IG}(\sigma_w^2 | \alpha_w, \beta_w) \quad (16)$$

2.3. Nonstationary blur prior

The prior for \mathbf{h} takes on the same structure as that for \mathbf{f} , but positivity is imposed on the pixels in \mathbf{h} itself by means of a rectified Gaussian. The set of distributions describing the blur prior are then:

$$p(\mathbf{h} | \mathbf{a}_h, \sigma_u, \mu_h) = \mathcal{N}^+(\mathbf{h} | \bar{\mathbf{h}}, \Sigma_h), \quad (17)$$

$$\text{where } \Sigma_h = \mathbb{E}[(\mathbf{h} - \bar{\mathbf{h}})(\mathbf{h} - \bar{\mathbf{h}})^T] = \mathbf{C}_h^{-1} \mathbf{Q}_u \mathbf{C}_h^{-T} \quad (18)$$

$$\text{and } \bar{\mathbf{h}} = \mathbf{C}_h^{-1} \boldsymbol{\eta}_h, \quad (19)$$

$$p(\mathbf{a}_{h_r} | \delta_{a_{h_r}}) = \mathcal{N}(\mathbf{a}_{h_r} | 0, \delta_{a_{h_r}} \mathbf{I}_{L_{a_h}}), \quad (20)$$

$$p(\mu_{h_r} | \delta_{\mu_h}) = \mathcal{N}^+(\mu_{h_r} | 0, \delta_{\mu_h}), \quad (21)$$

$$p(\sigma_{u_r}^2 | \alpha_{u_r}, \beta_{u_r}) = \mathcal{IG}(\sigma_{u_r}^2 | \alpha_{u_r}, \beta_{u_r}), \quad (22)$$

where all parameters are defined analogously to their counterparts in §2.1 and §2.2, with S_h partitioned into R_h blocks, denoted \mathcal{R}_{h_r} . Note that this is a very flexible definition of the blur model; it contains as a particular case the model used in [4], as well as any global autoregression.

3. BAYESIAN INFERENCE VIA MCMC

Inference in the Bayesian paradigm is based upon the posterior distribution of unknowns $\Theta = \{\mathbf{f}, \mathbf{a}_f, \sigma_v, \mu_f, \mathbf{h}, \mathbf{a}_h, \sigma_u, \mu_h, \sigma_w^2\}$, conditional on the observed data: $p(\Theta | \mathbf{g}) \propto p(\mathbf{g} | \Theta) p(\Theta)$.

With such a large parameter space, it is infeasible to perform direct optimisation or calculation of this function. Marginalisation of all nuisance variables is not tractable for this model. Methods of solution available thus include approximation via variational methods, or direct sampling of the parameter space using MCMC methods. We choose the second option, and use the Gibbs sampler with SA to explore the space. VB methods approximate the posterior by a separable distribution, whereas Gibbs sampling allows for full simulation of the posterior. Several issues that had to be overcome in practice will be discussed in this section, including: bulk-sampling of the image; positivity of the PSF, and improving convergence speed.

The Gibbs sampler [7, 9] draws samples $\theta^{(k)}$ at iteration k from the distribution of each $\theta \in \Theta$ in turn, conditional on the previously sampled values of all the other unknowns. These samples form a Markov chain, whose stationary distribution is the desired posterior.

These conditional PDFs, $p(\theta | \Theta_{\setminus \theta})$, only need be known up to a constant of proportionality; hence they are calculated simply by taking the product of the terms from $p(\mathbf{g} | \Theta) p(\Theta)$ that contain the variable θ of interest. This procedure results in the conditionals shown in Algorithm 1, where we also define the sets of image and blur parameters $\boldsymbol{\theta}_{f_r} = \{L_{r+}, \alpha_{v_r}, \beta_{v_r}, \delta_{\mu_f}, \delta_{a_{f_r}}, L_{a_f}\}$ and $\boldsymbol{\theta}_{h_r} = \{L_{h_{r+}}, \alpha_{u_r}, \beta_{u_r}, \delta_{\mu_h}, \delta_{a_{h_r}}, L_{a_h}\}$.

Sampling procedure The scalar Gaussian and IG distributions are easily sampled from standard pseudo-random number generators (the IG by the reciprocal of Gamma distributed samples). For the multivariate Gaussians, normally a Cholesky decomposition of the covariance matrix is found, but this is intractable for such a large matrix as $\text{cov}^k[\mathbf{f}]$. Instead, a conjugate gradients least squares (CGLS) solver is used [8]. For example, \mathbf{f} is sampled avoiding calculation of $\text{cov}^k[\mathbf{f}]$ and $\mathbb{E}^k[\mathbf{f}]$ by drawing $\boldsymbol{\nu} \leftarrow \mathcal{N}(\boldsymbol{\nu} | \mathbf{0}_{L+L+}, \mathbf{I}_{L+L+})$ and iteratively solving

$$\mathbf{M}^T \mathbf{M} \mathbf{x} = \mathbf{M}^T \mathbf{y} \quad (23)$$

Input: $\mathbf{g}, \mathbf{h}^{(0)}, \boldsymbol{\theta}_{f_r}, \boldsymbol{\theta}_{h_r}$
Output: $\hat{\mathbf{f}} = \frac{1}{K-k_0} \sum_{k=k_0}^K \mathbf{f}^{(k)}, \hat{\mathbf{h}} = \frac{1}{K-k_0} \sum_{k=k_0}^K \mathbf{h}^{(k)}$

- 1 **for** $k = 1 \dots n$ **do**
- 2 offset $\mathcal{R}_r, \forall r = 1 \dots R$, by $(i, j) \leftarrow \text{Uniform}(m_r)$;
- 3 $\mathbf{w}^{(k)} = \mathbf{g} - \mathbf{H} \mathbf{f}$;
- 4 $(\sigma_w^2)^{(k)} \leftarrow \mathcal{IG}(\sigma_w^2 | \alpha_w + \frac{L_{\pm}}{2}, \beta_w + \frac{1}{2} \mathbf{w}^T \mathbf{w})$;
- 5 **for** $r = 1 \dots R$: **do**
- 6 $\{(\sigma_{v_r}^2)^{(k)}, \mu_{f_r}^{(k)}, \mathbf{a}_{f_r}^{(k)}\}$
 $\leftarrow \text{ParamSampling}(r, \sigma_w^2, \mathbf{C}_{f_r}, \mathbf{f}_{r+}, \mu_{f_r}, \boldsymbol{\theta}_{f_r})$;
- 7 $(\text{cov}^k[\mathbf{f}])^{-1} = \mathbf{C}_f^T \mathbf{Q}_v^{-1} \mathbf{C}_f + \sigma_w^{-2} \mathbf{H}^T \mathbf{H}$;
- 8 $\mathbb{E}^k[\mathbf{f}] = \text{cov}^k[\mathbf{f}] (\sigma_w^{-2} \mathbf{H}^T \mathbf{g} + \mathbf{C}_f^T \mathbf{Q}_v^{-1} \boldsymbol{\eta}_f)$;
- 9 $\mathbf{f}^{(k)} \leftarrow \mathcal{N}(\mathbf{f} | \mathbb{E}^k[\mathbf{f}], \text{cov}^k[\mathbf{f}])$;
- 10 offset $\mathcal{R}_{h_r}, \forall r = 1 \dots R_h$, by $(i, j) \leftarrow \text{Uniform}(m_{r_h})$;
- 11 **for** $r = 1 \dots R_h$: **do**
- 12 $\{(\sigma_{u_r}^2)^{(k)}, \mu_{h_r}^{(k)}, \mathbf{a}_{h_r}^{(k)}\}$
 $\leftarrow \text{ParamSampling}(r, \sigma_w^2, \mathbf{C}_{h_r}, \mathbf{h}_{r+}, \mu_{h_r}, \boldsymbol{\theta}_{h_r})$;
- 13 $(\text{cov}^k[\mathbf{h}])^{-1} = \mathbf{C}_h^T \mathbf{Q}_v^{-1} \mathbf{C}_h + \sigma_w^{-2} \mathbf{F}_h^T \mathbf{F}_h$;
- 14 $\mathbb{E}^k[\mathbf{h}] = \text{cov}^k[\mathbf{h}] (\sigma_w^{-2} \mathbf{F}_h^T \mathbf{g} + \mathbf{C}_h^T \mathbf{Q}_v^{-1} \boldsymbol{\eta}_h)$;
- 15 $\mathbf{h}^{(k)} \leftarrow \mathcal{N}^+(\mathbf{h} | \mathbb{E}^k[\mathbf{h}], \text{cov}^k[\mathbf{h}])$;

Algorithm 1: Gibbs sampling for Blind Image Deconvolution.

The symbol \leftarrow means draw a sample from the RHS distribution. The indices k or $k-1$ of the quantities in the RHS of the equations have been dropped for clarity; it is implied that the most recently sampled value is used in each case.

$$\text{where } \mathbf{Q}_v^{-1} = \mathbf{L}^T \mathbf{L}, \mathbf{M} = \begin{bmatrix} \frac{\sigma_w^{-1} \mathbf{H}}{\mathbf{L} \mathbf{C}_f} \end{bmatrix}, \text{ and } \mathbf{y} = \begin{bmatrix} \frac{\sigma_w^{-1} \mathbf{g}}{\mathbf{L} \boldsymbol{\eta}_f} \end{bmatrix} + \boldsymbol{\nu}.$$

The same approach is used to sample \mathbf{a}_{f_r} and \mathbf{a}_{h_r} , though we make the approximation in $p(\mathbf{f} | \mathbf{a}_f, \sigma_v, \mu_f)$ of ignoring the dependency of the normalising term $\det|\Sigma_f^{-1}|$ on \mathbf{a}_f , resulting in an improper prior. This does not seem to adversely affect results however.

To sample the rectified Gaussians, rejection sampling is used. For \mathbf{h} , in order to use rejection sampling, it must be sampled pixel-wise, which means the more efficient CGLS solver is not possible. This requires the PDF of one pixel in the PSF, $\mathbf{h}_{[i]}$, conditional on the rest, $\mathbf{h}_{\setminus i}$:

$$p(\mathbf{h}_{[i]} | \mathbf{h}_{\setminus i}) \propto \mathcal{N}^+(\mathbf{h}_{[i]} | \mathbb{E}^k[\mathbf{h}]_{[i]} - \mathbf{N}_{[i,i]}^{-1} \mathbf{N}_{[i,\setminus i]} (\mathbf{h}_{\setminus i} - \mathbb{E}^k[\mathbf{h}]_{\setminus i}), \mathbf{N}_{[i,i]}^{-1})$$

where $\mathbf{N} = (\text{cov}^k[\mathbf{h}])^{-1}$, and the square brackets extract a sub-vector or entry of the matrix. A random update order is used to update the pixels, for $i = 1 \dots L_h$.

Simulated Annealing The pixel-wise sampling of \mathbf{h} results in slower convergence, or *mixing* of the Markov chain. SA [9, 7] is a popular method to accelerate the search for high probability regions of the parameter space. The posterior distribution is raised to a power $1/\tau$, where $\tau(k)$ is a temperature parameter, such that $p^{(k)}(\Theta | \mathbf{g}) = (p(\Theta | \mathbf{g}))^{1/\tau(k)}$. High temperatures result in the PDFs being less “peaky” and more uniform, avoiding the search getting trapped in local optima. Samples are gathered after burn-in once $\tau(k) \approx 1$. The distribution should be “cooled” sufficiently slowly, in analogy to annealing in metallurgy. We use a quasi-geometric annealing schedule: $\tau(k) = 1 + (\tau(0) - 1) \cdot \kappa^k$. The modification to the conditionals using the temperature parameter involves scal-

<p>Input: $r, \sigma_w^2, \mathbf{C}_{f_{r+}}, \mathbf{f}_{r+}, \mu_{f_r}, \boldsymbol{\theta}_{f_r}$</p> <p>Output: $\{(\sigma_{v_r}^2)^{(k)}, \mu_{f_r}^{(k)}, \mathbf{a}_{f_r}^{(k)}\}$</p> <ol style="list-style-type: none"> 1 $\mathbf{v}_r^{(k)} = \mathbf{C}_{f_{r+}} (\mathbf{f}_{r+} - \boldsymbol{\mu}_{f_{r+}})$; 2 $(\sigma_{v_r}^2)^{(k)} \leftarrow \mathcal{IG}(\sigma_{v_r}^2 \mid \alpha_{v_r} + \frac{L_r}{2}, \beta_{v_r} + \frac{1}{2} \mathbf{v}_r^T \mathbf{v}_r)$; 3 $\text{cov}^k[\mu_{f_r}] = \left(\ \mathbf{C}_{f_{r+}} \mathbf{1}_{L_{r+}}\ ^2 / \sigma_{v_r}^2 + \delta_{\mu_f}^{-1} \right)^{-1}$; 4 $\mathbb{E}^k[\mu_{f_r}] = \frac{\mathbf{1}_{L_{r+}}^T \mathbf{C}_{f_{r+}}^T \mathbf{C}_{f_{r+}} \mathbf{f}_{r+} / \sigma_{v_r}^2}{\ \mathbf{C}_{f_{r+}} \mathbf{1}_{L_{r+}}\ ^2 / \sigma_{v_r}^2 + \delta_{\mu_f}^{-1}}$; 5 $\mu_{f_r}^{(k)} \leftarrow \mathcal{N}^+(\mu_{f_r} \mid \mathbb{E}^k[\mu_{f_r}], \text{cov}^k[\mu_{f_r}])$; 6 $\text{cov}^k[\mathbf{a}_{f_r}] = (\mathbf{F}_{a_r}^{*T} \mathbf{F}_{a_r}^* / \sigma_{v_r}^2 + \delta_{a_{f_r}}^{-1} \mathbf{I}_{L_{a_f}})^{-1}$; 7 $\mathbb{E}^k[\mathbf{a}_{f_r}] = \text{cov}^k[\mathbf{a}_{f_r}] \sigma_{v_r}^{-2} \mathbf{F}_{a_r}^{*T} (\mathbf{f}_r - \boldsymbol{\mu}_{f_r})$; 8 $\mathbf{a}_{f_r}^{(k)} \leftarrow \mathcal{N}(\mathbf{a}_{f_r} \mid \mathbb{E}^k[\mathbf{a}_{f_r}], \text{cov}^k[\mathbf{a}_{f_r}])$; <p style="text-align: center;">Function ParamSampling</p>

ing the variance of the Gaussian distributions at iteration k by $\tau(k)$, whereas for the IG PDFs, we make the substitution $\alpha_\theta \rightarrow \frac{\alpha_\theta + 1}{\tau(k)} - 1$ and $\beta_\theta \rightarrow \beta_\theta / \tau(k)$. We constrain the maximum temperature for these distributions such that $\alpha_\theta > 1$, otherwise the samples will become unbounded.

Once the samples have been obtained, expectations may then be found to provide the desired parameter estimates, for example the minimum mean-squared error (MMSE) estimate, $\mathbb{E}[\mathbf{f}|\mathbf{g}] \approx \frac{1}{K-k_0} \sum_{k=k_0}^K \mathbf{f}^{(k)}$, where k_0 is chosen to allow for convergence during the *burn-in* period (see e.g. [9]). Other possibilities for inference with the samples include computing a maximum a posteriori (MAP) estimate, by cooling the temperature to zero, or to learn more about the posterior by looking at other statistics of the samples.



Fig. 2. Experimental results: (a) – (c) Exp. 1; (d) – (e) Exp. 2

4. RESULTS

We present two experiments using the Lena image, firstly synthetically blurred with a simulated out-of-focus blur of radius 12 (Exp. 1), and secondly with the image degraded by a real camera (Exp. 2). The blurred image in each case is shown in fig. 2. In Exp. 1, WGN was added to give a blurred-image SNR (BSNR) of 40dB. The observed image support S_g was 183×183 ; the PSF support S_h was

37×37 , giving $S_f = 219 \times 219$; the image and blur block sizes were $m_r = 8$ and $m_{r_h} = 3$ pixels, and the AR orders $P = 5$ and $P_h = 3$. The annealing used $\tau(0) = 2$ and $\kappa = 0.998$, over $K = 4000$ iterations, with $k_0 = 3000$. Due to the hierarchical model, the algorithm is fairly robust to different choices of the hyperparameters, although reasonable values will aid more rapid convergence. Due to the new segmentations at each iteration, lines 6 and 12 of Algorithm 1 were also repeated twice per main iteration to update the parameters more effectively. $\mathbf{h}(0)$ was chosen as a circular PSF of radius 10 (experiments with a Gaussian initialisation also converged successfully).

In Exp. 2, the image was displayed on a computer screen and photographed with the camera lens out-of-focus to produce a similar blurred image to that in Exp. 1 (in fact it is slightly more blurred). The green channel of the image only was used; this was bilinearly interpolated and downsampled, in order to avoid demosaicing. It is also important to work in a linear colour space, and convert to a gamma 2.2 space for final display. Despite the care taken to ensure a linear image, there is undoubtedly some residual non-linearity in the process, as can be seen from the results in fig. 2. The same parameters were used as in Exp. 1, although the annealing was modified to $\tau(0) = 3$, $\kappa = 0.999$, $K = 10000$, $k_0 = 9000$ to ensure better convergence. It is also possible that the PSF is spatially varying in this case, as more artefacts are visible away from the centre.

In order to calculate an improvement in signal-to-noise ratio (ISNR) value, another image shot under the same conditions but with the lens in focus was used as a reference (fig. 2(f)). This was aligned with the restored image and the images histogram matched.

5. CONCLUSIONS

Very flexible and locally adaptive image and blur models have been proposed. We have shown that MCMC methods can be used in BID to tackle the very challenging optimisation problems set by these models, as well as to gain insight into the quality of the obtained solutions. Future work will include better modelling of the AR parameter distributions, imposing positivity on the image, and achieving more rapid convergence of the Markov chain.

6. REFERENCES

- [1] P. Campisi and K. Egiazarian, Eds., *Blind image deconvolution: Theory and Applications*, CRC, 2007.
- [2] T. E. Bishop, R. Molina, and J. R. Hopgood, “Nonstationary blind image restoration using variational methods,” in *IEEE International Conference on Image Processing (ICIP)*, 2007.
- [3] T. E. Bishop and J. R. Hopgood, “Blind image restoration using a block-stationary signal model,” in *Proc. IEEE Conf. Acoustics, Speech and Signal Processing, ICASSP 2006.*, May 2006, vol. 2, pp. II–853–II–856.
- [4] R. Molina, J. Mateos, and A.K. Katsaggelos, “Blind deconvolution using a variational approach to parameter, image, and blur estimation,” *IEEE Trans. Image Process.*, vol. 15, no. 12, pp. 3715–3727, 2006.
- [5] D. Tzikas, A. Likas, and N. Galatsanos, “Variational Bayesian Blind Image Deconvolution With Student-t Priors,” in *IEEE International Conference on Image Processing (ICIP)*, 2007.
- [6] B.C. McCallum, “Blind deconvolution by simulated annealing,” *Optics Communications*, vol. 75, no. 2, pp. 101–105, 1990.
- [7] S. Geman and D. Geman, “Stochastic Relaxation, Gibbs distributions, and the Bayesian restoration of images,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-6, no. 6, pp. 721–741, 1984.
- [8] J. Kaipio and E. Somersalo, *Statistical and Computational Inverse Problems*, vol. 160 of *Applied Mathematical Sciences*, Springer, 2005.
- [9] Radford M. Neal, “Probabilistic inference using Markov chain Monte Carlo methods,” Tech. Rep. CRG-TR-93-1, Dept. of Computer Science, University of Toronto, University of Toronto, 1993.