

Parameter Estimation in Bayesian Super-Resolution Image Reconstruction from Low Resolution Rotated and Translated Images

Salvador Villena^{1,*}, Miguel Vega¹, Rafael Molina²,
and Aggelos K. Katsaggelos³

¹ Dpto. de Lenguajes y Sistemas Informáticos,
Universidad de Granada, Granada 18071, Spain

² Dpto. de Ciencias de la Computación e Inteligencia Artificial,
Universidad de Granada, Granada 18071, Spain

³ Department of Electrical Engineering and Computer Science,
Northwestern University, Evanston, IL 60208-3118, USA

Abstract. This paper deals with the problem of *high-resolution* (HR) image reconstruction, from a set of degraded, under-sampled, shifted and rotated images, utilizing the variational approximation within the Bayesian paradigm. The proposed inference procedure requires the calculation of the covariance matrix of the HR image given the LR observations and the unknown hyperparameters of the probabilistic model. Unfortunately the size and complexity of such matrix renders its calculation impossible, and we propose and compare three alternative approximations. The estimated HR images are compared with images provided by other HR reconstruction methods.

Keywords: High resolution images, Bayesian paradigm, Variational inference, covariance matrix calculation.

1 Introduction

We use the term Super-Resolution (SR) to denote the process of obtaining an HR image, or a sequence of HR images, from a set of LR images [1]. Following the Bayesian framework we focus in this paper on the reconstruction of HR images from a set of downsampled, rotated, and shifted LR images, (see [1] and the references therein, [2] and [3]).

In this paper we assume that the translation and rotation registration parameters are known or previously estimated and examine the difficulties in estimating the HR image, the noise, and the prior parameters in the Bayesian framework. Bayesian methods rely on image models that encapsulate prior image knowledge and avoid the ill-posedness of the image restoration problems. In this paper a prior model based on the ℓ_1 norm of vertical and horizontal first order differences

* This work was supported in part by the Comisión Nacional de Ciencia y Tecnología under contract TIC2007-65533 and the Consejería de Innovación, Ciencia y Empresa of the Junta de Andalucía under contract P07-TIC-02698.

of image pixel values is introduced. At a second level, hyperpriors on the model parameters are introduced and finally, the HR image as well as the parameters are estimated under the Bayesian paradigm, utilizing a variational approximation. The whole estimation process requires the calculation of the inverse of a high dimensional matrix which can not be easily calculated. Part of this paper is devoted to examining the adequacy of several approximate matrix inversion methods to our SR task.

The paper is organized as follows. In section 2 we discuss the Bayesian model and in section 3 the Bayesian inference we use. The different approximations to matrix inversion considered are described in section 4. Experimental results are described in section 5. Finally, section 6 concludes the paper.

2 Bayesian Modeling

Consider a set $\mathbf{g} = (\mathbf{g}_1^t, \dots, \mathbf{g}_U^t)^t$ of $U \geq 1$, observed LR images each with $P = N_1 \times N_2$ pixels. Our aim is to reconstruct a $p = M_1 \times M_2$ HR image \mathbf{f} , where $M_1 = L \times N_1$ and $M_2 = L \times N_2$, from the set \mathbf{g} of LR observed images using the Bayesian paradigm. We assume all images, \mathbf{f} and \mathbf{g}_q , $q = 1, \dots, U$, to be lexicographically ordered.

In this paper we assume that each LR observed image \mathbf{g}_q , for $q = 1, \dots, U$, is a noisy, downsampled, blurred, rotated through a known angle γ_q , and shifted by a known displacement \mathbf{d}_q , version of the HR image.

Let us now study the joint probability distribution $p(\Theta, \mathbf{f}, \mathbf{g})$ that is expressed, within the Hierarchical Bayesian paradigm (see [4]), in terms of the hyperprior model $p(\Theta)$ on the hyperparameters Θ , the prior model $p(\mathbf{f}|\Theta)$ and the degradation model $p(\mathbf{g}|\mathbf{f}, \Theta)$ as

$$p(\Theta, \mathbf{f}, \mathbf{g}) = p(\Theta)p(\mathbf{f}|\Theta)p(\mathbf{g}|\mathbf{f}, \Theta). \tag{1}$$

2.1 Prior Model

The prior model we use in this paper is

$$p(\mathbf{f}|\alpha^h, \alpha^v) = \frac{1}{Z(\alpha^h, \alpha^v)} \times \exp \left\{ - \sum_{i=1}^p [\alpha^h \|\Delta_i^h(\mathbf{f})\|_1 + \alpha^v \|\Delta_i^v(\mathbf{f})\|_1] \right\}, \tag{2}$$

where $\Delta_i^h(\mathbf{x})$ and $\Delta_i^v(\mathbf{x})$ represent the horizontal and vertical first order differences at pixel i , respectively, α^h and α^v are model parameters, and $Z(\alpha^h, \alpha^v)$ is the partition function that we approximate as

$$Z(\alpha^h, \alpha^v) \propto (\alpha^h \alpha^v)^{-p}. \tag{3}$$

2.2 Degradation Model

A 2-D image translation \mathbf{d} followed by a rotation through an angle γ is defined by

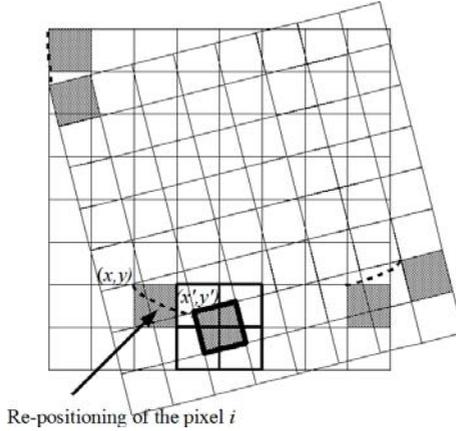


Fig. 1. Illustration of the global original and rotated grids

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \cos(\gamma) & -\sin(\gamma) \\ \sin(\gamma) & \cos(\gamma) \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \mathbf{d}, \tag{4}$$

where $(x'y')^t$ are the new coordinates and $(x\ y)^t$ the old ones.

When this geometric transformation is globally applied to a discrete image, the vertices of its corresponding centered pixel grid fall into subpixel regions of the rotated grid (see Fig. 1). Since we do not know the exact new pixel observations, an approximation, such as a nearest neighbor or a bilinear interpolation is usually applied, see [5], [2] and [3]. In this paper we use bilinear interpolation. Thus if we denote by \mathbf{f}' the discrete image obtained after applying a translation \mathbf{d} followed by a rotation with angle γ to the discrete image \mathbf{f} , we have $\mathbf{f}' = \mathbf{R}_{(\mathbf{d},\gamma)} \mathbf{f}$, where $\mathbf{R}_{(\mathbf{d},\gamma)}$ is a $p \times p$ real matrix.

The process to obtain the observed, displaced and rotated LR images \mathbf{g}_q , $q = 1, \dots, U$ from \mathbf{f} can thus be modeled as

$$\mathbf{g}_q = \mathbf{DHR}_{(\mathbf{d}_q,\gamma_q)} \mathbf{f} + \epsilon_q = \mathbf{C}_q \mathbf{f} + \epsilon_q, \tag{5}$$

where ϵ_q represents the acquisition noise, assumed to be additive white Gaussian with variance β^{-1} and \mathbf{C}_q is the $P \times p$ matrix

$$\mathbf{C}_q = \mathbf{DHR}_{(\mathbf{d}_q,\gamma_q)}, \tag{6}$$

where \mathbf{H} is a $p \times p$ matrix modeling sensor integration as a uniform blurring of size L , and \mathbf{D} is a $P \times p$ downsampling matrix.

Finally for the set of U observations $\mathbf{g} = (g_1^t, \dots, g_U^t)^t$ we have

$$p(\mathbf{g}|\mathbf{f}, \beta) \propto \beta^{\frac{UP}{2}} \exp \left[-\frac{\beta}{2} \sum_{q=1}^U \|\mathbf{g}_q - \mathbf{C}_q \mathbf{f}\|^2 \right]. \tag{7}$$

2.3 Hyperprior Model

Our prior knowledge on the different model parameters $\theta \in \Theta$, is modeled using

$$p(\Theta) = \prod_{\theta \in \Theta} p(\theta), \quad (8)$$

where $p(\theta)$ are gamma hyperpriors

$$p(\theta) = \Gamma(\theta | a_\theta^o, b_\theta^o), \quad \forall \theta \in \Theta. \quad (9)$$

The gamma distribution has the form

$$\Gamma(\theta | u, v) = \frac{v^u}{\Gamma(u)} \theta^{u-1} \exp[-v\theta], \quad (10)$$

where $\theta > 0$, $u > 0$ and $v > 0$ with mean $\mathbf{E}[\theta] = u/v$ and variance $\mathbf{var}[\theta] = u/v^2$.

Our set of model parameters is then $\Theta = (\alpha^h, \alpha^v, \beta)$, and the joint distribution is given by

$$p(\Theta, \mathbf{f}, \mathbf{g}) = p(\Theta) p(\mathbf{f} | \alpha^h, \alpha^v) p(\mathbf{g} | \mathbf{f}, \beta), \quad (11)$$

where $p(\Theta)$, $p(\mathbf{f} | \alpha^h, \alpha^v)$ and $p(\mathbf{g} | \mathbf{f}, \beta)$ are defined respectively in Eqs. (8), (2) and (7) above.

3 Bayesian Inference

Bayesian inference on the unknown HR image \mathbf{f} and hyperparameters given the observed \mathbf{g} is based on the posterior probability distribution

$$p(\Theta, \mathbf{f} | \mathbf{g}) = \frac{p(\Theta, \mathbf{f}, \mathbf{g})}{p(\mathbf{g})}. \quad (12)$$

Since $p(\Theta, \mathbf{f} | \mathbf{g})$ of Eq. (12) can not be found in closed form, because $p(\mathbf{g})$ can not be calculated analytically, we apply variational methods to approximate this distribution by the distribution $q(\Theta, \mathbf{f})$ minimizing the Kullback-Leibler (KL) divergence, which is given by [6,7]

$$\begin{aligned} C_{KL}(q(\Theta, \mathbf{f}) || p(\Theta, \mathbf{f} | \mathbf{g})) &= \int q(\Theta, \mathbf{f}) \log \left(\frac{q(\Theta, \mathbf{f})}{p(\Theta, \mathbf{f} | \mathbf{g})} \right) d\Theta d\mathbf{f} \\ &= \int q(\Theta, \mathbf{f}) \log \left(\frac{q(\Theta, \mathbf{f})}{p(\Theta, \mathbf{f}, \mathbf{g})} \right) d\Theta d\mathbf{f} + \text{const} = \mathcal{M}(q(\Theta, \mathbf{f}), \mathbf{g}) + \text{const}, \quad (13) \end{aligned}$$

and is always non negative and equal to zero only when $q(\Theta, \mathbf{f}) = p(\Theta, \mathbf{f} | \mathbf{g})$.

Due to the form of the prior proposed in Eq. (2) the above integral can not be evaluated, but we can however majorize the ℓ_1 prior by a function which renders the integral easier to calculate. The majorization to be applied here to our prior model is conceptually similar to the one applied in [8] to the TV prior.

Our prior can be rewritten in the more convenient form

$$p(\mathbf{f}|\alpha^h, \alpha^v) \propto (\alpha^h \alpha^v)^{p \times} \exp \left\{ - \sum_{i=1}^p \left[\alpha^h \sqrt{(\Delta_i^h(\mathbf{f}))^2} + \alpha^v \sqrt{(\Delta_i^v(\mathbf{f}))^2} \right] \right\}. \quad (14)$$

Let us consider the following inequality, also used in [9], which states that, for any $w \geq 0$ and $z > 0$

$$\sqrt{w} \leq \frac{w+z}{2\sqrt{z}}. \quad (15)$$

Let us define, for \mathbf{f} , \mathbf{u}^h and \mathbf{u}^v , where \mathbf{u}^h and \mathbf{u}^v are any p -dimensional vectors $\mathbf{u}^h \in (R^+)^p$, $\mathbf{u}^v \in (R^+)^p$ with components \mathbf{u}_i^h and \mathbf{u}_i^v , $i = 1, \dots, p$, the following functional

$$\begin{aligned} \mathbf{M}(\alpha^h, \alpha^v, \mathbf{f}, \mathbf{u}^h, \mathbf{u}^v) &= (\alpha^h \alpha^v)^{p \times} \\ &\exp \left\{ - \sum_{i=1}^p \left[\alpha^h \frac{(\Delta_i^h(\mathbf{f}))^2 + \mathbf{u}_i^h}{2\sqrt{\mathbf{u}_i^h}} + \alpha^v \frac{(\Delta_i^v(\mathbf{f}))^2 + \mathbf{u}_i^v}{2\sqrt{\mathbf{u}_i^v}} \right] \right\}. \end{aligned} \quad (16)$$

Now, using the inequality in Eq. (15) and comparing Eq. (16) with Eq. (14), we obtain $p(\mathbf{f}|\alpha^h, \alpha^v) \geq c \cdot \mathbf{M}(\alpha^h, \alpha^v, \mathbf{f}, \mathbf{u}^h, \mathbf{u}^v)$. As it will be shown later, vectors \mathbf{u}^h and \mathbf{u}^v are quantities that need to be computed and have an intuitive interpretation related to the unknown image \mathbf{f} . This leads to the following lower bound for the joint probability distribution

$$\begin{aligned} p(\Theta, \mathbf{f}, \mathbf{g}) &\geq c \cdot p(\Theta) \mathbf{M}(\alpha^h, \alpha^v, \mathbf{f}, \mathbf{u}^h, \mathbf{u}^v) p(\mathbf{g}|\mathbf{f}, \beta) \\ &= \mathbf{F}(\Theta, \mathbf{f}, \mathbf{g}, \mathbf{u}^h, \mathbf{u}^v), \end{aligned} \quad (17)$$

Hence, by defining

$$\tilde{\mathcal{M}}(q(\Theta, \mathbf{f}), \mathbf{g}, \mathbf{u}^h, \mathbf{u}^v) = \int q(\Theta, \mathbf{f}) \log \left(\frac{q(\Theta, \mathbf{f})}{\mathbf{F}(\Theta, \mathbf{f}, \mathbf{g}, \mathbf{u}^h, \mathbf{u}^v)} \right) d\Theta d\mathbf{f}, \quad (18)$$

and using Eq. (17), we obtain for $\mathcal{M}(q(\Theta, \mathbf{f}), \mathbf{g})$, defined in Eq. (13),

$$\mathcal{M}(q(\Theta, \mathbf{f}), \mathbf{g}) \leq \min_{\{\mathbf{u}^h, \mathbf{u}^v\}} \tilde{\mathcal{M}}(q(\Theta, \mathbf{f}), \mathbf{g}, \mathbf{u}^h, \mathbf{u}^v). \quad (19)$$

Therefore, by finding a sequence of distributions $\{q^k(\Theta, \mathbf{f})\}$ that monotonically decreases $\tilde{\mathcal{M}}(q(\Theta, \mathbf{f}), \mathbf{g}, \mathbf{u}^h, \mathbf{u}^v)$ for fixed \mathbf{u}^h and \mathbf{u}^v , a sequence of an ever decreasing upper bound of $C_{KL}(q(\Theta, \mathbf{f})||p(\Theta, \mathbf{f}|\mathbf{g}))$ is also obtained due to Eq. (13). Furthermore minimizing $\tilde{\mathcal{M}}(q(\Theta, \mathbf{f}), \mathbf{g}, \mathbf{u}^h, \mathbf{u}^v)$ with respect to \mathbf{u}^h and \mathbf{u}^v , generates vector sequences $\{\mathbf{u}^{h^k}\}$ and $\{\mathbf{u}^{v^k}\}$ that tightens the upper-bound for each distribution $q^k(\Theta, \mathbf{f})$. Therefore, the sequence $\{q^k(\Theta, \mathbf{f})\}$ is coupled with the sequences $\{\mathbf{u}^{h^k}\}$ and $\{\mathbf{u}^{v^k}\}$. We use the following iterative algorithm to approximate the posterior distribution of image and hyperparameters and tighten the upper-bound.

Algorithm 1. *Posterior image distribution and parameter estimation using*
 $q(\Theta, \mathbf{f}) = q(\Theta)q(\mathbf{f})$.

Given $q^1(\Theta)$, *an initial estimate of* $q(\Theta)$, *and* $\{\mathbf{u}^{h^1}, \mathbf{u}^{v^1}\} \in (R^+)^{2p}$, *for* $k = 1, 2, \dots$ *until a stopping criterion is met:*

1. *Find*

$$q^k(\mathbf{f}) = \arg \min_{q(\mathbf{f})} \tilde{\mathcal{M}}(q(\mathbf{f})q^k(\Theta), \mathbf{g}, \mathbf{u}^{h^k}, \mathbf{u}^{v^k}) \quad (20)$$

2. *Find*

$$\{\mathbf{u}^{h^{k+1}}, \mathbf{u}^{v^{k+1}}\} = \arg \min_{\{\mathbf{u}^h, \mathbf{u}^v\}} \tilde{\mathcal{M}}(q^k(\mathbf{f})q^k(\Theta), \mathbf{g}, \mathbf{u}^h, \mathbf{u}^v) \quad (21)$$

3. *Find*

$$q^{k+1}(\Theta) = \arg \min_{q(\Theta)} \tilde{\mathcal{M}}(q^k(\mathbf{f})q(\Theta), \mathbf{g}, \mathbf{u}^{h^{k+1}}, \mathbf{u}^{v^{k+1}}) \quad (22)$$

Set $q(\Theta) = \lim_{k \rightarrow \infty} q^k(\Theta)$, $q(\mathbf{f}) = \lim_{k \rightarrow \infty} q^k(\mathbf{f})$.

To calculate $q^k(\mathbf{f})$, we observe that differentiating $\tilde{\mathcal{M}}(q(\mathbf{f})q^k(\Theta), \mathbf{g}, \mathbf{u}^{h^k}, \mathbf{u}^{v^k})$ in Eq. (20) with respect to $q(\mathbf{f})$ and setting it equal to zero, we obtain

$$q^k(\mathbf{f}) = \mathcal{N}(\mathbf{f} \mid \mathbf{E}_{q^k(\Theta)}[\mathbf{f}], \mathbf{cov}_{q^k(\Theta)}[\mathbf{f}]), \quad (23)$$

with

$$\begin{aligned} \mathbf{cov}_{q^k(\Theta)}^{-1}[\mathbf{f}] &= \mathbf{E}_{q^k(\Theta)}[\beta] \sum_{q=1}^U \mathbf{C}_q^t \mathbf{C}_q + \mathbf{E}_{q^k(\Theta)}[\alpha^h] \Delta_h^t \mathbf{W}(\mathbf{u}^{h^k}) \Delta_h \\ &\quad + \mathbf{E}_{q^k(\Theta)}[\alpha^v] \Delta_v^t \mathbf{W}(\mathbf{u}^{v^k}) \Delta_v \end{aligned} \quad (24)$$

and $\mathbf{E}_{q^k(\Theta)}[\mathbf{f}]$ given by

$$\mathbf{E}_{q^k(\Theta)}[\mathbf{f}] = \mathbf{cov}_{q^k(\Theta)}[\mathbf{f}] \mathbf{E}_{q^k(\Theta)}[\beta] \sum_{q=1}^U \mathbf{C}_q^t \mathbf{g}_q. \quad (25)$$

In Eq. (24) Δ^h and Δ^v represent $p \times p$ convolution matrices associated respectively with the first order horizontal and vertical differences, and $\forall \mathbf{u} \in (R^+)^p$, $\mathbf{W}(\mathbf{u})$ is the diagonal $p \times p$ matrix with entries

$$\mathbf{W}(\mathbf{u})_{ii} = \frac{1}{\sqrt{\mathbf{u}_i}} \text{ for } i = 1, \dots, p. \quad (26)$$

The matrices $\mathbf{W}(\mathbf{u}^{h^k})$ and $\mathbf{W}(\mathbf{u}^{v^k})$ can be interpreted as spatial adaptivity matrices since they control the amount of horizontal and vertical smoothing at each pixel location depending on the strength of the intensity variation at that pixel, as expressed by the horizontal and vertical intensity gradients, respectively.

To calculate $\mathbf{u}^{d^{k+1}}$, for $d \in \{h, v\}$, we have from Eq. (21) that

$$\mathbf{u}^{d^{k+1}} = \arg \min_{\mathbf{u}} \sum_{i=1}^p \frac{\mathbf{E}_{q^k(\Theta)}[(\Delta_i^d(\mathbf{f}))^2] + \mathbf{u}_i}{\sqrt{\mathbf{u}_i}} \quad (27)$$

and consequently

$$\mathbf{u}_i^{d^{k+1}} = \mathbf{E}_{\mathbf{q}^k(\mathbf{f})} [(\Delta_i^d(\mathbf{f}))^2] \quad \text{for } i = 1, \dots, p. \quad (28)$$

It is clear from Eq. (28) that vectors $\mathbf{u}^{d^{k+1}}$, for $d \in \{h, v\}$, are functions of the spatial first order horizontal and vertical differences of the unknown image \mathbf{f} under the distribution $\mathbf{q}^k(\mathbf{f})$ and represent the local spatial activity of \mathbf{f} .

Finally, differentiating the right hand side of Eq. (22) with respect to $\mathbf{q}(\Theta)$ and setting it equal to zero we find that

$$\mathbf{q}^{k+1}(\Theta) \propto \exp \left\{ \mathbf{E}_{\mathbf{q}^k(\mathbf{f})} \left[\ln \mathbf{F}(\Theta, \mathbf{f}, \mathbf{g}, \mathbf{u}^{h^{k+1}}, \mathbf{u}^{v^{k+1}}) \right] \right\}. \quad (29)$$

Thus we obtain

$$\mathbf{q}^{k+1}(\Theta) = \mathbf{q}^{k+1}(\alpha^h) \mathbf{q}^{k+1}(\alpha^v) \mathbf{q}^{k+1}(\beta), \quad (30)$$

where $\mathbf{q}_h^{k+1}(\alpha^h)$, $\mathbf{q}_v^{k+1}(\alpha^v)$ and $\mathbf{q}^{k+1}(\beta)$ are respectively the gamma distributions

$$\mathbf{q}_d^{k+1}(\alpha^d) = \Gamma \left(\alpha^d | p + a_{\alpha^d}^o, \sum_i \sqrt{\mathbf{u}_i^{d^{k+1}}} + b_{\alpha^d}^o \right) \quad (31)$$

for $d \in \{h, v\}$ and

$$\mathbf{q}^{k+1}(\beta) = \Gamma \left(\beta | \frac{PU}{2} + a_\beta^o, \frac{\mathbf{E}_{\mathbf{q}^k(\mathbf{f})} \left[\sum_{q=1}^U \|\mathbf{g}_q - \mathbf{C}_q \mathbf{f}\|^2 \right]}{2} + b_\beta^o \right). \quad (32)$$

So we have

$$\mathbf{E}_{\mathbf{q}_d^{k+1}(\alpha^d)} [\alpha^d] = \frac{p + a_{\alpha^d}^o}{\sum_i \sqrt{\mathbf{u}_i^{d^{k+1}}} + b_{\alpha^d}^o}, \quad (33)$$

$$\mathbf{E}_{\mathbf{q}^{k+1}(\beta)} [\beta] = \frac{\frac{UP}{2} + a_\beta^o}{\frac{1}{2} \mathbf{E}_{\mathbf{q}^k(\mathbf{f})} \left[\sum_{q=1}^U \|\mathbf{g}_q - \mathbf{C}_q \mathbf{f}\|^2 \right] + b_\beta^o}. \quad (34)$$

It is possible to express the inverses of these means in the more meaningful forms

$$\frac{1}{\mathbf{E}_{\mathbf{q}_d^{k+1}(\alpha^d)} [\alpha^d]} = \gamma_{\alpha^d} \frac{1}{\alpha_o^d} + (1 - \gamma_{\alpha^d}) \frac{\sum_i \sqrt{\mathbf{u}_i^{d^{k+1}}}}{p} \quad (35)$$

and

$$\frac{1}{\mathbf{E}_{\mathbf{q}^{k+1}(\beta)} [\beta]} = \gamma_\beta \frac{1}{\beta_o} + (1 - \gamma_\beta) \frac{\mathbf{E}_{\mathbf{q}^k(\mathbf{f})} \left[\sum_{q=1}^U \|\mathbf{g}_q - \mathbf{C}_q \mathbf{f}\|^2 \right]}{UP}, \quad (36)$$

where

$$\gamma_{\alpha^d} = \frac{a_{\alpha^d}^o}{p + a_{\alpha^d}^o} \quad \text{and} \quad \gamma_\beta = \frac{a_\beta^o}{UP/2 + a_\beta^o}. \quad (37)$$

These mean values are convex linear combinations of the inverses of the means of the hyperpriors $\alpha_o^d = \frac{a_o^d}{b_o^d}$ and $\beta_o = \frac{a_\beta}{b_\beta}$, and their corresponding ML estimates. γ_{α^d} and γ_β take values in $[0, 1)$, and can be interpreted as the confidence on the parameter values.

Equation (25) can be solved iteratively utilizing the *Conjugate Gradient* (CG) method without the need of explicitly obtaining the full covariance matrix of Eq. (24), but estimation of \mathbf{u} in Eq. (28) and evaluation of Eq. (36) requires the evaluation of

$$\mathbf{E}_{q^k(\mathbf{f})} [(\Delta_i^d(\mathbf{f}))^2] = (\Delta_i^d(\mathbf{E}_{q^k(\mathbf{f})}[\mathbf{f}]))^2 + \mathbf{E}_{q^k(\mathbf{f})} [(\Delta_i^d(\mathbf{f} - \mathbf{E}_{q^k(\mathbf{f})}[\mathbf{f}]))^2], \quad (38)$$

where

$$\mathbf{E}_{q^k(\mathbf{f})} [(\Delta_i^d(\mathbf{f} - \mathbf{E}_{q^k(\mathbf{f})}[\mathbf{f}]))^2] = \text{tr}(\mathbf{cov}_{q^k(\mathbf{f})}[\mathbf{f}] \Delta_i^{d,t} \Delta_i^d), \quad (39)$$

and

$$\mathbf{E}_{q^k(\mathbf{f})} [\|\mathbf{g}_q - \mathbf{C}_q \mathbf{f}\|^2] = \|\mathbf{g}_q - \mathbf{C}_q \mathbf{E}_{q^k(\mathbf{f})}[\mathbf{f}]\|^2 + \text{tr}(\mathbf{cov}_{q^k(\mathbf{f})}[\mathbf{f}] \mathbf{C}_q^t \mathbf{C}_q), \quad (40)$$

respectively, and whose exact evaluation would require to evaluate the full covariance matrix. This problem can be solved utilizing different approximation methods described in the next section.

4 Approximate Matrix Inversion Methods

We have to calculate $\text{tr}[\mathbf{cov}_{q^k(\mathbf{f})}[\mathbf{f}]\mathbf{V}]$ where $\mathbf{cov}_{q^k(\mathbf{f})}[\mathbf{f}]^{-1}$ has been defined in Eq. (24), and \mathbf{V} is $\mathbf{V} = \Delta_i^{d,t} \Delta_i^d$, or $\mathbf{C}_q^t \mathbf{C}_q$ (See Eqs. (39) and (40)).

Based on the properties of the matrix $\mathbf{cov}_{q^k(\mathbf{f})}[\mathbf{f}]^{-1}$ we have selected and described here the following approximation methods to estimate $\mathbf{cov}_{q^k(\mathbf{f})}[\mathbf{f}]$: Jacobi or diagonal preconditioning, Circulant preconditioning, and *Factorized Sparse Approximate Inverse* (FSAI) [10]. The description of each of these methods has been completed with their corresponding approximations of $\text{tr}[\mathbf{cov}_{q^k(\mathbf{f})}[\mathbf{f}]\mathbf{V}]$.

4.1 Jacobi Preconditioning

Jacobi preconditioning approximates $\mathbf{cov}_{q^k(\mathbf{f})}[\mathbf{f}]$ by the inverse of its diagonal. This is the simplest and faster approximation. Using this approach we have

$$\text{tr}[\mathbf{cov}_{q^k(\mathbf{f})}[\mathbf{f}]\mathbf{V}] \approx \sum_{i=1}^p \mathbf{V}_{ii} \mathbf{cov}_{q^k(\mathbf{f})}[\mathbf{f}]_{ii}. \quad (41)$$

4.2 Circulant Preconditioning

A block semi-circulant approximation of $\mathbf{cov}_{q^k(\mathbf{f})}[\mathbf{f}]^{-1}$ and $\mathbf{C}_q^t \mathbf{C}_q$, can be defined with the help of $\mathbf{C}_q \approx \overline{\mathbf{C}}_q = \mathbf{D}\mathbf{H}\overline{\mathbf{R}}(\mathbf{d}_q, \gamma_q)$, where $\overline{\mathbf{R}}(\mathbf{d}_q, \gamma_q)$ is the convolution

matrix obtained by averaging the convolution filters associated to each row of $\mathbf{R}(\mathbf{d}_q, \gamma_q)$, that is,

$$\overline{\mathbf{R}}(\mathbf{d}_q, \gamma_q)_{i-1} = \frac{1}{p} \sum_{l=1}^p \mathbf{R}(\mathbf{d}_q, \gamma_q)_{\{(l+i-1) \bmod p\} l}. \quad (42)$$

Therefore, the matrix $\mathbf{cov}_{q^k(\mathbf{f})}[\mathbf{f}]^{-1}$ can be approximated by a matrix \mathbf{O} defined as

$$\begin{aligned} \mathbf{O} = \mathbf{E}_{q^k(\Theta)}[\beta] & \left(\sum_{q=1}^U \overline{\mathbf{C}}_q^t \overline{\mathbf{C}}_q \right) + \mathbf{E}_{q^k(\Theta)}[\alpha^h] z(\mathbf{u}^{h^k}) \Delta_h^t \Delta_h \\ & + \mathbf{E}_{q^k(\Theta)}[\alpha^v] z(\mathbf{u}^{v^k}) \Delta_v^t \Delta_v, \end{aligned} \quad (43)$$

with $z(\mathbf{u}) = \frac{1}{p} \sum_i \frac{1}{\sqrt{u_i}}, \forall \mathbf{u} \in (R^+)^p$.

Then, \mathbf{O} is a block semi-circulant approximation of $\mathbf{cov}_{q^k(\mathbf{f})}[\mathbf{f}]^{-1}$ and

$$\text{tr}[\mathbf{O}^{-1} \Delta_i^{dt} \Delta_i^d] = \frac{1}{p} \text{tr}[\mathbf{O}^{-1} \Delta^{dt} \Delta^d]$$

and $\text{tr}[\mathbf{O}^{-1} \overline{\mathbf{C}}_q^t \overline{\mathbf{C}}_q]$ can be calculated in the frequency domain.

4.3 FSAI Method

Finally we consider the Cholevsky factorization of the *symmetric definite positive* (SPD) matrix $\mathbf{cov}_{q^k(\mathbf{f})}[\mathbf{f}]^{-1} = \mathbf{L}\mathbf{L}^T$. The idea behind FSAI is to find a lower triangular matrix \mathbf{G} with \mathbf{Q} sparsity pattern (arrangement of nonzero elements in an sparse matrix that it must be specified a priori and such that $\mathbf{Q} \subseteq \{(i, j) : i \leq j\}$), that minimizes the Frobenius norm $\|\mathbf{I} - \mathbf{L}\mathbf{G}\|_F$, where the Frobenius norm of an $A_{N \times M}$ matrix is

$$\|A\|_F = \sqrt{\sum_{i=1}^N \sum_{j=1}^M |A_{ij}|^2}.$$

In [10] the following algorithm to find \mathbf{G} is proposed:

Algorithm 2

- Step 1: Compute a lower triangular matrix $\hat{\mathbf{G}}$ with sparsity pattern \mathbf{Q} , such that $(\hat{\mathbf{G}} \mathbf{cov}_{q^k(\mathbf{f})}[\mathbf{f}]^{-1})_{i,j} = \mathbf{I}_{i,j}, \forall (i, j) \in \mathbf{Q}$.
- Step 2: Calculate $\mathbf{D} = (\text{diag}(\hat{\mathbf{G}}))^{-1}$ and $\mathbf{G} = \mathbf{D}^{\frac{1}{2}} \hat{\mathbf{G}}$.

Note that in step 1 $\hat{\mathbf{G}}$ is computed by rows: each row requires the solution of a small local SPD linear system, the size of which is equal to the number of nonzero allowed in that row. A common choice for the sparsity pattern is to allow non-zeros in \mathbf{Q} only in positions corresponding to non-zeros in the lower triangular part of $\mathbf{cov}_{q^k(\mathbf{f})}[\mathbf{f}]^{-1}$ [11]. Finally we approximate $\text{tr}[\mathbf{cov}_{q^k(\mathbf{f})}[\mathbf{f}]\mathbf{V}]$ by $\text{tr}[\mathbf{G}\mathbf{V}\mathbf{G}^T] = \sum_{i=1}^p \sum_{j=1}^i \sum_{k=1}^i \mathbf{G}_{ij} \mathbf{V}_{jk} \mathbf{G}_{ik}$.

5 Experimental Results

A number of experiments have been carried out using synthetic LR images that allow us to measure the quality of the HR image reconstructed with the proposed Algorithm 1 (henceforth $\ell 1$) using different preconditioning techniques.

We show results for the image set of LR images obtained from the HR images in Fig. 2a and Fig. 3a. Sequences of 16 rotated, displaced and downsampled, by a factor of 4, images have been obtained and Gaussian noise of 30 dB and 40 dB was added. Reconstructions utilizing the different preconditioning methods have been performed and their quality has been numerically measured utilizing the peak signal-to-noise ratio (PSNR), and the *Structural Similarity Index Measure* (SSIM) defined in [12], whose maximal value, corresponding to equal images, is +1.

The proposed algorithm was ran until the criterion

$$\|\mathbf{f}^k - \mathbf{f}^{k-1}\|^2 / \|\mathbf{f}^{k-1}\|^2 < 10^{-4} \quad (44)$$

was satisfied, where \mathbf{f}^k denotes image point estimate for the k iteration step. Gamma hyperpriors parameter values a_θ^o and b_θ^o , $\forall \theta \in \Theta$ (see Eq. (9)) were determined experimentally for each image.

Table 1 shows a numeric comparison of the results obtained utilizing the different preconditioning methods and the reconstruction using a SAR prior. Figure 2b shows one of the 40 db LR observed images. Figure 2c shows the HR reconstruction obtained by using Jacobi's approximation (the other two approximations produced very similar reconstruction, see also table 1). Figure 2d displays the reconstruction using a SAR prior.

Finally, Fig. 3b shows one of the 30 dB LR observed images. Figure 3c shows the HR reconstruction obtained by using Jacobi's approximation and Fig. 3d displays the reconstruction using a SAR prior.

Table 1. PSNR and SSIM values of (2a) and (3a) images for 16 observed images

Image	Method	SNR 30 dB		SNR 40 dB	
		PSNR	SSIM	PSNR	SSIM
(2a)	Jacobi	24.1	0.76	29.3	0.90
	FSAI	23.5	0.75	28.6	0.89
	Circulant	23.3	0.79	28.5	0.91
	SAR	20.9	0.61	26.4	0.81
(3a)	Jacobi	31.0	0.89	34.4	0.94
	FSAI	30.8	0.88	34.3	0.94
	Circulant	30.7	0.88	34.5	0.93
	SAR	28.8	0.82	31.8	0.94

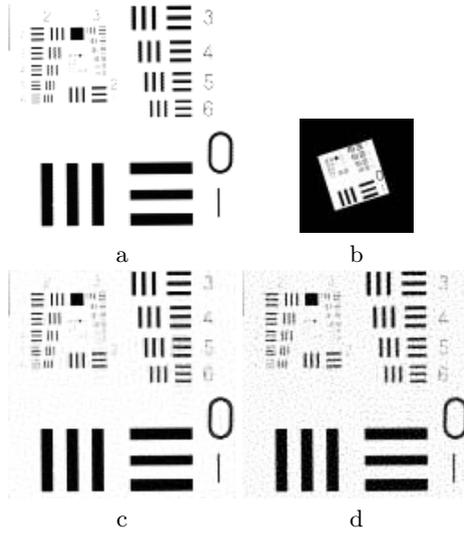


Fig. 2. (a) Original image, (b) one of the 16 40 dB LR observed images from the sequence, (c) HR reconstruction using Jacobi matrix approximation and (d) the reconstruction using SAR prior

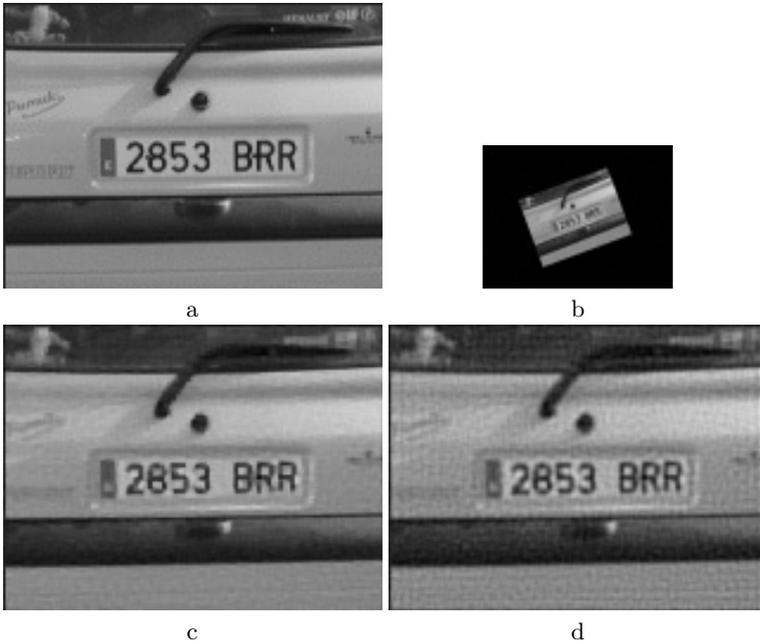


Fig. 3. (a) Original image, (b) One of the 16 30 dB LR observed images from the sequence, (c) HR reconstruction using Jacobi matrix approximation and (d) the reconstruction using SAR prior

6 Conclusions

The Bayesian high-resolution image reconstruction method proposed in [13] has been extended to consider arbitrary displacements and rotations. Several matrix inversion techniques have been considered in our algorithm implementation, from which similar results are obtained. The utilization of a prior based on the ℓ_1 norm of horizontal and vertical differences in images gives results better than the utilization of a SAR prior.

References

1. Katsaggelos, A., Molina, R., Mateos, J.: Super resolution of images and video. In: *Synthesis Lectures on Image, Video, and Multimedia Processing*. Morgan & Claypool (2007)
2. Tipping, M.E., Bishop, C.M.: Bayesian image super-resolution. In: Becker, S., Thrun, S., Obermayer, K. (eds.) *Advances in Neural Information Processing Systems*, vol. 15, pp. 1279–1286. MIT Press, Cambridge (2003)
3. Pickup, L.C., Capel, D.P., Roberts, S.J., Zisserman, A.: Bayesian methods for image super-resolution. *The Computer Journal* (2007)
4. Molina, R., Katsaggelos, A.K., Mateos, J.: Bayesian and regularization methods for hyperparameter estimation in image restoration. *IEEE Transactions on Image Processing* 8, 231–246 (1999)
5. Hardie, R.C., Barnard, K.J., Bognar, J.G., Armstrong, E.E., Watson, E.A.: High-resolution image reconstruction from a sequence of rotated and translated frames and its application to an infrared imaging system. *Society of Photo-Optical Instrumentation Engineers* 37(1), 247–260 (1998)
6. Kullback, S., Leibler, R.A.: On information and sufficiency. *Annals of Math. Stat.* 22, 79–86 (1951)
7. Kullback, S.: *Information Theory and Statistics*. Dover Publications, New York (1959)
8. Babacan, S., Molina, R., Katsaggelos, A.: Parameter estimation in tv image restoration using variational distribution approximation. *IEEE Trans. on Image Processing* 17(3), 326–339 (2008)
9. Bioucas-Dias, J., Figueiredo, M., Oliveira, J.: Total-variation image deconvolution: A majorization-minimization approach. In: *ICASSP 2006* (2006)
10. Kolotilina, L.Y., Yeremin, A.Y.: Factorized sparse approximate inverse preconditioning: I. theory. *SIAM J. Matrix Anal. Appl.*, 45–58 (1993)
11. Benzi, M.: Preconditioning techniques for large linear systems: A survey. *J. Comput. Phys.* 182, 418–477 (2002)
12. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: From error measurement to structural similarity. *IEEE Trans. on Img. Proc.* 13(4), 600–612 (2004)
13. Molina, R., Vega, M., Abad, J., Katsaggelos, A.K.: Parameter Estimation in Bayesian High-Resolution Image Reconstruction with Multisensors. *IEEE Transactions on Image Processing* 12(12), 1655–1667 (2003)