Bayesian Super-Resolution Image Reconstruction using an $\ell 1$ prior

Salvador Villena, Miguel Vega*
Dept. de Lenguajes y Sistemas Informáticos
Univ. de Granada, 18071 Granada, Spain
svillena@ugr.es, mvega@ugr.es

Rafael Molina
Dept. Ciencias de la Computación e Inteligencia Artificial
Univ. de Granada, 18071 Granada, Spain
rms@decsai.ugr.es

Aggelos K. Katsaggelos
Dept. of Electrical Engineering and Computer Science
Northwestern University, Evanston, Illinois 60208-3118
aggk@eecs.northwestern.edu

Abstract

This paper deals with the problem of high-resolution *(HR) image reconstruction, from a set of degraded, undersampled, shifted and rotated images, under the Bayesian paradigm, utilizing a variational approximation. 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 new prior based on the $\ell 1$ norm of vertical and horizontal first order differences of image pixel values is introduced and its parameters are estimated. The estimated HR images are compared with images provided by other HR reconstruction methods.*

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, arbitrarily rotated, and shifted LR images, a problem which has already received interest from the SR community (see [1] and the references therein), and in particular for the use of the Bayesian approach, see [11] and [9].

In this paper we assume that the registration parameters are known and examine the difficulties in estimating the unknown HR image and the model parameters within the Bayesian framework. The Bayesian approach in [10] for solving this problem can deal only with known translations. In this paper we extend the approach to deal with known rotations and translations, and also consider more realistic image priors.

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

2 Bayesian Modeling

Consider a set $\mathbf{g} = (\mathbf{g}_1^t, \ldots, \mathbf{g}_U^t)^t$ 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 low-resolution observed images using the Bayesian paradigm. We assume all images, $\mathbf{f}$ and $\mathbf{g}_q$, $q = 1, \ldots, U$, to be lexicographically ordered.

In this paper we consider that each LR observed image $\mathbf{g}_q$, for $q = 1, \ldots, U$, 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. Next we proceed to model the HR reconstruction problem using the Bayesian framework.

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

$$\mathrm{p}(\Theta, \mathbf{f}|\mathbf{g}) = \frac{\mathrm{p}(\Theta, \mathbf{f}, \mathbf{g})}{\mathrm{p}(\mathbf{g})}, \qquad (1)$$

where Θ represents the set of model parameters. Let us now study the joint probability distribution $\mathrm{p}(\Theta, \mathbf{f}, \mathbf{g})$ that can be expressed, within the Hierarchical Bayesian paradigm (see

*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-FQM-02701.

[8]), in terms of the hyperprior model $\mathrm{p}(\Theta)$, the prior model $\mathrm{p}(\mathbf{f}|\Theta)$ and the degradation model $\mathrm{p}(\mathbf{g}|\mathbf{f},\Theta)$ as

$$\mathrm{p}(\Theta,\mathbf{f},\mathbf{g}) = \mathrm{p}(\Theta)\mathrm{p}(\mathbf{f}|\Theta)\mathrm{p}(\mathbf{g}|\mathbf{f},\Theta)\,. \qquad (2)$$

2.1 Prior Model

The prior model we use is

$$\mathrm{p}(\mathbf{f}|\alpha^h,\alpha^v) = \frac{1}{Z(\alpha^h,\alpha^v)}\times \exp\left\{-\sum_{i=1}^{p}\left[\alpha^h \parallel \Delta_i^h(\mathbf{f}) \parallel_1 + \alpha^v \parallel \Delta_i^v(\mathbf{f}) \parallel_1\right]\right\}, \qquad (3)$$

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^h are model parameters, and $Z(\alpha^h,\alpha^v)$ is the partition function that we approximate as

$$Z(\alpha^h,\alpha^v) \propto (\alpha^h\alpha^v)^{-p}. \qquad (4)$$

The use in Eq. (3) of different parameters α^h and α^v for horizontal and vertical directions makes this model more adaptable to the image characteristics than the TV model that uses a single parameter.

2.2 Degradation Model

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

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \cos(\gamma) & -\sin(\gamma) \\ \sin(\gamma) & \cos(\gamma) \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \mathbf{d}\,, \qquad (5)$$

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], [11] and [9]. 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 $\mathbf{g}_q$, $q = 1,\ldots,U$ from $\mathbf{f}$ can thus be modeled as

$$\mathbf{g}_q = \mathbf{DHR}_{(\mathbf{d}_q,\gamma_q)}\mathbf{f} + \epsilon_q = \mathbf{C}_q\mathbf{f} + \epsilon_q\,, \qquad (6)$$

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

$$\mathbf{C}_q = \mathbf{DHR}_{(\mathbf{d}_q,\gamma_q)}, \qquad (7)$$

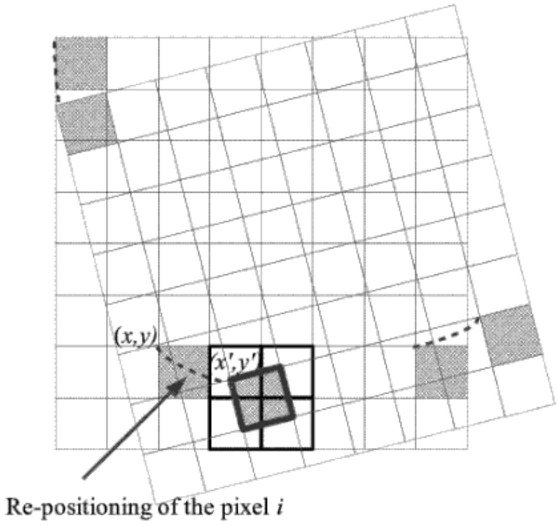

Figure 1. Illustration of the global original and rotated grids.

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 $\mathbf{g} = (\mathbf{g}_1^t,\ldots,\mathbf{g}_U^t)^t$ we have

$$p(\mathbf{g}|\mathbf{f},\beta) \propto \beta^{\frac{UP}{2}} \exp\left[-\frac{\beta}{2}\sum_{q=1}^{U}\|\mathbf{g}_q - \mathbf{C}_q\mathbf{f}\|^2\right]. \qquad (8)$$

2.3 Hyperprior Model

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

$$\mathrm{p}(\Theta) = \prod_{\theta\in\Theta}\mathrm{p}(\theta)\,, \qquad (9)$$

where $\mathrm{p}(\theta)$ are gamma hyperpriors

$$\mathrm{p}(\theta) = \Gamma(\theta|a_\theta^o, b_\theta^o)\,,\ \forall\,\theta\in\Theta. \qquad (10)$$

The gamma distribution has the form

$$\Gamma(\theta\mid u,v) = \frac{v^u}{\Gamma(u)}\theta^{u-1}\exp[-v\theta], \qquad (11)$$

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

$$\mathrm{p}(\Theta,\mathbf{f},\mathbf{g}) = \mathrm{p}(\Theta)\mathrm{p}(\mathbf{f}|\alpha^h,\alpha^v)\mathrm{p}(\mathbf{g}|\mathbf{f},\beta)\,, \qquad (12)$$

where $\mathrm{p}(\Theta)$, $\mathrm{p}(\mathbf{f}|\alpha^h,\alpha^v)$ and $\mathrm{p}(\mathbf{g}|\mathbf{f},\beta)$ are defined respectively in Eqs. (9), (3) and (8) above.

3 Bayesian Inference

Since $\mathrm{p}(\Theta, \mathbf{f}|\mathbf{g})$ of Eq. (1) can not be found in closed form, because $\mathrm{p}(\mathbf{g})$ can not be calculated analytically, we apply variational methods to approximate this distribution by the distribution $\mathrm{q}(\Theta, \mathbf{f})$ minimizing the Kullback-Leibler (KL) divergence, which is given by [7, 6]

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

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

Due to the form of the prior proposed in Eq. (3) the above integral can not be evaluated, but we can however majorize the $\ell 1$ 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 [3] to the TV prior.

Our prior can be rewritten in the more convenient form

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

Let us consider the following inequality, also used in [4], which states that, for any $w \geq 0$ and $z > 0$

$$\sqrt{w} \leq \frac{w+z}{2\sqrt{z}}. \tag{15}$$

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 $\mathbf{u}_i^v$, $i = 1, \ldots, p$, the following functional

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

Now, using the inequality in Eq. (15) and comparing Eq. (16) with Eq. (14), we obtain $\mathrm{p}(\mathbf{f}|\alpha^h, \alpha^v) \geq c \cdot \mathbf{M}(\alpha^h, \alpha^v, \mathbf{f}, \mathbf{u}^h, \mathbf{u}^v)$. As 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

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

Hence, by defining

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

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

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

Therefore, by finding a sequence of distributions $\{\mathrm{q}^k(\Theta, \mathbf{f})\}$ that monotonically decreases $\tilde{\mathcal{M}}(\mathrm{q}(\Theta, \mathbf{f}), \mathbf{g}, \mathbf{u}^h, \mathbf{u}^v)$ for fixed $\mathbf{u}^h$ and $\mathbf{u}^v$, a sequence of an ever decreasing upper bound of $C_{KL}(\mathrm{q}(\Theta, \mathbf{f})||\mathrm{p}(\Theta, \mathbf{f}|\mathbf{g}))$ is also obtained due to Eq. (13). However, also minimizing $\tilde{\mathcal{M}}(\mathrm{q}(\Theta, \mathbf{f}), \mathbf{g}, \mathbf{u}^h, \mathbf{u}^v)$ 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 $\mathrm{q}^k(\Theta, \mathbf{f})$. Therefore, the sequence $\{\mathrm{q}^k(\Theta, \mathbf{f})\}$ is coupled with the sequences $\{\mathbf{u}^{h^k}\}$ and $\{\mathbf{u}^{v^k}\}$. We use the following iterative algorithm to find such sequences.

Algorithm 1 *Posterior image distribution and parameter estimation using* $\mathrm{q}(\Theta, \mathbf{f}) = \mathrm{q}(\Theta)\mathrm{q}(\mathbf{f})$.
Given $\mathrm{q}^1(\Theta)$*, an initial estimate of* $\mathrm{q}(\Theta)$*, and* $\mathbf{u}^{h^1}$, $\mathbf{u}^{v^1} \in (R^+)^p$*, for k = 1, 2, ... until a stopping criterion is met:*

1. *Find*

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

2. *Find*

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

3. *Find*

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

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

To calculate $\mathrm{q}^k(\mathbf{f})$, we observe that differentiating the integral on the right-hand side of Eq. (20) with respect to $\mathrm{q}(\mathbf{f})$ and setting it equal to zero, we obtain

$$\mathrm{q}^k(\mathbf{f}) = \mathcal{N}\left(\mathbf{f} \mid \mathbf{E}_{q^k(\mathbf{f})}[\mathbf{f}], \mathbf{cov}_{q^k(\mathbf{f})}[\mathbf{f}]\right), \tag{23}$$

with

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

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

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

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

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

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

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

and consequently

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

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 $\mathrm{q}^k(\mathbf{f})$ and represent the local spatial activity of $\mathbf{f}$.

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

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

Thus we obtain

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

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

$$\mathrm{q}_d^{k+1}(\alpha^d) = \Gamma\left(\alpha^d | p + a^o_{\alpha^d}, \sum_i \sqrt{\mathbf{u}_i^{d^{k+1}}} + b^o_{\alpha^d}\right) \tag{31}$$

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

$$\mathrm{q}^{k+1}(\beta) = \Gamma\left(\beta | \frac{PU}{2} + a^o_\beta, \frac{\mathbf{E}_{\mathrm{q}^k(\mathbf{f})}\left[\sum_{q=1}^U \|\mathbf{g}_q - \mathbf{C}_q\mathbf{f}\|^2\right]}{2} + b^o_\beta\right). \tag{32}$$

So we have

$$\mathbf{E}_{\mathrm{q}_d^{k+1}(\alpha^d)}\left[\alpha^d\right] = \frac{p + a^o_{\alpha^d}}{\sum_i \sqrt{\mathbf{u}_i^{d^{k+1}}} + b^o_{\alpha^d}}, \tag{33}$$

$$\mathbf{E}_{\mathrm{q}^{k+1}(\beta)}\left[\beta\right] = \frac{\frac{UP}{2} + a^o_\beta}{\frac{1}{2}\mathbf{E}_{\mathrm{q}^k(\mathbf{f})}\left[\sum_{q=1}^U \|\mathbf{g}_q - \mathbf{C}_q\mathbf{f}\|^2\right] + b^o_\beta}. \tag{34}$$

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

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

and

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

where

$$\gamma_{\alpha^d} = \frac{a^o_{\alpha^d}}{p + a^o_{\alpha^d}} \quad \text{and} \quad \gamma_\beta = \frac{a^o_\beta}{UP/2 + a^o_\beta}. \tag{37}$$

These mean values are convex linear combinations of the inverses of the means of the hyperpriors $\alpha_o^d = \frac{a^0_{\alpha^d}}{b^0_{\alpha^d}}$ and $\beta_o = \frac{a^0_\beta}{b^0_\beta}$, 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) requires the evaluation of

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

whose exact evaluation would require to evaluate the full covariance matrix. This problem is solved here utilizing the following approximation

$$\mathbf{cov}_{q^k(\mathbf{f})}[\mathbf{f}] \approx \mathbf{O}^{-1} \tag{39}$$

where

$$\begin{aligned}\mathbf{O} = {} & \mathbf{E}_{q^k(\Theta)}[\beta]\left(\sum_{q=1}^U \overline{\mathbf{C}}_q^t \overline{\mathbf{C}}_q\right) \\ + {} & \mathbf{E}_{q^k(\Theta)}[\alpha^h] z(\mathbf{u}^{h^k})\Delta_h^t \Delta_h \\ + {} & \mathbf{E}_{q^k(\Theta)}[\alpha^v] z(\mathbf{u}^{v^k})\Delta_v^t \Delta_v,\end{aligned} \tag{40}$$

with

$$z(\mathbf{u}) = \frac{1}{P}\sum_i \frac{1}{\sqrt{\mathbf{u}_i}}, \forall \mathbf{u} \in (R^+)^P, \tag{41}$$

and

$$\overline{\mathbf{C}}_q = \mathbf{D}\mathbf{H}\overline{\mathbf{R}}_{(\mathbf{d}_q, \gamma_q)}, \tag{42}$$

where $\overline{\mathbf{R}}_{(\mathbf{d}_q, \gamma_q)}$ is a block circulant matrix with circulant blocks whose first column is given by

$$\overline{\mathbf{R}}_{(\mathbf{d}_q, \gamma_q)\ i\ 1} = \frac{1}{p}\sum_{l=1}^{p} \mathbf{R}_{(\mathbf{d}_q, \gamma_q)\ \{(l+i-1) \bmod p\}\ l}. \tag{43}$$

Now, using this approximation, we have

$$\mathbf{E}_{\mathrm{q}^k(\mathbf{f})}\left[\left(\Delta_i^d(\mathbf{f}-\mathbf{E}_{\mathrm{q}^k(\mathbf{f})}[\mathbf{f}])\right)^2\right] \approx \frac{1}{p}\mathbf{tr}\left(\mathbf{O}^{-1}\Delta^{d\,t}\Delta^d\right), \tag{44}$$

that can be evaluated in the Fourier domain.

Finally, to calculate $\mathbf{E}_{\mathrm{q}^k(\mathbf{f})}\left[\sum_{q=1}^{U}\|\mathbf{g}_q-\mathbf{C}_q\mathbf{f}\|^2\right]$ in Eq. (32), after applying the same approximation, we have for $q = 1,\ldots,U$

$$\mathbf{E}_{\mathrm{q}^k(\mathbf{f})}\left[\|\,\mathbf{g}_q-\mathbf{C}_q\mathbf{f}\,\|^2\right] \approx \|\,\mathbf{g}_q-\mathbf{C}_q\mathbf{E}_{\mathrm{q}^k(\mathbf{f})}[\mathbf{f}]\,\|^2 + \mathbf{tr}\left(\mathbf{O}^{-1}\overline{\mathbf{C}}_q^t\overline{\mathbf{C}}_q\right). \tag{45}$$

4 Experiments

A number of experiments have been carried out in order to evaluate the performance of the proposed Algorithm 1 (henceforth *ℓ1*), compared with the models which use either a TV prior (see [2]) (henceforth *TV*) or a SAR prior (see [10]) (henceforth *SAR*) instead of the proposed $\ell 1$ prior, using in both cases the same variational method.

We show results for the image set of Fig. 2. Sequences of 11 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 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 exactly equal images, is $+1$.

The proposed algorithm was ran until the criterion

$$\|\mathbf{f}^k-\mathbf{f}^{k-1}\|^2/\|\mathbf{f}^{k-1}\|^2 < 10^{-4} \tag{46}$$

was satisfied, where $\mathbf{f}^k$ denotes image point estimate for the k iteration step. When for a given parameter $\theta \in \Theta$ a reliable hint above its value, say θ_o, is available, Gamma hyperpriors parameter values $a_\theta^o = \theta_o b_\theta^o$ and $b_\theta^o \to \infty$, that makes $\mathbf{E}[\theta] \to \theta_o$ and $\mathbf{var}[\theta] \to 0$ (see Eqs. (10-11)) could be utilized. When dealing, as in this section, with synthetic images whose noise variance is known this value can be used. The neighborhood of the values $\alpha_o^d = (\sum_i(\Delta_i^d(\mathbf{f}))^2/p)^{-1}$ for $d \in \{h,v\}$, being $\mathbf{f}$ the original image, have been explored to find the best PSNR value.

Table 1 shows a numeric comparison of the results obtained utilizing the different methods. Fig. 3 shows a detail of one of the 40 dB observations of image Fig. 2b and of their restorations. Finally, Fig. 4 shows one of the 30 dB observations of image Fig. 2d and their different restorations. It is observed that the restored images by the $\ell 1$ method are at least as good as the images estimated by the others methods, both in terms of visual quality and the objective quality metrics used.

5 Conclusion

A new Bayesian super-resolution method for the reconstruction of HR images from a set of translated and rotated, noisy LR images which utilizes a prior on the $\ell 1$ norm of horizontal and vertical differences in images has been proposed. The new method compares favorably with other methods in the literature.

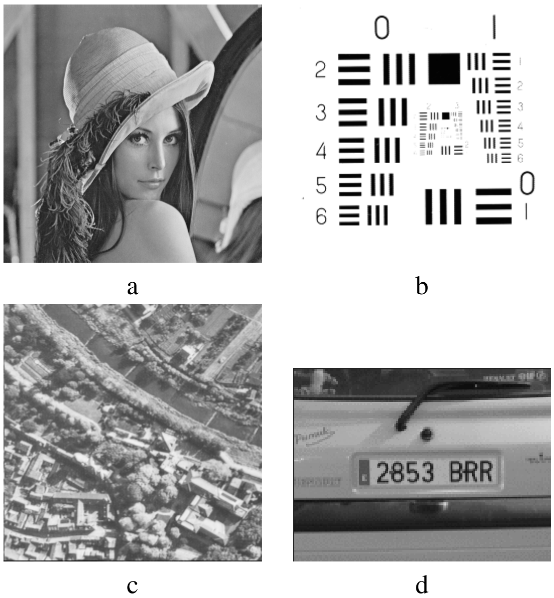

Figure 2. Image set utilized in the experiments: (a) 512×512 Lena image, (b) 256×256 numbers pattern, (c) aerial view and (d) 160×120 car plate.

References

[1] A.K. Katsaggelos, R. Molina, and J. Mateos, *Super resolution of images and video*, Synthesis Lectures on Image, Video, and Multimedia Processing, Morgan & Claypool, 2007.

[2] D. Babacan, R. Molina, and A. Katsaggelos, "Total variation super resolution using a variational approach", In *IEEE International Conference on Image Processing (2008)*, volume I, pp. 641–644, 2008.

[3] S. Babacan, R. Molina, and A. Katsaggelos, "Parameter estimation in tv image restoration using variational distribution approximation", *IEEE Trans. on Image Processing*, 17(3), March 2008, pp. 326–339.

[4] J. Bioucas-Dias, M. Figueiredo, and J. Oliveira, "Total-variation image deconvolution: A majorization-minimization approach", In *ICASSP'2006*, 2006.

[5] R. C. Hardie, K. J. Barnard, J. G. Bognar, E. E. Armstrong, and E. A. Watson, "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), 1998, pp. 247–260.

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

[7] S. Kullback and R. A. Leibler, "On information and sufficiency", *Annals of Math. Stat.*, 22, 1951, pp. 79–86.

[8] R. Molina, A. K. Katsaggelos, and J. Mateos, "Bayesian and regularization methods for hyperparameter estimation in im-

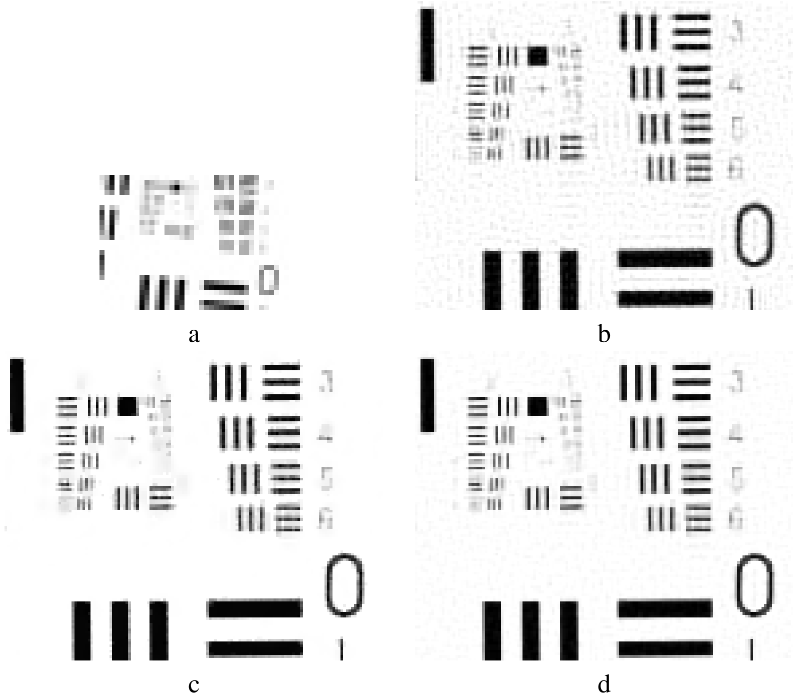

Figure 3. (a) Details of one of the 11 40 dB LR observed images from the sequence of image Fig. 2b, and of their restorations using the (b) SAR , (c) TV and (d) $\ell1$ algorithms.

Table 1. Values of PSNR and SSIM for different reconstructions of the images in Fig. 2.

		SNR 30 dB		SNR 40 dB	
Image	Method	PSNR	SSIM	PSNR	SSIM
(a)	SAR	33.0	0.87	35.1	0.91
	TV	33.5	0.87	35.5	0.91
	$\ell1$	33.7	0.89	35.7	0.92
(b)	SAR	23.3	0.78	27.0	0.84
	TV	25.4	0.89	30.0	0.93
	$\ell1$	25.4	0.93	30.6	0.95
(c)	SAR	26.8	0.85	29.2	0.90
	TV	26.7	0.84	28.7	0.89
	$\ell1$	27.2	0.86	29.3	0.91
(d)	SAR	28.5	0.83	32.1	0.89
	TV	29.7	0.85	32.3	0.90
	$\ell1$	30.0	0.88	33.6	0.94

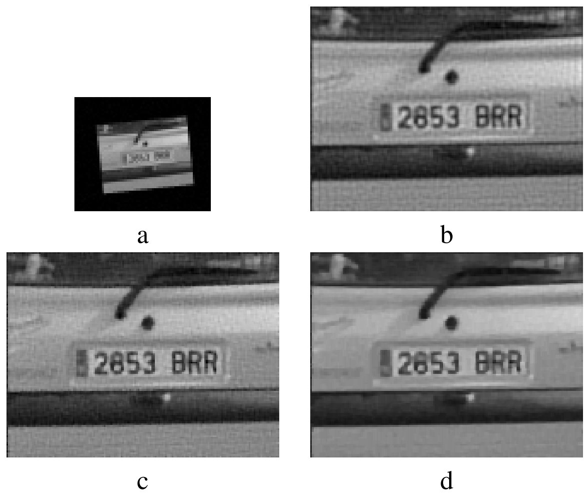

Figure 4. (a) One of the 11 30 dB LR observed images of the sequence from image Fig. 2d, and their restorations using the (b) SAR , (c) TV and (d) $\ell1$ algorithms.

age restoration", *IEEE Transactions on Image Processing*, 8, 1999, pp. 231–246.

[9] L. C. Pickup, D. P. Capel, S. J. Roberts, and A. Zisserman, "Bayesian methods for image super-resolution", *The Computer Journal*, 12(1), 2009, pp. 101–113.

[10] R. Molina, M. Vega, J. Abad and A.K. Katsaggelos, "Parameter Estimation in Bayesian High-Resolution Image Reconstruction with Multisensors", *IEEE Transactions on Image Processing*, 12(12), December 2003, pp. 1655–1667.

[11] M. Tipping and C. Bishop. Bayesian image super-resolution, 2003.

[12] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error measurement to structural similarity", *IEEE Trans. on Img Proc.*, 13(4), 2004, pp. 600–612.