

L1 PRIOR MAJORIZATION IN BAYESIAN IMAGE RESTORATION

Miguel Vega^a, Rafael Molina^b, and Aggelos K. Katsaggelos^c

- a) Dept. Ciencias de la Computación e I. A., Univ. de Granada, 18071 Granada, Spain.
b) Dept. de Lenguajes y Sistemas Informáticos, Univ. de Granada, 18071 Granada, Spain.
c) Dept. of Electrical Engineering and Computer Science,
Northwestern University, Evanston, Illinois 60208-3118.
rms@decsai.ugr.es, mvega@ugr.es, aggk@eecs.northwestern.edu

ABSTRACT

Bayesian methods rely on image priors that encapsulate prior image knowledge and avoid the ill-posedness of image restoration problems. In this paper a new prior based on the l1 norm of vertical and horizontal first order differences of image pixel values is introduced and its parameters are estimated. The results obtained from its application studied and compared with the ones provided by other methods in the literature.

1. INTRODUCTION

In many application areas such as remote sensing, medical imaging, astrophysics and digital photography, the acquired images are degraded and therefore need to be restored. Image degradations may be due to the finite resolution of the acquisition instruments, limited exposure times, adverse environmental conditions, or the motion between the camera and the scene.

A standard formulation of the image degradation model is given in lexicographical form by [1]

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}, \quad (1)$$

where the $P \times 1$ vectors \mathbf{x} , \mathbf{y} , and \mathbf{n} represent, respectively, the original image, the available noisy and blurred image, and the white Gaussian noise with independent elements of variance $\sigma_n^2 = \beta^{-1}$, and \mathbf{H} represents the known blurring matrix. The images are assumed to be of size $m \times n$, with $P = m \times n$. The restoration process aims to find an estimate of \mathbf{x} from \mathbf{y} and \mathbf{H} and some knowledge about \mathbf{n} and possibly \mathbf{x} .

A number of approaches to the restoration problem have been developed (see, for example [1, 2, 3] and the references therein). Bayesian image restoration techniques are based on the study of the posterior probability distribution of the unknown image \mathbf{x} given the available data. In the Bayesian framework a prior image model on \mathbf{x} is introduced, whose aim is to encapsulate our prior image knowledge and consequently to avoid the ill-posedness of the image restoration problem. The selection of this Bayesian prior model is a critical issue. Prior models imposing image smoothness, like the *Continuous Auto-Regression* (CAR) or the *Simultaneous Auto-Regression* (SAR) image models (see [4, 3]), are known to over-smooth edge regions. More sophisticated priors have been proposed based on wavelets [5, 6, 7], *Compound Gauss Markov Random*

This work has been supported by the "Comisión Nacional de Ciencia y Tecnología" under contract TIN2007-65533 and the Consejería de Innovación, Ciencia y Empresa of the Junta de Andalucía under contract P07-TIC-02698.

Fields (CGMRF) [8, 9], *Total-Variation* (TV) [10, 11, 12, 13], or *Fields of Experts* (FoE) [14, 15].

In this paper we present a new prior based on the l1 norm of the horizontal and vertical differences between image pixel values. For this model the image and hyperparameter values are estimated within the Bayesian framework, utilizing variational distribution approximations. Variational approximations not only provide point estimates, but also probability distribution estimates, thus allowing to avoid local minima (See [16, 17]) and to properly take into account the available prior knowledge on hyperparameter values [18].

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

Our Bayesian inference on the unknown image \mathbf{x} and hyperparameters given the observed image \mathbf{y} will be based on the posterior probability distribution

$$p(\Theta, \mathbf{x} | \mathbf{y}) = \frac{p(\Theta, \mathbf{x}, \mathbf{y})}{p(\mathbf{y})}, \quad (2)$$

where Θ represents the set of model parameters. The approximation of this posterior probability distribution, that can not be expressed in closed form, by applying variational methods is explained in section 3. Let us now study the joint probability distribution $p(\Theta, \mathbf{x}, \mathbf{y})$ that can be expressed, within the Hierarchical Bayesian paradigm (see [19]), in terms of the hyperprior model $p(\Theta)$, the prior model $p(\mathbf{x} | \Theta)$ and the degradation model $p(\mathbf{y} | \mathbf{x}, \Theta)$ as

$$p(\Theta, \mathbf{x}, \mathbf{y}) = p(\Theta)p(\mathbf{x} | \Theta)p(\mathbf{y} | \mathbf{x}, \Theta). \quad (3)$$

2.1. Hyperprior Model

Our prior knowledge on the different model parameters $\theta \in \Theta$ has been modeled with the help of gamma hyperpriors

$$p(\theta) = \Gamma(\theta | a_\theta^o, b_\theta^o), \quad \forall \theta \in \Theta. \quad (4)$$

The gamma distribution has the form

$$\Gamma(\theta | u, v) = \frac{v^u}{\Gamma(u)} \theta^{u-1} \exp[-v\theta], \quad (5)$$

where $\theta > 0$, $u > 0$ and $v > 0$ with mean and variance

$$\mathbf{E}[\theta] = u/v, \quad \mathbf{var}[\theta] = u/v^2 \quad (6)$$

(u and v are referred to as scale and precision parameters, respectively).

2.2. Prior Model

The prior model we use is

$$p(\mathbf{x}|\alpha^h, \alpha^v) = \frac{1}{Z(\alpha^h, \alpha^v)} \times \exp \left\{ -\sum_{i=1}^P \left[\alpha^h \|\Delta_i^h(\mathbf{x})\|_1 + \alpha^v \|\Delta_i^v(\mathbf{x})\|_1 \right] \right\}, \quad (7)$$

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

$$Z(\alpha^h, \alpha^v) \propto (\alpha^h \alpha^v)^{-P}. \quad (8)$$

The value $Z(\alpha^h, \alpha^v) = (\alpha^h \alpha^v)^{-P}$ corresponds to the approximation when $p(\mathbf{x}|\alpha^h, \alpha^v)$ is separately integrable in every coordinate. This approximation of the partition function is similar in form to the one defined in [11] for the TV prior. An obvious special case of this model, that has been considered in the experimental section, is obtained when $\alpha^h = \alpha^v = \alpha$.

2.3. Degradation Model

Given the degradation model of Eq. (1), the distribution of the observed image \mathbf{y} given \mathbf{x} is

$$p(\mathbf{y}|\mathbf{x}, \beta) \propto \beta^{\frac{P}{2}} \exp \left[-\frac{1}{2} \beta \|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2 \right]. \quad (9)$$

Although the estimation of the parameter β could have been easily incorporated into our method, we assume here that the noise variance is known or estimated in advance (see for instance [19]).

Our set of model parameters is then

$$\Theta = (\alpha^h, \alpha^v), \quad (10)$$

and the joint distribution is given by

$$p(\Theta, \mathbf{x}, \mathbf{y}) = p(\alpha^h) p(\alpha^v) p(\mathbf{x}|\alpha^h, \alpha^v) p(\mathbf{y}|\mathbf{x}). \quad (11)$$

3. BAYESIAN INFERENCE

Since $p(\Theta, \mathbf{x}|\mathbf{y})$ in Eq. (2) can not be found in closed form, because $p(\mathbf{y})$ can not be calculated analytically, we apply variational methods to approximate this distribution by the distribution $q(\Theta, \mathbf{x})$ minimizing the Kullback-Leibler (KL) divergence, that is given by [20, 21]

$$\begin{aligned} C_{KL}(q(\Theta, \mathbf{x})||p(\Theta, \mathbf{x}|\mathbf{y})) &= \\ & \int q(\Theta, \mathbf{x}) \log \left(\frac{q(\Theta, \mathbf{x})}{p(\Theta, \mathbf{x}|\mathbf{y})} \right) d\Theta d\mathbf{x} \\ &= \int q(\Theta, \mathbf{x}) \log \left(\frac{q(\Theta, \mathbf{x})}{p(\Theta, \mathbf{x}, \mathbf{y})} \right) d\Theta d\mathbf{x} + \text{const} \\ &= \mathcal{M}(q(\Theta, \mathbf{x})) + \text{const}, \end{aligned} \quad (12)$$

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

Due to the form of the prior proposed in Eq. (7) the above integral is difficult to evaluate, but we can however majorize the l1 prior

by a function which renders the integral easier to calculate. The majorization to be applied here to our prior model highly resembles the one applied in [13] to the TV prior.

Our prior can be rewritten in the more convenient form

$$p(\mathbf{x}|\alpha^h, \alpha^v) \propto (\alpha^h \alpha^v)^P \times \exp \left\{ -\sum_{i=1}^P \left[\alpha^h \sqrt{(\Delta_i^h(\mathbf{x}))^2} + \alpha^v \sqrt{(\Delta_i^v(\mathbf{x}))^2} \right] \right\}. \quad (13)$$

Let us consider the following inequality, also used in [11], which states that, for any $w \geq 0$ and $z > 0$

$$\sqrt{w} \leq \frac{w+z}{2\sqrt{z}}. \quad (14)$$

Let us define, for \mathbf{x} , \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

$$\mathbf{M}(\alpha^h, \alpha^v, \mathbf{x}, \mathbf{u}^h, \mathbf{u}^v) = (\alpha^h \alpha^v)^P \times \exp \left\{ -\sum_{i=1}^P \left[\alpha^h \frac{(\Delta_i^h(\mathbf{x}))^2 + \mathbf{u}_i^h}{2\sqrt{\mathbf{u}_i^h}} + \alpha^v \frac{(\Delta_i^v(\mathbf{x}))^2 + \mathbf{u}_i^v}{2\sqrt{\mathbf{u}_i^v}} \right] \right\}. \quad (15)$$

Now, using the inequality in Eq. (14) and comparing Eq. (15) with Eq. (13), we obtain $p(\mathbf{x}|\alpha^h, \alpha^v) \geq c \cdot \mathbf{M}(\alpha^h, \alpha^v, \mathbf{x}, \mathbf{u}^h, \mathbf{u}^v)$. 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{x} . This leads to the following lower bound for the joint probability distribution

$$\begin{aligned} p(\Theta, \mathbf{x}, \mathbf{y}) &\geq c \cdot p(\alpha^h) p(\alpha^v) \mathbf{M}(\alpha^h, \alpha^v, \mathbf{x}, \mathbf{u}^h, \mathbf{u}^v) p(\mathbf{y}|\mathbf{x}) \\ &= \mathbf{F}(\Theta, \mathbf{x}, \mathbf{y}, \mathbf{u}^h, \mathbf{u}^v), \end{aligned} \quad (16)$$

Hence, by defining

$$\tilde{\mathcal{M}}(q(\Theta, \mathbf{x}), \mathbf{u}^h, \mathbf{u}^v) = \int q(\Theta, \mathbf{x}) \log \left(\frac{q(\Theta, \mathbf{x})}{\mathbf{F}(\Theta, \mathbf{x}, \mathbf{y}, \mathbf{u}^h, \mathbf{u}^v)} \right) d\Theta d\mathbf{x}, \quad (17)$$

and using Eq. (16), we obtain for $\mathcal{M}(q(\Theta, \mathbf{x}))$, defined in Eq. (12), what follows

$$\mathcal{M}(q(\Theta, \mathbf{x})) \leq \min_{\{\mathbf{u}^h, \mathbf{u}^v\}} \tilde{\mathcal{M}}(q(\Theta, \mathbf{x}), \mathbf{u}^h, \mathbf{u}^v). \quad (18)$$

Therefore, by finding a sequence of distributions $\{q^k(\Theta, \mathbf{x})\}$ that monotonically decrease $\tilde{\mathcal{M}}(q(\Theta, \mathbf{x}), \mathbf{u}^h, \mathbf{u}^v)$ for fixed \mathbf{u}^h and \mathbf{u}^v , a sequence of an ever decreasing upper bound of $C_{KL}(q(\Theta, \mathbf{x})||p(\Theta, \mathbf{x}|\mathbf{y}))$ is also obtained due to Eq. (12). However, also minimizing $\mathcal{M}(q(\Theta, \mathbf{x}), \mathbf{u}^h, \mathbf{u}^v)$ with respect to \mathbf{u}^h and \mathbf{u}^v , generates vector sequences $\{\mathbf{u}^{h,k}\}$ and $\{\mathbf{u}^{v,k}\}$ that tighten the upper-bound for each distribution $q^k(\Theta, \mathbf{x})$. Therefore, the sequence $\{q^k(\Theta, \mathbf{x})\}$ 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*

$$q(\Theta, \mathbf{x}) = q(\Theta)q(\mathbf{x}).$$

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

1. Find

$$q^k(\mathbf{x}) = \arg \min_{q(\mathbf{x})} \int q(\mathbf{x})q^k(\Theta) \times \log \left(\frac{q(\mathbf{x})q^k(\Theta)}{\mathbf{F}(\Theta, \mathbf{x}, \mathbf{y}, \mathbf{u}^{h^k}, \mathbf{u}^{v^k})} \right) d\mathbf{x}d\Theta \quad (19)$$

2. Find

$$\{\mathbf{u}^{h^{k+1}}, \mathbf{u}^{v^{k+1}}\} = \arg \min_{\{\mathbf{u}^h, \mathbf{u}^v\}} \int q^k(\mathbf{x})q^k(\Theta) \times \log \left(\frac{q^k(\mathbf{x})q^k(\Theta)}{\mathbf{F}(\Theta, \mathbf{x}, \mathbf{y}, \mathbf{u}^h, \mathbf{u}^v)} \right) d\mathbf{x}d\Theta \quad (20)$$

3. Find

$$q^{k+1}(\Theta) = \arg \min_{q(\Theta)} \int q^k(\mathbf{x})q(\Theta) \times \log \left(\frac{q^k(\mathbf{x})q(\Theta)}{\mathbf{F}(\Theta, \mathbf{x}, \mathbf{y}, \mathbf{u}^{h^{k+1}}, \mathbf{u}^{v^{k+1}})} \right) d\mathbf{x}d\Theta \quad (21)$$

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

To calculate $q^k(\mathbf{x})$, we observe that differentiating the integral on the right-hand side of Eq. (19) with respect to $q(\mathbf{x})$ and setting it equal to zero, we obtain

$$q^k(\mathbf{x}) = \mathcal{N}(\mathbf{x} | \mathbf{E}_{q^k(\mathbf{x})}[\mathbf{x}], \mathbf{cov}_{q^k(\mathbf{x})}[\mathbf{x}]), \quad (22)$$

with

$$\mathbf{cov}_{q^k(\mathbf{x})}^{-1}[\mathbf{x}] = \beta \mathbf{H}^t \mathbf{H} + \left(\mathbf{E}_{q^k(\Theta)}[\alpha^h] \Delta_h^t \mathbf{W}(\mathbf{u}^{h^k}) \Delta_h + \mathbf{E}_{q^k(\Theta)}[\alpha^v] \Delta_v^t \mathbf{W}(\mathbf{u}^{v^k}) \Delta_v \right) \quad (23)$$

and

$$\mathbf{E}_{q^k(\mathbf{x})}[\mathbf{x}] = \beta \mathbf{cov}_{q^k(\mathbf{x})}[\mathbf{x}] \mathbf{H}^t \mathbf{y}. \quad (24)$$

In Eq. (23) Δ^h and Δ^v represent $p \times p$ convolution matrices associated with the first order horizontal and vertical differences, respectively, and, $\forall \mathbf{u} \in (R^+)^P$, $\mathbf{W}(\mathbf{u})$ is the diagonal $P \times P$ matrix with entries $\mathbf{W}(\mathbf{u})_{ii} = \frac{1}{\sqrt{u_i}}$ for $i = 1, \dots, P$. 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. That is, for pixels with high spatial activity around the horizontal direction, the corresponding entries of $W(\mathbf{u}^{h^k})$ are very small, which means that no horizontal smoothness is enforced, while for pixels in a flat region the corresponding entries of $W(\mathbf{u}^{h^k})$ are very large, which means that smoothness is enforced; the same stands for the vertical direction.

To calculate $\{\mathbf{u}^{h^{k+1}}, \mathbf{u}^{v^{k+1}}\}$ we have from Eq. (20) that

$$\mathbf{u}^{h^{k+1}} = \arg \min_{\mathbf{u}} \sum_{i=1}^P \frac{\mathbf{E}_{q^k(\mathbf{x})}[(\Delta_i^h(\mathbf{x}))^2] + \mathbf{u}_i}{\sqrt{\mathbf{u}_i}} \quad (25)$$

and

$$\mathbf{u}^{v^{k+1}} = \arg \min_{\mathbf{u}} \sum_{i=1}^P \frac{\mathbf{E}_{q^k(\mathbf{x})}[(\Delta_i^v(\mathbf{x}))^2] + \mathbf{u}_i}{\sqrt{\mathbf{u}_i}}, \quad (26)$$

and consequently

$$\mathbf{u}_i^{h^{k+1}} = \mathbf{E}_{q^k(\mathbf{x})} \left[(\Delta_i^h(\mathbf{x}))^2 \right], \quad (27)$$

and

$$\mathbf{u}_i^{v^{k+1}} = \mathbf{E}_{q^k(\mathbf{x})} \left[(\Delta_i^v(\mathbf{x}))^2 \right], \quad (28)$$

for $i = 1, \dots, P$.

It is clear from Eqs. (27) and (28) that vectors $\mathbf{u}^{h^{k+1}}$ and $\mathbf{u}^{v^{k+1}}$ are respectively functions of the spatial first order horizontal and vertical differences of the unknown image \mathbf{x} under the distribution $q^k(\mathbf{x})$ and represent the local spatial activity of \mathbf{x} .

Finally, differentiating the right hand side of Eq. (21) with respect to $q(\Theta)$ and setting it equal to zero we find that

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

Thus we obtain $q^{k+1}(\Theta) = q^{k+1}(\alpha^h)q^{k+1}(\alpha^v)$, where $q^{k+1}(\alpha^h)$ and $q^{k+1}(\alpha^v)$ are respectively the gamma distributions

$$q^{k+1}(\alpha^h) = \Gamma \left(\alpha^h | P + a_{\alpha^h}^o, \sum_i \sqrt{\mathbf{u}_i^{h^{k+1}}} + b_{\alpha^h}^o \right) \quad (30)$$

and

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

The following point estimates for α^h

$$\mathbf{E}_{q^{k+1}(\alpha^h)}[\alpha^h] = \frac{P + a_{\alpha^h}^o}{\sum_i \sqrt{\mathbf{u}_i^{h^{k+1}}} + b_{\alpha^h}^o}, \quad (32)$$

and for α^v

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

can be utilized. It is possible to express the inverses of these point estimates in the more meaningful way

$$\frac{1}{\mathbf{E}_{q^{k+1}(\alpha^h)}[\alpha^h]} = \gamma_{\alpha^h} \frac{1}{\alpha_o^h} + (1 - \gamma_{\alpha^h}) \frac{\sum_i \sqrt{\mathbf{u}_i^{h^{k+1}}}}{P} \quad (34)$$

and

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

as convex linear combinations between the inverses of the given values $\alpha_o^h = \frac{a_{\alpha^h}^o}{b_{\alpha^h}^o}$ and $\alpha_o^v = \frac{a_{\alpha^v}^o}{b_{\alpha^v}^o}$, and their corresponding ML estimates. It is mentioned here that γ_{α^h} and γ_{α^v} take values into $[0, 1]$, and can be interpreted as the confidence on the given parameter values, and are given by

$$\gamma_{\alpha^h} = \frac{a_{\alpha^h}^o}{P + a_{\alpha^h}^o} \quad \text{and} \quad \gamma_{\alpha^v} = \frac{a_{\alpha^v}^o}{P + a_{\alpha^v}^o}. \quad (36)$$

In section 2.2 reference was made to the particularization of our prior model corresponding to $\alpha^h = \alpha^v = \alpha$. In this particular case we would have $q^{k+1}(\Theta) = q^{k+1}(\alpha)$, where

$$q^{k+1}(\alpha) = \Gamma \left(\alpha | 2P + a_{\alpha}^o, \sum_i \left(\sqrt{\mathbf{u}_i^{h^{k+1}}} + \sqrt{\mathbf{u}_i^{v^{k+1}}} \right) + b_{\alpha}^o \right), \quad (37)$$

with the point estimate for α

$$\frac{1}{\mathbf{E}_{q^{k+1}(\alpha)}[\alpha]} = \gamma_\alpha \frac{1}{\alpha_o} + (1 - \gamma_\alpha) \frac{\sum_i \left(\sqrt{\mathbf{u}_i^{h^{k+1}}} + \sqrt{\mathbf{u}_i^{v^{k+1}}} \right)}{2P}, \quad (38)$$

with

$$\alpha_o = \frac{a_\alpha^o}{b_\alpha^o} \quad \text{and} \quad \gamma_\alpha = \frac{a_\alpha^o}{2P + a_\alpha^o}. \quad (39)$$

Equation (24) can be solved iteratively utilizing the *Conjugate Gradient* (CG) method without the need to explicitly obtain the full covariance matrix of Eq. (23), but estimation of \mathbf{u}^h and \mathbf{u}^v in Eqs (27) and (28) requires the evaluation of

$$\begin{aligned} \mathbf{E}_{q^k(\mathbf{x})} \left[\Delta_i^h(\mathbf{x})^2 \right] &= (\Delta_i^h(\mathbf{E}_{q^k(\mathbf{x})}[\mathbf{x}]))^2 \\ &+ \mathbf{E}_{q^k(\mathbf{x})} \left[\Delta_i^h(\mathbf{x} - \mathbf{E}_{q^k(\mathbf{x})}[\mathbf{x}])^2 \right] \end{aligned} \quad (40)$$

and

$$\begin{aligned} \mathbf{E}_{q^k(\mathbf{x})} \left[\Delta_i^v(\mathbf{x})^2 \right] &= (\Delta_i^v(\mathbf{E}_{q^k(\mathbf{x})}[\mathbf{x}]))^2 \\ &+ \mathbf{E}_{q^k(\mathbf{x})} \left[\Delta_i^v(\mathbf{x} - \mathbf{E}_{q^k(\mathbf{x})}[\mathbf{x}])^2 \right]. \end{aligned} \quad (41)$$

Exact evaluation of these expressions would require the evaluations of the full covariance matrix, which constitutes a formidable problem, to be sorted here utilizing the same kind of approximation that in [13]

$$\begin{aligned} \text{cov}_{q^k(\mathbf{x})}[\mathbf{x}] &\approx \mathbf{B}^{-1} = \left\{ \beta \mathbf{H}^t \mathbf{H} + \right. \\ &\left. \left(\mathbf{E}_{q^k(\Theta)}[\alpha^h] z(\mathbf{u}^{h^k}) \Delta_h^t \Delta_h + \mathbf{E}_{q^k(\Theta)}[\alpha^v] z(\mathbf{u}^{v^k}) \Delta_v^t \Delta_v \right) \right\}^{-1}, \end{aligned} \quad (42)$$

with

$$z(\mathbf{u}) = \frac{1}{P} \sum_i \frac{1}{\sqrt{\mathbf{u}_i}}, \quad \forall \mathbf{u} \in (R^+)^P. \quad (43)$$

Now we have

$$\mathbf{E}_{q^k(\mathbf{x})} \left[\Delta_i^h(\mathbf{x} - \mathbf{E}_{q^k(\mathbf{x})}[\mathbf{x}])^2 \right] \approx \frac{1}{P} \text{tr}(\mathbf{B}^{-1} \Delta_h^t \Delta_h), \quad (44)$$

and

$$\mathbf{E}_{q^k(\mathbf{x})} \left[\Delta_i^v(\mathbf{x} - \mathbf{E}_{q^k(\mathbf{x})}[\mathbf{x}])^2 \right] \approx \frac{1}{P} \text{tr}(\mathbf{B}^{-1} \Delta_v^t \Delta_v). \quad (45)$$

4. EXPERIMENTS

A number of experiments have been carried out in order to evaluate the performance of the proposed Algorithm 1 (thereafter *ALG1*), and its particularization for $\alpha^h = \alpha^v = \alpha$ (*ALG2*), compared with the TV-based algorithm in [13] (*TV*) and the algorithm in [19] (*SAR*). In [13] novel Bayesian variational algorithms were proposed that favorably compare with other state of the art TV-based algorithms. The algorithm in [19] models image distribution by a SAR model and simultaneously estimates image, noise variance and SAR prior hyperparameter.

The image set utilized in the experiments is shown at Fig. 1. Motion blurring with a displacement of 9 pixels have been applied to the original images and Gaussian noise of 20 dB, 30 dB and 40 dB added. Restoration quality has been numerically measured utilizing the improvement in signal-to-noise ratio (ISNR), defined as $\text{ISNR} = 10 \log_{10}(\|\mathbf{x} - \mathbf{y}\|^2 / \|\mathbf{x} - \hat{\mathbf{x}}\|^2)$, where \mathbf{x} , \mathbf{y} and $\hat{\mathbf{x}}$ are the original, observed and reconstructed images respectively. We

have also included image quality measures utilizing the *Structural Similarity Index Measure* (SSIM) defined in [22], whose maximal value, corresponding to exactly equal images, is +1.

The proposed algorithms were initialized utilizing the observations and were ran until the criterion $\|\mathbf{x}^k - \mathbf{x}^{k-1}\|^2 / \|\mathbf{x}^{k-1}\|^2 < 10^{-6}$ was satisfied, where \mathbf{x}^k denotes an image point estimate at the k -th iteration step. The Gamma hyperprior parameter values a_θ^o and b_θ^o , $\forall \theta \in \Theta$ (see Eq. (4)) were determined experimentally for each image.

Table 1 shows a numeric comparison of the results obtained utilizing the different methods and Fig. 2 details of the different restorations of the 30 dB SNR observed image in Fig. 1a. Although differences between figures of merit shown in table 1 are small, it is claimed that the use of the l1 norm allows to obtain better results in general, better if a different prior parameter for horizontal and vertical differences is considered. In some image regions, like the one shown in Fig. 2, the improvement is visually appreciable.

5. CONCLUSION

A new Bayesian method for the restoration of blurred and noisy images with a prior based on the l1 norm of horizontal and vertical differences in images has been proposed. The new method compares favorably to state of the art restoration methods in the literature.

6. REFERENCES

- [1] A.K. Katsaggelos, Ed., *Digital Image Restoration*, New York, Springer-Verlag, 1991.
- [2] M. R. Banham and A. K. Katsaggelos, "Digital image restoration," *IEEE Signal Proces. Mag.*, vol. 14, no. 2, pp. 24–41, Feb. 1997.
- [3] R. Molina, J. Núñez, F.J. Cortijo, and J. Mateos, "Image restoration in astronomy. a bayesian perspective," *IEEE Signal Processing Magazine*, vol. 18, no. 2, pp. 11–29, 2001.
- [4] B. D. Ripley, *Spatial Statistics*, pp. 88 – 90, Wiley, 1981.
- [5] D. Donoho, "Nonlinear solution of linear inverse problems by wavelet-vaguelette decompositions," *Journal of Applied and Computational Harmonic Analysis*, vol. 1, pp. 100–115, 1995.
- [6] M. R. Banham and A. K. Katsaggelos, "Spatially adaptive waveletbased multiscale image restoration," *IEEE Trans. on Image Processing*, vol. 5, pp. 619–634, 1996.
- [7] M. Figueiredo and R. Nowak, "An EM algorithm for wavelet-based image restoration," *IEEE Trans. on Image Processing*, vol. 12, no. 8, pp. 906–916, 2003.
- [8] R. Molina, A.K. Katsaggelos, J. Mateos, A. Hermoso, and C.A. Segall, "Restoration of severely blurred high range images using stochastic and deterministic relaxation algorithms in compound gauss markov random fields," *Pattern Recognition*, vol. 33, no. 4, pp. 555–571, 2000.
- [9] R. Molina, J. Mateos, A.K. Katsaggelos, and M. Vega, "Bayesian multichannel image restoration using compound gauss-markov random fields," *IEEE Transactions on Image Processing*, vol. 12, no. 12, pp. 1642– 1654, December 2003.
- [10] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D*, pp. 259–268, 1992.

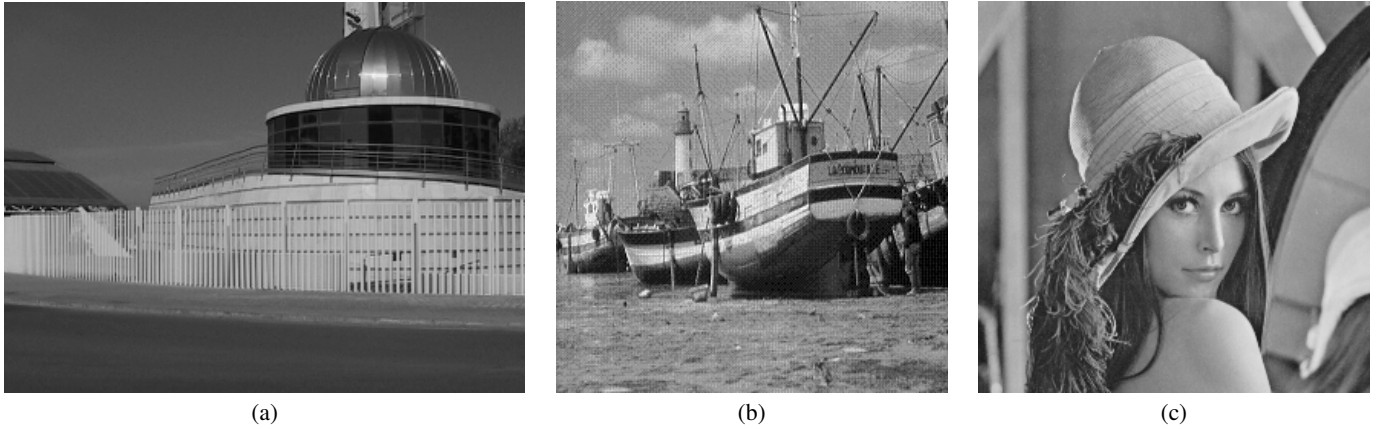


Fig. 1. Image set for the experiments: 320×240 observatory of the "Parque de las Ciencias de Granada", and 256×256 boat and Lena images.

Table 1. Values of ISNR and SSIM for the observatory, boat and Lena images degraded by a horizontal displacement blur of 9 pixels

SNR(dB)	Method	observatory		boat		Lena	
		ISNR	SSIM	ISNR	SSIM	ISNR	SSIM
20	SAR	0.6	0.63	2.1	0.62	3.2	0.74
	TV	0.8	0.67	2.4	0.65	3.3	0.68
	ALG1	1.4	0.68	2.5	0.65	3.3	0.67
	ALG2	0.8	0.69	2.3	0.64	3.2	0.67
30	SAR	3.1	0.76	4.7	0.78	5.9	0.83
	TV	3.4	0.78	5.9	0.83	6.5	0.82
	ALG1	4.3	0.82	6.0	0.83	6.7	0.82
	ALG2	3.5	0.79	6.0	0.83	6.6	0.82
40	SAR	7.3	0.89	8.7	0.91	9.8	0.91
	TV	7.5	0.89	10.3	0.93	10.2	0.90
	ALG1	8.3	0.91	10.7	0.93	10.6	0.90
	ALG2	7.7	0.89	10.8	0.93	10.7	0.90

- [11] J. Bioucas-Dias, M. Figueiredo, and J. Oliveira, "Total-variation image deconvolution: A majorization-minimization approach," in *ICASSP'2006*, 2006.
- [12] J. Bioucas-Dias, M. Figueiredo, and J. Oliveira, "Adaptive bayesian/total-variation image deconvolution: A majorization-minimization approach," in *EUSIPCO'2006*, Sep 2006.
- [13] S.D. Babacan, R. Molina, and A.K. Katsaggelos, "Parameter estimation in tv image restoration using variational distribution approximation," *IEEE Trans. on Image Processing*, vol. 17, no. 3, pp. 326–339, March 2008.
- [14] Stefan Roth and Michael J. Black, "Fields of experts: A framework for learning image priors," in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun 2005, vol. 2, pp. 860–867.
- [15] G. Chantas, N. P. Galatsanos, A. Likas, and M. Saunders, "Variational bayesian image restoration based on a product of t -distributions image prior," *IEEE Trans. on Image Processing*, vol. 17, no. 10, pp. 1795–1805, 2008.
- [16] J. Miskin, *Ensemble Learning for Independent Component Analysis*, Ph.D. thesis, Astrophysics Group, University of Cambridge, 2000.
- [17] D. Tzikas, A. Likas, and N. Galatsanos, "Life after the EM algorithm: The variational approximation for Bayesian inference," *IEEE Signal Processing Magazine* to appear, 2008.
- [18] R. Molina, J. Mateos, and A.K. Katsaggelos, "Blind deconvolution using a variational approach to parameter, image, and blur estimation," *IEEE Trans. on Image Processing*, vol. 15, no. 12, pp. 3715–3727, December 2006.
- [19] R. Molina, A. K. Katsaggelos, and J. Mateos, "Bayesian and regularization methods for hyperparameter estimation in image restoration," *IEEE Trans. on Img. Proc.*, vol. 8, no. 2, pp. 231–246, 1999.
- [20] S. Kullback and R. A. Leibler, "On information and sufficiency," *Annals of Math. Stat.*, vol. 22, pp. 79–86, 1951.
- [21] S. Kullback, *Information Theory and Statistics*, New York, Dover Publications, 1959.
- [22] 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.*, vol. 13, no. 4, pp. 600–612, 2004.

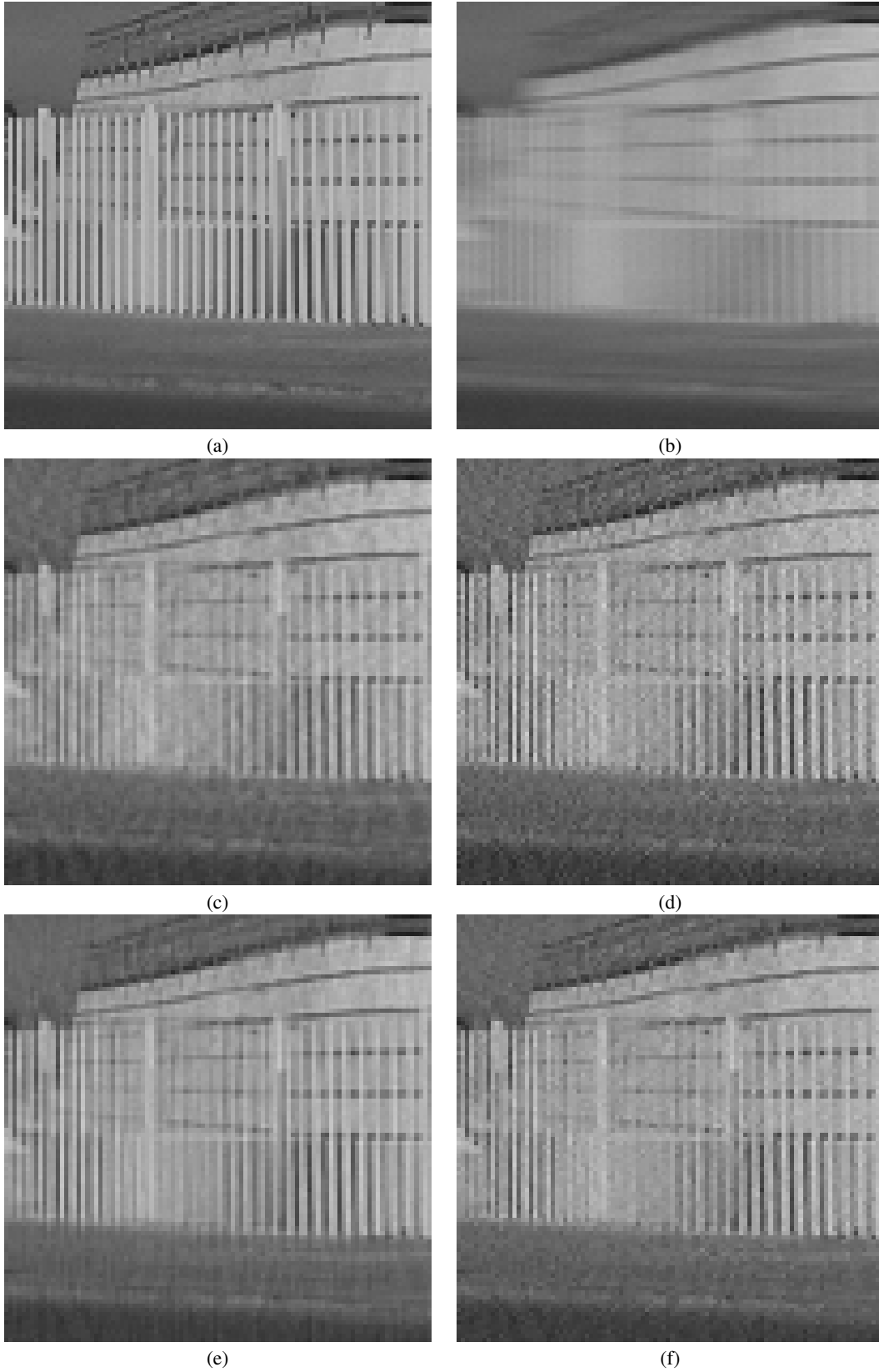


Fig. 2. 100×100 detail of (a) the original image in Fig. 1a ; (b) blurred image with SNR=30 dB; its restorations utilizing (c) SAR , (d) TV , (e) ALG1 and (f) ALG2 algorithms.