

Astronomical Image Restoration Using Variational Methods and Model Combination

Miguel Vega^{1,*}

Dept. de Lenguajes y Sistemas Informáticos, Univ. de Granada, 18071 Granada, Spain

Javier Mateos

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

Rafael Molina

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

Aggelos K. Katsaggelos

*Dept. of Electrical Engineering and Comp. Sci., Northwestern Univ., Evanston, Illinois
60208-3118*

Abstract

In this work we develop a variational framework for the combination of several prior models in Bayesian image restoration and apply it to astronomical images. Since each combination of a given observation model and a prior model produces a different posterior distribution of the underlying image, the use of variational posterior distribution approximation on each posterior will produce as many posterior approximations as priors we want to combine. A unique approximation is obtained here by finding the distribution on the unknown image given the observations that minimizes a linear convex combination of the Kullback-Leibler divergences associated with each poste-

*Corresponding author

Email addresses: mvega@ugr.es (Miguel Vega), jmd@decsai.ugr.es (Javier Mateos), rms@decsai.ugr.es (Rafael Molina), aggk@eecs.northwestern.edu (Aggelos K. Katsaggelos)

¹This work has been supported by the “Consejería de Innovación, Ciencia y Empresa of the Junta de Andalucía” under contract P07-TIC-02698 and by the “Ministerio de Ciencia e Innovación” under contract TIN2010-15137.

rior distribution. We find this distribution in closed form and also relate the proposed approach to other prior combination methods in the literature. Experimental results on both synthetic images and on real astronomical images validate the proposed approach.

Keywords: Model Combination, Bayesian Methods, Variational Methods, Astronomical Image Processing

1. Introduction

As explained in [3, 21], the field of digital image restoration has a quite long history that began in the 1950s with the space program. The first images of the Earth, Moon and planet Mars were, at that time, of unimaginable resolution. However, the images were obtained under major technical difficulties such as vibrations, bad pointing, motion due to spinning, etc. These difficulties resulted, in most cases, in medium to large degradations that could be scientifically and economically devastating. The need to retrieve as much information as possible from such degraded images was the aim of the early efforts to adapt the one-dimensional signal processing algorithms to images, creating a new field that is today known as Digital Image Restoration and Reconstruction. However, for a long time image restoration was considered as a luxury in fields such as optical astronomy.

In 1990 something happened which changed the situation of image restoration in optical astronomy. After the launch of the \$2 billion Hubble Space Telescope (HST) an impossible mistake was discovered in the main mirror. The mirror had a severe problem of spherical aberration because it was polished with the help of a faulty device and checked with the same faulty device. Thus, the checking was perfectly coherent with the polishing but the curvature of the mirror was wrong. Since a single minute of observing telescope time cost about \$100.000, any effort to improve the images was cheap. Since then Astronomy has been an area of important developments and applications of image restoration methods.

In the early 2000s, the papers [21],[30] provided reviews of image restoration in Astronomy, the first paper with an emphasis on Bayesian modeling and inference while the second emphasized the multiresolution approach and wavelets as its mathematical framework. The wavelet framework applied to astronomical image restoration was further developed in the book [27] (see also [28]). Concepts such as Compressed Sensing and Sparse Image Process-

ing (see [7, 25, 29]) have also been applied to imaging inverse problems in Astronomy. However, while in the field of image restoration the combination of prior images models has received some recent attention, to our knowledge, no work in this area has been reported in the astronomical community.

In image restoration there have been several recent attempts to combine image priors (see [9], [24] and [31]). In [9] a Student's-t *Product of Experts (PoE)* image prior model was proposed and learnt only from the observations. Furthermore, the introduction of Bayesian inference methodology, based on the constrained variational approximation, allowed to bypass the difficulty of evaluating the normalization constant of the PoE. PoE priors were learnt in [24] and [31] using a large training set of images and also stochastic sampling methods.

A combination of the TV image model proposed in [1] and the PoE model of [9] has been very recently proposed in [8]. This model combination may be considered a spatially adaptive version of the TV model which furthermore, as the method in [9], has the ability to enforce simultaneously a number of different properties on the image.

In this paper, we apply a novel variational Bayesian methodology to combine prior models in image restoration. While the methodology can be developed more generally we present it here for simplicity for the combination of a sparse and a non-sparse image priors. We also show that this methodology applied to image restoration produces, as a special case, the model developed in [10].

The paper is organized as follows. In section 2 the Bayesian modeling for image restoration is presented. Section 3 describes the variational approximation of the posterior distribution of the unknown image and how inference is performed. Section 4 presents experimental results and section 5 concludes the paper.

2. Bayesian Modeling

Let us assume that \mathbf{x} , the unknown astronomical image of size $p = m \times n$ we would have observed under ideal conditions, is expressed as a column vector by lexicographically ordering the pixels in the image. The observed image \mathbf{y} of size $p = m \times n$ is also expressed as a column vector by lexicographically ordering the pixels in the image.

In this work, we adopt a hierarchical Bayesian framework consisting of two stages (see for example [17]). The first stage is used to model the acquisition

process and the unknown image \mathbf{x} . The observation \mathbf{y} is a random vector with conditional distribution $p(\mathbf{y}|\mathbf{x}, \beta)$. For the unknown image \mathbf{x} we have M models which we want to combine. They are denoted by $p_i(\mathbf{x}|\gamma_i)$ for $i = 1, \dots, M$. These distributions depend on additional parameters β and $\boldsymbol{\gamma} = \{\gamma_1, \dots, \gamma_M\}$ (called *hyperparameters*), which are modeled by assigning *hyperprior* distributions in the second stage of the hierarchical model. Let us now describe those probability distributions. Without lack of generality in the following we will present the combination of two image models, i.e., $M = 2$.

2.1. Observation Model

We assume in this paper that the observed \mathbf{y} is given by the expression

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

where $\boldsymbol{\nu}$ is the acquisition noise, assumed to be white Gaussian with covariance $\beta^{-1}\mathbf{I}$, which is widely used since it produces good restorations (see [12, 18, 20, 22]). Astronomical images often include significant photon (Poisson) noise (see [4, 14, 21]), but we have not included here such type of noise for simplicity. However the variational approach to be utilized in this paper can be applied to Poisson noise as well (see [19] for details). In Eq. (1) \mathbf{H} represents the blurring operator, which in the case of astronomical observations from Earth telescopes is mainly due to the presence of atmospheric turbulence. Then we obtain for the observation model the following normal distribution

$$p(\mathbf{y}|\mathbf{x}, \beta) = \mathcal{N}(\mathbf{H}\mathbf{x}, \beta^{-1}\mathbf{I}). \quad (2)$$

2.2. Image Models

As we have already explained in the introduction, in this paper we combine a sparse prior, the prior model proposed in [33] based on the ℓ_1 norm of the horizontal and vertical first order differences, and a non-sparse one, the simultaneous autoregression (SAR) model [23]. Note that the idea of combining sparse and non-sparse models has also been proposed in other contexts, see for instance [26]. We first consider the prior model based on the ℓ_1 norm which, as the TV prior model [1], is very effective in preserving edges while imposing smoothness. It is defined as

$$p_1(\mathbf{x}|\boldsymbol{\gamma}_1) \propto (\alpha^h \alpha^v)^{\frac{p}{4}} \times \exp \left\{ - \sum_{i=1}^p [\alpha^h \|\Delta_i^h(\mathbf{x})\|_1 + \alpha^v \|\Delta_i^v(\mathbf{x})\|_1] \right\}, \quad (3)$$

where $\Delta_i^h(\mathbf{x})$ and $\Delta_i^v(\mathbf{x})$ represent the horizontal and vertical first order differences at pixel i , respectively, $\boldsymbol{\gamma}_1 = \{\alpha^h, \alpha^v\}$, and α^h and α^v are the horizontal and vertical model parameters.

We also consider the SAR prior, defined as

$$p_2(\mathbf{x}|\gamma_2) \propto \gamma_2^{\frac{p}{2}} \exp\left\{-\frac{\gamma_2}{2} \|\mathbf{C}\mathbf{x}\|^2\right\}, \quad (4)$$

where \mathbf{C} is the Laplacian operator. This prior is expected to preserve image textures better than the ℓ_1 prior.

Notice that in principle we could have considered a prior model of the form

$$p(\mathbf{x}|\boldsymbol{\gamma}_1, \gamma_2) = \frac{1}{Z(\boldsymbol{\gamma}_1, \gamma_2)} \exp\left\{-\sum_{i=1}^p [\alpha_1^h \|\Delta_i^h(\mathbf{x})\|_1 + \alpha_1^v \|\Delta_i^v(\mathbf{x})\|_1] - \frac{\gamma_2}{2} \|\mathbf{C}\mathbf{x}\|^2\right\}, \quad (5)$$

but since there is no known approximation to the partition function $Z(\boldsymbol{\gamma}_1, \gamma_2)$, the estimation of the parameters would be impossible for this prior model (see however [11] in the context of model learning).

2.3. Hyperpriors on the Hyperparameters

Our prior knowledge on the different model parameters, $\boldsymbol{\gamma}_1$, γ_2 and β has been modeled with the help of the gamma hyperpriors

$$p(\omega) = \Gamma(\omega|a_\omega^o, b_\omega^o) = \frac{(b_\omega^o)^{a_\omega^o}}{\Gamma(a_\omega^o)} \omega^{a_\omega^o-1} \exp[-b_\omega^o \omega], \quad (6)$$

where $\omega > 0$ denotes a hyperparameter, and $a_\omega^o > 0$ and $b_\omega^o > 0$ are the shape and inverse scale parameters, respectively. Their mean and variance are

$$\mathbf{E}[\omega] = a_\omega^o/b_\omega^o, \quad \mathbf{var}[\omega] = a_\omega^o/b_\omega^{o2}. \quad (7)$$

Finally, combining Eqs. (2) and (6) with the different prior models we obtain the joint probability distributions

$$p_i(\mathbf{y}, \Omega, \gamma_i) = p(\mathbf{y}|\mathbf{x}, \beta)p(\beta)p_i(\mathbf{x}|\gamma_i)p(\gamma_i), \quad \text{for } i = 1, \dots, M, \quad (8)$$

where $\Omega = \{\mathbf{x}, \beta\}$.

3. Bayesian Inference and Variational Approximation of the Posterior Distribution

Let us denote by $\Theta = \{\Omega, \gamma_1, \gamma_2\}$ the set of all unknowns. The Bayesian inference is based on the posterior distribution $p(\Theta | \mathbf{y})$, that we now approximate, utilizing variational methods, by the factorizable distribution minimizing the linear convex combinations of Kullback-Leibler (KL) divergence functions

$$\hat{q}(\Theta) = \underset{q(\Theta)}{\operatorname{argmin}} \sum_{i=1}^M \lambda_i C_{KL}(q(\Omega)q(\gamma_i) \parallel p_i(\mathbf{y}, \Omega, \gamma_i)) \quad (9)$$

where $\lambda_i \geq 0$, $\sum_i \lambda_i = 1$,

$$q(\Omega) = q(\mathbf{x})q(\beta), \quad (10)$$

$$q(\Theta) = q(\Omega) \prod_{i=1}^M q(\gamma_i), \quad (11)$$

and the Kullback-Leibler (KL) divergences [13] are given by

$$C_{KL}(q(\Omega)q(\gamma_i) \parallel p_i(\mathbf{y}, \Omega, \gamma_i)) = \int q(\Omega)q(\gamma_i) \log \left(\frac{q(\Omega)q(\gamma_i)}{p_i(\mathbf{y}, \Omega, \gamma_i)} \right) d\Omega d\gamma_i. \quad (12)$$

The estimation of $\{\lambda_i\}$ will not be addressed in this paper.

Notice that

$$\int q(\Omega)q(\gamma_i) \log \left(\frac{q(\Omega)q(\gamma_i)}{p_i(\mathbf{y}, \Omega, \gamma_i)} \right) d\Omega d\gamma_i = \int q(\Theta) \log \left(\frac{q(\Omega)q(\gamma_i)}{p_i(\mathbf{y}, \Omega, \gamma_i)} \right) d\Theta, \quad (13)$$

so expression (9) can be rewritten in the more compact form

$$\hat{q}(\Theta) = \underset{q(\Theta)}{\operatorname{argmin}} \int q(\Theta) \log \left(\frac{q(\Omega)}{p(\mathbf{y}|\mathbf{x}, \beta)p(\beta)} \prod_{i=1}^M \left[\frac{q(\gamma_i)}{p_i(\mathbf{x}|\gamma_i)p(\gamma_i)} \right]^{\lambda_i} \right) d\Theta. \quad (14)$$

Unfortunately, we can not directly tackle the minimization of (14) because of the ℓ_1 image prior $p_1(\mathbf{x}|\gamma_1)$ of Eq. (3). In earlier work with ℓ_1 priors (see [32, 33]), this difficulty was overcome by resorting to majorization-minimization (MM) approaches, which is also the method adopted in this

paper. The MM approach was first introduced in the image processing field in [6, 5] as an approximation to the TV regularization problem in denoising.

The main principle of the MM approach is to find a bound of the joint distribution in (8) which makes the minimization of (14) tractable. Let us first consider the following functional $M(\mathbf{x}, \mathbf{u}^h, \mathbf{u}^v, \alpha^h, \alpha^v)$ with p -dimensional vectors $\mathbf{u}^h \in (R^+)^p$, $\mathbf{u}^v \in (R^+)^p$, with components $\mathbf{u}^h(i)$ and $\mathbf{u}^v(i)$, $i = 1, \dots, p$

$$M(\mathbf{x}, \mathbf{u}^h, \mathbf{u}^v, \alpha^h, \alpha^v) = (\alpha^h \alpha^v)^{p/4} \times \exp \left\{ - \sum_{i=1}^p \left[\frac{\alpha^h (\Delta_i^h(\mathbf{x}))^2 + \mathbf{u}^h(i)}{\sqrt{\mathbf{u}^h(i)}} + \frac{\alpha^v (\Delta_i^v(\mathbf{x}))^2 + \mathbf{u}^v(i)}{\sqrt{\mathbf{u}^v(i)}} \right] \right\}. \quad (15)$$

The auxiliary variables \mathbf{u}^h and \mathbf{u}^v are quantities that need to be computed and have, as will be shown later, an intuitive interpretation related to the unknown image \mathbf{x} . It can be shown that the functional $M(\mathbf{x}, \mathbf{u}^h, \mathbf{u}^v, \alpha^h, \alpha^v)$ is a lower bound of the image prior $p_1(\mathbf{x}|\gamma_1)$, that is,

$$p_1(\mathbf{x}|\gamma_1) \geq M(\mathbf{x}, \mathbf{u}^h, \mathbf{u}^v, \alpha^h, \alpha^v). \quad (16)$$

This lower bound can be used to find a lower bound for the joint distribution, for $i = 1$, in (8)

$$p_1(\mathbf{y}, \Omega, \gamma_1) \geq p(\mathbf{y}|\mathbf{x}, \beta) p(\beta) M(\mathbf{x}, \mathbf{u}^h, \mathbf{u}^v, \alpha^h, \alpha^v) p(\gamma_1) = F(\mathbf{y}, \mathbf{u}^h, \mathbf{u}^v, \Omega, \gamma_1), \quad (17)$$

which results in an upper bound of the KL distance as

$$C_{KL}(q(\Omega)q(\gamma_1) \parallel p_1(\mathbf{y}, \Omega, \gamma_1)) \leq C_{KL}(q(\Omega)q(\gamma_1) \parallel F(\mathbf{y}, \mathbf{u}^h, \mathbf{u}^v, \Omega, \gamma_1)). \quad (18)$$

It can be shown (see [32]) that the minimization of $C_{KL}(q(\Omega)q(\gamma_1) \parallel p_1(\mathbf{x}, \Omega, \gamma_1))$ can be replaced by the minimization of its upper bound (18), as minimizing this bound with respect to the unknowns and the auxiliary variables \mathbf{u}^h and \mathbf{u}^v in an alternating fashion results in closer bounds at each iteration. The bound in (18) is quadratic and therefore it is easy to evaluate analytically. A detailed study of the convergence of the used MM approach is beyond the scope of this paper and we refer the reader to [5, 6] for detail. See also [2] on the quality of this estimated posterior distribution as well as on the proximity of the estimated posterior to the true posterior distribution.

To calculate $q(\gamma_i)$ we only have to look at the only divergence where that distribution is present. So we can write

$$q(\gamma_1) = \text{const} \times \exp \left(\langle \log F(\mathbf{y}, \mathbf{u}^h, \mathbf{u}^v, \Omega, \gamma_1) \rangle_{q(\Omega)} \right), \quad (19)$$

where the bound in (18) has been utilized, and $\mathbf{E}_{q(\Omega)}[\cdot] = \langle \cdot \rangle_{q(\Omega)}$ (we will however remove the subscript $q(\Omega)$ when the used distribution is clear from the context). Similarly,

$$q(\gamma_2) = \text{const} \times \exp \left(\langle \log p_2(\mathbf{y}, \Omega, \gamma_2) \rangle_{q(\Omega)} \right). \quad (20)$$

To calculate the rest of the unknown distributions $q(\theta)$ with $\theta \in \Omega$ we have to look at all the divergences. So we obtain

$$q(\theta) = \text{const} \times \exp \left(\langle \log [p(\mathbf{y}|\mathbf{x}, \beta)p(\beta) \right. \\ \left. [M(\mathbf{x}, \mathbf{u}^h, \mathbf{u}^v, \alpha^h, \alpha^v) p(\gamma_1)]^{\lambda_1} [p_2(\mathbf{x}|\gamma_2) p(\gamma_2)]^{1-\lambda_1} \rangle_{q(\Omega_\theta)} \right), \quad (21)$$

where Ω_θ denotes the set Ω with the hyperparameter θ removed and we take into account that $\lambda_2 = 1 - \lambda_1$.

From Eq. (21), the distributions $q(\mathbf{x})$ can be found as the multivariate Gaussian

$$q(\mathbf{x}) = \mathcal{N} \left(\mathbf{x} \mid \mathbf{E}_{q(\mathbf{x})}[\mathbf{x}], \mathbf{cov}_{q(\mathbf{x})}[\mathbf{x}] \right), \quad (22)$$

with

$$\{\mathbf{cov}_{q(\mathbf{x})}[\mathbf{x}]\}^{-1} = \langle \beta \rangle \mathbf{H}^t \mathbf{H} + (1 - \lambda_1) \langle \gamma_2 \rangle \mathbf{C}^t \mathbf{C} \\ + \lambda_1 \left(\langle \alpha^h \rangle \Delta^{h^t} W(\mathbf{u}^h) \Delta^h + \langle \alpha^v \rangle \Delta^{v^t} W(\mathbf{u}^v) \Delta^v \right), \quad (23)$$

and

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

where Δ^h and Δ^v represent $p \times p$ convolution matrices associated with the first order horizontal and vertical differences, respectively, and $W(\mathbf{u}^h)$ and $W(\mathbf{u}^v)$ are a $p \times p$ diagonal matrices of the form $W(\mathbf{u}^d) = \text{diag} \left(\mathbf{u}^d(i)^{-\frac{1}{2}} \right)$, for $i = 1, \dots, p$, $d = h, v$. The inverse covariance in Eq. (23) is a generalization

of the result in [10], which proposed a hierarchical spatially adaptive combination of image priors, corresponding to first order differences along different directions. These priors were found all to contribute with the same weight to the resulting covariance matrix. In the convex combination in Eq. (23), the weights of the two priors can take different values, thus allowing its adaptability to different scenarios.

The matrices $W(\mathbf{u}^h)$ and $W(\mathbf{u}^v)$ in Eq. (23) can be interpreted as spatial adaptivity matrices since they control the amount of 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. Their elements are calculated as

$$\mathbf{u}^d(i) = \langle (\Delta_i^d(\mathbf{x}))^2 \rangle, \quad (25)$$

for $d = h, v$.

Finally, the distributions of the hyperparameters $q(\boldsymbol{\gamma}_1) = q(\alpha^h)q(\alpha^v)$, $q(\gamma_2)$ and $q(\beta)$ are found from Eqs. (19), (20) and (21) as the gamma distributions

$$q(\alpha^d) \propto (\alpha^d)^{\frac{p}{4}-1+a_{\alpha^d}^o} \exp \left[-\alpha^d (b_{\alpha^d}^o + \sum_i \sqrt{\mathbf{u}^d(i)}) \right], \quad (26)$$

for $d = h, v$,

$$q(\gamma_2) \propto \gamma_2^{\frac{p}{2}-1+a_{\gamma_2}^o} \exp \left[-\gamma_2 \left(b_{\gamma_2}^o + \frac{\langle \|\mathbf{C}\mathbf{x}\|^2 \rangle}{2} \right) \right], \quad (27)$$

and

$$q(\beta) \propto \beta^{\frac{p}{2}-1+a_{\beta}^o} \exp \left[-\beta \left(b_{\beta}^o + \frac{\langle \|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2 \rangle}{2} \right) \right]. \quad (28)$$

The proposed algorithm is summarized below in Algorithm 1.

The following point estimates for α^d , for $d = h, v$, and for γ_2 and β can be utilized

$$\mathbf{E}_{q(\alpha^d)} [\alpha^d] = \frac{p/4 + a_{\alpha^d}^o}{\sum_i \sqrt{\mathbf{u}_i^d} + b_{\alpha^d}^o}, \quad (29)$$

$$\mathbf{E}_{q(\gamma_2)} [\gamma_2] = \frac{p + 2a_{\gamma_2}^o}{\langle \|\mathbf{C}\mathbf{x}\|^2 \rangle + 2b_{\gamma_2}^o}, \quad (30)$$

$$\mathbf{E}_{q(\beta)} [\beta] = \frac{p + 2a_{\beta}^o}{\langle \|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2 \rangle + 2b_{\beta}^o}. \quad (31)$$

Algorithm 1 Variational Bayesian Image Restoration

Calculate initial estimates of the original image and hyperparameters

while convergence criterion is not met **do**

1. Estimate the image \mathbf{x} using Eq. (24).
 2. Compute spatial adaptivity vectors \mathbf{u}_b^h and \mathbf{u}_b^v using Eq. (25).
 3. Estimate the distributions of the hyperparameters γ_1 , γ_2 and β using Eqs. (26), (27) and (28).
-

4. Experimental Results

A number of experiments have been performed with the proposed method (henceforth referred as *ALG1*) using several synthetically degraded and real astronomical images and PSFs. When the parameter λ_1 in Eq. (9) is set equal to zero, the contribution of the prior model in Eq. (3) vanishes, and the prior in Eq. (4) becomes responsible of the restoration; this model case will be referred as *SAR*. When we set $\lambda_1 = 1$, we arrive to the opposite situation in which the contribution of the model in Eq. (4) vanishes. In this case our model becomes the one in [33], and will be referred to as ℓ_1 . Our goal in this section is to study the effects of model combination, i.e., to assess if applying ALG1, for intermediate λ_1 values between 0 (*SAR*) and 1 (ℓ_1), it is possible to achieve some benefit in image restoration, particularly when restoring astronomical images. When the images are not blurred and the only degradation is additive noise, we have also included comparisons with the results obtained applying *Hard Thresholding (HT)* of the *Undecimated Wavelet Transform (UWT)*, using the Symmlet 4 Conjugate Mirror Filter (see [29, 30]).

We begin dealing with synthetic images in our first experiment, in which the 512×512 *Lena* and the 256×256 *cameraman* images have been blurred and Gaussian noise of different variances added, to obtain degraded images of *Blurred Signal to Noise Ratio (BSNR)* of 20 dB, 30 dB and 40 dB. Three kinds of blur have been considered: $\mathbf{H} = \mathbf{I}$, i.e., *no blur (nb)*, convolution with a *Gaussian kernel of variance 10 (gauss10)* and *motion blur corresponding to a horizontal displacement of 9 pixels (moh9)*.

The obtained results for HT (for the nb case), SAR, ℓ_1 , and ALG1 methods, have been numerically compared to the original images in term of the *Increase in Signal to Noise Ratio (ISNR)* with respect to the degraded images, and of the *Structural Similarity Index Measure (SSIM)* [34], an index which

takes values in the range $[-1, 1]$, where the higher the value the greater the similarity between images. Three noise realizations have been used in each experiment, and the obtained means and errors for ISNR and SSIM for the Lena and cameraman images are shown in tables 1 and 2 respectively.

In this first experiment, where the noise variance β as well as the original image \mathbf{x}_{orig} are known, the hyperpriors in Eq. (6) can reflect the perfect knowledge of a given parameter ω having a value equal to ω_T , by setting the distribution parameters $a_\omega^o \rightarrow \infty$ and $b_\omega^o \rightarrow \infty$, while $a_\omega^o/b_\omega^o \rightarrow \omega_T$. From equations (29) and (30) we take

$$\frac{a_{\alpha^d}^o}{b_{\alpha^d}^o} = \frac{p/4}{\sum_i \Delta_i^d(\mathbf{x}_{orig})} \quad (32)$$

for $d = h, v$, and

$$\frac{a_{\gamma_2}^o}{b_{\gamma_2}^o} = \frac{p}{\|\mathbf{C}\mathbf{x}_{orig}\|^2}. \quad (33)$$

We run the proposed algorithm until the criterion $\|\mathbf{E}_{q^k(\mathbf{x})}[\mathbf{x}] - \mathbf{E}_{q^{k-1}(\mathbf{x})}[\mathbf{x}]\|^2 / \|\mathbf{E}_{q^{k-1}(\mathbf{x})}[\mathbf{x}]\|^2 < 5 \times 10^{-4}$ is satisfied, which usually takes place after fewer than five iterations. The experiments have been performed on a desktop PC with an Intel(R) Core(TM) i7 860 2.80 GHz processor. When processing the Lena image with the gauss10 blur, the SAR algorithm takes 0.13 s for initialization and 0.05 s per iteration, the ℓ_1 algorithm takes 0.25 s for initialization and 4.5 s per iteration, and finally ALG1 0.26 s for initialization and 5.7 s per iteration. The HT method takes a total time of 66 s for the same image for the no blur case.

For the proposed ALG1, the interval $\lambda_1 \in [0, 1]$ has been explored using a step of 0.1 and the value corresponding to the highest ISNR value has been selected and shown in Tables 1 and 2. As it can be observed in Tables 1 and 2, in most of the cases the value of λ_1 corresponding to the highest ISNR is an intermediate value, which indicates that the combination of the two priors produced better results than using any of the priors alone. The proposed ALG1 performs also better than HT in all considered cases, except for the cameraman image with no blur and noise of 20 dB, where the proposed method obtains mean ISNR and SSIM values slightly lower than HT. In all other cases, the proposed algorithm outperforms HT. Certainly the differences in favor of model combination in term of ISNR and SSIM, even though appreciable in general, specially for higher noise intensities, are

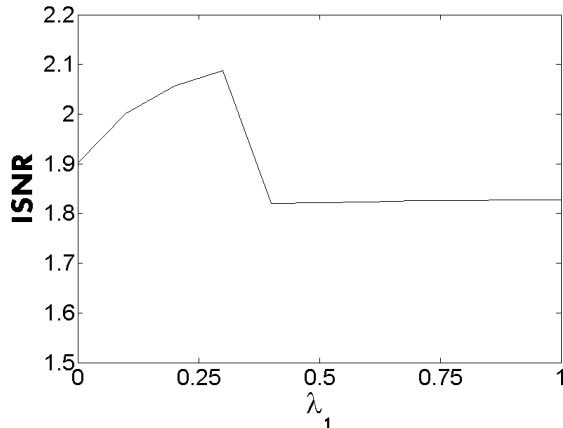


Figure 1: Variation with the value of λ_1 of the ISNR obtained with the proposed method in the restoration of the Lena image for gauss10 blur at 30 dB.

not numerically very significant. As an example, Figure 1 shows a plot of the variation, with the value of λ_1 , of the ISNR obtained with the proposed method in the reconstruction of the Lena image, for gauss10 blur at 30 dB.

Figure 2(a) shows the observed Lena image for gauss10 blur at 30 dB, and the obtained restorations using the SAR method in Figure 2(b), using the ℓ_1 method in Figure 2(c) and, finally, in Figure 2(d) using the proposed method for $\lambda_1 = 0.3$, the value resulting in the higher ISNR. The SAR restoration preserves textures better than ℓ_1 (see the top and the ribbon of the hat of Lena in figures 2(b) and 2(c)), but produces a good amount of ringing, while ℓ_1 preserves better edges and produces virtually no ringing, but at the expense of an oversmoothing of some regions, like the eyes of Lena in Figure 2(c). It can be observed in Figure 2(d) how some of the most pleasing features of both the SAR and ℓ_1 restorations have been preserved by the model combination in the proposed method.

For the second experiment, an astronomical image has been considered: the 512×512 image from the impact of the Comet Shoemaker-Levy 9 with Jupiter, acquired using a CCD camera at the William Hershel telescope through a narrow-band interference filter centered at the methane band at 892 nm, on July 18th, 1994, depicted in Fig. 3(a).

Although there is no exact expression describing the shape of the PSF for images taken from ground based telescopes, previous studies [15], [22] have



Figure 2: (a) Observed Lena image for gauss blur at 30 dB; (b) Restoration using SAR method; (c) Restoration using ℓ_1 method; (d) Restoration using the proposed method for $\lambda_1 = 0.3$.

Table 1: Mean and error values of ISNR and SSIM for the Lena image for different blurs and noises.

BSNR	Method	nb		gauss10		moh9	
		ISNR	SSIM	ISNR	SSIM	ISNR	SSIM
20 dB	HT	2.822	0.9276				
		± 0.0006	$\pm 10^{-4}$				
	SAR	2.427	0.9186	2.263	0.7677	3.627	0.7857
		± 0.008	$\pm 10^{-4}$	± 0.014	$\pm 9 \cdot 10^{-4}$	± 0.018	± 0.002
	$\ell 1$	1.7	0.90	2.460	0.7885	3.867	0.8267
		± 0.5	± 0.01	± 0.017	$\pm 7 \cdot 10^{-4}$	± 0.018	$\pm 6 \cdot 10^{-4}$
	ALG1	2.963	0.9279	2.75	0.7873	4.0	0.811
		± 0.017	$\pm 4 \cdot 10^{-4}$	± 0.07	± 0.004	± 0.3	± 0.024
	λ_1		0.4		0.3		0.2
30 dB	HT	-1.67	0.9720				
		± 0.022	$\pm 8 \cdot 10^{-5}$				
	SAR	0.4285	0.9842	1.902	0.7873	6.02	0.8556
		± 0.0013	$\pm 6 \cdot 10^{-5}$	± 0.003	$\pm 3 \cdot 10^{-4}$	± 0.03	± 0.0015
	$\ell 1$	0.43	0.9842	1.827	0.7875	6.2	0.865
		± 0.16	$\pm 7 \cdot 10^{-4}$	± 0.003	$\pm 3 \cdot 10^{-4}$	± 1.3	± 0.025
	ALG1	0.4825	0.9845	2.09	0.7925	6.880	0.8794
		± 0.0034	$\pm 6 \cdot 10^{-5}$	± 0.08	± 0.0011	± 0.017	± 0.0012
	λ_1		0.1		0.3		0.8
40 dB	HT	-1.710	0.9971				
		± 0.0025	$\pm 9 \cdot 10^{-6}$				
	SAR	0.0461	0.9982	1.8934	0.7921	9.509	0.9179
		± 0.0013	$\pm 5 \cdot 10^{-6}$	$\pm 6 \cdot 10^{-4}$	$\pm 4 \cdot 10^{-5}$	± 0.018	$\pm 3 \cdot 10^{-4}$
	$\ell 1$	-0.070	0.9981	1.844	0.7906	10.2	0.926
		± 0.0025	$\pm 6 \cdot 10^{-6}$	± 0.003	$\pm 2.5 \cdot 10^{-5}$	± 0.8	± 0.019
	ALG1	0.0503	0.9982	1.8934	0.7921	10.82	0.9396
		± 0.0017	$\pm 5 \cdot 10^{-6}$	$\pm 6 \cdot 10^{-4}$	$\pm 4 \cdot 10^{-5}$	± 0.12	± 0.0014
	λ_1		0.1		0		0.5

suggested the following radially symmetric approximation for the PSF

$$h(r) = \left(1 + \frac{r^2}{R^2}\right)^{-\delta}, \quad (34)$$

with $\delta = 3$ and $R = 3.5$ pixels.

In this experiment we know neither the noise variance value, nor the values of the γ parameters, and only the observed image \mathbf{y} is available. The hyperpriors of Eq. (6) can reflect also this lack of knowledge about the value of a given parameter ω , by using hyperparameter values $a_\omega^o \rightarrow 0$ and $b_\omega^o \rightarrow 0$, while $a_\omega^o/b_\omega^o > 0$. Due to the lack of knowledge, the hyperparameter initial values ω^0 used in ALG1 can be of importance. We have taken in all cases the following values: $\alpha^{d0} = \frac{p/4}{\sum_i \Delta_i^d(\mathbf{y})}$ for $d = h, v$, $\gamma_2^0 = \frac{p}{\|\mathbf{C}\mathbf{y}\|^2}$ and $\beta^0 = \frac{p}{\|(\mathbf{I}-\mathbf{H})\mathbf{y}\|^2}$, if $\mathbf{H} \neq \mathbf{I}$ or $\beta^0 = 1$ otherwise.

Algorithm 1 has been first applied with $\lambda_1 = 0$ (SAR) obtaining the restoration depicted in Figure 3(b). The value obtained for γ_2 , that we

Table 2: Mean and error values of ISNR and SSIM for the cameraman image for different blurs and noises.

BSNR	Method	nb		gauss10		moh9	
		ISNR	SSIM	ISNR	SSIM	ISNR	SSIM
20 dB	HT	2.329	0.9293				
		± 0.015	1.7×10^{-4}				
	SAR	1.109	0.8403	1.28	0.6320	3.038	0.6280
		± 0.017	$\pm 3 \times 10^{-4}$	± 0.01	± 0.0014	± 0.018	± 0.0012
	$\ell 1$	1.3	0.836	1.504	0.6839	3.489	0.751
		± 0.6	± 0.025	± 0.011	± 0.0015	± 0.014	± 0.004
	ALG1	2.0	0.87	1.77	0.688	3.6	0.72
		± 0.8	± 0.04	± 0.05	± 0.005	± 0.3	± 0.06
	λ_1		0.4		0.2		0.3
30 dB	HT	0.12	0.9754				
		± 0.03	8×10^{-5}				
	SAR	0.145	0.9708	1.1355	0.6646	5.67	0.724
		± 0.001	$\pm 1.4 \times 10^{-4}$	± 0.0019	$\pm 5 \times 10^{-4}$	± 0.05	± 0.003
	$\ell 1$	0.45	0.9730	1.1543	0.6766	7.3	0.86
		± 0.01	$\pm 1.9 \times 10^{-4}$	$\pm 6 \times 10^{-4}$	$\pm 6 \times 10^{-4}$	± 0.3	± 0.03
	ALG1	0.868	0.9775	1.25	0.702	7.3	0.80
		± 0.025	$\pm 3 \times 10^{-4}$	± 0.11	± 0.006	± 0.4	± 0.05
	λ_1		0.1		0.2		0.9
40 dB	HT	-1.61	0.9947				
		± 0.03	7×10^{-5}				
	SAR	0.016	0.9968	1.1347	0.6777	9.82	0.8355
		± 0.001	$\pm 2.3 \times 10^{-5}$	$\pm 7 \times 10^{-4}$	$\pm 9 \times 10^{-5}$	± 0.05	$\pm 5 \times 10^{-4}$
	$\ell 1$	0.002	0.9968	1.04	0.6763	11.6	0.931
		± 0.014	$\pm 1.7 \times 10^{-5}$	± 0.05	± 0.0016	± 1.4	± 0.023
	ALG1	0.093	0.9969	1.20	0.707	11.71	0.935
		± 0.008	$\pm 1.8 \times 10^{-5}$	± 0.16	± 0.007	± 0.27	± 0.005
	λ_1		0.4		0.2		0.2

denote as γ_{2SAR} , was 6.57×10^{-4} and the value for β , denoted as β_{SAR} , was 0.0015. Later Algorithm 1 has been applied with $\lambda_1 = 1$ ($\ell 1$) obtaining the restoration presented in Figure 3(c) and the resulting values $\alpha_{\ell 1}^h = 0.0040$, $\alpha_{\ell 1}^v = 0.0038$, and $\beta_{\ell 1} = 0.0015$, for α^h , α^v , and $\beta_{\ell 1}$, respectively. Finally Algorithm 1 was run, for different λ_1 values, assuming perfect knowledge of the other parameter values and using the hyperparameter values $\frac{a_{\alpha^d}^o}{b_{\alpha^d}^o} = \alpha_{\ell 1}^d$, for $d = h, v$, $\frac{a_{\gamma_2}^o}{b_{\gamma_2}^o} = \gamma_{2SAR}$ and $\frac{a_{\beta}^o}{b_{\beta}^o} = \frac{1}{2}(\beta_{SAR} + \beta_{\ell 1})$. The best restoration with the proposed method, depicted in Figure 3(d), was obtained with a value of $\lambda_1 = 0.1$.

The restoration of the image of the impact of the comet Shoemaker-Levy with Jupiter, depicted in Figure 3(a), using the proposed method (see Figure 3(d)), appears better than the SAR restoration, presented in Figure 3(b), which appears a bit noisy, and the $\ell 1$ restoration in Figure 3(c). Edges are

better recovered in the ALG1 and ℓ_1 restorations, see the three impacts in the lower part of Jupiter, but the restoration with ALG1 does not exhibit the general oversmoothing of the ℓ_1 restoration.

For the third experiment, we consider the 512×512 astronomical image from the 8th field of the Alhambra survey (see [16]) depicted in Figure 4(a). This image was acquired at the Calar Alto observatory, using the CCD1.1 of the wide field optical LAICA camera, with the Alhambra filter system which covers the region from 3500\AA to 9700\AA . We used the PSF defined in Eq. (34) with the same parameters δ and R as in the second experiment. The procedure to determine the parameters used in the second experiment has also been utilized here. The obtained values were: $\gamma_{2SAR} = 2.2 \cdot 10^{-6}$, $\beta_{SAR} = 1.9 \cdot 10^{-6}$, $\alpha_{l_1}^h = 0.0042$, $\alpha_{l_1}^v = 0.0169$, and $\beta_{l_1} = 1.5 \cdot 10^{-6}$. Algorithm 1 has been applied with $\lambda_1 = 0$ (SAR) obtaining the restoration depicted in Figure 4(b), with $\lambda_1 = 1$ (ℓ_1) obtaining the restoration presented in Figure 4(c), and finally the restoration with the proposed method, depicted in Figure 4(d), was obtained with $\lambda_1 = 0.7$.

The SAR restoration depicted in Figure 4(b) exhibit an aliasing effect, while the ℓ_1 restorations depicted in Figure 4(c), and the restoration using the proposed method, depicted in Figure 4(d), have a better visual quality. The images depicted in Figures 4 have been obtained by rescaling the full range of each image to the interval $[0, 1]$, and then saturating the 1% of the data at low and high intensities.

5. Conclusions

We have presented a new method for the restoration of astronomical images that combines the information provided by two priors: a sparse prior, based on the ℓ_1 norm of the horizontal and vertical differences between image pixel values, and a non-sparse one. The proposed methodology is based on the search of the distribution of the original image given the observations, that minimizes a linear convex combination of the KL divergences associated with each pair of observation and prior models. Based on the presented experimental results, combining information from different priors using the proposed methodology can achieve better reconstructions than utilizing only one prior.

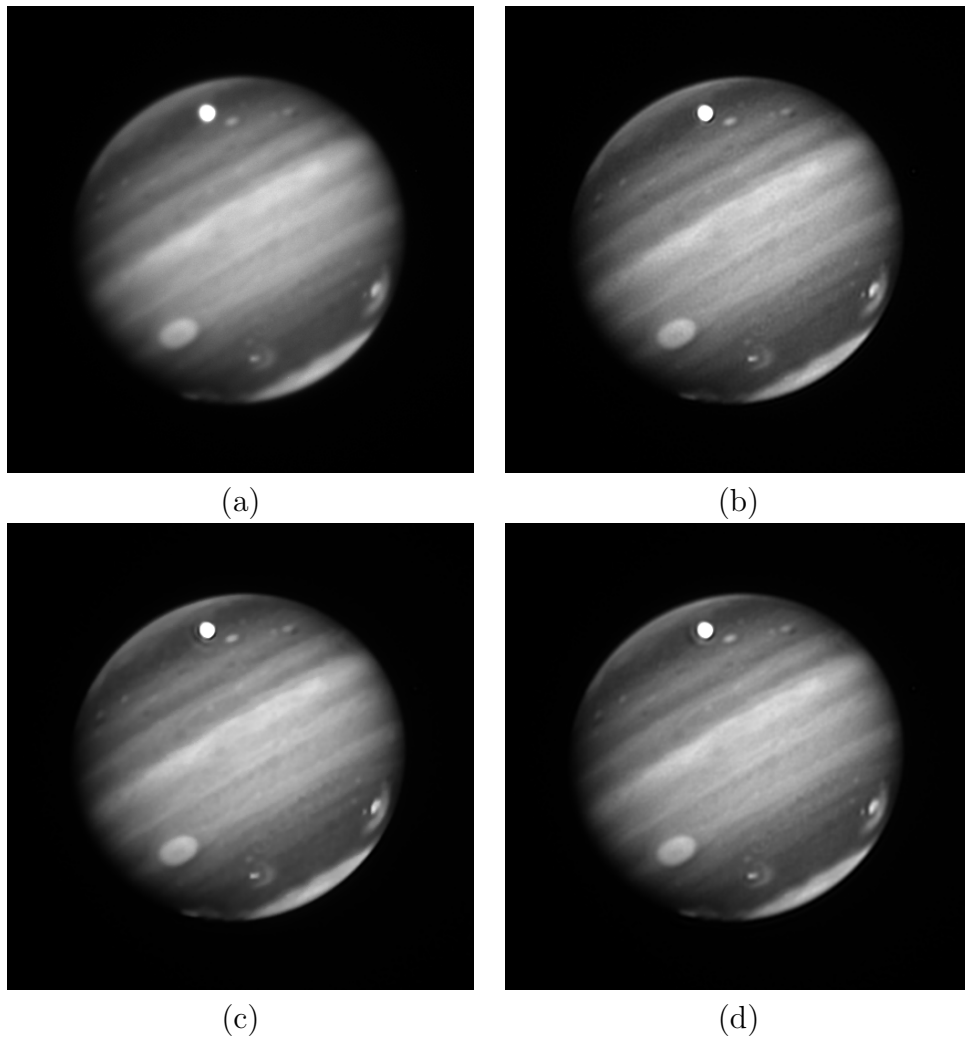


Figure 3: (a) Observed image for the impact of the Comet Shoemaker-Levy 9 with Jupiter; (b) Restoration using SAR method; (c) Restoration using ℓ_1 method; (d) Restoration using the proposed method for $\lambda_1 = 0.1$.

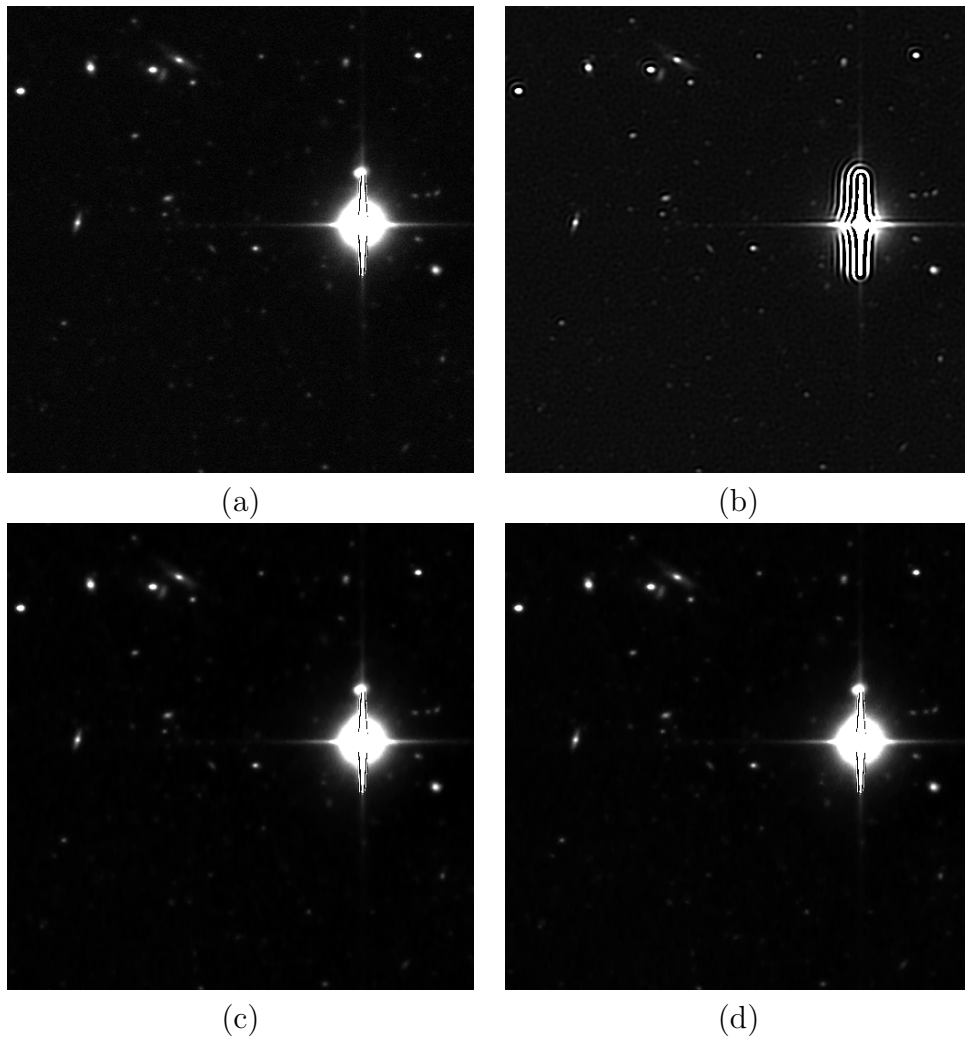


Figure 4: (a) Observed 512×512 image region of 8th field of the Alhambra survey (see [16]); (b) Restoration using SAR method; (c) Restoration using ℓ_1 method; (d) Restoration using the proposed method for $\lambda_1 = 0.7$.

Acknowledgment

We acknowledge Dr. J. Perea from the *Instituto de Astrofísica de Andalucía del Consejo Superior de Investigaciones Científicas* who kindly provided us with the astronomical images used in this paper.

References

- [1] D. Babacan, R. Molina, A. Katsaggelos, Parameter estimation in TV image restoration using variational distribution approximation, *IEEE Trans. on Image Processing* 17 (2008) 326–339.
- [2] D. Babacan, R. Molina, A. Katsaggelos, Variational Bayesian blind deconvolution using a total variation prior, *IEEE Trans. on Image Processing* 18 (2009) 12–26.
- [3] M.R. Banham, A.K. Katsaggelos, Digital image restoration, *IEEE Signal Processing Magazine* 14 (1997) 24–41.
- [4] F. Benvenuto, A.L. Camera, C. Theys, A. Ferrari, H. Lantéri, M. Bertero, The study of an iterative method for the reconstruction of images corrupted by Poisson and Gaussian noise, *Inverse Problems* 24 (2008) 035016 1–20.
- [5] J. Bioucas-Dias, M. Figueiredo, J. Oliveira, Adaptive Bayesian/Total-Variation image deconvolution: A majorization-minimization approach, in: *EUSIPCO 2006*.
- [6] J. Bioucas-Dias, M. Figueiredo, J. Oliveira, Total variation-based image deconvolution: a majorization-minimization approach, in: *ICASSP 2006*, volume 2, p. II.
- [7] J. Bobin, J.L. Starck, R. Ottensamer, Compressed sensing in astronomy, *IEEE Journal of Selected Topics in Signal Processing* 2 (2008) 718–726.
- [8] G. Chantas, N. Galatsanos, R. Molina, A. Katsaggelos, Variational Bayesian image restoration with a spatially adaptive product of total variation image priors, *IEEE Trans. on Image Processing* 19 (2010) 351–362.

- [9] G. Chantas, N.P. Galatsanos, A. Likas, M. Saunders, Variational Bayesian image restoration based on a product of t-distributions image prior, *IEEE Trans. on Image Processing* 17 (2008) 1795–1805.
- [10] J. Chantas, N.P. Galatsanos, A. Likas, Bayesian restoration using a new nonstationary edge-preserving image prior, *IEEE Trans. on Image Processing* 15 (2006) 2987–2997.
- [11] G. Hinton, Training products of experts by minimizing contrastive divergence, *Neural Computation* 14 (2002) 1771–1800.
- [12] A.K. Katsaggelos, M.G. Kang, Spatially adaptive iterative algorithm for the restoration of astronomical images, *Inter. Jour. of Imaging Systems and Technology*, (Special issue on "Image Reconstruction and Restoration in Astronomy") 6 (1995) 305–313.
- [13] S. Kullback, *Information Theory and Statistics*, Dover Publications, Mineola, N.Y., 1959.
- [14] H. Lantéri, C. Theys, Restoration of astrophysical images: the case of Poisson data with additive Gaussian noise, *EURASIP J. Appl. Signal Process.* 2005 (2005) 2500–2513.
- [15] A.F.J. Moffat, A theoretical investigation of focal stellar images in the photographic emulsion and application to photographic photometry, *Astron. Astrophys.* 3 (1969) 454–461.
- [16] M. Moles, N. Benítez, J.A.L. Aguerri, E.J. Alfaro, T. Broadhurst, J. Cabrera-Caño, F.J. Castander, J. Cepa, M. Cerviño, D. Cristobal-Hornillos, A. Fernández-Soto, R.M. González Delgado, L. Infante, I. Márquez, V.J. Martínez, J. Masegosa, A. del Olmo, J. Perea, F. Prada, J.M. Quintana, S.F. Sánchez, The ALHAMBRA survey: A large area multimediuum-band optical and near-infrared photometric survey, *The Astronomical Journal* 136 (2008) 1325.
- [17] R. Molina, A.K. Katsaggelos, J. Mateos, Bayesian and regularization methods for hyperparameter estimation in image restoration, *IEEE Trans. on Image Processing* 8 (1999) 231–246.

- [18] R. Molina, A.K. Katsaggelos, J. Mateos, J. Abad, Compound gauss-markov random fields for astronomical image restoration, *Vistas in Astronomy* (Special issue on "Vision Modeling and Information Coding") 40 (1997) 539–546.
- [19] R. Molina, A. Lopez, J. Martin, A. Katsaggelos, Variational posterior distribution approximation in Bayesian emission tomography reconstruction using a gamma mixture prior, in: *VISAPP 2007*, pp. 165–173.
- [20] R. Molina, J. Mateos, J. Abad, N. Pérez de la Blanca, A. Molina, F. Moreno, Bayesian image restoration in astronomy. applications to images of the recent collision of comet Shoemaker-Levy 9 with Jupiter, *Inter. Jour. of Imaging Systems and Technology* (Special issue on "Image Reconstruction and Restoration in Astronomy") 6 (1995) 370–375.
- [21] R. Molina, J. Núñez, F. Cortijo, J. Mateos, Image restoration in astronomy. A Bayesian perspective, *IEEE Signal Processing Magazine* 18 (2001) 11–29.
- [22] R. Molina, B.D. Ripley, Using spatial models as priors in astronomical image analysis, *J. Applied Statistic* 16 (1989) 193–206.
- [23] B.D. Ripley, *Spatial Statistics*, Wiley, 1981.
- [24] S. Roth, M. Black, Fields of experts: a framework for learning image priors, in: *IEEE CVPR 2005*, volume 2, pp. 860–867.
- [25] J.L. Starck, J. Bobin, Astronomical data analysis and sparsity: From wavelets to compressed sensing, *Proceedings of the IEEE* 98 (2010) 1021–1030.
- [26] J.L. Starck, M. Elad, D. Donoho, Image decomposition via the combination of sparse representation and a variational approach, *IEEE Trans. on Image Processing* 14 (2005) 1570–1582.
- [27] J.L. Starck, F. Murtagh, *Astronomical Image and Data Analysis*, Springer-Verlag, 2006.
- [28] J.L. Starck, F. Murtagh, A. Bijaoui, *Image Processing and Data Analysis: The Multiscale Approach*, Cambridge University Press, 1998.

- [29] J.L. Starck, F. Murtagh, J. Fadili, Sparse Image and Signal Processing: Wavelets, Curvelets, Morphological Diversity, Cambridge University Press, 2010.
- [30] J.L. Starck, E. Pantin, F. Murtagh, Deconvolution in astronomy: a review, Publications of the Astronomical Society of the Pacific 114 (2002) 1051–1069.
- [31] D. Sun, W.K. Cham, Postprocessing of low bit-rate block dct coded images based on a fields of experts prior, IEEE Trans. on Image Processing 16 (2007) 2743–2751.
- [32] M. Vega, J. Mateos, R. Molina, A.K. Katsaggelos, Super resolution of multispectral images using l1 image models and interband correlations, in: IEEE MLSP 2009, Grenoble (France), pp. 1–6.
- [33] M. Vega, R. Molina, A.K. Katsaggelos, l1 prior majorization in Bayesian image restoration, in: 16th IEEE DSP 2009, Santorini (Greece), pp. 1–6.
- [34] Z. Wang, A.C. Bovik, H.R. Sheikh, E.P. Simoncelli, Image quality assessment: From error visibility to structural similarity, IEEE Trans. on Image Processing 13 (2004) 600–612.