Journal of Systems Engineering and Electronics
Vol. 28, No. 6, December 2017, pp.1236 – 1247

Astronomical image restoration using variational Bayesian blind deconvolution

Xiaoping Shi, Rui Guo*, Yi Zhu, and Zicai Wang

Control and Simulation Center, Harbin Institute of Technology, Harbin 150080, China

Abstract: An algorithm is presented for image prior combinations based blind deconvolution and applied to astronomical images. Using a hierarchical Bayesian framework, the unknown original image and all required algorithmic parameters are estimated simultaneously. Through utilization of variational Bayesian analysis, approximations of the posterior distributions on each unknown are obtained by minimizing the Kullback-Leibler (KL) distance, thus providing uncertainties of the estimates during the restoration process. Experimental results on both synthetic images and real astronomical images demonstrate that the proposed approaches compare favorably to other state-of-the-art reconstruction methods.

Keywords: blind deconvolution, variational Bayesian, model combination, astronomical image processing.

DOI: 10.21629/JSEE.2017.06.21

1. Introduction

As explained in [1,2], the digital image restoration techniques have improved enormously since 1950s with the space program. The program was mainly responsible for restoring images in optical astronomy which were, at that time, of unimaginable resolution but held untold scientific benefits. However, the images obtained from the various space missions were subject to major technical difficulties such as substandard imaging environments, bad pointing, the spinning and tumbling of the spacecraft, etc. These difficulties resulted, in most cases, in blurring degradations to the acquired images that could be scientifically and economically devastated. The extreme need for the ability to retrieve as much meaningful information as possible from degraded images was encountered. As a result, it was not long before some of classical signal processing approaches were adapted to images, finding their way into the new field described as digital image restoration. Not surprisingly, astronomical imaging is one of the important developments and primary applications of digital image restoration.

A great deal of methods have been proposed to tackle the image restoration in astronomy. The paper [3] provides reviews of the major approaches including theoretical expectations and computer simulations. A wavelet framework based on Bayesian modeling and inference was applied to astronomical image restoration and further expanded in the book [4]. Compressed sensing and sparse image processing based on sparsity [5 – 7] have been introduced to astronomical data analysis. However, while the variational Bayesian blind deconvolution has received more attention in the field of image processing, to our knowledge, few attempts for this method have been made in the astronomical image restoration.

In classical image restoration methods, it seeks an estimate of the original image assuming the blur function is given and the degradation process is known. In contrast, blind deconvolution deals with the much more challenging and realistic problem in the area of astronomical image restoration, where little exact information is available about the degradation. In general, the degradation process for astronomical images is nonlinear and spatially varying, which is hard to be inferred only from the observed images. Nevertheless, most of the literatures assume that the observed astronomical image results from a linear spatially invariant (LSI) system, which is not suitable for many real image processing situations in astronomy.

There are two main categories to the blind deconvolution problem according to the stage at which we identify the degradation process. In the a priori degradation identification methods, the degradation process is previously identified from the observed images, and then used in one of the classical image restoration algorithms [8]. Another category of methods identify both the original image and degradation process simultaneously [9 – 11]. In these methods, prior knowledge about the original image, degradation, and noise is introduced in the estimation process, and the Bayesian framework allows the prior models on the unknown parameters are described in a system-

Manuscript received July 29, 2016.
*Corresponding author.

atic way, which is often employed in blind deconvolution. To obtain approximations to the posterior distributions of the unknown parameters, variational methods can be applied to the blind deconvolution problem with the use of the Kullback–Leibler (KL) divergence [12]. The variational Bayesian blind deconvolution has been utilized in [13–15].

In this paper, we propose to use a novel variational Bayesian methodology for the blind deconvolution problem by incorporating the combination of prior models in image restoration. A combination of the sparse total variation (TV) or $l1$ norm and the non-sparse simultaneous auto regressive (SAR) image prior models is systematically included in a hierarchical Bayesian framework, along with the observed images, the unknown original image, and the motion or blur parameters. Using a variational Bayesian analysis, the original image and all algorithmic parameters are jointly estimated and do not require user supervision. Moreover, the proposed Bayesian framework based blind deconvolution provides uncertainties of the estimates during the restoration process, which helps to prevent error propagation and improves robustness.

The paper is organized as follows. Section 2 provides the hierarchical Bayesian paradigm, along with the prior models for the observation, the image, the motion and the hyperparameter. The variational approximation method based on the Bayesian inference is described in Section 3. Section 4 presents experimental results and Section 5 concludes the paper.

2. Hierarchical Bayesian paradigm

In digital image processing, an observed image can be obtained from the two-dimensional convolution of the original image with a linear space-invariant blur or a motion model, plus the additive noise. That is

$$g(x,y) = f(x,y) * h(x,y) + n(x,y) = \sum_{(n,m)\in \boldsymbol{S}_h} f(n,m)h(x-n,y-m) + n(x,y), \quad (x,y) \in \boldsymbol{S}_f \tag{1}$$

where $*$ denotes the 2-D linear convolution operator, $g(x,y)$, $f(x,y)$, $h(x,y)$, and $n(x,y)$ represent the observed image, the original image, the blur or motion, and the additive noise, respectively. $\boldsymbol{S}_f \subset \mathbf{R}^2$ is the support of the image, and $\boldsymbol{S}_h \subset \mathbf{R}^2$ is the support of the model for blur or motion. Recovering the original image $f(x,y)$ becomes the process of the deconvolution of the $h(x,y)$ from the observed image characteristics.

The image degradation model of (1) is also often expressed as a discrete linear system, which leads to a matrix-vector form, that is

$$\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f} + \boldsymbol{n} \tag{2}$$

by lexicographically ordering the vectors $\boldsymbol{g}$, $\boldsymbol{f}$, and $\boldsymbol{n}$. The observation noise $\boldsymbol{n}$ generally originates during image acquisition, processing, or transmission, and is commonly assumed to be white Gaussian noise uncorrelated with the image. $\boldsymbol{H}$ represents a block circulant with circulant blocks (BCCB) matrix [16], which allows (1) to be expressed in the discrete frequency domain with impulse response $\boldsymbol{h}$. Thus (2) can also be written as $\boldsymbol{g} = \boldsymbol{F}\boldsymbol{h} + \boldsymbol{n}$, where $\boldsymbol{F}$ is a matrix formed from the image data and $\boldsymbol{h}$ becomes a vector parameterizing the model for blur or motion. We assume the observed image $\boldsymbol{g}$ and the original image $\boldsymbol{f}$ consist of N and PN pixels, respectively, where the integer $P > 1$ is the factor of increase in resolution. The blind deconvolution problem for this matrix-vector form is equivalent to estimating $\boldsymbol{f}$ and $\boldsymbol{h}$ given $\boldsymbol{g}$ using prior knowledge about $\boldsymbol{f}$, $\boldsymbol{h}$ and $\boldsymbol{n}$.

According to the principle of the Bayesian philosophy, the original image $\boldsymbol{f}$ and the model for blur or motion $\boldsymbol{h}$ are assigned corresponding prior probability density functions (PDFs) $p_i(\boldsymbol{f}|\alpha_i)$, for $i = 1,\ldots,m$, and $p(\boldsymbol{h})$, which incorporate our knowledge about the imaging process and the nature of the image into the estimation process. The parameter m is the number of models which we want to combine. The conditional distribution $p(\boldsymbol{g}|\boldsymbol{f},\boldsymbol{h},\beta)$ is termed the likelihood of the observations. These parameters α_i and β in above distributions are termed hyperparameters and assumed unknown. To alleviate the ill-posed problem that these parameters are unknown, an additional modeling stage for hyperprior distributions can be incorporated to model the prior knowledge of the hyperparameters, leading to a hierarchical Bayesian paradigm. This abstraction generally increases robustness to errors which result from the uncertainty in the observed data. This hierarchical Bayesian paradigm on the form of joint distribution is defined as

$$p(\alpha_i,\beta,\boldsymbol{f},\boldsymbol{h},\boldsymbol{g}) = p(\boldsymbol{g}|\boldsymbol{f},\boldsymbol{h},\beta)p(\boldsymbol{f}|\alpha_i)p(\boldsymbol{h})p(\alpha_i)p(\beta) \tag{3}$$

and inference is based on $p(\alpha_i,\beta,\boldsymbol{f},\boldsymbol{h}|\boldsymbol{g})$.

The hyperparameters in the hierarchical Bayesian paradigm are crucial, which have to be addressed when dealing with inference for image restoration problems. The first problem is that the definition of $p(\alpha_i)$ and $p(\beta)$ can deal with the case of known hyperparameters in degenerate distributions, and with more likely situation that little knowledge about these hyperparameters is available. The second problem to be solved is to determine how the hyperparameters allow the inference to carry out, and commonly used methods include the evidence analysis in [17] and the empirical analysis in [18]. The third problem to be considered is how to decide the way of calculating the posterior distribution $p(\alpha_i, \beta, \boldsymbol{f}, \boldsymbol{h}|\boldsymbol{g})$.

In the following subsections we first provide the various prior models for the individual distributions appearing in this paper, then study the estimation of these unknown parameters using inference models under the hierarchical Bayesian paradigm.

(i) First stage: prior models on the observation, image and motion

For the observation noise $\boldsymbol{n}$, we assume that a flexible model for zero mean independent white Gaussian noise is used, specifying the likelihood of the observed image with independent elements of variance $\sigma_n^2 = \beta^{-1}$. Thus we have

$$p(\boldsymbol{g}|\boldsymbol{f}, \boldsymbol{h}, \beta) \propto \beta^{N/2} \exp\left[-\frac{\beta}{2}\|\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}\|^2\right]. \quad (4)$$

Similarly, we can obtain another form of observation model using the $N \times N$ convolution matrix $\boldsymbol{F}$ and rewrite (4) as

$$p(\boldsymbol{g}|\boldsymbol{f}, \boldsymbol{h}, \beta) \propto \beta^{N/2} \exp\left[-\frac{\beta}{2}\|\boldsymbol{g} - \boldsymbol{F}\boldsymbol{h}\|^2\right]. \quad (5)$$

The incorporation of accurate image models $p_i(\boldsymbol{f}|\alpha_i)$ introduces better quality of the estimated original image $\boldsymbol{f}$ and improves the accuracy of other unknowns, constraining the possible solutions of the prior knowledge to the most probable ones. Several important considerations while choosing the image prior refer to the typical descriptions of the natural images and the analytical tractability of the Bayesian inference. These generally limit the possible choices.

In this paper we adopt the SAR prior model [19], the prior model based on the $l1$ norm of horizontal and vertical first-order differences (f.o.d.) [20] and the TV prior model [21]. The SAR model is a common non-sparse prior from the class of Gaussian models with quadratic energy, which generally overpenalizes the image edges. It is defined by

$$p_1(\boldsymbol{f}|\alpha_1) \propto \alpha_1^{\frac{PN}{2}} \exp\left[-\frac{\alpha_1}{2}\|\mathbb{C}\boldsymbol{f}\|^2\right] \quad (6)$$

where $\mathbb{C}$ denotes the discrete Laplacian operator and α_1 denotes the hyperparameter in this prior model. To preserve the image edges by not overpenalizing outliers in the image gradient distribution, two popular sparse priors below are widely used. The prior model based on $l1$ norm is given by

$$p_2(\boldsymbol{f}|\alpha_2) \propto (\alpha_2^h \alpha_2^v)^{\frac{PN}{4}} \cdot$$
$$\exp\left\{-\sum_{i=1}^{PN}[\alpha_2^h\|\Delta_i^h(\boldsymbol{f})\|_1 + \alpha_2^v\|\Delta_i^v(\boldsymbol{f})\|_1]\right\} \quad (7)$$

where the operators $\Delta_i^h(\boldsymbol{f})$ and $\Delta_i^v(\boldsymbol{f})$ denote the horizontal and vertical f.o.d., respectively, at the pixel i with $\alpha_2 = \{\alpha_2^h, \alpha_2^v\}$, being α_2^h and α_2^v the model hyperparameters. The TV prior model is given by

$$p_3(\boldsymbol{f}|\alpha_3) = \alpha_3^{\frac{PN}{2}} \cdot$$
$$\exp\left[-\frac{\alpha_3}{2}\sum_{i=1}^{PN}\sqrt{(\Delta_i^h(\boldsymbol{f}))^2 + (\Delta_i^v(\boldsymbol{f}))^2}\right] \quad (8)$$

where α_3 denotes the hyperparameters of this prior model. The most important difference between TV and $l1$ lies in the $l1$ prior model with two hyperparameters α_2^h and α_2^v providing more possible direction dependence of the edge strength.

For the model for $\boldsymbol{h}$ in this paper, we assume that the blur is known and only consider a parametric motion model consisting of translational and rotational information modeled as $\boldsymbol{h} = (\theta, x, y)^{\mathrm{T}}$, where θ is the rotation angle, and x and y are the horizontal and vertical displacements corresponding to a reference frame. However, the motion parameters estimated solely based on the observed images make estimation errors unavoidable and cause significant drawbacks in blind deconvolution. Therefore, we treat the motion parameters as stochastic variables and model their uncertainty using the following Gaussian distributions:

$$p(h) = \mathbb{N}(\boldsymbol{h}|\tilde{\boldsymbol{h}}^p, \boldsymbol{\Lambda}^p), \quad (9)$$

where $\tilde{\boldsymbol{h}}^p$ denotes the initial estimate of $\boldsymbol{h}$ using some conventional registration algorithms [22,23], and $\boldsymbol{\Lambda}^p$ denotes the prior covariance matrix. The two parameters introduce prior knowledge about the motion information into the estimation process.

(ii) Second stage: hyperpriors on the hyperparameters

The hyperparameters α_1, α_2, α_3 and β are important for the performance of the blind deconvolution when they are unknown and need to be simultaneously estimated by introducing a second stage in the hierarchical Bayesian paradigm. To obtain tractable Bayesian inference, conjugate hyperprior distributions are incorporated to carry out and calculate the posterior distribution $p(\alpha_i, \beta, \boldsymbol{f}, \boldsymbol{h}|\boldsymbol{g})$.

The generally used hypeprior distribution is the uninformative prior model, given by

$$p(\omega) = \text{const}. \quad (10)$$

However, little information is available about its prior knowledge, instead we employ Gamma distributions, which are more flexible hyperprior distributions, given by

$$p(\omega) = \Gamma(\omega; a_\omega^o, b_\omega^o) = \frac{(b_\omega^o)^{a_\omega^o}\omega^{a_\omega^o - 1}}{\Gamma(a_\omega^o)}\exp[-b_\omega^o\omega] \quad (11)$$

where ω denotes a hyperparameter, a_ω^o is the scale parameter and b_ω^o is the shape parameter, both of which are assumed to be known and make the estimation process rely on the prior knowledge about the hyperparameters. The Gamma distribution has the following mean, variance, and mode:

$$\mathrm{E}[\omega] = a_\omega^o / b_\omega^o, \quad \mathrm{Var}[\omega] = a_\omega^o/(b_\omega^o)^2,$$
$$\mathrm{Mode}[\omega] = a_\omega^o - 1/b_\omega^o. \tag{12}$$

3. Variational Bayesian inference

For all unknowns in the previous section, including hyperparameters, image and motion, we denote by $\Theta = \Theta_i = (\alpha_i, \Omega) = (\alpha_i, \beta, \boldsymbol{f}, \boldsymbol{h})$ for clarity. As mentioned above, the Bayesian inference is based on the posterior distribution

$$p(\Theta_i|\boldsymbol{g}) = p(\alpha_i, \beta, \boldsymbol{f}, \boldsymbol{h}|\boldsymbol{g}) = \frac{p(\alpha_i, \beta, \boldsymbol{f}, \boldsymbol{h}, \boldsymbol{g})}{p(\boldsymbol{g})}. \tag{13}$$

There are many existing methods in the literature that can be utilized to obtain estimates of the unknowns, such as the evidence-based analysis [24], the empirical-based approach [25], and the expectation-maximization (EM) algorithm [26]. Among these approaches, the EM algorithm can guarantee convergence to local maxima of the posterior distribution which is obtained from the marginal distribution to provide estimates of the unknowns. However, in many applications, the posterior distribution $p(\Theta_i|\boldsymbol{g})$ cannot be found, since

$$p(\boldsymbol{g}) = \int\int\int\int p(\alpha_i, \beta, \boldsymbol{f}, \boldsymbol{h}, \boldsymbol{g}) \mathrm{d}\boldsymbol{h} \mathrm{d}\boldsymbol{f} \mathrm{d}\beta \mathrm{d}\alpha_i \tag{14}$$

cannot be calculated analytically. Therefore, an approximation of a simpler tractable distribution $q(\Theta)$ using the variational Bayesian analysis [27] is proposed. Good posterior distribution approximation $q(\Theta)$ will be found by decreasing the KL divergence between $q(\Theta_i)$ and $p(\Theta_i|\boldsymbol{g})$, given by

$$C_{\mathrm{KL}}(q(\Theta_i)||p(\Theta_i|\boldsymbol{g})) =$$
$$\int q(\Theta_i) \lg(\frac{q(\Theta_i)}{p(\Theta_i|\boldsymbol{g})}) \mathrm{d}\Theta_i =$$
$$\int q(\Theta_i) \lg(\frac{q(\Theta_i)}{p(\Theta_i, \boldsymbol{g})}) \mathrm{d}\Theta_i + \mathrm{const} \tag{15}$$

which is always nonnegative and equal to zero if $q(\Theta_i)$ and $p(\Theta_i|\boldsymbol{g})$ coincide, which corresponds to the EM algorithm. To decrease the complexity of computation, analytical solutions for the family of distributions $q(\Theta_i)$ obtain a factorized form utilizing the mean field approximation [28,29], given by

$$q(\Theta_i) = q(\Omega) q(\alpha_i) = q(\boldsymbol{f}) q(\boldsymbol{h}) q(\beta) q(\alpha_i). \tag{16}$$

Next, we assume a vector parameter $\Upsilon \in \Theta$ and denote by Θ_Υ the subset of Θ with Υ removed. Each distribution of these parameters can be estimated through an iterative procedure using the factorization. During each iteration solution process, the distribution of the parameter Υ can be obtained by the knowledge about other parameters Θ_Υ, given by

$$q(\Upsilon) = \arg\min_{q(\Upsilon)} C_{\mathrm{KL}}(q(\Theta_\Upsilon) q(\Upsilon) || p(\Theta|\boldsymbol{g})) =$$
$$\mathrm{const} \times \exp(\langle \lg p(\Theta, \boldsymbol{g}) \rangle_{q(\Theta_\Upsilon)}) \tag{17}$$

where $\langle \bullet \rangle_{q(\Theta_\Upsilon)}$ denotes the expected value relating to the distribution $q(\Theta_\Upsilon)$.

For the model combination, we propose two configurations which are between quadratic SAR and non-quadratic $l1$ or TV image priors. Unfortunately, we cannot directly utilize the prior models $p_i(\boldsymbol{f}|\alpha_i)$ for $i = 2, 3$, within variational Bayesian methods due to the intractable performance of the KL distance in (15). In this paper, majorization-minimization (MM) [30] approaches can be proposed to tackle the problem by obtaining bounds to the joint distribution, which make the minimization of (15) tractable. Lower bounds to the distribution in (3), for $i = 2, 3$, can be found as follows:

$$M_2(\alpha_2, \boldsymbol{f}, w_2) = (\alpha_2^h \alpha_2^v)^{\frac{PN}{4}} \cdot$$
$$\exp\left\{ -\sum_{i=1}^{PN} \left[\alpha_2^h \frac{(\Delta_i^h(\boldsymbol{f}))^2 + \omega_{2i}^h}{2\sqrt{\omega_{2i}^h}} + \right.\right.$$
$$\left.\left. \alpha_2^v \frac{(\Delta_i^v(\boldsymbol{f}))^2 + \omega_{2i}^v}{2\sqrt{\omega_{2i}^v}} \right] \right\} \tag{18}$$

and

$$M_3(\alpha_3, \boldsymbol{f}, w_3) = (\alpha_3)^{\frac{PN}{2}} \cdot$$
$$\exp\left\{ -\frac{\alpha_3}{2} \sum_{i=1}^{PN} \left[\frac{(\Delta_i^h(\boldsymbol{f}))^2 + (\Delta_i^v(\boldsymbol{f}))^2 + \omega_{3i}}{\sqrt{\omega_{3i}}} \right] \right\} \tag{19}$$

where PN-dimensional auxiliary variables ω_{2i}^h, $\omega_{2i}^v \omega_{3i} \in (\mathbf{R}^+)^{PN}$ for $i = 1, \ldots, PN$, play the role of local adaptability factors, needing to be computed and lower the smoothing effect of the function.

We use the following geometric-arithmetic mean inequality, which is expressed as

$$\sqrt{ab} \leqslant \frac{a+b}{2} \Rightarrow \sqrt{a} \leqslant \frac{a+b}{2\sqrt{b}}. \tag{20}$$

We can find these functions in (18) and (19) are lower bounds of the image priors $p_i(\boldsymbol{f}|\alpha_i)$, for $i = 2, 3$, that is

$$p_i(\boldsymbol{f}|\alpha_i) \geqslant M(\alpha_i, \boldsymbol{f}, w_i). \tag{21}$$

These lower bounds can result in lower bounds for the joint distributions for $i=2,3$ in (3), respectively, given by

$$p(\Theta_i,\boldsymbol{g}) \geqslant p(\boldsymbol{g}|\Omega)M_i(\alpha_i,\boldsymbol{f},\boldsymbol{w}_i)p(\boldsymbol{h})p(\alpha_i)p(\beta) = F(\Theta_i,\boldsymbol{g},w_i), \tag{22}$$

and consequently, an upper bound of the KL distance in (14) is given by

$$C_{\mathrm{KL}}(q(\Theta_i)||p(\Theta_i|\boldsymbol{g})) \leqslant C_{\mathrm{KL}}(q(\Theta_i)||F(\Theta_i,\boldsymbol{g},w_i)). \tag{23}$$

Therefore, the minimization of this upper bound can be utilized instead of the minimization in (15), since the form of the inequality in (23) allows the unknowns and the auxiliary variable w lead to an alternating optimization strategy, where closer bounds at each iteration revise the estimates of the unknown distributions using the current estimates of other distributions. Also this upper bound is quadratic and can be evaluated analytically. Utilizing this bound, we can calculate $q(\alpha_i)$, for $i=1,2,3$, using the standard solution of the variational Bayesian methods in (17). Thus we can obtain

$$q(\alpha_1) \propto \exp(\langle \lg p_1(\Omega,\alpha_1,\boldsymbol{g})\rangle_{q(\Omega)}) \tag{24}$$

and

$$q(\alpha_i) \propto \exp(\langle \lg F_i(\Theta_i,\boldsymbol{g},w_i)\rangle_{q(\Omega)}),\ \text{for } i=2,3 \tag{25}$$

where $\langle * \rangle_{q(\Omega)}$ denotes the desired estimation value of $*$ based on the $q(\Omega)$ distribution. Next, we consider all the divergences to obtain the distributions for the rest of the unknowns $q(\xi)$ with $\xi \in \Omega$, given by

$$q_c(\xi) \propto \exp(\langle \lg[p(\boldsymbol{g}|\Omega)p(\boldsymbol{h})p(\beta)\cdot [M_{i_c}(\alpha_{i_c},\boldsymbol{f},w_{i_c})p(\alpha_{i_c})]^{\lambda_{i_c}}\cdot [p_1(\boldsymbol{f}|\alpha_1)p(\alpha_1)]^{1-\lambda_{i_c}}]\rangle_{q(\Theta_{i_c\xi})}) \tag{26}$$

which includes the two configurations $c \in \{l1, \mathrm{TV}\}$, for $i_c=2$ when $c=l1$ and $i_c=3$ when $c=\mathrm{TV}$, and $\Theta_{i_c\xi}$ denotes the set of parameters with ξ removed.

(i) Estimation of the original image and registration parameters distributions

From (26), the distribution function $q(\boldsymbol{f})$ of TV and $l1$ can be found as

$$q_{\mathrm{TV}}(\boldsymbol{f}) \propto \exp\left[-\frac{1}{2}\left(\lambda_3\langle\alpha_3\rangle\sum_i \frac{(\Delta_i^h(\boldsymbol{f}))^2+(\Delta_i^v(\boldsymbol{f}))^2+w_{3i}}{\sqrt{w_{3i}}}+ (1-\lambda_3)\langle\alpha_1\rangle\|\mathbb{C}\boldsymbol{f}\|^2+\langle\beta\rangle\langle\|\boldsymbol{g}-\boldsymbol{H}\boldsymbol{f}\|^2\rangle_{q(\boldsymbol{h})}\right)\right] \tag{27}$$

and

$$q_{l1}(\boldsymbol{f}) \propto \exp\left[-\frac{1}{2}\left(\lambda_2\left\{\langle\alpha_2^h\rangle\sum_i \frac{(\Delta_i^h(\boldsymbol{f}))^2+w_{2i}^h}{\sqrt{w_{2i}^h}}+ \langle\alpha_2^v\rangle\sum_i \frac{(\Delta_i^v(\boldsymbol{f}))^2+w_{2i}^v}{\sqrt{w_{2i}^v}}\right\}+(1-\lambda_2)\langle\alpha_1\rangle\|\mathbb{C}\boldsymbol{f}\|^2+ \langle\beta\rangle\langle\|\boldsymbol{g}-\boldsymbol{H}\boldsymbol{f}\|^2\rangle_{q(\boldsymbol{h})}\right)\right]. \tag{28}$$

Next, we obtain the distribution of the motion model $\boldsymbol{h}$ using

$$q(\boldsymbol{h}) \propto \exp\left[-\frac{1}{2}(\langle\beta\rangle\langle\|\boldsymbol{g}-\boldsymbol{H}\boldsymbol{f}\|^2\rangle_{q(\boldsymbol{f})}+ (\boldsymbol{h}-\tilde{\boldsymbol{h}}^p)^{\mathrm{T}}(\boldsymbol{\Lambda}^p)^{-1}(\boldsymbol{h}-\tilde{\boldsymbol{h}}^p))\right]. \tag{29}$$

These distributions are related to the expectation values $\langle\|\boldsymbol{g}-\boldsymbol{H}\boldsymbol{f}\|^2\rangle_{q(\boldsymbol{f})}$ and $\langle\|\boldsymbol{g}-\boldsymbol{H}\boldsymbol{f}\|^2\rangle_{q(\boldsymbol{h})}$, which are not easy to calculate since $\boldsymbol{H}$ in (2) is always with nonlinear form in many astronomical applications. Therefore, the motion model $\boldsymbol{h}$ will be expanded by its first-order Taylor series $\boldsymbol{O}_r(\tilde{\boldsymbol{h}})$, for $r=1,\dots,3$, around the mean value $\tilde{\boldsymbol{h}}=\langle\boldsymbol{h}\rangle=(\tilde{\theta},\tilde{x},\tilde{y})^{\mathrm{T}}$ of the distribution $q(\boldsymbol{h})$, and $\boldsymbol{h}$ can be replaced by

$$\boldsymbol{h} \approx \tilde{\boldsymbol{h}}+[\boldsymbol{O}_1(\tilde{\boldsymbol{h}}),\boldsymbol{O}_2(\tilde{\boldsymbol{h}}),\boldsymbol{O}_3(\tilde{\boldsymbol{h}})](\boldsymbol{h}-\tilde{\boldsymbol{h}}), \tag{30}$$

resulting in the following approximation:

$$\langle\|\boldsymbol{g}-\boldsymbol{H}\boldsymbol{f}\|^2\rangle_{q(\boldsymbol{h})} \approx \|\boldsymbol{g}-\boldsymbol{H}\boldsymbol{f}\|^2+ \sum_{i=1}^{3}\sum_{j=1}^{3}\langle\beta\rangle\xi_{ij}\boldsymbol{f}^{\mathrm{T}}\boldsymbol{O}_i(\tilde{\boldsymbol{h}})^{\mathrm{T}}\boldsymbol{O}_j(\tilde{\boldsymbol{h}})\boldsymbol{f} \tag{31}$$

where for $i,j=1,\dots,3$, ξ_{ij} denotes the essential factor of the 3×3 covariance matrix $\boldsymbol{\Lambda}^p$ of the posterior $q(\boldsymbol{h})$ in (29).

Substituting (31) into (27) and (28), these posterior distributions $q_c(\boldsymbol{f})$ result in the multivariate Gaussian

$$q_c(\boldsymbol{f}) = \mathbb{N}(\boldsymbol{f}|\mathrm{E}_{q_c(\boldsymbol{f})}[\boldsymbol{f}],\mathrm{cov}_{q_c(\boldsymbol{f})}[\boldsymbol{f}]) \tag{32}$$

with mean parameter

$$\mathrm{E}_{q_c(\boldsymbol{f})}[\boldsymbol{f}] = \mathrm{cov}_{q_c(\boldsymbol{f})}[\boldsymbol{f}][\langle\beta\rangle(\tilde{\boldsymbol{h}})^{\mathrm{T}}\boldsymbol{g}] \tag{33}$$

and inverse covariance

$$\mathrm{cov}^{-1}_{q_{\mathrm{TV}}(\boldsymbol{f})}[\boldsymbol{f}] = \langle\beta\rangle(\tilde{\boldsymbol{h}})^{\mathrm{T}}\tilde{\boldsymbol{h}}+ \sum_{i=1}^{3}\sum_{j=1}^{3}\langle\beta\rangle\xi_{ij}\boldsymbol{O}_i(\tilde{\boldsymbol{h}})^{\mathrm{T}}\boldsymbol{O}_j(\tilde{\boldsymbol{h}})+ \lambda_3\langle\alpha_3\rangle(\Delta^{\boldsymbol{h}^{\mathrm{T}}}\boldsymbol{W}(w_3)\Delta^{\boldsymbol{h}}+\Delta^{\boldsymbol{v}^{\mathrm{T}}}\boldsymbol{W}(w_3)\Delta^{\boldsymbol{v}})+ (1-\lambda_3)\langle\alpha_1\rangle\mathbb{C}^{\mathrm{T}}\mathbb{C} \tag{34}$$

for $c=$ TV, and

$$\mathrm{cov}^{-1}_{q_{l1}(\boldsymbol{f})}[\boldsymbol{f}] = \langle\beta\rangle(\tilde{\boldsymbol{h}})^{\mathrm{T}}\tilde{\boldsymbol{h}} + \sum_{i=1}^{3}\sum_{j=1}^{3}\langle\beta\rangle\xi_{ij}\boldsymbol{O}_i(\tilde{\boldsymbol{h}})^{\mathrm{T}}\boldsymbol{O}_j(\tilde{\boldsymbol{h}}) +$$

$$\lambda_2(\langle\alpha_2^h\rangle\Delta^{h^{\mathrm{T}}}\boldsymbol{W}(w_2^h)\Delta^{\boldsymbol{h}} + \langle\alpha_2^v\rangle\Delta^{v^{\mathrm{T}}}\boldsymbol{W}(w_2^v)\Delta^v) +$$

$$(1-\lambda_2)\langle\alpha_1\rangle\mathbb{C}^{\mathrm{T}}\mathbb{C} \tag{35}$$

for $c = l1$. It is clear from the equations above that Δ^h and Δ^v denote the $PN \times PN$ convolution matrices related respectively to the horizontal and vertical f.o.d., and $W(w)_{ii} = \dfrac{1}{\sqrt{w_i}}$, for $i = 1, \ldots, PN$, represents a $PN \times PN$ diagonal matrix with the essential factors w_i, introducing the local spatial activity into the estimation process of the original image. Therefore, the matrix $\boldsymbol{W}(w)$ can be used as space adaptation matrices by controlling the spatial smoothing applied at different locations.

The following expressions for the auxiliary vector w_i are calculated as

$$w_{2i}^h = \mathrm{E}_{q_{l1}(\boldsymbol{f})}[(\Delta_i^h(\boldsymbol{f}))^2],$$

$$w_{2i}^v = \mathrm{E}_{q_{l1}(\boldsymbol{f})}[(\Delta_i^v(\boldsymbol{f}))^2] \tag{36}$$

$$w_{3i} = \mathrm{E}_{q_{\mathrm{TV}}(\boldsymbol{f})}[(\Delta_i^h(\boldsymbol{f}))^2 + (\Delta_i^v(\boldsymbol{f}))^2]. \tag{37}$$

Next, substituting (30) again into the following approximation, we obtain

$$\langle\|\boldsymbol{g} - \boldsymbol{Hf}\|^2\rangle_{q(\boldsymbol{f})} \approx \|\boldsymbol{g} - \tilde{\boldsymbol{h}}\mathrm{E}_{q_c(\boldsymbol{f})}[\boldsymbol{f}] - \boldsymbol{\Upsilon}(\boldsymbol{h} - \tilde{\boldsymbol{h}})\|^2 +$$

$$\mathrm{tr}[(\tilde{\boldsymbol{h}})^{\mathrm{T}}\tilde{\boldsymbol{h}}\mathrm{cov}_{q_c(\boldsymbol{f})}] + 2\Theta^{\mathrm{T}}(\boldsymbol{h} - \tilde{\boldsymbol{h}}) +$$

$$(\boldsymbol{h} - \tilde{\boldsymbol{h}})^{\mathrm{T}}\boldsymbol{\Psi}(\boldsymbol{h} - \tilde{\boldsymbol{h}}) \tag{38}$$

which can model the distribution $q(\boldsymbol{h})$ of (29) as the Gaussian distribution $q(\boldsymbol{h}) = N(\boldsymbol{h}|\langle\boldsymbol{h}\rangle, \boldsymbol{\Lambda})$ with two variables

$$\langle\boldsymbol{h}\rangle = \boldsymbol{\Lambda}[(\boldsymbol{\Lambda}^p)^{-1}\tilde{\boldsymbol{h}}^p + \langle\beta\rangle(\boldsymbol{\Gamma}\tilde{\boldsymbol{h}} + \boldsymbol{\Psi}\tilde{\boldsymbol{h}} + \boldsymbol{Q} - \boldsymbol{\Phi})] \tag{39}$$

and

$$\boldsymbol{\Lambda}^{-1} = (\boldsymbol{\Lambda}^p)^{-1} + \langle\beta\rangle(\boldsymbol{\Psi} + \boldsymbol{\Gamma}). \tag{40}$$

In (38)–(40) above, $\boldsymbol{\Upsilon} = [\boldsymbol{O}_1(\tilde{\boldsymbol{h}})\mathrm{E}_{q_c(\boldsymbol{f})}[\boldsymbol{f}], \boldsymbol{O}_2(\tilde{\boldsymbol{h}})\mathrm{E}_{q_c(\boldsymbol{f})}[\boldsymbol{f}], \boldsymbol{O}_3(\tilde{\boldsymbol{h}})\mathrm{E}_{q_c(\boldsymbol{f})}[\boldsymbol{f}]]$, $\boldsymbol{Q}$ and $\boldsymbol{\Phi}$ denote 3×1 vectors with components $\boldsymbol{\Phi}_i = \mathrm{tr}[(\tilde{\boldsymbol{h}})^{\mathrm{T}}\boldsymbol{O}_i(\tilde{\boldsymbol{h}})\mathrm{cov}_{q_c(\boldsymbol{f})}]$ and $\boldsymbol{Q}_i = (\boldsymbol{g} - \tilde{\boldsymbol{h}}\mathrm{E}_{q_c(\boldsymbol{f})}[\boldsymbol{f}])^{\mathrm{T}}\boldsymbol{\Upsilon}_i$, respectively, for $i = 1, 2, 3$, and $\boldsymbol{\Psi}$ and $\boldsymbol{\Gamma}$ denote 3×3 matrices with components $\boldsymbol{\Psi}_{ij} = \mathrm{tr}[\boldsymbol{O}_i(\tilde{\boldsymbol{h}})^{\mathrm{T}}\boldsymbol{O}_j(\tilde{\boldsymbol{h}})\mathrm{cov}_{q_c(\boldsymbol{f})}[\boldsymbol{f}]]$, and $\boldsymbol{\Gamma}_{ij} = \boldsymbol{\Upsilon}_i^{\mathrm{T}}\boldsymbol{\Upsilon}_j$, respectively, for $i, j = 1, 2, 3$.

(ii) Estimation of the hyperparameter distributions

Finally, the distributions of the hyperparameters α_i and β are obtained as Gamma distributions. Using (26), the β is expressed as

$$q(\beta) \propto \beta^{\frac{N}{2}-1+a_\beta^0}\cdot$$

$$\exp\left[-\beta\left(b_\beta^0 + \frac{\mathrm{E}_{q_c(f)}[\|\boldsymbol{g} - \boldsymbol{Hf}\|^2]}{2}\right)\right]. \tag{41}$$

Using (24) we obtain the distribution for α_1, expressed as

$$q(\alpha_1) \propto \alpha_1^{\frac{PN}{2}-1+a_{\alpha_1}^0}\cdot$$

$$\exp\left[-\alpha_1\left(b_{\alpha_1}^0 + \frac{\mathrm{E}_{q_c(\boldsymbol{f})}[\|\mathbb{C}\boldsymbol{f}\|^2]}{2}\right)\right] \tag{42}$$

and from (25) we obtain

$$q(\alpha_2^h) \propto (\alpha_2^h)^{\frac{PN}{4}-1+a_{\alpha_2^h}^0}\cdot$$

$$\exp\left[-\alpha_2^h\left(b_{\alpha_2^h}^0 + \sum_{i=1}^{PN}\sqrt{w_{2i}^h}\right)\right], \tag{43}$$

$$q(\alpha_2^v) \propto (\alpha_2^v)^{\frac{PN}{4}-1+a_{\alpha_2^v}^0}\cdot$$

$$\exp\left[-\alpha_2^v\left(b_{\alpha_2^v}^0 + \sum_{i=1}^{PN}\sqrt{w_{2i}^v}\right)\right], \tag{44}$$

for α_2, and for α_3

$$q(\alpha_3) \propto (\alpha_3)^{\frac{PN}{2}-1+a_{\alpha_3}^0}\cdot$$

$$\exp\left[-\alpha_3\left(b_{\alpha_3}^0 + \sum_{i=1}^{PN}\sqrt{w_{3i}}\right)\right]. \tag{45}$$

The proposed iterative algorithm is summarized below in Algorithm 1 and its flowchart is given in Fig. 1.

Algorithm 1 Variational Bayesian blind deconvolution using SAR and $l1$ or TV model combinations.

Calculate initial estimates of original image, motion model and hyperparameters.

while convergence criterion is not met **do**

(i) Estimate the image distribution using (26).

(ii) Calculate the vector w_2 using (36), for $c = l1$, or w_3 using (37) for $c =$ TV.

(iii) Estimate the motion model using (39).

(iv) Estimate the distributions of the hyperparameters β using (41), α_1 using (42), and α_2 using (43) and (44), for $c = l1$, or α_3 using (45), for $c =$ TV.

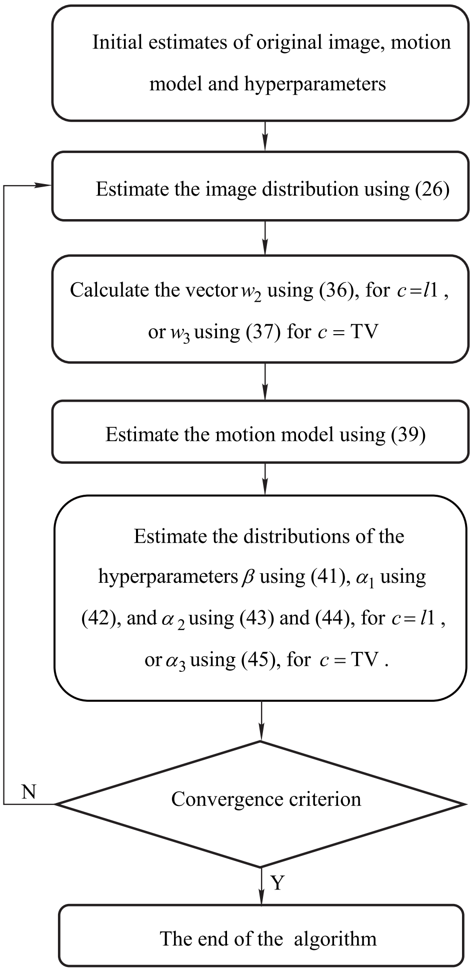


Fig. 1 Flowchart of Algorithm 1

4. Experimental results

A number of experiments will be conducted with the method proposed in this section, using synthetically degraded images and sequential astronomical images. Our goal is to evaluate the performance of the different algorithms mentioned in this paper and their original image restoration quality in the application of astronomy, particularly when studying the effects of model combination in Algorithm 1 (ALG1). In synthetic experiments, the obtained results for the restored original image are compared quantitatively in terms of the peak signal-to-noise ratio (PSNR), which is defined as

$$\mathrm{PSNR} = 10\lg\frac{NP}{\|\hat{\boldsymbol{f}}-\boldsymbol{f}\|^2} \tag{46}$$

where $\hat{\boldsymbol{f}}$ and $\boldsymbol{f}$ denote the estimated and original images, respectively, and pixel values in both images are taken in the range $[0,1]$.

In all experiments reported below, the initial values of ALG1 are chosen as follows: the initial parameters in the iterative motion estimation process based on (39) and (40) are estimated from the observed images using the standard Lucas-Kanade method [31], and the stopping criterion $\|\boldsymbol{g}-\boldsymbol{F}\boldsymbol{h}^i\|^2 > \|\boldsymbol{g}-\boldsymbol{F}\boldsymbol{h}^{i-1}\|^2$ has been used, where $\boldsymbol{h}^i$ and $\boldsymbol{h}^{i-1}$ are the motion parameters estimated at the ith and $(i-1)$th iterations, respectively. The maximum number of iterations is set to 100. The original image estimate $\boldsymbol{f}^0$ is then initialized using bicubic interpolation of the first observation. The remaining parameters have been initialized to the following values: $a^o_{\alpha_l} = b^o_{\alpha_l} = 0$, $a^o_\beta = b^o_\beta = 0$, $w^h_{2i} = (\Delta^h_i(\boldsymbol{f}^0))^2$, $w^v_{2i} = (\Delta^v_i(\boldsymbol{f}^0))^2$, $w_{3i} = (\Delta^h_i(\boldsymbol{f}^0))^2 + (\Delta^v_i(\boldsymbol{f}^0))^2$, $\alpha^h_2 = PN/\left(2\sum_i^{PN}\sqrt{w^h_{2i}}\right)$, $\alpha^v_2 = PN/\left(2\sum_i^{PN}\sqrt{w^v_{2i}}\right)$, $\alpha_3 = PN/\left(2\sum_i^{PN}\sqrt{w_{3i}}\right)$, $\beta = N/\|\boldsymbol{g}-\boldsymbol{H}\boldsymbol{f}^0\|^2$.

Many parameters in Steps 2 – 4 of Algorithm 1 are estimated using traces of different matrix products associated with the covariance matrix $\mathrm{cov}_{q_c(\boldsymbol{f})}[\boldsymbol{f}]$. The Jacobi approximation will be applied here since this covariance matrix has no exact form, and it can provide the best balance between precision and efficiency. The convergence criterion in this paper adopts $\|\boldsymbol{f}^n-\boldsymbol{f}^{n-1}\|^2/\|\boldsymbol{f}^{n-1}\|^2 < 10^{-5}$, where $\boldsymbol{f}^n$ and $\boldsymbol{f}^{n-1}$ are the image estimates at the nth and $(n-1)$th iterations, respectively.

All the unknowns of our problem in the proposed ALG1 are automatically calculated except for the λ parameter, which determines the different contribution of the prior model combination we use. Two combinations denoted by $l1$-SAR and TV-SAR are considered in the experiments, which correspond to the combination of the SAR and one of $l1$ norm and TV based model priors. For $l1$-SAR with $\lambda_{l1} = \{1-\lambda_2, \lambda_2, 0\}$, when the parameter λ_2 is set equal to zero, the role of the $l1$ norm vanishes, and the SAR model becomes solely responsible for the restoration, while the combination for $\lambda_2 = 1$ creats a completely opposite situation in which our model becomes the $l1$ norm. For TV-SAR with $\lambda_{\mathrm{TV}} = \{1-\lambda_3, 0, \lambda_3\}$, we obtain TV from the degenerate combination when $\lambda_3 = 1$. When $\lambda_3 = 0$ we obtain again the SAR model.

In this section, the results obtained using the following methods are compared: bicubic interpolation (denoted by BBC), iterated back-projection in [32] (denoted by IBP), which is based on the back-projection reconstruction method, the robust super-resolution method in [33] (denoted by RSR), which is based on a robust median estimator, projection onto convex sets (denoted by POCS) and the methods in this paper (denoted by SAR, $l1$, TV, TV-SAR and $l1$-SAR).

We deal with synthetic images in our first experiment, in which the set of three images of 128×128 pixels of size, depicted in Fig. 2, have been warped, blured and down-sampled by a factor $\sqrt{p} = 2$ to generate three sequences of six low-resolution images, then additive white Gaussian noise at signal to noise ratio (SNR) levels of 10 dB, 20 dB, 30 dB and 40 dB is added. For the warping, consisting of both translation and rotation, the following translation vectors have been chosen as

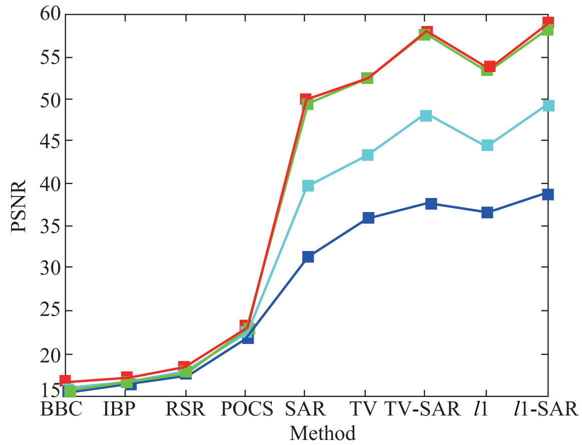
$$\begin{pmatrix} 0.0 \\ 0.0 \end{pmatrix}, \begin{pmatrix} 0.0 \\ 0.5 \end{pmatrix}, \begin{pmatrix} 0.5 \\ 0.0 \end{pmatrix}, \begin{pmatrix} 1.0 \\ 0.0 \end{pmatrix}, \begin{pmatrix} 0.0 \\ 1.0 \end{pmatrix}$$

and the rotation angles are $(0°, 3°, -3°, 5°, -5°)$, respectively. We simulate each SNR level with 20 different noise realizations, and mean PSNR values and standard deviations of these experiments are reported. For IBP, RSR and POCS algorithms, they employ a great deal of unknown algorithmic parameters that need to be tuned or supervised, and the maximum PSNR value can be obtained by searching for the algorithmic parameters exhaustively during their iterations. This tuning process will be troublesome and time-consuming. For ALG1, there is no need for parameter tuning and user supervision since all required algorithmic parameters of ALG1 are estimated automatically.

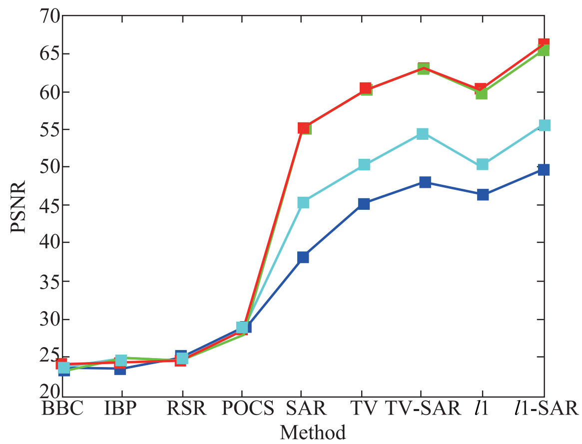
(a) Synthetic image-1 (b) Synthetic image-2 (c) Synthetic image-3

Fig. 2 Set of 128×128 images used in the synthetic experiments

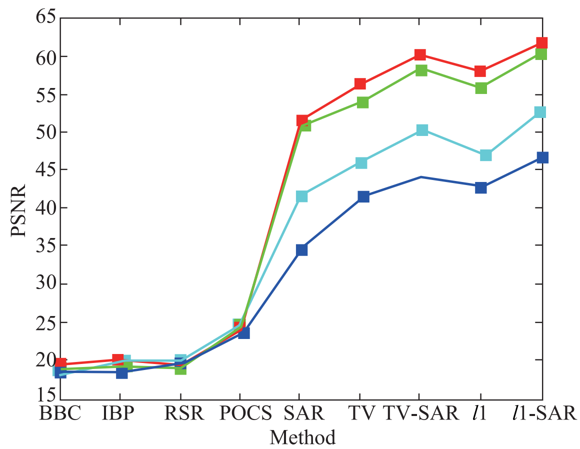
Table 1 shows the mean PSNR values with standard deviations of the restored images corresponding to the different algorithms, and Fig. 3 shows the plots of the mean PSNR values. The proposed ALG1 for SAR, $l1$, TV, TV-SAR and $l1$-SAR models result in better reconstruction performance in terms of PSNR than the other ones. It is also clear that all algorithms behave better than bicubic interpolation. Regarding the difference in performance of each model separately, $l1$ norm and TV based models give better results than the SAR model, $l1$ provides slightly higher quality than TV. The two proposed model combinations TV-SAR and $l1$-SAR obtain similar PSNR values and give better results than SAR, $l1$ and TV alone. The λ_2 and λ_3 values need to be exhaustively searched between the value space $[0, 1]$, with a precision of 0.05, to maximize the PSNR values for the TV-SAR and $l1$-SAR model combinations. The adopted λ_2 and λ_3 values are also shown in Table 1. Details for estimated original images are shown in Fig. 4, Fig. 5, and Fig. 6, for the SNR=10 dB case. Obviously ALG1 provides the most visually enhanced restorations with significantly reduced ringing artifacts and much sharper edges compared to the other methods.



(a) Plots of synthetic image-1



(b) Plots of synthetic image-2



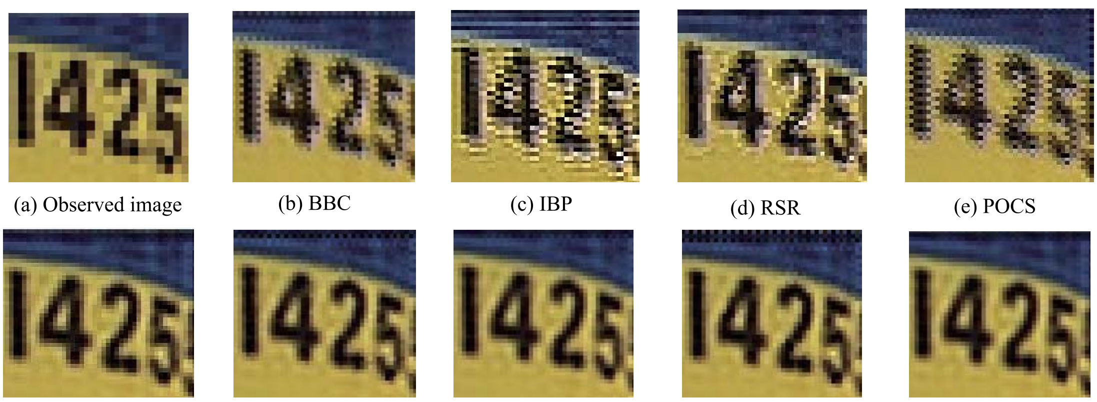
(c) Plots of synthetic image-3

: 10 dB; : 20 dB; : 30 dB; : 40 dB.

Fig. 3 Plots of the mean PSNR values corresponding to the different methods and noise levels, for the images in Fig. 1

Table 1 Mean PSNR values and standard deviations of the estimated original images in Fig. 1 at different SNR levels, along with λ_2 and λ_3 values obtained from the exhaustive search PSNR maximization

SNR	Method	a		b		c	
		PSNR	λ_2/λ_3	PSNR	λ_2/λ_3	PSNR	λ_2/λ_3
10 dB	BBC	15.46+0.05		23.33+0.03		18.15+0.02	
	IBP	16.22+0.08		23.26+0.03		18.24+0.01	
	RSR	17.11+0.13		24.73+0.09		19.49+0.07	
	POCS	21.43+0.02		28.49+0.06		23.34+0.07	
	SAR	31.18+0.06		38.23+0.14		34.44+0.17	
	TV	35.78+0.11		45.01+0.12		41.59+0.11	
	TV-SAR	37.52+0.03	0.90	47.92+0.07	0.90	44.01+0.08	0.90
	$l1$	36.78+0.07		46.30+0.07		42.94+0.09	
	$l1$-SAR	38.78+0.14	0.75	49.48+0.06	0.80	46.51+0.04	0.90
20 dB	BBC	15.99+0.04		23.66+0.03		18.48+0.06	
	IBP	16.43+0.13		24.17+0.14		19.72+0.14	
	RSR	17.39+0.21		24.25+0.12		19.72+0.15	
	POCS	22.12+0.09		28.72+0.09		24.24+0.06	
	SAR	39.51+0.17		45.39+0.03		41.48+0.03	
	TV	43.41+0.14		49.99+0.04		45.83+0.03	
	TV-SAR	48.27+0.21	0.80	54.58+0.16	0.75	50.37+0.02	0.80
	$l1$	44.10+0.07		50.22+0.12		46.66+0.18	
	$l1$-SAR	49.33+0.05	0.80	55.35+0.07	0.75	52.50+0.12	0.75
30 dB	BBC	16.03+0.11		23.95+0.06		19.07+0.11	
	IBP	16.49+0.02		24.40+0.04		19.00+0.07	
	RSR	17.44+0.05		24.67+0.04		19.12+0.06	
	POCS	22.39+0.15		27.62+0.07		23.92+0.04	
	SAR	49.24+0.11		55.28+0.08		50.96+0.04	
	TV	52.29+0.08		60.02+0.17		54.30+0.06	
	TV-SAR	57.73+0.03	0.75	63.39+0.13	0.35	58.45+0.21	0.15
	$l1$	53.01+0.14		59.58+0.11		55.87+0.11	
	$l1$-SAR	58.22+0.17	0.60	65.06+0.06	0.40	60.34+0.04	0.15
40 dB	BBC	16.44+0.10		24.03+0.21		19.36+0.15	
	IBP	16.78+0.17		24.10+0.17		19.91+0.04	
	RSR	17.94+0.06		24.32+0.08		19.32+0.07	
	POCS	22.71+0.02		28.22+0.09		24.02+0.11	
	SAR	49.74+0.02		55.02+0.09		51.44+0.12	
	TV	52.37+0.11		59.93+0.03		56.37+0.04	
	TV-SAR	57.90+0.06	0.55	63.07+0.11	0.10	59.89+0.03	0.05
	$l1$	53.44+0.07		60.03+0.06		57.98+0.07	
	$l1$-SAR	59.01+0.07	0.45	66.10+0.05	0.10	61.88+0.07	0.05



(a) Observed image (b) BBC (c) IBP (d) RSR (e) POCS

(f) SAR (g) TV (h) TV-SAR with λ_2=0.90 (i) $l1$ (j) $l1$-SAR with λ_3=0.75

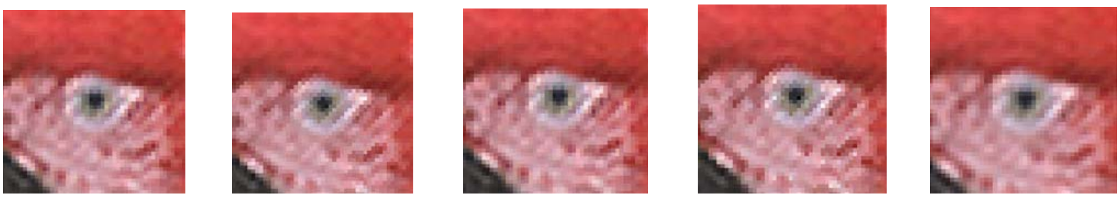
Fig. 4 Details for results of estimated original image-1 from different methods in the case when SNR=10 dB

(a) Observed image (b) BBC (c) IBP (d) RSR (e) POCS

(f) SAR (g) TV (h) TV-SAR with λ_2=0.90 (i) $l1$ (j) $l1$-SAR with λ_3=0.80

Fig. 5 **Details for results of estimated original image-2 from different methods in the case when SNR=10 dB**

(a) Observed image (b) BBC (c) IBP (d) RSR (e) POCS



(f) SAR (g) TV (h) TV-SAR with λ_2=0.90 (i) $l1$ (j) $l1$-SAR with λ_3=0.90

Fig. 6 **Details for results of estimated original image-3 from different methods in the case when SNR=10 dB**

Next we study the performance of the proposed ALG1 on real-image datasets. Two data sets will been used here, the set of real astronomical images taken from Canon EOS 70D camera and TW-CMV20K camera (see Fig. 7), corresponding to the different methods in this paper have been applied to image restoration in astronomy by a factor $P = 2$. Fig. 8 and Fig. 9 show the estimated original images obtained using BBC, IBP, RSR and POCS algorithms and the proposed variational Bayesian blind deconvolution, for the 20 observed images of the moon. The superior quality of the restoration obtained by our proposed methods is evident, with sharper edges and fewer ringing artifacts than the rest of methods.

(a) Canon EOS 70D camera (b) TW-CMV20K camera

Fig. 7 **Set of real astronomical images taken from different cameras**

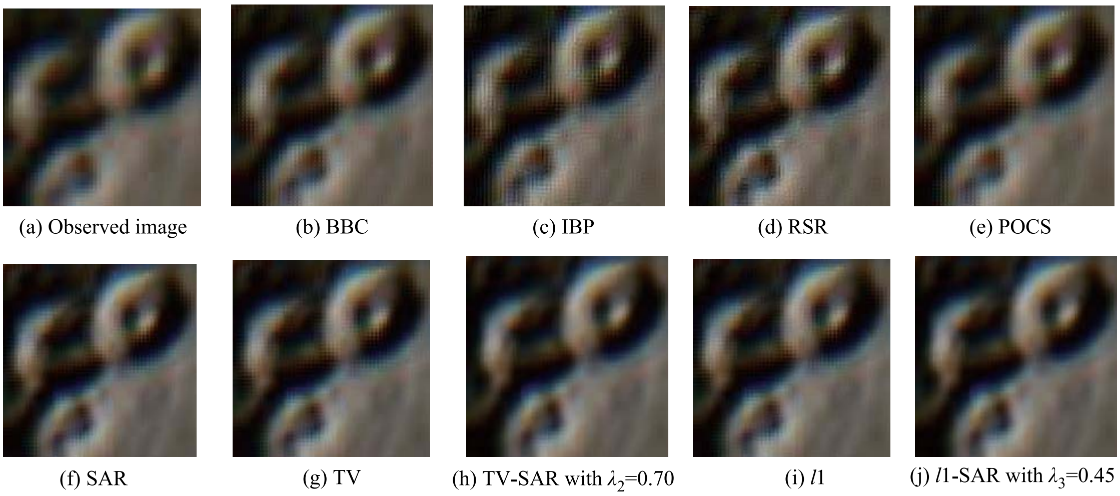


Fig. 8 Estimated original images from 20 observed observations captured with Canon EOS 70D camera

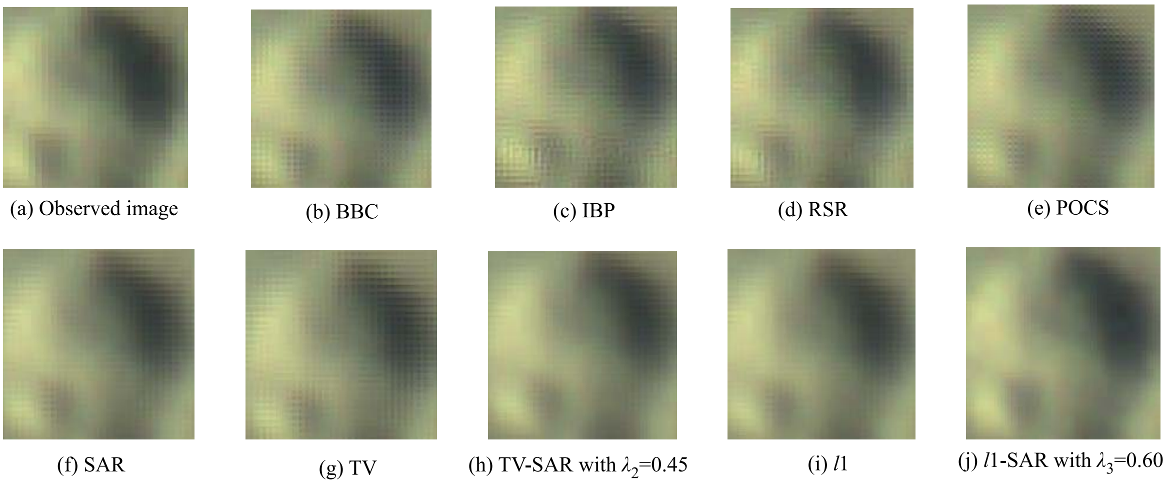


Fig. 9 Estimated original images from 20 observed observations captured with TW-CMV20K camera

5. Conclusions

In this paper, a novel variational Bayesian blind deconvolution methodology for the combination of sparse and non-sparse image priors, has been applied to the image restoration in astronomy from rotated and displaced observed astronomical images. The blind deconvolution problem is formulated using a hierarchical Bayesian paradigm, and Bayesian inference utilized finds the original image given the observations by decreasing the KL divergences associated with each pair of observation and prior models. Based on the presented experimental results, the original image estimates obtained by the proposed method compare favorably with the images provided by other state of the art reconstruction methods.

References

[1] M. R. Banham, A. K. Katsaggelos. Digital image restoration. *IEEE Signal Processing Magazine*, 1997, 14(2): 24 – 41.

[2] R. Molina, J. Nunez, F. J. Cortijo, et al. Image restoration in astronomy: a Bayesian perspective. *IEEE Signal Processing Magazine*, 2001, 18(2): 11 – 29.

[3] J. L. Starck, E. Pantin, F. Murtagh. Deconvolution in astronomy: a review. *Publications of the Astronomical Society of the Pacific*, 2002, 114(800): 1051 – 1069.

[4] J. L. Starck, F. Murtagh. *Astronomical image and data analysis*. New York: Springer-Verlag, 2006.

[5] J. Bobin, J. L. Starck, R. Ottensamer. Compressed sensing in astronomy. *IEEE Journal of Selected Topics in Signal Processing*, 2008, 2(5): 718 – 726.

[6] J. L. Starck, J. Bobin. Astronomical data analysis and sparsity: from wavelets to compressed sensing. *Proceedings of the IEEE*, 2010, 98(6): 1021 – 1030.

[7] J. L. Starck, F. Murtagh, J. M. Fadili. *Sparse image and signal processing: wavelets, curvelets, morphological diversity*. Cambridge, UK: Cambridge University Press, 2010.

[8] R. Fergus, B. Singh, A. Hertzmann, et al. Removing camera shake from a single photograph. *ACM Trans. on Graphics*, 2006, 25(3): 87 – 794.

[9] T. F. Chan, C. K. Wong. Total variation blind deconvolution.

IEEE Trans. on Image Processing, 1998, 7(3): 370 – 375.

[10] J. H. Money, S. H. Kang. Total variation minimizing blind deconvolution with shock filter reference. *Image and Vision Computing*, 2008, 26(2): 302 – 314.

[11] R. Molina, J. Mateos, A. K. Katsaggelos. Blind deconvolution using a variational approach to parameter, image, and blur estimation. *IEEE Trans. on Image Processing*, 2006, 15(12): 3715 – 3727.

[12] S. Kullback. *Information theory and statistics*. New York: Dover, 1997.

[13] A. C. Likas, N. P. Galatsanos. A variational approach for Bayesian blind image deconvolution. *IEEE Trans. on Signal Processing*, 2004, 52(8): 2222 – 2233.

[14] J. Miskin, D. J. C. MacKay. *Ensemble learning for blind image separation and deconvolution*. New York: Springer-Verlag, 2000.

[15] S. D. Babacan, R. Molina, A. K. Katsaggelos. Variational Bayesian blind deconvolution using a total variation prior. *IEEE Trans. on Image Processing*, 2009, 18(1): 12 – 26.

[16] A. K. Jain. *Fundamentals of digital image processing*. New Jersey: Prentice Hall, 1989.

[17] R. Molina, A. K. Katsaggelos, J. Mateos. Bayesian and regularization methods for hyperparameter estimation in image restoration. *IEEE Trans. on Image Processing*, 1999, 8(2): 231 – 246.

[18] R. Molina. On the hierarchical Bayesian approach to image restoration: applications to astronomical images. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 1994, 16(11): 1122 – 1128.

[19] S. Villena, M. Vega, S. D. Babacan, et al. Bayesian combination of sparse and non sparse priors in image superresolution. *Digital Signal Processing*, 2013, 23(2): 530 – 541.

[20] S. Villena, M. Vega, R. Molina, et al. Bayesian superresolution image reconstruction using an $l1$ prior. *Proc. of the 6th International Symposium on Image and Signal Processing and Analysis*, 2009: 152 – 157.

[21] S. D. Babacan, R. Molina, A. Katsaggelos. Variational Bayesian super resolution. *IEEE Trans. on Image Processing*, 2011, 20(4): 984 – 999.

[22] B. D. Lucas, T. Kanade. An iterative image registration technique with an application to stereo vision. *Proc. of the 7th International Joint Conference on Artificial Intelligence*, 1981: 121 – 130.

[23] P. Vandewalle, S. Susstrunk, M. Vetterli. A frequency domain approach to registration of aliased images with application to superresolution. *EURASIP Journal on Applied Signal Processing*, 2006, 2006(1): 1 – 14.

[24] M. E. Tipping, C. M. Bishop. Bayesian image superresolution. *Proc. of Advances in Neural Information Processing Systems*, 2003: 1303 – 1310.

[25] L. C. Pickup, D. P. Capel, S. J. Roberts, et al. Bayesian methods for image super-resolution. *The Computer Journal*, 2009, 52(1): 101 – 113.

[26] A. D. Dempster, N. M. Laird, D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, 1977, 39(1): 1 – 37.

[27] C. Bishop. *Pattern recognition and machine learning*. New York: Springer, 2006.

[28] M. J. Beal. Variational algorithms for approximate Bayesian inference. London, UK: University College London, 2003.

[29] J. Miskin. Ensemble learning for independent component analysis. Cambridge, UK: University of Cambridge, 2000.

[30] M. A. T. Figueiredo, J. M. Bioucas-Dias, R. D. Nowak. Majorization-minimization algorithms for wavelet-based image restoration. *IEEE Trans. on Image Processing*, 2007, 16(12): 2980 – 2991.

[31] B. D. Lucas, T. Kanade. An iterative image registration technique with an application to stereo vision. *Proc. of the Imaging Understanding Workshop*, 1981: 121 – 130.

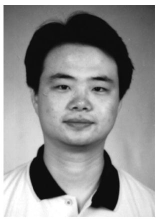
[32] M. Irani, S. Peleg. Improving resolution by image registration. *Graphical Models and Image Processing*, 1991, 53(3): 231 – 239.

[33] A. Zomet, A. Rav-Acha, S. Peleg. Robust super-resolution. *Proc. of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2001: 645 – 650.

Biographies

Xiaoping Shi was born in 1965. He received his Bachelor Physics degree from Harbin Institute of Technology in 1988 and Ph.D. degree from Harbin Institute of Technology in 1994. He became a Ph.D. supervisor of control science and engineering in 2003, and is currently a professor at the Control and Simulation Center, in Harbin Institute of Technology. His areas of research interest are image processing, computer vision, spacecraft intelligent control and complex system simulation.
E-mail: sxp@hit.edu.cn

Rui Guo was born in 1989. He received his Bachelor Physics degree from Northeast Petroleum University in 2012 and master degree from Northeast Petroleum University in 2015. He is currently a Ph.D. student of control science and engineering at the Control and Simulation Center, in Harbin Institute of Technology. His research interests focus on image processing (super resolution image reconstruction), and computer vision.
E-mail: ailahm@126.com



Yi Zhu was born in 1972. He received his Bachelor Physics degree from Harbin Institute of Technology in 1994 and Ph.D. degree from Harbin Institute of Technology in 2006. He is currently an associate professor of control science and engineering at the Control and Simulation Center, in Harbin Institute of Technology. His research interests are image restoration and computer vision.
E-mail: zhuyi@hit.edu.cn

Zicai Wang was born in 1932. He graduated from Harbin Institute of Technology in 1957 and is currently a professor of the Control and Simulation Center in Harbin Institute of Technology. He became an Academician of Chinese Academy of Engineering in 2001 and is vice president of China Institute of System Simulation. His research interests are system simulation and fuzzy and neural network theory.
E-mail: wzc@hit.edu.cn