

Bayesian Combination of Sparse and non Sparse Priors in Image Super Resolution[☆]

S. Villena^{a,*}, M. Vega^a, S. D. Babacan^b, R. Molina^c, A.K. Katsaggelos^d

^a*Dept. de Lenguajes y Sistemas Informáticos, Universidad de Granada, 18071 Granada, Spain*

^b*Beckman Institute University of Illinois at Urbana-Champaign. USA*

^c*Dept. de Ciencias de la Computación e I. A. Universidad de Granada, 18071 Granada, Spain*

^d*Dept. of Electrical Engineering and Computer Science, Northwestern University, Evanston, Illinois 60208-3118*

Abstract

In this paper the application of image prior combinations to the Bayesian *Super Resolution (SR)* image registration and reconstruction problem is studied. Two sparse image priors, a *Total Variation (TV)* prior and a prior based on the ℓ_1 norm of horizontal and vertical *first order differences (f.o.d.)*, are combined with a non-sparse *Simultaneous Auto Regressive (SAR)* prior. Since, for a given observation model, each prior produces a different posterior distribution of the underlying *High Resolution (HR)* image, the use of variational approximation will produce as many posterior approximations as priors we want to combine. A unique approximation is obtained here by finding the distribution on the HR image given the observations that minimizes a linear convex combination of *Kullback-Leibler (KL)* divergences. We find this distribution in closed form. The estimated HR images are compared with the ones obtained by other SR reconstruction methods.

Keywords: Super Resolution, total variation, variational methods, parameter estimation, Bayesian methods.

1. Introduction

Image SR is an active research field that studies the process of obtaining an HR image from a set of degraded *Low Resolution (LR)* images (see [1, 2] for

[☆]This work was supported in part by the Comisión Nacional de Ciencia y Tecnología under contract TIN2010-15137 and the Spanish research programme Consolider Ingenio 2010: MIPRCV (CSD2007-00018).

*Corresponding author

Email addresses: svillena@ugr.es (S. Villena), mvega@ugr.es (M. Vega), dbabacan@illinois.edu (S. D. Babacan), rms@decsai.ugr.es (R. Molina), aggk@eecs.northwestern.edu (A.K. Katsaggelos)

a review). The basic principle in SR is that changes in LR images caused by the blur and the camera (and/or scene) motion provide additional information that can be utilized to reconstruct the HR image. Usually SR methods include two parts: registration, where the motion between LR images is estimated, and image reconstruction, where the HR image is recovered from the LR images.

In the Bayesian framework a prior model on the HR image to be reconstructed is introduced, its aim is to encapsulate our prior image knowledge and consequently to avoid the ill-posedness of the image reconstruction problem. The selection of this Bayesian prior model is a critical issue. Prior models imposing image smoothness, like the *Conditional Auto Regressive (CAR)* or SAR image models (see [3]), are known to oversmooth edge regions. More sophisticated priors based on wavelets [4], TV [5], or the ℓ_1 norm of horizontal and vertical f.o.d. [6], have been proposed. However, these priors lead to over-smooth non edge regions.

Combining image priors is an interesting way to take advantage of their behavior. While in image restoration there have been several attempts to combine image priors [7–9], no such attempts have been made in the SR literature apart from our conference paper [10], from which the present paper grows. In [9] a Student’s *t Product of Experts (PoE)* image prior model was proposed and learnt from the observations. In [8] the PoE prior was learnt using a large training set of images and also stochastic sampling methods. A combination of the TV image prior model and the PoE model of [9] has been recently proposed in [11]. This method can be considered a spatially adaptive version of the TV model which furthermore, as the method in [9], has the ability to simultaneously enforce different properties on the image.

In this paper, a combination of the sparse TV and ℓ_1 , and the non sparse SAR image prior models is applied to SR. The rationale of this modelling is to benefit from the ability of the sparse priors to recover image edges, and at the same time to avoid their tendency to oversmooth inner regions by combining them with an smoothness promoting prior model.

Accurate registration of displaced and rotated images, is vital in SR image reconstruction. There are two major approaches to registration in SR, which differ in the stage where registration is performed. In the first approach the motion parameters are previously estimated from the observed LR images, in a preprocessing step, and then used in a separate image estimation process (see [12–15]). The limited accuracy inherent to HR registration from LR images is a shortcoming of this first approach. The second approach is to alternate between HR image registration and HR image estimation (see [5, 16–25]).

All the HR reconstruction methods proposed in this work depend on model parameters, usually called hyperparameters. Hyperparameter estimation is also a critical issue, which has been studied in previous SR papers, see for instance [5, 6, 10, 21, 25]. In this paper the entire SR problem, that is HR image reconstruction, registration, and hyperparameter estimation, is approached from a Bayesian perspective. All the unknown, i.e., the HR image and the displacements and rotations, are systematically included in a hierarchical Bayesian model. Using a variational Bayesian analysis, sparse and no sparse HR image

prior models can be combined. The proposed framework provides uncertainties of the estimates during the restoration process, which helps to prevent error-propagation and improves robustness. All required algorithmic parameters are estimated along with the HR image and the motion parameters, and therefore the proposed algorithms do not require user supervision.

The rest of this paper is organized as follows. Section 2 provides the mathematical model for the LR image acquisition process. We provide the description of the hierarchical Bayesian framework modeling the unknowns in Section 3. The inference procedure to develop the proposed methods is presented in Section 4. We demonstrate the effectiveness of the proposed methods with experimental results in Section 5 and conclusions are drawn in Section 6.

2. Problem Formulation

We assume that the imaging process generates L LR images \mathbf{y}_k , $k = 1, \dots, L$, from the HR image \mathbf{x} . The LR images \mathbf{y}_k consist of $N = N_h \times N_v$ pixels (where N_h and N_v are the observations pixels number in horizontal and vertical, respectively) and the HR image \mathbf{x} of PN pixels, where the integer $P > 1$ is the factor of increase in resolution. In this paper we adopt the matrix-vector notation such that the images \mathbf{y}_k and \mathbf{x} are arranged as $N \times 1$ and $PN \times 1$ vectors, respectively. The imaging process introduces shifting, blurring and downsampling, which is modeled as

$$\mathbf{y}_k = \mathbf{A}\mathbf{H}_k\mathbf{C}(\mathbf{s}_k)\mathbf{x} + \mathbf{n}_k = \mathbf{B}_k(\mathbf{s}_k)\mathbf{x} + \mathbf{n}_k, \quad (1)$$

with the system $N \times PN$ matrix $\mathbf{B}_k = \mathbf{A}\mathbf{H}_k\mathbf{C}(\mathbf{s}_k)$, where \mathbf{A} is the $N \times PN$ downsampling matrix, \mathbf{H}_k is the $PN \times PN$ blurring matrix, $\mathbf{C}(\mathbf{s}_k)$ is the $PN \times PN$ warping matrix generated by the motion vector \mathbf{s}_k , and \mathbf{n}_k is the $N \times 1$ acquisition noise. Note that the matrices \mathbf{H}_k and $\mathbf{C}(\mathbf{s}_k)$ and the noise \mathbf{n}_k can be different for each LR image \mathbf{y}_k .

In this work, we assume that the blurring matrices \mathbf{H}_k are known and we consider a motion model consisting of translational and rotational motion, so that $\mathbf{s}_k = (\theta_k, c_k, d_k)^t$, where θ_k is the rotation angle, and c_k and d_k are the horizontal and vertical translations of the k^{th} HR image with respect to the reference frame \mathbf{x} .

The effects of downsampling, blurring, and warping are combined into the system matrix $\mathbf{B}_k(\mathbf{s}_k)$, from which each row maps the pixels of the HR image \mathbf{x} to a given pixel in the LR image \mathbf{y}_k . Given Eq. (1), the SR problem can be expressed as the search of an estimate of the HR image \mathbf{x} from the set of LR images $\{\mathbf{y}_k\}$ using our prior knowledge about $\{\mathbf{C}(\mathbf{s}_k)\}$, $\{\mathbf{n}_k\}$, and \mathbf{x} .

3. Hierarchical Bayesian Model

In this work, we adopt a hierarchical Bayesian framework consisting of two stages. The first stage is used to model the acquisition process, the unknown HR image \mathbf{x} and the motion vectors $\{\mathbf{s}_k\}$. For the unknown \mathbf{x} we have m models

which we want to combine. They are denoted by $p_i(\mathbf{x}|\alpha_i)$ for $i = 1, \dots, m$. Prior distributions $p(\mathbf{s}_k)$ are assigned to the unknowns \mathbf{s}_k , for $k = 1, \dots, L$. The observation $\mathbf{y} = \{\mathbf{y}_k\}$ is also a random process with the corresponding *conditional* distribution $p(\mathbf{y}|\mathbf{x}, \{\mathbf{s}_k\}, \{\beta_k\})$. These distributions depend on additional parameters α_i and $\{\beta_k\}$ (called *hyperparameters*), which are modeled by assigning *hyperprior* distributions in the second stage of the hierarchical model.

In the following subsections we provide the description of the individual distributions used to model the unknowns.

3.1. Observation Model

Using the model in Eq. (1) and assuming that \mathbf{n}_k is zero-mean white Gaussian noise with inverse variance (precision) β_k , the conditional distribution of the LR image \mathbf{y}_k is given by

$$p(\mathbf{y}_k|\mathbf{x}, \mathbf{s}_k, \beta_k) \propto \beta_k^{N/2} \exp \left[-\frac{\beta_k}{2} \|\mathbf{y}_k - \mathbf{B}_k(\mathbf{s}_k)\mathbf{x}\|^2 \right]. \quad (2)$$

Assuming statistical independence of the noise among the LR image acquisitions, the conditional probability of the set of LR images \mathbf{y} given \mathbf{x} can be expressed as

$$p(\mathbf{y}|\mathbf{x}, \{\mathbf{s}_k\}, \{\beta_k\}) = \prod_{k=1}^L p(\mathbf{y}_k|\mathbf{x}, \mathbf{s}_k, \beta_k). \quad (3)$$

The independent Gaussian model in Eq. (3) is used in most of the existing super resolution methods [18, 21, 22, 26].

Let us now explicitly state the form of the warping matrices $\mathbf{C}(\mathbf{s}_k)$ of Eq. (1). We denote the coordinates of the reference HR grid by (u, v) and the coordinates of the k^{th} warped HR grid, after applying $\mathbf{C}(\mathbf{s}_k)$ to \mathbf{x} , by (u_k, v_k) . Let us also define

$$\begin{aligned} \Delta u_k &= u_k - u = u \cos(\theta_k) - v \sin(\theta_k) + c_k - u \\ \Delta v_k &= v_k - v = u \sin(\theta_k) + v \cos(\theta_k) + d_k - v. \end{aligned}$$

Note that the coordinates (u_k, v_k) generally correspond to fractional values, and therefore the HR image value at pixel (u_k, v_k) in the k^{th} HR grid has to be calculated using resampling (see Figure 1). As in [27], we incorporate bilinear interpolation to approximate the HR image value at (u_k, v_k) using the four neighboring HR image values $x_{tl(s_k)}$, $x_{tr(s_k)}$, $x_{bl(s_k)}$, and $x_{br(s_k)}$ (see black bold remarked pixels in Figure 1), which are the pixels at the top-left, top-right, bottom-left and bottom-right locations of the pixel at (u_k, v_k) , respectively.

Let us denote by $(a_k(s_k), b_k(s_k))^T$ the vector difference between the pixel position at (u_k, v_k) and the pixel at its top-left position in the reference HR grid, that is,

$$\begin{aligned} a_k(s_k) &= \Delta u_k - \text{floor}(\Delta u_k) \\ b_k(s_k) &= \Delta v_k - \text{ceil}(\Delta v_k). \end{aligned}$$

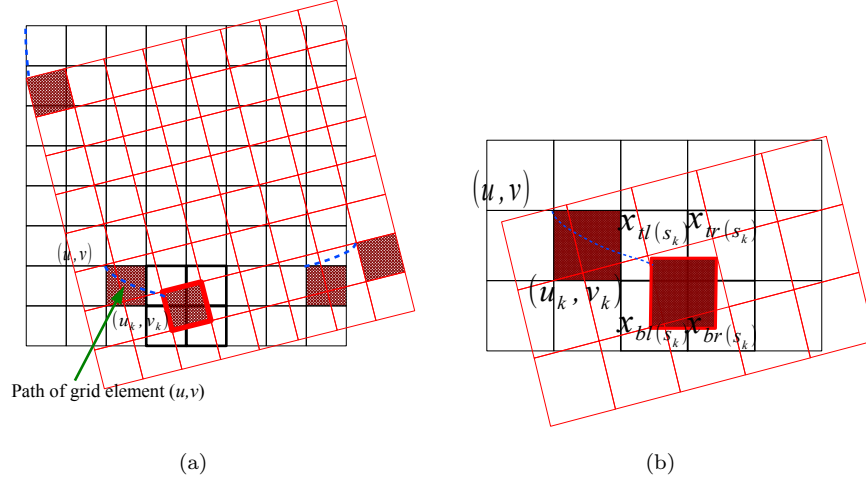


Figure 1: (a) HR image grid (in black) and the k^{th} image grid (in red). (b) Detailed view of (a), with the pixel notation used for the bilinear interpolation of grid element (u_k, v_k) .

Using bilinear interpolation, the warped image $\mathbf{C}(s_k)\mathbf{x}$ can be approximated as (see [27] for details)

$$\begin{aligned} \mathbf{C}(s_k)\mathbf{x} \approx & \mathbf{D}_{\mathbf{b}_k(s_k)}(\mathbf{I} - \mathbf{D}_{\mathbf{a}_k(s_k)})\mathbf{L}_{\mathbf{bl}(s_k)}\mathbf{x} + (\mathbf{I} - \mathbf{D}_{\mathbf{b}_k(s_k)})\mathbf{D}_{\mathbf{a}_k(s_k)}\mathbf{L}_{\mathbf{tr}(s_k)}\mathbf{x} \\ & + (\mathbf{I} - \mathbf{D}_{\mathbf{b}_k(s_k)})(\mathbf{I} - \mathbf{D}_{\mathbf{a}_k(s_k)})\mathbf{L}_{\mathbf{tl}(s_k)}\mathbf{x} + \mathbf{D}_{\mathbf{b}_k(s_k)}\mathbf{D}_{\mathbf{a}_k(s_k)}\mathbf{L}_{\mathbf{br}(s_k)}\mathbf{x}, \end{aligned} \quad (4)$$

where $\mathbf{D}_{\mathbf{a}_k(s_k)}$ and $\mathbf{D}_{\mathbf{b}_k(s_k)}$ denote diagonal matrices with the vectors $\mathbf{a}_k(s_k)$ and $\mathbf{b}_k(s_k)$ in their diagonal, respectively. The matrices $\mathbf{L}_{\mathbf{z}}$ with $\mathbf{z} \in \{\mathbf{bl}(s_k), \mathbf{br}(s_k), \mathbf{tl}(s_k), \mathbf{tr}(s_k)\}$ are constructed in such a way that the product $\mathbf{L}_{\mathbf{z}}\mathbf{x}$ produces pixels at the top-left, top-right, bottom-left and bottom-right locations of (u_k, v_k) , respectively.

3.2. Image Models

The quality of the estimated HR image as well as the accuracy in the estimates of other unknowns depends on the incorporation of accurate image models. In this paper the SAR prior model (see [3]), the prior model based in the ℓ_1 norm over horizontal and vertical f.o.d. (see [6]) and the TV prior model (see [5]) are used.

The SAR model is an smooth prior with quadratic energy which is well known to oversmooth edge regions (see [3]). It is defined by

$$p_1(\mathbf{x}|\alpha_1) \propto \alpha_1^{\frac{PN}{2}} \exp \left\{ -\frac{\alpha_1}{2} \|\mathcal{C}\mathbf{x}\|^2 \right\}, \quad (5)$$

where \mathcal{C} is the Laplacian operator and α_1 the hyperparameter of this prior model.

The ℓ_1 norm prior model is defined as

$$p_2(\mathbf{x}|\alpha_2) \propto (\alpha_2^h \alpha_2^v)^{\frac{PN}{4}} \exp \left\{ - \sum_{i=1}^{PN} [\alpha_2^h \|\Delta_i^h(\mathbf{x})\|_1 + \alpha_2^v \|\Delta_i^v(\mathbf{x})\|_1] \right\}, \quad (6)$$

where $\Delta_i^h(\mathbf{x})$ and $\Delta_i^v(\mathbf{x})$ represent the horizontal and vertical f.o.d., respectively, for the pixel i , $\alpha_2 = \{\alpha_2^h, \alpha_2^v\}$, being α_2^h and α_2^v the model hyperparameters.

The TV prior model is defined as

$$p_3(\mathbf{x}|\alpha_3) \propto \alpha_3^{\frac{PN}{2}} \exp \left\{ - \frac{\alpha_3}{2} \sum_{i=1}^{PN} \sqrt{(\Delta_i^h(\mathbf{x}))^2 + (\Delta_i^v(\mathbf{x}))^2} \right\}, \quad (7)$$

where α_3 is the hyperparameters of this prior model. Both TV, and ℓ_1 , are sparse priors, very effective in preserving edges. The main difference between TV and ℓ_1 is the presence of two hyperparameters α_2^h and α_2^v , in the ℓ_1 prior. These two parameters allow to adapt the model to possible direction dependence of the edge strength.

3.3. Modeling the Uncertainties in the Registration Parameters

Let us denote by $\bar{\mathbf{s}}_k^p$ the estimate of \mathbf{s}_k obtained from LR observations in a preprocessing step, using conventional registration algorithms, such as the one reported in [28]. As mentioned earlier, these estimates are in general inaccurate, which lowers the image restoration quality. Therefore, we model the motion parameters as stochastic variables following Gaussian distributions with *a priori* means set equal to the preliminary motion parameters $\bar{\mathbf{s}}_k^p$, that is,

$$p(\mathbf{s}_k) = \mathcal{N}(\mathbf{s}_k | \bar{\mathbf{s}}_k^p, \mathbf{\Xi}_k^p) \quad (8)$$

with $\mathbf{\Xi}_k^p$ the *a priori* covariance matrix. The parameters $\bar{\mathbf{s}}_k^p$ and $\mathbf{\Xi}_k^p$ incorporate prior knowledge about the motion parameters into the estimation procedure. If such knowledge is not available, $\bar{\mathbf{s}}_k^p$ and $(\mathbf{\Xi}_k^p)^{-1}$ can be set equal to zero, which makes the observations solely responsible for the estimation process. Similar models utilizing Gaussian distributions to model the uncertainty in preliminary motion parameters have also been used in some existing algorithms [18, 21, 26], but with different inference methods.

3.4. Hyperpriors on the Hyperparameters

The hyperparameters α_1 , α_2 , α_3 and $\{\beta_k\}$ are crucial for the performance of the SR algorithm. For their modeling, we employ Gamma distributions

$$p(\omega) = \Gamma(\omega | a_\omega^o, b_\omega^o) = \frac{(b_\omega^o)^{a_\omega^o}}{\Gamma(a_\omega^o)} \omega^{a_\omega^o-1} \exp[-b_\omega^o \omega], \quad (9)$$

where $\omega > 0$ denotes a hyperparameter, and $a_\omega^o > 0$ and $b_\omega^o > 0$ are the shape and scale parameters, respectively. The hyperpriors are chosen as Gamma distributions since they are conjugate priors for the Gaussian distribution.

Finally, combining Eqs. (3), (5), (6), (7), (8) & (9) we obtain

$$p_l(\Theta_l, \mathbf{y}) = p_l(\mathbf{x}|\alpha_l)p(\alpha_l) \prod_{k=1}^L [p(\mathbf{y}_k|\mathbf{x}, \mathbf{s}_k, \beta_k)p(\beta_k)p(\mathbf{s}_k)], \quad (10)$$

which for $l \in \{1, 2, 3\}$, denotes the different joint probability distributions corresponding to the SAR, ℓ_1 and TV prior models respectively. In Eq. (10) $\Theta_l = \{\Omega, \alpha_l\}$, $\Omega = \{\mathbf{x}, \{\mathbf{s}_k\}, \{\beta_k\}\}$, $p(\alpha_2) = p(\alpha_2^h)p(\alpha_2^v)$ and the conditional probability for the set of L LR observations, defined in Eq. (3), has been used.

4. Variational Bayesian Inference

Let us denote the set of all unknowns by $\Phi = \{\Omega, \{\alpha_l\}\}$. In this paper, Bayesian inference is based on the posterior distribution $p(\Phi|\mathbf{y})$ of Φ given the observed \mathbf{y} . We propose here to approximate this posterior distribution by the distribution minimizing the following linear convex combination of m KL divergence measures

$$\hat{q}(\Phi) = \underset{q(\Phi)}{\operatorname{argmin}} \sum_{l=1}^m \lambda_l C_{KL}(q(\Omega)q(\alpha_l)||p_l(\Theta_l|\mathbf{y})), \quad (11)$$

where m is the number of image priors to be combined in our model, $\lambda_l \geq 0$ for $l = 1, \dots, m$ and $\sum_{l=1}^m \lambda_l = 1$. In Eq. (11) the posteriors $p(\Theta_l|\mathbf{y})$, for $l = 1, \dots, m$ are given by

$$p(\Theta_l|\mathbf{y}) = \frac{p(\Theta_l, \mathbf{y})}{p(\mathbf{y})}, \quad (12)$$

where the joint distribution $p(\Theta_l, \mathbf{y})$, corresponding to the different prior models, has been defined in Eq. (10), $q(\Phi) = q(\Omega) \prod_{l=1}^m q(\alpha_l)$, $q(\Omega) = q(\mathbf{x}) \prod_{k=1}^L q(\beta_k)$, and the KL divergences are defined as

$$C_{KL}(q(\Omega)q(\alpha_l)||p_l(\Theta_l|\mathbf{y})) = \int q(\Omega)q(\alpha_l) \log \left(\frac{q(\Omega)q(\alpha_l)}{p_l(\Theta_l, \mathbf{y})} \right) d\Omega d\alpha_l + \text{const.} \quad (13)$$

Notice that since $p_l(\Theta_l|\mathbf{y})$ can not be found because $p(\mathbf{y})$ can not be calculated analytically, we apply variational methods to approximate this posterior distribution (see [29]). The KL divergence is always non negative and zero if $p_l(\Theta_l|\mathbf{y})$ and $q_l(\Theta_l)$ coincide, so by decreasing the KL divergence we are looking for good posterior distribution approximations.

The estimation of $\boldsymbol{\lambda} = \{\lambda_1, \lambda_2, \dots, \lambda_m\}$ will not be addressed in this paper, but we will show experimentally that non-degenerate combinations of the divergences (those with more than one coefficient $\lambda_l \neq 0$) may provide better reconstructions than degenerate ones.

Taking into account that

$$\int q(\Omega)q(\alpha_l) \log \left(\frac{q(\Omega)q(\alpha_l)}{p_l(\Theta_l, \mathbf{y})} \right) d\Omega d\alpha_l = \int q(\Phi) \log \left(\frac{q(\Omega)q(\alpha_l)}{p_l(\Theta_l, \mathbf{y})} \right) d\Phi, \quad (14)$$

Eq. (11) can be written in the more compact form:

$$\hat{q}(\Phi) = \underset{q(\Phi)}{\operatorname{argmin}} \int q(\Phi) \log \left(\frac{q(\Omega)}{p(\mathbf{y}|\Omega) \prod_{k=1}^L p(\beta_k)} \prod_{l=1}^m \left[\frac{q(\alpha_l)}{p_l(\mathbf{x}|\alpha_l)p(\alpha_l)} \right]^{\lambda_l} \right) d\Phi. \quad (15)$$

In this paper two model configurations are studied. We first consider the combination between SAR and ℓ_1 prior models, and afterward the combination between SAR and TV. The two configurations are denoted as Φ_c , for $c \in \{\ell_1, TV\}$, with $\lambda_{\ell_1} = \{1 - \lambda_2, \lambda_2, 0\}$ and $\lambda_{TV} = \{1 - \lambda_3, 0, \lambda_3\}$. The configuration corresponding to the combination between the three considered prior models has not been considered, because TV and ℓ_1 are very similar, differing only on the adaptability of the ℓ_1 prior to the possible direction dependent strength of the edges.

Unfortunately, we can not directly tackle the minimization of Eq. (15) because of the prior models $p_l(\mathbf{x}|\alpha_l)$ for $l = 2, 3$, defined in Eqs. (6) and (7), respectively. In earlier work with ℓ_1 prior model (see [6]) and with TV prior model (see [5]), this difficulty was overcome by resorting to majorization-minimization (MM) approaches, which is also the path followed here, in this paper.

In the MM approach lower bounds to the joint distributions of Eq. (10) are found, which make the minimization of Eq. (11) tractable. Lower bounds to the $p_l(\Theta_l, \mathbf{y})$ distributions of Eq. ((10)), for $l = 2, 3$ can be found as follows. Let us first consider the functionals

$$M_2(\alpha_2, \mathbf{x}, \mathbf{w}_2) = (\alpha_2^h \alpha_2^v)^{\frac{PN}{4}} \exp \left\{ - \sum_{i=1}^{PN} \left[\alpha_2^h \frac{(\Delta_i^h(\mathbf{x}))^2 + w_{2i}^h}{2\sqrt{w_{2i}^h}} + \alpha_2^v \frac{(\Delta_i^v(\mathbf{x}))^2 + w_{2i}^v}{2\sqrt{w_{2i}^v}} \right] \right\} \quad (16)$$

and

$$M_3(\alpha_3, \mathbf{x}, \mathbf{w}_3) = (\alpha_3)^{\frac{PN}{2}} \exp \left\{ - \frac{\alpha_3}{2} \sum_{i=1}^{PN} \left[\frac{(\Delta_i^h(\mathbf{x}))^2 + (\Delta_i^v(\mathbf{x}))^2 + w_{3i}}{\sqrt{w_{3i}}} \right] \right\}, \quad (17)$$

with PN dimensional vectors $\mathbf{w}_2^h, \mathbf{w}_2^v, \mathbf{w}_3 \in (R^+)^{PN}$, with components $w_{2i}^h, w_{2i}^v, w_{3i}$ for $i = 1, \dots, PN$, and with $\alpha_2 = \{\alpha_2^h, \alpha_2^v\}$ and $\mathbf{w}_2 = \{\mathbf{w}_2^h, \mathbf{w}_2^v\}$. In Eqs (16) and (17), w_{2i}^h, w_{2i}^v , and w_{3i} act as local adaptability factors: the greater their values, the lower the smoothing effect of the functional. We will see in Eqs (33) and (34), that these vectors have large values in pixel locations with high f.o.d. values, i.e. near image edges, that is, where we want the smoothing effect of the prior to be lower.

It can be shown by applying, in Eqs (16) and (17), the inequality $\sqrt{z} \leq \frac{z+w}{2\sqrt{w}}$, $\forall z \geq 0, w > 0$, (details can be found in [5, 6]) that these functionals are lower bounds of the image priors $p_l(\mathbf{x}|\alpha_l)$, for $l = 2, 3$, in Eqs. (6) and (7) respectively. These lower bounds can be used to find lower bounds for the respective joint distributions, that is, for $l = 2, 3$ we have

$$p_l(\Theta_l, \mathbf{y}) \geq p(\mathbf{y}|\Omega) \mathbf{M}_l(\alpha_l, \mathbf{x}, \mathbf{w}_l) p(\alpha_l) \prod_{k=1}^L [p(\beta_k) p(\mathbf{s}_k)] = \mathbf{F}_l(\Theta_l, \mathbf{y}, \mathbf{w}_l), \quad (18)$$

and consequently, an upper bound to the integral in Eq. (15) is then obtained.

The minimization in Eq. (15) can then be replaced by the minimization of its upper bound, since minimizing this bound with respect to the unknowns and the auxiliary variable \mathbf{w}_l in an alternating fashion results in closer bounds at each iteration. This bound is quadratic and therefore it can be evaluated analytically.

Before we proceed to calculate the posterior approximation, we first observe that to calculate $q(\alpha_l)$, $l = 1, 2, 3$, we only have to look at the only divergence where that distribution is present. So we can write

$$q(\alpha_1) \propto \exp \left(\langle \log p_1(\Omega, \alpha_1, \mathbf{y}) \rangle_{q(\Omega)} \right), \quad (19)$$

and

$$q(\alpha_l) \propto \exp \left(\langle \log \mathbf{F}_l(\Theta_l, \mathbf{y}, \mathbf{w}_l) \rangle_{q(\Omega)} \right), \text{ for } l = 2, 3, \quad (20)$$

where $\langle \star \rangle_{q(\Omega)}$ denotes the expected value of \star using the $q(\Omega)$ distribution, i.e. $\langle \star \rangle_{q(\Omega)} = E_{q(\Omega)}[\star]$. In what follows, we will use $\langle \star \rangle$ for simplicity, making the distribution $q(\Omega)$ being utilized clear from the context.

However, to calculate the distributions for the rest of the unknowns $q(\xi)$, $\xi \in \Omega$ we have to take into account all the divergences. We obtain

$$q_c(\xi) \propto \exp \left(\left\langle \log \left[p(\mathbf{y}|\Omega) \prod_{k=1}^L [p(\beta_k) p(\mathbf{s}_k)] \right. \right. \right. \\ \left. \left. \left. [\mathbf{M}_{l_c}(\alpha_{l_c}, \mathbf{x}, \mathbf{w}_{l_c}) p(\alpha_{l_c})]^{\lambda_{l_c}} [p_1(\mathbf{x}|\alpha_1) p(\alpha_1)]^{1-\lambda_{l_c}} \right] \right\rangle_{q(\Phi_{c\xi})} \right), \quad (21)$$

for the two configurations $c \in \{\ell 1, TV\}$, where $l_c = 2$ for $c = \ell 1$ and $l_c = 3$ for $c = TV$. In Eq. (21) $\Phi_{c\xi}$, for $c \in \{\ell 1, TV\}$, denote the set of variables Φ_c with ξ removed.

4.1. Estimation of the HR Image and Registration Parameters Distributions

In order to obtain the distribution function $q_c(\mathbf{x})$ we use Eq. (21) resulting in the multivariate Gaussian

$$q_{\ell 1}(\mathbf{x}) \propto \exp \left[-\frac{1}{2} \left(\lambda_2 \left\{ \langle \alpha_2^h \rangle \sum_i \frac{(\Delta_i^h(\mathbf{x}))^2 + w_{2i}^h}{\sqrt{w_{2i}^h}} + \langle \alpha_2^v \rangle \sum_i \frac{(\Delta_i^v(\mathbf{x}))^2 + w_{2i}^v}{\sqrt{w_{2i}^v}} \right\} \right. \right. \\ \left. \left. + (1 - \lambda_2) \langle \alpha_1 \rangle \|\mathcal{C}\mathbf{x}\|^2 + \sum_k \langle \beta_k \rangle \langle \|\mathbf{y}_k - \mathbf{B}_k(\mathbf{s}_k)\mathbf{x}\|^2 \rangle_{q(\mathbf{s}_k)} \right) \right] \quad (22)$$

for $c = \ell 1$, and for $c = TV$

$$q_{TV}(\mathbf{x}) \propto \exp \left[-\frac{1}{2} \left(\lambda_3 \langle \alpha_3 \rangle \sum_i \frac{(\Delta_i^h(\mathbf{x}))^2 + (\Delta_i^v(\mathbf{x}))^2 + w_{3i}}{\sqrt{w_{3i}}} \right. \right. \\ \left. \left. + (1 - \lambda_3) \langle \alpha_1 \rangle \|\mathcal{C}\mathbf{x}\|^2 + \sum_k \langle \beta_k \rangle \langle \|\mathbf{y}_k - \mathbf{B}_k(\mathbf{s}_k)\mathbf{x}\|^2 \rangle_{q(\mathbf{s}_k)} \right) \right]. \quad (23)$$

Also, from Eq. (21), we can find for the registration parameters, the distribution

$$q(\mathbf{s}_k) \propto \exp \left[-\frac{1}{2} \left(\langle \beta_k \rangle \langle \|\mathbf{y}_k - \mathbf{B}_k(\mathbf{s}_k)\mathbf{x}\|^2 \rangle_{q(\mathbf{x})} \right. \right. \\ \left. \left. + (\mathbf{s}_k - \bar{\mathbf{s}}_k^p)^t (\boldsymbol{\Xi}_k^p)^{-1} (\mathbf{s}_k - \bar{\mathbf{s}}_k^p) \right) \right]. \quad (24)$$

The explicit form of these distributions depends on the expectation values $\langle \|\mathbf{y}_k - \mathbf{B}_k(\mathbf{s}_k)\mathbf{x}\|^2 \rangle_{q(\mathbf{s}_k)}$ and $\langle \|\mathbf{y}_k - \mathbf{B}_k(\mathbf{s}_k)\mathbf{x}\|^2 \rangle_{q(\mathbf{x})}$. These calculations are not easy since $\mathbf{C}(\mathbf{s}_k)$, in Eq. (1), is nonlinear with respect to \mathbf{s}_k . Therefore, we expand $\mathbf{C}(\mathbf{s}_k)$ using its first-order Taylor series around the mean value $\bar{\mathbf{s}}_k = \langle \mathbf{s}_k \rangle = (\bar{\theta}_k, \bar{c}_k, \bar{d}_k)^T$ of the distribution $q(\mathbf{s}_k)$, of Eq. (24), resulting in

$$\mathbf{C}(\mathbf{s}_k) \approx \mathbf{C}(\bar{\mathbf{s}}_k) + [\mathbf{N}_1(\bar{\mathbf{s}}_k), \mathbf{N}_2(\bar{\mathbf{s}}_k), \mathbf{N}_3(\bar{\mathbf{s}}_k)] (\mathbf{s}_k - \bar{\mathbf{s}}_k), \quad (25)$$

with

$$[\mathbf{N}_1(\bar{\mathbf{s}}_k)\mathbf{x}, \mathbf{N}_2(\bar{\mathbf{s}}_k)\mathbf{x}, \mathbf{N}_3(\bar{\mathbf{s}}_k)\mathbf{x}] = \\ [(\mathbf{P}_1(\bar{\mathbf{s}}_k)\mathbf{M}_1(\bar{\mathbf{s}}_k) + \mathbf{P}_2(\bar{\mathbf{s}}_k)\mathbf{M}_2(\bar{\mathbf{s}}_k)), \mathbf{M}_1(\bar{\mathbf{s}}_k), \mathbf{M}_2(\bar{\mathbf{s}}_k)]. \quad (26)$$

In Eq. (26),

$$\mathbf{M}_1(\bar{\mathbf{s}}_k) = (\mathbf{I} - \mathbf{D}_{\mathbf{b}_k(\mathbf{s}_k)}) (\mathbf{L}_{\mathbf{tr}(s_k)} - \mathbf{L}_{\mathbf{tl}(s_k)}) + \mathbf{D}_{\mathbf{b}_k(\mathbf{s}_k)} (\mathbf{L}_{\mathbf{br}(s_k)} - \mathbf{L}_{\mathbf{bl}(s_k)}) \\ \mathbf{M}_2(\bar{\mathbf{s}}_k) = (\mathbf{I} - \mathbf{D}_{\mathbf{a}_k(\mathbf{s}_k)}) (\mathbf{L}_{\mathbf{bl}(s_k)} - \mathbf{L}_{\mathbf{tl}(s_k)}) + \mathbf{D}_{\mathbf{a}_k(\mathbf{s}_k)} (\mathbf{L}_{\mathbf{br}(s_k)} - \mathbf{L}_{\mathbf{tr}(s_k)}),$$

$\mathbf{P}_1(\bar{\mathbf{s}}_k) = -[\mathbf{D}_{\mathbf{u}} \sin(\bar{\theta}_k) + \mathbf{D}_{\mathbf{v}} \cos(\bar{\theta}_k)]$ and $\mathbf{P}_2(\bar{\mathbf{s}}_k) = [\mathbf{D}_{\mathbf{u}} \cos(\bar{\theta}_k) - \mathbf{D}_{\mathbf{v}} \sin(\bar{\theta}_k)]$, where $\mathbf{D}_{\mathbf{u}}$ and $\mathbf{D}_{\mathbf{v}}$ are diagonal matrices whose diagonals are the vectors \mathbf{u} and \mathbf{v} , respectively.

Using Eq. (25), $\mathbf{B}_k(\mathbf{s}_k)$ of Eq. (1) can be approximated by

$$\mathbf{B}_k(\mathbf{s}_k) = \mathbf{A}\mathbf{H}_k\mathbf{C}(\mathbf{s}_k) \approx \mathbf{B}_k(\bar{\mathbf{s}}_k) + [\mathbf{O}_{k1}(\bar{\mathbf{s}}_k), \mathbf{O}_{k2}(\bar{\mathbf{s}}_k), \mathbf{O}_{k3}(\bar{\mathbf{s}}_k)](\mathbf{s}_k - \bar{\mathbf{s}}_k), \quad (27)$$

with $\mathbf{O}_{kr}(\bar{\mathbf{s}}_k) = \mathbf{A}\mathbf{H}_k\mathbf{N}_r(\bar{\mathbf{s}}_k)$, for $r = 1, \dots, 3$, and we obtain

$$\begin{aligned} < \|\mathbf{y}_k - \mathbf{B}_k(\mathbf{s}_k)\mathbf{x}\|^2 >_{\mathbf{q}(\mathbf{s}_k)} \approx \|\mathbf{y}_k - \mathbf{B}(\bar{\mathbf{s}}_k)\mathbf{x}\|^2 + \\ & \sum_{i=1}^3 \sum_{j=1}^3 < \beta_k > \xi_{kij} \mathbf{x}^t \mathbf{O}_{ki}(\bar{\mathbf{s}}_k)^t \mathbf{O}_{kj}(\bar{\mathbf{s}}_k) \mathbf{x}, \end{aligned} \quad (28)$$

where for $i, j = 1, \dots, 3$, ξ_{kij} are the elements of the 3×3 covariance matrix Ξ_k of the posterior $\mathbf{q}(\mathbf{s}_k)$, of Eq. (24).

Substituting Eq. (28) into Eqs. (22) and (23), we obtain for the posterior distribution $\mathbf{q}_c(\mathbf{x})$ the multivariate Gaussian

$$\mathbf{q}_c(\mathbf{x}) = \mathcal{N}(\mathbf{x} | \mathbb{E}_{\mathbf{q}_c(\mathbf{x})}[\mathbf{x}], \text{cov}_{\mathbf{q}_c(\mathbf{x})}[\mathbf{x}]), \quad (29)$$

with mean

$$\mathbb{E}_{\mathbf{q}_c(\mathbf{x})}[\mathbf{x}] = \text{cov}_{\mathbf{q}_c(\mathbf{x})}[\mathbf{x}] \left[\sum_k < \beta_k > \mathbf{B}_k(\bar{\mathbf{s}}_k)^t \mathbf{y}_k \right], \quad (30)$$

and inverse covariance

$$\begin{aligned} \text{cov}_{q_{\ell 1}(\mathbf{x})}^{-1}[\mathbf{x}] &= \sum_k < \beta_k > \mathbf{B}_k(\bar{\mathbf{s}}_k)^t \mathbf{B}_k(\bar{\mathbf{s}}_k) + \sum_k \sum_{i=1}^3 \sum_{j=1}^3 < \beta_k > \xi_{kij} \mathbf{O}_{ki}(\bar{\mathbf{s}}_k)^t \mathbf{O}_{kj}(\bar{\mathbf{s}}_k) \\ &+ \lambda_2 \left(< \alpha_2^h > \Delta^h{}^t \mathbf{W}(w_2^h) \Delta^h + < \alpha_2^v > \Delta^v{}^t \mathbf{W}(w_2^v) \Delta^v \right) + (1 - \lambda_2) < \alpha_1 > \mathcal{C}^t \mathcal{C}, \end{aligned} \quad (31)$$

for $c = \ell 1$, and

$$\begin{aligned} \text{cov}_{q_{TV}(\mathbf{x})}^{-1}[\mathbf{x}] &= \sum_k < \beta_k > \mathbf{B}_k(\bar{\mathbf{s}}_k)^t \mathbf{B}_k(\bar{\mathbf{s}}_k) + \sum_k \sum_{i=1}^3 \sum_{j=1}^3 < \beta_k > \xi_{kij} \mathbf{O}_{ki}(\bar{\mathbf{s}}_k)^t \mathbf{O}_{kj}(\bar{\mathbf{s}}_k) \\ &+ \lambda_3 < \alpha_3 > \left(\Delta^h{}^t \mathbf{W}(w_3) \Delta^h + \Delta^v{}^t \mathbf{W}(w_3) \Delta^v \right) + (1 - \lambda_3) < \alpha_1 > \mathcal{C}^t \mathcal{C}, \end{aligned} \quad (32)$$

for $c = TV$. In the equations above, Δ^h and Δ^v represent the $PN \times PN$ convolution matrices associated respectively with the horizontal and vertical f.o.d., and $\mathbf{W}(\mathbf{w})$, $\forall \mathbf{w} \in (R^+)^{PN}$, is a $PN \times PN$ diagonal matrix with elements $\mathbf{W}(\mathbf{w})_{ii} = \frac{1}{\sqrt{w_i}}$, for $i = 1, \dots, PN$. These $\mathbf{W}(\mathbf{w})$ matrices can be interpreted as space adaptation matrices.

The following expressions are obtained for the \mathbf{w} parameters

$$w_{2i}^h = \mathbb{E}_{q_{\ell 1}(\mathbf{x})}[(\Delta_i^h(\mathbf{x}))^2], \quad w_{2i}^v = \mathbb{E}_{q_{\ell 1}(\mathbf{x})}[(\Delta_i^v(\mathbf{x}))^2], \quad (33)$$

$$w_{3i} = \mathbb{E}_{q_{TV}(\mathbf{x})}[(\Delta_i^h(\mathbf{x}))^2 + (\Delta_i^v(\mathbf{x}))^2]. \quad (34)$$

Now, we use again Eq. (27), to obtain the following approximation

$$\begin{aligned} < \| \mathbf{y}_k - \mathbf{B}_k(\mathbf{s}_k)\mathbf{x} \|^2 >_{\mathbf{q}(\mathbf{x})} \approx \| \mathbf{y}_k - \mathbf{B}(\bar{\mathbf{s}}_k)\mathbf{E}_{\mathbf{q}_c(\mathbf{x})}[\mathbf{x}] - \mathbf{\Upsilon}_k(\mathbf{s}_k - \bar{\mathbf{s}}_k) \|^2 \\ + \text{trace} [\mathbf{B}(\bar{\mathbf{s}}_k)^t \mathbf{B}(\bar{\mathbf{s}}_k) \text{cov}_{\mathbf{q}_c(\mathbf{x})}] + 2\mathbf{\Phi}_k^t(\mathbf{s}_k - \bar{\mathbf{s}}_k) + (\mathbf{s}_k - \bar{\mathbf{s}}_k)^t \mathbf{\Psi}_k(\mathbf{s}_k - \bar{\mathbf{s}}_k) \end{aligned} \quad (35)$$

which allows us to express the distribution $\mathbf{q}(\mathbf{s}_k)$ of Eq. (24) as the Gaussian $\mathbf{q}(\mathbf{s}_k) = \mathcal{N}(\mathbf{s}_k | <\mathbf{s}_k>, \mathbf{\Xi}_k)$ with parameters

$$<\mathbf{s}_k> = \mathbf{\Xi}_k \left[(\mathbf{\Xi}_k^p)^{-1} \bar{\mathbf{s}}_k^p + <\beta_k> (\mathbf{\Gamma}_k \bar{\mathbf{s}}_k + \mathbf{\Psi}_k \bar{\mathbf{s}}_k + \mathbf{Q}_k - \mathbf{\Phi}_k) \right], \quad (36)$$

and

$$\mathbf{\Xi}_k^{-1} = (\mathbf{\Xi}_k^p)^{-1} + <\beta_k> (\mathbf{\Psi}_k + \mathbf{\Gamma}_k). \quad (37)$$

In Eqs. (35), (36) and (37) above,

$$\mathbf{\Upsilon}_k = [\mathbf{O}_{k1}(\bar{\mathbf{s}}_k)\mathbf{E}_{\mathbf{q}_c(\mathbf{x})}[\mathbf{x}], \mathbf{O}_{k2}(\bar{\mathbf{s}}_k)\mathbf{E}_{\mathbf{q}_c(\mathbf{x})}[\mathbf{x}], \mathbf{O}_{k3}(\bar{\mathbf{s}}_k)\mathbf{E}_{\mathbf{q}_c(\mathbf{x})}[\mathbf{x}]],$$

$\mathbf{\Phi}_k$ and \mathbf{Q}_k are 3×1 vectors with elements $\Phi_{ki} = \text{trace} [\mathbf{B}(\bar{\mathbf{s}}_k)^t \mathbf{O}_{ki}(\bar{\mathbf{s}}_k) \text{cov}_{\mathbf{q}_c(\mathbf{x})}]$, and $Q_{ki} = (\mathbf{y}_k - \mathbf{B}(\bar{\mathbf{s}}_k)\mathbf{E}_{\mathbf{q}_c(\mathbf{x})}[\mathbf{x}])^t \mathbf{\Upsilon}_{ki}$, respectively, for $i = 1, 2, 3$, and $\mathbf{\Psi}_k$ and $\mathbf{\Gamma}_k$ are 3×3 matrices with elements $\Psi_{kij} = \text{trace} [\mathbf{O}_{ki}(\bar{\mathbf{s}}_k)^t \mathbf{O}_{kj}(\bar{\mathbf{s}}_k) \text{cov}_{\mathbf{q}_c(\mathbf{x})}[\mathbf{x}]]$, and $\Gamma_{kij} = \mathbf{\Upsilon}_{ki}^t \mathbf{\Upsilon}_{kj}$, respectively, for $i, j = 1, 2, 3$.

An interesting observation is that this registration method is a generalized stochastic version of the Lucas-Kanade registration algorithm [28] as applied to the super resolution problem. The classical Lucas-Kanade method can be obtained as a special case of Eq. (36) by setting the matrix $\mathbf{\Psi}_k$ equal to zero. This matrix incorporates the uncertainty of the image estimate \mathbf{x} into the motion estimation procedure.

4.2. Estimation of the hyperparameter distributions

Finally we obtain the distributions for the hyperparameters $\{\alpha_l\}$ and $\{\beta_k\}$, which are found to be Gamma distributions. For the $\{\beta_k\}$ hyperparameters, using Eq. (21), we obtain

$$\mathbf{q}(\beta_k) \propto \beta_k^{\frac{N}{2}-1+a_{\beta_k}^0} \exp \left[-\beta_k \left(b_{\beta_k}^0 + \frac{\mathbf{E}_{\mathbf{q}_c(\mathbf{x})} [\|\mathbf{y}_k - \mathbf{B}_k(\mathbf{s}_k)\mathbf{x}\|^2]}{2} \right) \right] \quad (38)$$

Using Eq. (19) we obtain for α_1 the distribution

$$\mathbf{q}(\alpha_1) \propto \alpha_1^{\frac{PN}{2}-1+a_{\alpha_1}^0} \exp \left[-\alpha_1 \left(b_{\alpha_1}^0 + \frac{\mathbf{E}_{\mathbf{q}_c(\mathbf{x})} [\|\mathcal{C}\mathbf{x}\|^2]}{2} \right) \right], \quad (39)$$

and from Eq. (20) we obtain

$$\mathbf{q}(\alpha_2^h) \propto (\alpha_2^h)^{\frac{PN}{4}-1+a_{\alpha_2^h}^0} \exp \left[-\alpha_2^h \left(b_{\alpha_2^h}^0 + \sum_{i=1}^{PN} \sqrt{w_{2i}^h} \right) \right], \quad (40)$$

$$\mathbf{q}(\alpha_2^v) \propto (\alpha_2^v)^{\frac{PN}{4}-1+a_{\alpha_2^v}^0} \exp \left[-\alpha_2^v \left(b_{\alpha_2^v}^0 + \sum_{i=1}^{PN} \sqrt{w_{2i}^v} \right) \right], \quad (41)$$

for α_2 , and finally for α_3

$$q(\alpha_3) \propto (\alpha_3)^{\frac{PN}{2}-1+a_{\alpha_3}^0} \exp \left[-\alpha_3 \left(b_{\alpha_3}^0 + \sum_{i=1}^{PN} \sqrt{w_{3i}} \right) \right]. \quad (42)$$

We summarize below the proposed iterative SR algorithm 1, which comprises HR image estimation, registration and estimation of the model hyperparameters. Two model configurations are considered, the combination between SAR and ℓ_1 prior models, for $c = \ell_1$, and the combination between SAR and TV, for $c = TV$.

Algorithm 1 SR and registration using SAR and ℓ_1 or TV model combinations

Require: Initial values for HR image, registration parameters and hyperparameters.

- 1: **while** convergence criterion is not met **do**
 - 2: Estimate the image distribution using Eq. (21).
 - 3: Calculate the vectors \mathbf{w}_2^h and \mathbf{w}_2^v using Eq. (33), for $c = \ell_1$, or \mathbf{w}_3 using Eq. (34) for $c = TV$.
 - 4: Estimate the registration parameters using Eq. (36).
 - 5: Estimate the hyperparameters $\{\beta_k\}$ using Eq. (38), α_1 using Eq. (40), and α_2 using Eqs. (40) and (41), for $c = \ell_1$, or α_3 for $c = TV$, using Eq.(42).
-

5. Experimental Results

A number of experiments have been run, with synthetic and real images, in order to evaluate the performance of the proposed method, both in terms of HR image restoration quality and of registration accuracy, some of them will be described in this section. The two most widely used SR sensor integration models are uniform and Gaussian functions. In this paper we have used, for \mathbf{H}_k in Eq. (1), a 3×3 uniform PSF.

We use in this section, as convergence criterion for Algorithm 1 $\frac{\|\mathbf{x}^j - \mathbf{x}^{j-1}\|^2}{\|\mathbf{x}^{j-1}\|^2} < 10^{-5}$, where \mathbf{x}^j and \mathbf{x}^{j-1} are the image estimates at the j^{th} and $(j-1)^{th}$ iterations, respectively. Step 4 of Algorithm 1 is an iterative registration process based on Eqs. (36) and (37), for which the stopping criterion $\|\mathbf{y}_k - \mathbf{A}\mathbf{H}_k\mathbf{C}(\mathbf{s}_k^i)\hat{\mathbf{x}}\|^2 > \|\mathbf{y}_k - \mathbf{A}\mathbf{H}_k\mathbf{C}(\mathbf{s}_k^{i-1})\hat{\mathbf{x}}\|^2$, has been applied, where \mathbf{s}_k^i and \mathbf{s}_k^{i-1} are the motion parameters estimated at the i^{th} and $(i-1)^{th}$ iterations, respectively. This iterative registration process has been limited to a maximum of 35 iterations. Setting $a_\omega^o = 0$ and $b_\omega^o = 0$ in Eq.(9), for $\omega \in \{\{\alpha_l\}, \{\beta_k\}\}$, we have used non-informative prior in all the experiments. In Algorithm 1 initialization, a bicubic interpolation of the first observation \mathbf{y}_1 , has been used as initial value \mathbf{x}^0 for the HR image. The remaining parameters have been initialized to the following values: $w_{2i}^h = (\Delta_i^h(\mathbf{x}^0))^2$, $w_{2i}^v = (\Delta_i^v(\mathbf{x}^0))^2$, $w_{3i} = (\Delta_i^h(\mathbf{x}^0))^2 + (\Delta_i^v(\mathbf{x}^0))^2$,

$\alpha_2^h = PN/(2 \sum_i^{PN} \sqrt{w_{2i}^h})$, $\alpha_2^v = PN/(2 \sum_i^{PN} \sqrt{w_{2i}^v})$, $\alpha_3 = PN/(2 \sum_i^{PN} \sqrt{w_{3i}})$ and $\beta_k = N/\|\mathbf{y}_k - \mathbf{B}_k(\mathbf{s}_k)\mathbf{x}^0\|^2$.

The parameter estimations in steps 3-5 of Algorithm 1, which have been described in section 4, require the evaluation of traces of different matrix products involving the covariance matrix $\text{cov}_{\mathbf{q}_c(\mathbf{x})}[\mathbf{x}]$. As this covariance matrix can not be obtained in exact form, an approximation has to be applied. In our previous paper [30], an analysis of approximations to the covariance matrix was performed. The Jacobi approximation, which has been adopted in this paper, offered the best tradeoff between precision and efficiency.

The proposed Algorithm 1 allows the determination of all the unknowns of our problem, except for the $\boldsymbol{\lambda}$ parameter, which determines the relative contribution of the different prior models we are combining. Two variants of the proposed Algorithm 1 have been considered, the first, for $c = \ell 1$, corresponds to the combination of the SAR and $\ell 1$ norm based model priors (denoted by $\ell 1$ -SAR), and the second, for $c = TV$, corresponds to the combination of the SAR and TV model priors (denoted by TV -SAR). For $\ell 1$ -SAR $\boldsymbol{\lambda}_{\ell 1} = \{1 - \lambda_2, \lambda_2, 0\}$ and there are two degenerate combinations of interest. When $\lambda_2 = 0$ only the SAR model is used, and this combination will be denoted as *SARREG*, while the combination for $\lambda_2 = 1$, will be denoted as $\ell 1$. $\ell 1$ coincides with the model proposed in [6], except for image registration which we incorporate in this paper. For $c = TV$, $\boldsymbol{\lambda}_{TV} = \{1 - \lambda_3, 0, \lambda_3\}$. We denote as TV the degenerate combination obtained when $\lambda_3 = 1$. When $\lambda_3 = 0$ we obtain again the *SARREG* combination.

In this section we compare the results obtained using *SARREG*, $\ell 1$, TV , TV -SAR and $\ell 1$ -SAR, with the ones obtained by bicubic interpolation (denoted by *BBC*), the robust SR method in [13] (denoted by *ZMT*), based on backprojection with median filtering, the robust SR method in [14] (denoted by *RSR*), based on bilateral TV filters and the SR method in [23] (denoted by *SDK*), based on multidimensional kernel regression.

Let us consider first the experiments with synthetic images, which are based on the set of four images of 120×120 pixels of size, depicted in Figure 2. Sequences of five LR images have been generated from the image set through warping, blurring and downsampling by a factor $\sqrt{P} = 2$. For the warping, the following motion vectors have been used for the images in the sequence: $\mathbf{s}_1 = (0.0^\circ, 0.0, 0.0)^t$, $\mathbf{s}_2 = (3.0^\circ, 0.0, 0.5)^t$, $\mathbf{s}_3 = (-3.0^\circ, 0.5, 0.0)^t$, $\mathbf{s}_4 = (5.0^\circ, 1.0, 0.0)^t$ and $\mathbf{s}_5 = (-5.0^\circ, 0.0, 1.0)^t$. A 3×3 uniform PSF has been used for the blur. Finally additive white Gaussian noise with *Signal to Noise Ratio (SNR)* levels between 10 dB and 40 dB, has been added to the LR observations. We conducted simulations with 3 different noise realizations at each SNR level. Motion errors have also been simulated, corrupting the original translation parameters with white Gaussian noise with standard deviation of 1, and the rotation parameters with noise uniformly distributed in $[-2^\circ, 2^\circ]$.

A numerical comparison between the original and reconstructed HR images, obtained with the different methods, has been performed in terms of the peak

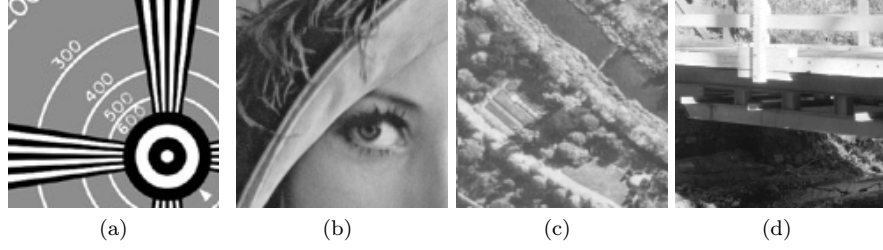


Figure 2: Set of 120×120 images used in the synthetic experiments.

signal-to-noise ratio (PSNR), defined as

$$\text{PSNR} = 10 \log_{10} \frac{NP}{\|\hat{\mathbf{x}} - \mathbf{x}\|^2}, \quad (43)$$

where $\hat{\mathbf{x}}$ and \mathbf{x} are the estimated and original HR images, respectively, and pixel values have been normalized to lie in the interval $[0, 1]$.

Figure 3.a-d shows, for the images in Figures 2.a-d, plots of the PSNR values corresponding to the different methods, except for the SDK method, and for all noise levels. The SDK method [23], which reconstructs video sequences, assumes differentiability in the temporal direction, to which a first order Taylor expansion is applied. We believe it is unfair to compare the SDK method in this experiment since there are abrupt changes between the images in the sequence.

The proposed methods SARREG, ℓ_1 , TV, TV-SAR and ℓ_1 -SAR, behave better in terms of PSNR than the other ones. Regarding the various models separately, the sparse ℓ_1 and TV give better results than the non-sparse SARREG and in most cases, ℓ_1 gives higher PSNR values than TV. The two proposed model combinations ℓ_1 -SAR and TV-SAR obtain similar PSNRs and perform better than ℓ_1 , TV and SAR alone.

Table 1 shows the λ_2 values maximizing PSNR, for the ℓ_1 -SAR model combination, and the λ_3 values for TV-SAR, in the present experiment, for the different images and noise levels. These λ_2 and λ_3 values have been obtained by exhaustive search of the value space $[0, 1]$, with a precision of 0.05. Figure 4, for example, shows the PSNR as a function of λ_2 for ℓ_1 -SAR combination and as a function of λ_3 for the TV-SAR combination for the reconstruction of the HR image in Figure 2.c, at 20 dB. For PSNR as a function of λ_2 and λ_3 , curves similar in shape to the ones shown in Figure 4, have been obtained for the rest of the images, and experiments.

In the second experiment, in order to compare our results also with the ones obtained using the SDK method, a smooth temporal sequence has been considered. For the image in Figure 2.b, the following rotations and translations were used: $\mathbf{s}_1 = (0.0^\circ, 0.0, 0.0)^t$, $\mathbf{s}_2 = (0.1^\circ, 0.3, 0.1)^t$, $\mathbf{s}_3 = (0.2^\circ, 0.4, 0.2)^t$, $\mathbf{s}_4 = (0.3^\circ, 0.5, 0.3)^t$ and $\mathbf{s}_5 = (0.4^\circ, 0.6, 0.4)^t$. A 3×3 uniform PSF has been used for the blur. Finally additive white Gaussian noise with SNR levels between 5 dB and 40 dB, has been added to the LR observations. We conducted

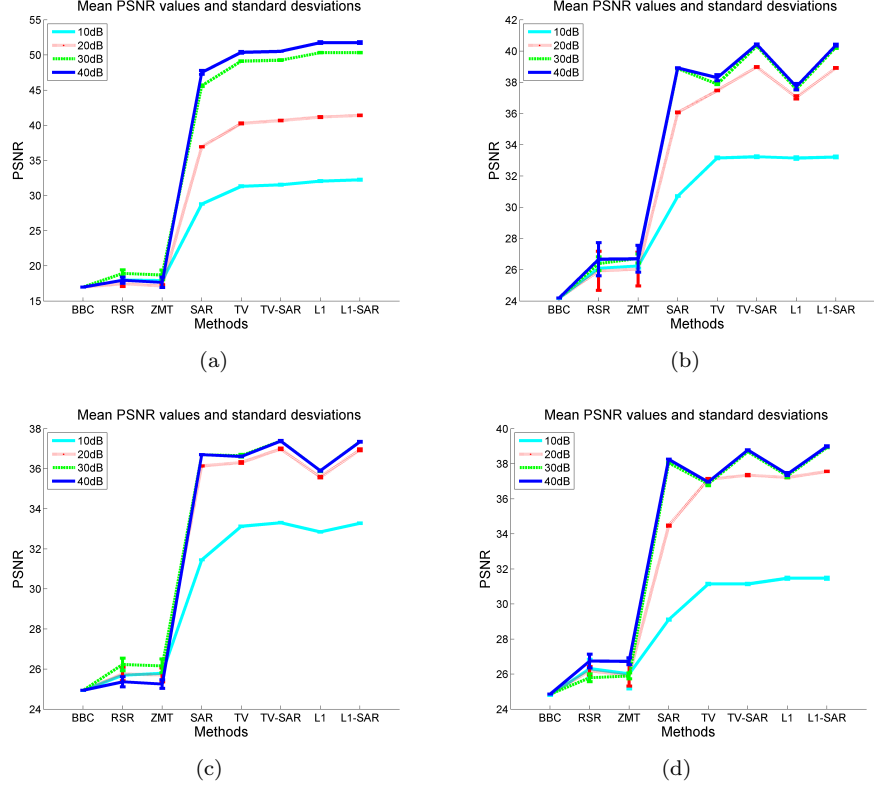


Figure 3: Mean PSNR values, and standard deviations, corresponding to the different methods and noise levels, for the images in Figures 2.a-d.

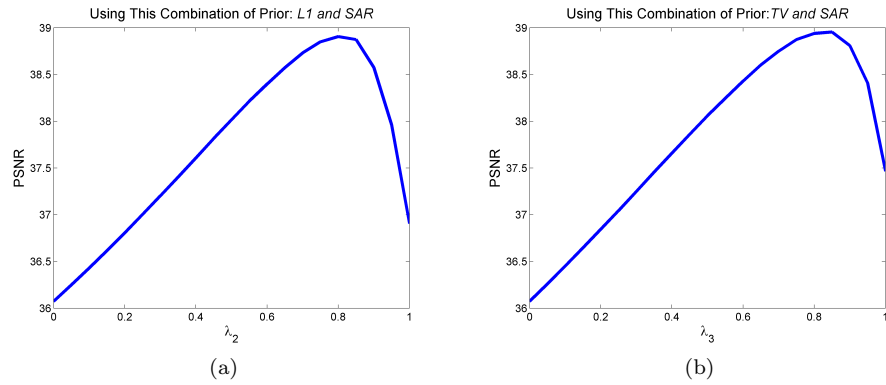


Figure 4: PSNR of the reconstruction of the HR image in Figure 2.c, from 5 LR observations with 20 dB, as a function of λ_2 for ℓ_1 -SAR and λ_3 for TV-SAR.

Table 1: λ_2 and λ_3 values obtained from the exhaustive search PSNR maximization, for the different noise levels and images in Figure 2.

| Image | SNR | 10 dB | 20 dB | 30 dB | 40 dB |
|---------|-------------|-------|-------|-------|-------|
| Fig 2.a | λ_2 | 0.95 | 0.85 | 0.85 | 0.55 |
| | λ_3 | 0.80 | 0.85 | 0.65 | 0.45 |
| Fig 2.b | λ_2 | 0.90 | 0.75 | 0.30 | 0.00 |
| | λ_3 | 0.85 | 0.75 | 0.30 | 0.00 |
| Fig 2.c | λ_2 | 0.90 | 0.80 | 0.10 | 0.00 |
| | λ_3 | 0.90 | 0.85 | 0.05 | 0.00 |
| Fig 2.d | λ_2 | 1.00 | 0.95 | 1.00 | 0.90 |
| | λ_3 | 0.95 | 0.90 | 0.95 | 0.55 |

simulations with 3 different noise realizations at each SNR level. Motion errors have also been simulated, corrupting the original translation parameters with white Gaussian noise with standard deviation of 1, and the rotation parameters with noise uniformly distributed in $[-2^\circ, 2^\circ]$. Figure 5.a shows a plot of the PSNR values corresponding to the different methods and noise levels. As it can be observed in Figure 5.a, the proposed methods SARREG, ℓ_1 , TV, TV-SAR and ℓ_1 -SAR, behave better in terms of PSNR than the other ones.

For this experiment, a comparison of the efficiency of the different methods in terms of the CPU time on an @Intel(R) Core(TM) i7CPU 950 at 3.07 GHz processor, has been performed. Figure 5.b, shows CPU plots for the different noise levels, our proposed methods are more time consuming than the other considered methods, with the exception of the SDK method. In the case of the proposed TV-SAR and ℓ_1 -SAR methods the execution times shown in Figure 5.b, corresponds to a given λ_m value. The exhaustive search process, multiplies these times by the number of explored λ_m values.

In a third experiment, a comparison has been performed between the registration accuracy of the proposed methods, and the one obtained utilizing the *Lucas-Kanade* (LK) [28] registration method. A comparison with the *Vandewalle* (VAN) method [15] was also performed, its performance is not shown because its registration accuracy is considerably worse. LK was applied to the HR images resulting from bilinear interpolations of the observations. In this experiment, in order to illustrate the robustness of the registration process, together with the image sequence of the first experiment, two more five images sequences have been considered, with motion vectors $\mathbf{s}_1 = (0.0^\circ, 0.0, 0.0)^t$, $\mathbf{s}_2 = (-6.4^\circ, 1.3, -0.6)^t$, $\mathbf{s}_3 = (3.2^\circ, -2.6, -4.1)^t$, $\mathbf{s}_4 = (4.0^\circ, 3.5, 2.6)^t$ and $\mathbf{s}_5 = (-7.2^\circ, 0.5, 6.2)^t$, for the first sequence, and $\mathbf{s}_1 = (0.0^\circ, 0.0, 0.0)^t$, $\mathbf{s}_2 = (-8.0^\circ, 2.0, -1.0)^t$, $\mathbf{s}_3 = (3.0^\circ, -3.0, 5.0)^t$, $\mathbf{s}_4 = (-4.0^\circ, 4.0, 2.0)^t$ and $\mathbf{s}_5 = (5.0^\circ, 1.0, 2.0)^t$ for the second sequence. In this case motion errors have also been simulated, corrupting the original translation parameters with white Gaussian noise with standard deviation of 1, and the rotation parameters with noise uniformly distributed in $[-2^\circ, 2^\circ]$.

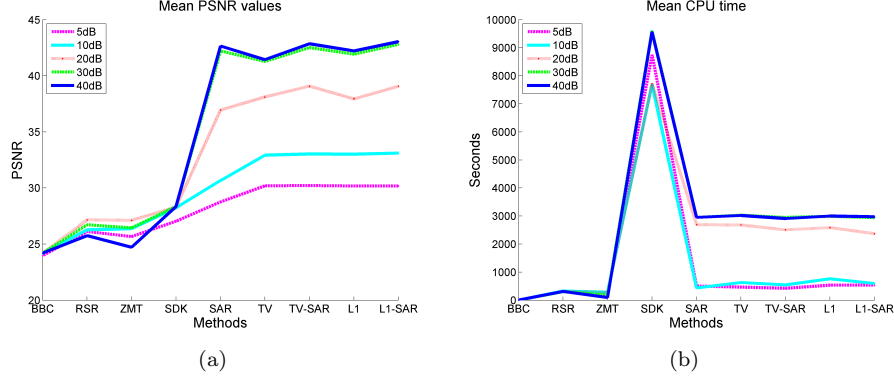


Figure 5: (a) Mean PSNR values, (b) mean CPU time corresponding to the different methods and noise levels, for the image in Figure 2.b.

The accuracy of the registration has been measured in terms of the absolute registration errors $|\mathbf{s}_k - \langle \mathbf{s}_k \rangle|$, between the estimated parameters $\langle \mathbf{s}_k \rangle$ and their true values \mathbf{s}_k , which in this case are known. Figure 6 shows the error values for the registration of the HR image in Figure 2.b, for the different methods, at different noise levels. From the results shown in Figure 6, the proposed methods outperform to the LK method.

Let us finally study the performance of the proposed methods on real observations. Two data sets have been used, the set of real observations in [31], and image sequences captured with a Canon IXUS700 camera, to which the different methods have been applied to enhance their resolution by a factor $\sqrt{P} = 4$.

Figure 7 shows the HR reconstructions obtained using BBC, RSR, ZMT and the proposed methods, for the first 20 LR images of the “text” sequence in [31]. The superior quality of the HR reconstructions obtained by our proposed methods is evident in Figure 7. We note, for example, that the gray background is cleaner in the reconstructions obtained using the model combinations TV-SAR, for $\lambda_3 = 0.8$, of Figure 7.g, and ℓ_1 -SAR, for $\lambda_2 = 0.4$ of Figure 7.h, than in the ℓ_1 reconstruction of Figure 7.d.

Finally, for the sequence of 19 observations captured with a Canon Ixus700 camera, Figure 8 shows the HR reconstructions obtained using the different methods. Once again, the reconstructions obtained using model combinations, shown in Figures 8.g & h, are superior to the ones obtained using the rest of methods under comparison.

6. Conclusions

In this paper, a novel variational Bayesian methodology for the combination of sparse and non-sparse image priors, has been applied to SR image reconstruction from rotated and displaced LR images. The entire SR problem, that

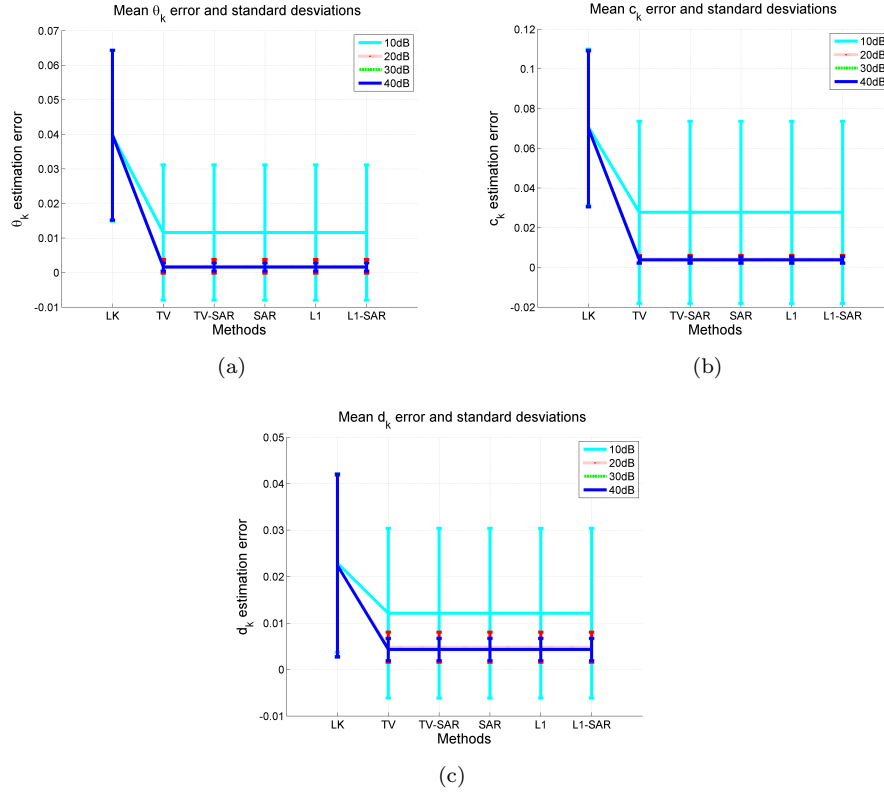


Figure 6: Mean absolute motion error, and standard deviations, of the estimated values for (a) the rotation angle θ_k , (b) the horizontal displacement c_k , and (c) the vertical displacement d_k , corresponding to the different methods and noise levels, for the image in Figure 2.b.



Figure 7: HR images obtained from the first 20 LR frames of the *text* sequence using: (a) BBC, (b) RSR, (c) ZMT, (d) SAR, (e) TV, (f) ℓ_1 , (g) TV-SAR with $\lambda_3 = 0.8$ and (h) ℓ_1 -SAR with $\lambda_2 = 0.4$.



Figure 8: HR images obtained from 19 LR observations captured with a Canon Ixus700 camera, using: (a) BBC, (b) RSR, (c) ZMT, (d) SAR, (e) TV, (f) ℓ_1 , (g) TV-SAR with $\lambda_3 = 0.5$ and (h) ℓ_1 -SAR with $\lambda_2 = 0.5$.

is, the joint HR image reconstruction, registration and parameter estimation, has been studied from a Bayesian perspective. For the combination of different image prior models, the Bayesian inference utilized finds the HR image given the observations, which minimizes a linear convex combination of KL divergences. We have found this distribution in closed form. The HR image estimates obtained by the proposed method compare favorably with the images provided by other state of the art SR reconstruction methods. Future work will address the estimation of the weights assigned to each KL divergence in their convex combination.

References

- [1] A. K. Katsaggelos, R. Molina, J. Mateos, *Super Resolution of Images and Video*, Morgan and Claypool, 2007.
- [2] P. Milanfar, *Super-Resolution Imaging*, Digital Imaging and Computer Vision, Taylor&Francis/CRC Press (Ed.), 2010.
- [3] R. Molina, J. Núñez, F. J. Cortijo, J. Mateos, Image restoration in Astronomy. A Bayesian perspective, *IEEE Signal Processing Magazine* 18 (2) (2001) 11–29.
- [4] M. Banham, A. Katsaggelos, Spatially adaptive wavelet based image restoration, *IEEE Trans. on Image Processing* 12 (8) (1996) 619–634.
- [5] S. D. Babacan, R. Molina, A. Katsaggelos, Variational bayesian super resolution, *IEEE Transactions on Image Processing* 20 (4) (2011) 984 – 999.
- [6] S. Villena, M. Vega, R. Molina, A. K. Katsaggelos, Bayesian super-resolution image reconstruction using an l1 prior, in: 6th ISPA Conference, Salzburg (Austria), 2009, pp. 152–157.
- [7] S. Roth, M. J. Black, Fields of experts: A framework for learning image priors, in: In CVPR, 2005, pp. 860–867.
- [8] D. Sun, W.-K. Cham, Postprocessing of low bit-rate block dct coded images based on a fields of experts prior, *IEEE Transactions on Image Processing* 16 (11) (2007) 2743–2751.
- [9] G. Chantas, N. P. Galatsanos, A. Likas, M. Saunders, Variational bayesian image restoration based on a product of t-distributions image prior, *IEEE Transaction on Image Processing* 17 (10) (October 2008) 1795–1805.
- [10] S. Villena, M. Vega, D. Babacan, R. Molina, A. Katsaggelos, Using the kullback-leibler divergence to combine image priors in super-resolution image reconstruction, in: *IEEE International Conference on Image Processing*, Hong-Kong (China), 2010, pp. 893–896.

- [11] G. Chantas, N. Galatsanos, R. Molina, A. K. Katsaggelos, Variational bayesian image restoration with a spatially adaptive product of total variation image priors, *IEEE Transactions on Image Processing* 19 (2) (2010) 351–362.
- [12] M. Elad, A. Feuer, Restoration of a single superresolution image from several blurred, noisy, and undersampled measured images, *IEEE Transactions on Image Processing* 6 (1997) 1646–1658.
- [13] A. Zomet, A. Rav-Acha, S. Peleg, Robust super-resolution, in: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2001)*, 2001, pp. 645–650.
- [14] S. Farsiu, M. D. Robinson, M. Elad, P. Milanfar, Fast and robust multi-frame super resolution, *IEEE Transactions on Image Processing* 13 (10) (Oct. 2004) 1327–1344.
- [15] P. Vandewalle, L. Sbaiz, J. Vandewalle, M. Vetterli, Super-Resolution from Unregistered and Totally Aliased Signals Using Subspace Methods, *IEEE Transactions on Signal Processing* 55 (7, Part 2) (2007) 3687–3703.
- [16] R. Hardie, K. Barnard, E. Armstrong, Joint MAP registration and high-resolution image estimation using a sequence of undersampled images, *IEEE Transactions on Image Processing* 6 (12) (1997) 1621–1633.
- [17] M. Ng, J. Koo, N. Bose, Constrained total least-squares computations for high-resolution image reconstruction with multisensors, *International Journal of Imaging Systems and Technology* 12 (1) (2002) 35–42.
- [18] M. E. Tipping, C. M. Bishop, Bayesian image super-resolution, in: *Advances in Neural Information Processing Systems 15 (NIPS)*, MIT Press, 2003, pp. 1303–1310.
- [19] F. Šroubek, J. Flusser, Multichannel blind deconvolution of spatially misaligned images, *IEEE Transactions on Image Processing* 7 (2005) 45–53.
- [20] N. Woods, N. Galatsanos, A. Katsaggelos, Stochastic methods for joint registration, restoration, and interpolation of multiple undersampled images, *IEEE Transactions on Image Processing* 15 (2006) 201–213.
- [21] L. C. Pickup, D. P. Capel, S. J. Roberts, A. Zisserman, Bayesian methods for image super-resolution, *The Computer Journal* 52 (1) (2009) 101–113.
- [22] A. Kanemura, S.-I. Maeda, S. Ishii, Superresolution with compound Markov random fields via the variational EM algorithm, *Neural Networks* 22 (7) (2009) 1025 – 1034.
- [23] H. Takeda, P. Milanfar, M. Protter, M. Elad, Superresolution without explicit subpixel motion estimation, *IEEE Transactions on Image Processing* 18 (9) (2009) 1958–1975.

- [24] M. Protter, M. Elad, H. Takeda, P. Milanfar, Generalizing the nonlocal-means to super-resolution reconstruction, *IEEE Transactions on Image Processing* 18 (1) (2009) 36–51.
- [25] M. V. Zibetti, F. S. Bazán, J. Mayer, Estimation of the parameters in regularized simultaneous super-resolution, *Pattern Recognition Letters* 32 (1) (2011) 69–78.
- [26] E. S. Lee, M. G. Kang, Regularized adaptive high-resolution image reconstruction considering inaccurate subpixel registration, *IEEE Transactions on Image Processing* 12 (7) (2003) 826–837.
- [27] Y. He, K. H. Yap, L. Chen, L. P. Chau, A nonlinear least square technique for simultaneous image registration and super-resolution, *IEEE Transactions on Image Processing* 16 (11) (2007) 2830–2841.
- [28] B. Lucas, T. Kanade, An iterative image registration technique with an application to stereo vision, in: *Proceedings of Imaging Understanding Workshop*, 1981, pp. 121–130.
- [29] C. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006.
- [30] S. Villena, M. Vega, R. Molina, A. K. Katsaggelos, Parameter estimation in bayesian super-resolution image reconstruction from low resolution rotated and translated images, in: *11th ACIVS Conference, Bordeaux (France)*, 2009, pp. 188–199.
- [31] S. Farsiu, *MDSP resolution enhancement software*, University of California at Santa Cruz (2004).