2010 IEEE International Workshop on Machine Learning for Signal Processing (MLSP 2010)
August 29 – September 1, 2010, Kittilä, Finland

IMAGE PRIOR COMBINATION IN SUPER-RESOLUTION IMAGE REGISTRATION & RECONSTRUCTION

Salvador Villena[a], Miguel Vega[a], S. Derin Babacan[b], Rafael Molina[c], and Aggelos K. Katsaggelos[b]

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

ABSTRACT

In this paper the application of image prior combinations to the Bayesian *Super Resolution (SR)* image registration and reconstruction problem is studied. A sparse image prior based on the horizontal and vertical first order differences is combined with a non-sparse 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 posterior distribution approximation on each posterior will produce as many posterior approximations as priors we want to combine. A unique approximation is obtained here by finding the distribution on the HR image given the observations that minimize a linear convex combination of the *Kullback-Leibler (KL)* divergences associated with each posterior distribution. We find this distribution in closed form. The estimated HR images are compared with images provided by other SR reconstruction 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 [10, 12] reviews). The basic principle in SR is that changes in the 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 from the set of LR observations.

While in image restoration there have been several recent attempts to combine image priors (see [16], [18] and [4]), no such attempts have been made in the SR literature. In [4] a Student's-t *Product of Experts (PoE)* image prior model was proposed and learnt only from the observations. Furthermore the introduction of Bayesian inference methodology, based on the constrained variational approximation, allowed to bypass the difficulty of evaluating the normalization constant of the PoE. PoE priors were learnt in [16] and [18] using a large training set of images and also stochastic sampling methods.

A combination of the TV image model proposed in [1] and the PoE model of [4] has been very recently proposed in [3]. This model combination may be considered a spatially adaptive version of the TV model which furthermore, as the method in [4], has the ability to enforce simultaneously a number of different properties on the image.

In this paper, we apply to SR image reconstruction a novel variational Bayesian methodology for the combination of the sparse prior, proposed in [20], based on the $\ell 1$ norm of the horizontal and vertical differences between image pixel values, and a non-sparse SAR prior.

SR image reconstruction is very sensitive to the quality of the registration parameter estimation. Some methods perform registration at the LR level in a previous step, see [6, 21]. This may induce registration imprecisions. It was shown in [7] (see also [8]) that the simultaneous reconstruction of the HR image and registration, at the HR level, gives superior results. In [19], [14] and [15], the errors in motion and blur parameters are assumed to follow Gaussian distributions. In [19], the HR image is marginalized out from the joint distribution and the motion and blur parameters are estimated from this marginal distribution. A major disadvantage of this method is that the marginalization of the HR image requires the utilization of a Gaussian image prior, which overpenalizes strong image edges and therefore reduces the quality of the estimated HR image. In [14] and [15], this problem is overcome by marginalizing the motion and blur parameters, and employing a Huber prior to model the HR image. In this paper registration is performed at the HR level, jointly with reconstruction, within a full Bayesian framework.

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 which develops the proposed method is presented in Section 4. We demonstrate the effectiveness of the proposed approach with experimental results in Section 5. Finally, section 6 concludes the paper.

2. PROBLEM FORMULATION

The imaging process is assumed to have generated L LR images $\mathbf{y}_k$, $k = 1, \ldots, L$, from the HR image $\mathbf{x}$. The LR images $\mathbf{y}_k$ and the HR image $\mathbf{x}$ consist of N and PN pixels, respectively, where the integer $P > 1$ is the factor of increase in resolution. In this paper we adopt the matrix-vector notation such that images $\mathbf{y}_k$ and $\mathbf{x}$ are arranged as $N \times 1$ and $PN \times 1$ vectors, respectively. The imaging process introduces warping, blurring and downsampling,

This work was supported in part by the Comisión Nacional de Ciencia y Tecnología under contract TIC2007-65533 and the Consejería de Innovación, Ciencia y Empresa of the Junta de Andalucía under contract P07-FQM-02701.

978-1-4244-7876-7/10/$26.00 ©2010 IEEE

which is modeled as

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

where $\mathbf{A}$ is the $N \times PN$ downsampling matrix, $\mathbf{H}_k$ the $PN \times PN$ blurring matrix, $\mathbf{C}(\mathbf{s}_k)$ 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. In this work, we assume that the blurring $\mathbf{H}_k$ matrices are known. The effects of downsampling, blurring, and warping can be combined into a single $N \times PN$ system matrix $\mathbf{B}(\mathbf{s}_k)$. Given (1), the super resolution problem is to find an estimate of the HR image $\mathbf{x}$ from the set of LR images $\{\mathbf{y}_k\}$ using prior knowledge about $\{\mathbf{n}_k\}$ and $\mathbf{x}$.

In this work we assume that the motion vectors $\mathbf{s}_k$ are not known, so they have to be estimated along with the HR image $\mathbf{x}$. 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 respectively the horizontal and vertical translations of the k^{th} HR image with respect to the reference frame $\mathbf{x}$.

3. HIERARCHICAL BAYESIAN MODELS

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

Using the model in (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

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

Assuming statistical independence of the noise among the LR image acquisitions, the conditional probability of the set of LR images $\mathbf{y} = \{\mathbf{y}_k\}$, given $\mathbf{x}$, $\{\mathbf{s}_k\}$ and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_L)$ can be expressed as

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

Let us now explicitly state the form of the matrices $\mathbf{C}(\mathbf{s}_k)$. 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)

$$u_k = u\cos(\theta_k) - v\sin(\theta_k) + c_k \quad (4)$$
$$v_k = u\sin(\theta_k) + v\cos(\theta_k) + d_k\,. \quad (5)$$

Let us denote the displacements between the grids by $\Delta(u_k,v_k)^T = (u, v)^T - (u_k, v_k)^T$. The vector difference between the pixel at (u_k, v_k) and the pixel at its top-left position in the reference HR grid is denoted by $(a_k(s_k), b_k(s_k))^T$, that is,

$$a_k(s_k) = \Delta u_k - \mathrm{floor}(\Delta u_k)\,, \quad (6)$$
$$b_k(s_k) = \Delta v_k - \mathrm{floor}(\Delta v_k)\,. \quad (7)$$

The $PN \times 1$ vectors $\mathbf{a}_k(\mathbf{s}_k)$ and $\mathbf{b}_k(\mathbf{s}_k)$ contain the displacement values for all pixels of the rotated image.

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. Using bilinear interpolation, the warped image $\mathbf{C}(\mathbf{s}_k)\mathbf{x}$ can be approximated as (see [8] for details)

$$\mathbf{C}(\mathbf{s}_k)\mathbf{x} \approx \sum_{\mathbf{z}\in\mathbf{nb}(s_k)} \mathcal{W}_{\mathbf{z}}\mathbf{L}_{\mathbf{z}}\mathbf{x}\,, \quad (8)$$

where $\mathbf{nb}(s_k) = \{\mathbf{bl}(s_k), \mathbf{br}(s_k), \mathbf{tl}(s_k), \mathbf{tr}(s_k)\}$, represent the pixels at the bottom-left, bottom-right, top-left and top-right locations, respectively. In Eq. (8), the matrices $\mathbf{L}_{\mathbf{z}}$ and $\mathcal{W}_{\mathbf{z}}$ with $\mathbf{z} \in \mathbf{nb}(s_k)$, are constructed in such a way that the product $\mathbf{L}_{\mathbf{z}}\mathbf{x}$ produces pixels at the corresponding top-left, top-right, bottom-left and bottom-right locations, respectively, and the matrices $\mathcal{W}_{\mathbf{z}}$ are their weights in the bilinear interpolation, that is, $\mathcal{W}_{\mathbf{br}(s_k)} = \mathbf{D}_{\mathbf{b}_k(\mathbf{s}_k)}\mathbf{D}_{\mathbf{a}_k(\mathbf{s}_k)}$, $\mathcal{W}_{\mathbf{bl}(s_k)} = \mathbf{D}_{\mathbf{b}_k(\mathbf{s}_k)}(\mathbf{I} - \mathbf{D}_{\mathbf{a}_k(\mathbf{s}_k)})$, $\mathcal{W}_{\mathbf{tl}(s_k)} = (\mathbf{I}-\mathbf{D}_{\mathbf{b}_k(\mathbf{s}_k)})(\mathbf{I}-\mathbf{D}_{\mathbf{a}_k(\mathbf{s}_k)})$ and $\mathcal{W}_{\mathbf{tr}(s_k)} = (\mathbf{I}-\mathbf{D}_{\mathbf{b}_k(\mathbf{s}_k)})\mathbf{D}_{\mathbf{a}_k(\mathbf{s}_k)}$ where $\mathbf{D}_{\mathbf{a}_k(\mathbf{s}_k)}$ and $\mathbf{D}_{\mathbf{b}_k(\mathbf{s}_k)}$ denote diagonal matrices with the vectors $\mathbf{a}_k(\mathbf{s}_k)$ and $\mathbf{b}_k(\mathbf{s}_k)$ at their diagonal, respectively.

As we have already explained in the introduction, in this paper we are combining a sparse prior, the $\ell 1$ model [20], and a non-sparse one, the simultaneous autoregression (SAR) model [13]. Note that the idea of combining sparse and non-sparse models has also been proposed in other contexts, see for instance [17]. The $\ell 1$ prior, which is very effective at preserving edges while imposing smoothness is defined by

$$\mathrm{p}_1(\mathbf{x}|\alpha_1) = \frac{1}{Z(\alpha_1)} \times \exp\left\{-\sum_{i=1}^{PN}\left[\alpha_1^h \parallel \Delta_i^h(\mathbf{x}) \parallel_1 + \alpha_1^v \parallel \Delta_i^v(\mathbf{x}) \parallel_1\right]\right\}, \quad (9)$$

where $\Delta_i^h(\mathbf{x})$ and $\Delta_i^v(\mathbf{x})$ represent the horizontal and vertical first order differences at pixel i, respectively, $\alpha_1 = \{\alpha_1^h, \alpha_1^v\}$ are model parameters, and $Z(\alpha_1)$ is the partition function that we approximate as

$$Z(\alpha_1) \propto (\alpha_1^h\alpha_1^v)^{-\frac{PN}{4}}. \quad (10)$$

The use in Eq. (9) of different parameters α_1^h and α_1^v for horizontal and vertical directions makes this model more adaptable to the image characteristics than the TV model that uses a single parameter.

We also consider the SAR prior, defined as

$$\mathrm{p}_2(\mathbf{x}|\alpha_2) \propto \alpha_2^{\frac{PN}{2}} \exp\left\{-\frac{\alpha_2}{2} \|\mathbf{Q}\mathbf{x}\|^2\right\}, \quad (11)$$

where $\mathbf{Q}$ is the Laplacian operator. This prior is expected to preserve image textures better than the $\ell 1$ prior.

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

$$\mathrm{p}(\mathbf{x}|\alpha_1, \alpha_2) = \frac{1}{Z(\alpha_1, \alpha_2)} \exp\left\{-\sum_{i=1}^{PN}\left[\alpha_1^h \parallel \Delta_i^h(\mathbf{x}) \parallel_1 + \alpha_1^v \parallel \Delta_i^v(\mathbf{x}) \parallel_1\right] - \frac{\alpha_2}{2} \|\mathbf{Q}\mathbf{x}\|^2\right\}, \quad (12)$$

but since there is no known approximation to the partition function $Z(\alpha_1, \alpha_2)$, the estimation of the parameters would be impossible for this prior model $\mathrm{p}(\mathbf{x}|\alpha_1, \alpha_2)$ (see however [9] in the context of model learning).

Let us denote by $\bar{\mathbf{s}}_k^p$ the estimate of $\mathbf{s}_k$ obtained from LR observations in a preprocessing step, using registration algorithms, such as the ones reported in [11]. 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,

$$\mathrm{p}(\mathbf{s}_k) = \mathcal{N}(\mathbf{s}_k|\bar{\mathbf{s}}_k^p, \boldsymbol{\Xi}_k^p)\,, \tag{13}$$

with $\boldsymbol{\Xi}_k^p$ the *a priori* covariance matrix. The parameters $\bar{\mathbf{s}}_k^p$ and $\boldsymbol{\Xi}_k^p$ incorporate prior knowledge about the motion parameters into the estimation procedure. If such knowledge is not available, $\bar{\mathbf{s}}_k^p$ and $(\boldsymbol{\Xi}_k^p)^{-1}$ can be set equal to zero, which makes the observations solely responsible for the estimation process.

The hyperparameters $\{\alpha_i\}$ and $\{\beta_k\}$ are crucial in determining the performance of the SR algorithm. For their modeling, we employ Gamma distributions, that is,

$$\mathrm{p}(\omega) = \Gamma(\omega|a_\omega^o, b_\omega^o) = \frac{(b_\omega^o)^{a_\omega^o}}{\Gamma(a_\omega^o)}\omega^{a_\omega^o - 1}\exp\left[-b_\omega^o\omega\right], \tag{14}$$

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

Finally, combining (3), (13) and (14), with the two different prior models we obtain the joint probability distributions

$$\mathrm{p}_i(\mathbf{y}, \Omega, \alpha_i) = \mathrm{p}(\mathbf{y}|\mathbf{x}, \{\mathbf{s}_k\}, \boldsymbol{\beta})\mathrm{p}_i(\mathbf{x}|\alpha_i)\,\mathrm{p}(\alpha_i)\prod_{k=1}^{L}\mathrm{p}(\beta_k)\mathrm{p}(\mathbf{s}_k)\,, \tag{15}$$

for $i = 1, 2$, where $\Omega = \{\mathbf{x}, \{\mathbf{s}_k\}, \boldsymbol{\beta}\}$.

4. VARIATIONAL BAYESIAN INFERENCE

Let us denote the set of all unknowns by $\Theta = \{\Omega, \boldsymbol{\alpha}\}$, where $\boldsymbol{\alpha} = (\alpha_1, \alpha_2)$. Bayesian inference is based on the posterior distribution $\mathrm{p}(\Theta \mid \mathbf{y})$. We propose here to approximate this distribution by the distribution minimizing the following linear convex combination of KL divergence measures

$$\hat{\mathrm{q}}(\Theta) = \underset{\mathrm{q}(\Theta)}{\operatorname{argmin}} \sum_{i=1}^{2}\lambda_i C_{KL}(\mathrm{q}(\Omega)\mathrm{q}(\alpha_i) \parallel \mathrm{p}_i(\Omega, \alpha_i|\mathbf{y})) \tag{16}$$

where $\lambda_i \geq 0$, $\lambda_1 + \lambda_2 = 1$,

$$\mathrm{q}(\Omega) = \mathrm{q}(\mathbf{x})\prod_{k=1}^{L}\mathrm{q}(\beta_k)\mathrm{q}(\mathbf{s}_k)\,, \quad \mathrm{q}(\Theta) = \mathrm{q}(\Omega)\prod_{i=1}^{2}\mathrm{q}(\alpha_i)\,, \tag{17}$$

and the KL divergences are given by

$$C_{KL}(\mathrm{q}(\Omega)\mathrm{q}(\alpha_i) \parallel \mathrm{p}_i(\Omega, \alpha_i|\mathbf{y})) = \int \mathrm{q}(\Omega)\mathrm{q}(\alpha_i)\log\left(\frac{\mathrm{q}(\Omega)\mathrm{q}(\alpha_i)}{\mathrm{p}_i(\mathbf{y}, \Omega, \alpha_i)}\right)d\Omega d\alpha_i + \text{const}\,. \tag{18}$$

The estimation of λ_1 and λ_2 will not be addressed in this paper, but we will show experimentally that a non-degenerate combination of divergences, $\lambda_1, \lambda_2 > 0$, provides a better reconstruction than a degenerate one.

Taking into account that

$$\int \mathrm{q}(\Omega)\mathrm{q}(\alpha_i)\log\left(\frac{\mathrm{q}(\Omega)\mathrm{q}(\alpha_i)}{\mathrm{p}_i(\mathbf{y}, \Omega, \alpha_i)}\right)d\Omega d\alpha_i = \int \mathrm{q}(\Theta)\log\left(\frac{\mathrm{q}(\Omega)\mathrm{q}(\alpha_i)}{\mathrm{p}_i(\mathbf{y}, \Omega, \alpha_i)}\right)d\Theta\,, \tag{19}$$

expression (16) can be rewritten in the more compact form as

$$\hat{\mathrm{q}}(\Theta) = \underset{\mathrm{q}(\Theta)}{\operatorname{argmin}} \int \mathrm{q}(\Theta)\log\left(\frac{\mathrm{q}(\Omega)}{\mathrm{p}(\mathbf{y}|\mathbf{x}, \{\mathbf{s}_k\}, \boldsymbol{\beta})} \times \left[\prod_{k=1}^{L}\frac{1}{\mathrm{p}(\beta_k)\mathrm{p}(\mathbf{s}_k)}\right]\prod_{i=1}^{2}\left[\frac{\mathrm{q}(\alpha_i)}{\mathrm{p}_i(\mathbf{x}|\alpha_i)\mathrm{p}(\alpha_i)}\right]^{\lambda_i}\right)d\Theta\,. \tag{20}$$

Unfortunately, we can not directly tackle the minimization of (20) because of the $\ell 1$ image prior $\mathrm{p}_1(\mathbf{x}|\alpha_1)$ of Eq. (9). This difficulty is overcome in this paper by resorting to the majorization-minimization (MM) approach (see [1]).

The main principle of the MM approach is to find a bound of the joint distribution in (15) which makes the minimization of (20) tractable. Let us first consider the following functional $\mathrm{M}(\alpha_1, \mathbf{x}, \mathbf{w})$, where $\mathbf{w} = \{\mathbf{u}^h, \mathbf{u}^v\}$ and $\mathbf{u}^h \in (R^+)^{PN}$ and $\mathbf{u}^v \in (R^+)^{PN}$ with components $\mathbf{u}_i^h$ and $\mathbf{u}_i^v$, $i = 1, \ldots, PN$,

$$\mathrm{M}(\alpha_1, \mathbf{x}, \mathbf{w}) \propto (\alpha_1^h\alpha_1^v)^{\frac{PN}{4}} \times \exp\left\{-\sum_{i=1}^{PN}\left[\alpha_1^h\frac{(\Delta_i^h(\mathbf{x}))^2 + \mathbf{u}_i^h}{2\sqrt{\mathbf{u}_i^h}} + \alpha_1^v\frac{(\Delta_i^v(\mathbf{x}))^2 + \mathbf{u}_i^v}{2\sqrt{\mathbf{u}_i^v}}\right]\right\}. \tag{21}$$

It can be shown (details can be found in [1]) that the functional $\mathrm{M}(\alpha_1, \mathbf{x}, \mathbf{w})$ is a lower bound of the image prior $\mathrm{p}_1(\mathbf{x}|\alpha_1)$, that is,

$$\mathrm{p}_1(\mathbf{x}|\alpha_1) \quad \geq \quad \mathrm{M}(\alpha_1, \mathbf{x}, \mathbf{w}). \tag{22}$$

This lower bound can be used to find a lower bound for the joint distribution, for $i = 1$, in (15), that is,

$$\mathrm{p}_1(\mathbf{y}, \Omega, \alpha_1) \geq \mathrm{p}(\mathbf{y}|\mathbf{x}, \boldsymbol{\beta})\prod_{k=1}^{L}\mathrm{p}(\beta_k)\mathrm{M}(\alpha_1, \mathbf{x}, \mathbf{w})\,\mathrm{p}(\alpha_1) = \mathrm{F}(\Omega, \alpha_1, \mathbf{w}, \mathbf{y})\,, \tag{23}$$

which results in an upper bound of the KL distance as

$$C_{KL}(\mathrm{q}(\Omega, \alpha_1) \parallel \mathrm{p}_1(\mathbf{y}, \Omega, \alpha_1)) \leq C_{KL}(\mathrm{q}(\Omega, \alpha_1) \parallel \mathrm{F}(\Omega, \alpha_1, \mathbf{w}, \mathbf{y})). \tag{24}$$

The minimization of $C_{KL}(\mathrm{q}(\Omega, \alpha_1) \parallel \mathrm{p}_1(\mathbf{y}, \Omega, \alpha_1))$ can then be replaced by the minimization of its upper bound (24), since minimizing this bound with respect to the unknowns and the auxiliary variable $\mathbf{w}$ in an alternating fashion results in closer bounds at each iteration. The bound in (24) is quadratic and therefore it can be evaluated analytically.

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

$$\mathrm{q}(\alpha_1) \quad = \quad \text{const} \times \exp\left(\langle \log F(\Omega, \alpha_1, \mathbf{w}, \mathbf{y})\rangle_\Omega\right)\,, \tag{25}$$

$$\mathrm{q}(\alpha_2) \quad = \quad \text{const} \times \exp\left(\langle \log \mathrm{p}_2(\mathbf{y}, \Omega, \alpha_2)\rangle_\Omega\right)\,, \tag{26}$$

where $\mathrm{E}_{\mathrm{q}(\Omega)}[\cdot] = <\cdot>_\Omega$.

Furthermore to calculate the rest of the unknown distributions $\mathrm{q}(\xi)$, $\xi \in \Omega$ we have to take into account both divergences. We

obtain

$$\mathrm{q}(\xi) = \mathrm{const} \times \exp\left(\left\langle \log\left[\mathrm{p}(\mathbf{y}|\mathbf{x},\{\mathbf{s}_k\},\boldsymbol{\beta})\prod_{k=1}^{L}\mathrm{p}(\beta_k)\mathrm{p}(\mathbf{s}_k)\right.\right.\right.$$
$$\left.\left.\left.\times\left[\mathrm{M}(\alpha_1,\mathbf{x},\mathbf{w})\,\mathrm{p}(\alpha_1)\right]^{\lambda_1}\left[\mathrm{p}_2(\mathbf{x}|\alpha_2)\mathrm{p}(\alpha_2)\right]^{1-\lambda_1}\right]\right\rangle_{\Theta_\xi}\right), \quad (27)$$

where Θ_ξ denotes the set Θ with ξ removed. In the following, the subscript of the expected value will be removed when it is clear from the context.

Let us now calculate the approximation. From Eq. (27), $\mathrm{q}(\mathbf{x})$ is the multivariate Gaussian

$$\mathrm{q}(\mathbf{x}) \propto \exp\left(-\frac{1}{2}\lambda_1\left\{<\alpha_1^h>\sum_i\frac{(\Delta_i^h(\mathbf{x}))^2+u_i^h}{\sqrt{u_i^h}}\right.\right.$$
$$\left.+<\alpha_1^v>\sum_i\frac{(\Delta_i^v(\mathbf{x}))^2+u_i^v}{\sqrt{u_i^v}}\right\}-\frac{(1-\lambda_1)}{2}<\alpha_2>\|\mathbf{Q}\mathbf{x}\|^2$$
$$\left.-\frac{1}{2}\sum_k<\beta_k>\mathrm{E}_{\mathbf{s}_k}\left[\|\,\mathbf{y}_k-\mathbf{A}\mathbf{H}_k\mathbf{C}(\mathbf{s}_k)\mathbf{x}\,\|^2\right]\right). \quad (28)$$

Also from Eq. (27) the posterior distribution approximation $\mathrm{q}(\mathbf{s}_k)$ is found as

$$\mathrm{q}(\mathbf{s}_k) \propto \exp\left(-\frac{1}{2}<\beta_k>\mathrm{E}_{\mathbf{x}}\left[\|\,\mathbf{y}_k-\mathbf{A}\mathbf{H}_k\mathbf{C}(\mathbf{s}_k)\mathbf{x}\,\|^2\right]\right.$$
$$\left.-\frac{1}{2}\left(\mathbf{s}_k-\bar{\mathbf{s}}_k^p\right)^T\left(\boldsymbol{\Xi}_k^p\right)^{-1}\left(\mathbf{s}_k-\bar{\mathbf{s}}_k^p\right)\right). \quad (29)$$

The explicit form of $\mathrm{q}(\mathbf{x})$ and $\mathrm{q}(\mathbf{s}_k)$, of Eqs. (28) and (29), depend on the expectations $\mathrm{E}_{\mathbf{s}_k}\left[\|\,\mathbf{y}_k-\mathbf{A}\mathbf{H}_k\mathbf{C}(\mathbf{s}_k)\mathbf{x}\,\|^2\right]$ and $\mathrm{E}_{\mathbf{x}}\left[\|\,\mathbf{y}_k-\mathbf{A}\mathbf{H}_k\mathbf{C}(\mathbf{s}_k)\mathbf{x}\,\|^2\right]$. These calculations are not easy since $\mathbf{C}(\mathbf{s}_k)\mathbf{x}$ is nonlinear with respect to $\mathbf{s}_k$. Therefore, we expand $\mathbf{C}(\mathbf{s}_k)\mathbf{x}$ using its first-order Taylor series around the mean value $<\mathbf{s}_k>=\bar{\mathbf{s}}_k=(\bar{\theta}_k,\bar{c}_k,\bar{d}_k)^T$ of the distribution $\mathrm{q}(\mathbf{s}_k)$. This first order expansion is a reasonable approximation under the assumption that $\mathbf{s}_k$ is close to $<\mathbf{s}_k>$, that is, in the region where most of the mass of the distribution is. Proceeding in this fashion, we obtain the following approximation of $\mathbf{C}(\mathbf{s}_k)\mathbf{x}$

$$\mathbf{C}(\mathbf{s}_k)\mathbf{x} \approx \mathbf{C}(\bar{\mathbf{s}}_k)\mathbf{x}+$$
$$\left(\mathbf{O}_{k1}(\bar{\mathbf{s}}_k)\mathbf{x},\mathbf{O}_{k2}(\bar{\mathbf{s}}_k)\mathbf{x},\mathbf{O}_{k3}(\bar{\mathbf{s}}_k)\mathbf{x}\right)\left(\mathbf{s}_k-\bar{\mathbf{s}}_k\right), \quad (30)$$

where

$$\left(\mathbf{O}_{k1}(\bar{\mathbf{s}}_k)\mathbf{x},\mathbf{O}_{k2}(\bar{\mathbf{s}}_k)\mathbf{x},\mathbf{O}_{k3}(\bar{\mathbf{s}}_k)\mathbf{x}\right)=$$
$$\sum_{\mathbf{z}\in\mathbf{nb}(\bar{s}_k)}\left(\mathcal{V}_{a\mathbf{z}}\mathbf{L}_{\mathbf{z}}\mathbf{x},\mathcal{V}_{b\mathbf{z}}\mathbf{L}_{\mathbf{z}}\mathbf{x}\right)\begin{pmatrix}-\mathbf{u}\sin(\bar{\theta}_k)-\mathbf{v}\cos(\bar{\theta}_k),\mathbf{1},\mathbf{0}\\ \mathbf{u}\cos(\bar{\theta}_k)-\mathbf{v}\sin(\bar{\theta}_k),\mathbf{0},\mathbf{1}\end{pmatrix} \quad (31)$$

with $\mathcal{V}_{a\mathbf{br}(\bar{s}_k)}=-\mathcal{V}_{a\mathbf{bl}(\bar{s}_k)}=\mathbf{D}_{\mathbf{b}_k(\bar{\mathbf{s}}_k)}$, $\mathcal{V}_{a\mathbf{tr}(\bar{s}_k)}=-\mathcal{V}_{a\mathbf{tl}(\bar{s}_k)}=(\mathbf{I}-\mathbf{D}_{\mathbf{b}_k(\bar{\mathbf{s}}_k)})$, and $\mathcal{V}_{b\mathbf{bl}(\bar{s}_k)}=-\mathcal{V}_{b\mathbf{tl}(\bar{s}_k)}=(\mathbf{I}-\mathbf{D}_{\mathbf{a}_k(\bar{\mathbf{s}}_k)})$, and $\mathcal{V}_{b\mathbf{br}(\bar{s}_k)}=-\mathcal{W}_{\mathbf{tr}(s_k)}=\mathbf{D}_{\mathbf{a}_k(\bar{\mathbf{s}}_k)}$. The allowed extension for this conference paper prevents showing here the final expressions obtained for the expectation values.

Notice that instead of the first-order expansion in (30), a second-order expansion can be employed on the error term $\|\,\mathbf{y}_k-\mathbf{A}\mathbf{H}_k\,\mathbf{C}(\mathbf{s}_k)\mathbf{x}\,\|^2$ (see [14] and [15] for the use of this approximation).

Table 1. Mean PSNRs with standard deviations provided by the SR algorithms at different SNR levels for the Lena image

SNR	10dB	15dB	20dB
Bicubic	24.24 $\pm1.0\ 10^{-4}$	24.27 $\pm1.0\ 10^{-4}$	24.27 $\pm1.0\ 10^{-4}$
ZMT	26.57 ±0.42	26.06 ±0.16	26.63 ±0.78
RSR	26.47 $\pm7.0\ 10^{-3}$	25.87 ±0.20	26.55 ±0.80
TV	33.90 $\pm1.0\ 10^{-3}$	36.44 $\pm2.5\ 10^{-3}$	39.67 $\pm5.0\ 10^{-3}$
SAR	32.37 $\pm2.0\ 10^{-3}$	35.24 $\pm4\ 10^{-3}$	39.74 $\pm7.0\ 10^{-4}$
L1	33.97 $\pm1.0\ 10^{-3}$	36.12 ±0.018	39.72 $\pm4.0\ 10^{-3}$
ALG1(λ_1)	34.35 $\pm3.0\ 10^{-4}$	37.72 ±0.001	41.10 $\pm1.0\ 10^{-4}$
λ_1	0.9	0.8	0.75

Furthermore, the elements of $\mathbf{u}^d$, for $d\in\{h,v\}$, in Eq. (21) are equal to

$$u_i^d = \mathrm{E}_{\mathbf{x}}[(\Delta_i^d(\mathbf{x}))^2]. \quad (32)$$

In the last step of the algorithm, the distributions $\mathrm{q}(\alpha_1^d)$, for $d\in\{h,v\}$, $\mathrm{q}(\alpha_2)$ and $\mathrm{q}(\beta_k)$ are the Gamma distributions, given by

$$\mathrm{q}(\alpha_1^d) \propto {\alpha_1^d}^{\frac{PN}{4}-1+a^o_{\alpha_1^d}}\exp\left[-\alpha_1^d(b^o_{\alpha_1^d}+\sum_i\sqrt{u_i^d})\right], \quad (33)$$

$$\mathrm{q}(\alpha_2) \propto {\alpha_2}^{\frac{PN}{2}-1+a^o_{\alpha_2}}\exp\left[-\alpha_2\left(b^o_{\alpha_2}+\frac{\mathrm{E}_{\mathbf{x}}\left[\|\mathbf{Q}\mathbf{x}\|^2\right]}{2}\right)\right], \quad (34)$$

$$\mathrm{q}(\beta_k) \propto \beta_k^{\frac{N}{2}-1+a^o_\beta}\exp\left[-\beta_k(b^o_\beta+\frac{\mathrm{E}_{\mathbf{x},\mathbf{s}_k}\left[\|\,\mathbf{y}_k-\mathbf{B}(\mathbf{s}_k)\mathbf{x}\,\|^2\right]}{2})\right] \quad (35)$$

In Eq. (35) appears the expectation value

$$\mathrm{E}_{\mathbf{x},\mathbf{s}_k}\left[\|\,\mathbf{y}_k-\mathbf{B}(\mathbf{s}_k)\mathbf{x}\,\|^2\right]=\mathrm{E}_{\mathbf{x}}\left[\mathrm{E}_{\mathbf{s}_k}\left[\|\,\mathbf{y}_k-\mathbf{B}(\mathbf{s}_k)\mathbf{x}\,\|^2\right]\right] \quad (36)$$

which, as the ones above mentioned, is not easy to evaluate, and also benefits from the expansion of Eq. (30).

The proposed algorithm is summarized below in Algorithm 1.

Algorithm 1 Variational Bayesian Super Resolution

Calculate initial estimates of the HR image and hyperparameters
while convergence criterion is not met **do**
1. Estimate the HR image distribution using Eq. (28).
2. Compute spatial adaptivity vector $\mathbf{u}^d$, for $d\in\{h,v\}$, using Eq. (32).
3. Estimate the distribution of the registration parameters using Eq. (29).
4. Estimate the distributions of the hyperparameters α_1, α_2 and $\{\beta_k\}$ using Eqs. (33), (34) and (35).

5. EXPERIMENTAL RESULTS

In this section, we evaluate the performance of the proposed algorithm. The quality of the restored HR image is measured quantitatively by the peak signal-to-noise ratio (PSNR), which is defined

Table 2. Mean PSNRs with standard deviations provided by the SR algorithms at different SNR levels for the disk image

SNR	10dB	15dB	20dB
Bicubic	17.61 $\pm 1.0\ 10^{-4}$	17.61 $\pm 1.0\ 10^{-4}$	17.61 $\pm 1.0\ 10^{-4}$
ZMT	19.3 ± 0.5	18.69 ± 0.03	18.7 b± 0.3
RSR	20.06 ± 0.08	18.75 ± 0.28	18.75 ± 0.28
TV	35.99 $\pm 2.0\ 10^{-4}$	40.45 ± 0.01	45.70 $\pm 2.0\ 10^{-3}$
SAR	31.54 $\pm 1.0\ 10^{-3}$	35.81 $\pm 2.0\ 10^{-4}$	39.76 $\pm 4.0\ 10^{-3}$
L1	37.11 $\pm 4.0\ 10^{-4}$	41.49 $\pm 9.0\ 10^{-4}$	45.70 $\pm 6.0\ 10^{-4}$
ALG1(λ_1)	37.27 $\pm 3.0\ 10^{-4}$	42.40 $\pm 5.0\ 10^{-4}$	46.32 ± 0.0043
λ_1	0.95	0.85	0.85

Table 3. Mean PSNRs with standard deviations provided by the SR algorithms at different SNR levels for the Lena image with a downsampling factor of four

SNR	10dB	15dB	20dB
Bicubic	19.19 $\pm 2.0\ 10^{-4}$	19.18 $\pm 1.0\ 10^{-4}$	19.19 $\pm 1.0\ 10^{-4}$
ZMT	23.71 0.18	23.57 0.16	24.01 0.03
RSR	23.37 $pm 0.18$	23.34 $pm 0.12$	23.72 $pm 0.09$
TV	28.16 $\pm 6.0\ 10^{-3}$	31.27 $\pm 3.0\ 10^{-3}$	33.53 $\pm 4.0\ 10^{-3}$
SAR	28.15 $\pm 3.0\ 10^{-3}$	29.62 $\pm 2.0\ 10^{-3}$	30.38 $\pm 1.0\ 10^{-4}$
L1	28.70 $\pm 4.0\ 10^{-3}$	31.69 $\pm 1.5\ 10^{-2}$	33.91 $\pm 2.0\ 10^{-4}$
ALG1(λ_1)	29.49 $\pm 4.0\ 10^{-3}$	32.45 $\pm 1.5\ 10^{-3}$	34.95 $\pm 6.0\ 10^{-4}$
λ_1	0.8	0.9	0.95

as $\mathrm{PSNR} = 10 \log_{10} \frac{NP}{\|\hat{\mathbf{x}}-\mathbf{x}\|^2}$, where $\hat{\mathbf{x}}$ and $\mathbf{x}$ are the estimated and original HR images, respectively, and pixel values in both images are normalized to lie in the interval $[0, 1]$. We also provide examples of the estimated HR images to assess their visual quality.

In all experiments reported below, the initial values of Algorithm 1 are chosen as follows: The HR image estimate is initialized using the *average image* [15], which is an oversmooth estimate of the HR image obtained using the LR images as $\mathbf{x}_a = \mathbf{S}^{-1} \sum_{k=1}^{L} \mathbf{B}(\mathbf{s}_k)^T \mathbf{y}_k$, where $\mathbf{S}$ is a diagonal matrix with the column sums of $\mathbf{B}_k$ as its elements. Note that this initial estimate is calculated very efficiently, and it generally increases the robustness of the algorithm to noise.

The inverse covariance matrices $(\mathbf{\Xi}_k^p)^{-1}$ are set equal to zero matrices, that is, no prior information is utilized about the uncertainty of motion vectors. The covariance matrices in Algorithm 1 are initially set equal to zero. The rest of the algorithm parameters are automatically calculated from the initial HR image estimate using the algorithmic steps provided in Algorithm 1. As convergence criterion we used $\|\mathbf{x}^n - \mathbf{x}^{n-1}\|^2 / \|\mathbf{x}^{n-1}\|^2 < 10^{-5}$, where $\mathbf{x}^n$ and $\mathbf{x}^{n-1}$ are the image estimates at the n-th and $(n-1)$-st iterations, respectively.

When we set $\lambda_1 = 0$ in Eq. (28), our prior model becomes a *SAR*. On the other hand, when we set $\lambda_1 = 1$ in Eq. (28), our models coincides with the Variational Super-Resolution method in [20] (denoted by *VSRℓ1*). The optimal value of λ_1, for our Algorithm 1, is found here experimentally, and the obtained reconstruction will be denoted by *ALG1*(λ_1). In this section, we evaluate the performance of the proposed algorithm *ALG1*(λ_1) in comparison with the following methods: 1) Bicubic interpolation, 2) the robust SR method in [21] (denoted by *ZMT*), which is based on backprojection with median filtering, 3) the robust SR method in [6] (denoted by *RSR*), which is based on bilateral TV priors, 4) *SAR*, 5) Variational Super-Resolution method in [2] (denoted by *VSR*) and 6) *VSRℓ1*.

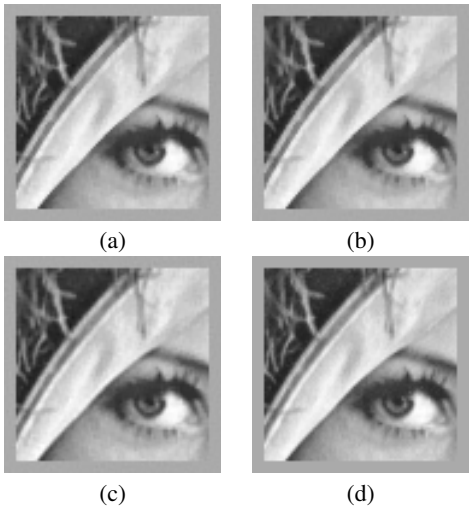

Fig. 1. Example estimated HR images from different SR methods in the case when SNR=15dB. Results of (a) *VSR* ($\mathrm{PSNR} = 36.39\mathrm{dB}$), (b) *VSR$\ell$1* ($\mathrm{PSNR} = 35.97\mathrm{dB}$), (c) *ALG1*(0.85) ($\mathrm{PSNR} = 37.67\mathrm{dB}$) and (d) the original image.

We generated sets of 5 synthetic LR images from 80×80 fragments of the HR Lena image, and for the disc image in [5], through warping, blurring and downsampling by a factor of 2. The warping consisted of translations of $(0,0)^t$, $(0,0.5)^t$, $(0.5,0)^t$, $(1,0)^t$ and $(0,1)^t$ pixels respectively, and rotations of angles $(0^\circ, 3^\circ, -3^\circ, 5^\circ, -5^\circ)$. As blur we used a 3×3 uniform PSF. The LR images obtained after the warping, blurring and downsampling operations are further degraded by additive white Gaussian noise at SNR levels of 10 dB, 15 dB, and 20 dB.

We conducted simulations with 3 different noise realizations at each SNR level, and the average PSNR and standard deviations of these experiments are shown in Table 1 for the Lena image, and in Table 2 for the disc image. Registration parameters were estimated with errors for the displacements c_k, d_k, and rotation angles θ_k, that were less than 0.02 and 0.008° for 10 dB, 0.003 and 0.001° for 15 dB, and 0.001 and 0.0008° for 20 dB. As expected, all SR algorithms result in better reconstructions than bicubic interpolation. It is also clear that the proposed method provides the best performance among all methods across all noise levels.

Examples of HR restorations are shown in Fig. (1) for the SNR = 15 dB degradation of the Lena image. It is clear that *VSR*, *VSRℓ1* and the proposed method provide the most visually enhanced restorations with significantly reduced ringing artifacts and much sharper edges compared to the other methods.

Finally, Fig. (2) shows the variation of the PSNR of the restoration obtained with *ALG1*(λ_1) when λ_1 changes from $\lambda_1 = 0$ (SAR) to $\lambda_1 = 1$ (VSRℓ1), reaching its maximum value at $\lambda_1 = 0.80$, for the Lena image with SNR=15 dB. This is an example of non-degenerate combination of divergences providing better reconstructions than degenerate ones.

We also performed simulations, for a downsampling factor

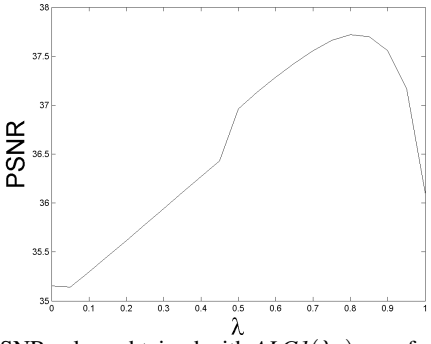

Fig. 2. PSNR values obtained with *ALG1*(λ_1), as a function of λ_1, for the SNR=15 dB Lena image.

of four, with the Lena image. We generated sets of 16 synthetic LR images through warping, blurring and downsampling by a factor of 4.The warping consisted of translations within the interval $[-(1,1)^t,(1,1)^t]$ pixels, and rotations within the interval $[-5^\circ,5^\circ]$. As blur we used a 5×5 uniform PSF. The LR images obtained after the warping, blurring and downsampling operations are further degraded by additive white Gaussian noise at SNR levels of 10 dB, 15 dB, and 20 dB. The obtained numerical results, at these SNR levels are shown in Table 3.

6. CONCLUSIONS

In this paper, we apply to SR image reconstruction of rotated and displaced LR images, a novel variational Bayesian methodology for the combination of image priors. A sparse prior, based on the $\ell 1$ norm of the horizontal and vertical differences between image pixel values, and a non-sparse one have been combined. Estimation of the registration parameter values is performed at the HR level, jointly with reconstruction. The methodology is based on finding the distribution on the HR image given the observations, which minimizes a linear convex combination of the KL divergences associated with each pair of observation and prior models. We have found this distribution in closed form. The HR image estimates obtained from the proposed method compare favorably with the images provided by other SR reconstruction methods. Their superiority over the results obtained when each prior is used independently has also been established experimentally. Future work will address the estimation of the weights assigned to each KL divergence in the convex combination.

7. REFERENCES

[1] S. D. Babacan, R. Molina, and A. K. Katsaggelos. Parameter estimation in TV image restoration using variational distribution approximation. *IEEE Trans. Image Process.*, 17(3):326–339, March 2008.

[2] S. D. Babacan, R. Molina, and A. K. Katsaggelos. Total variation super resolution using a variational approach. In *IEEE ICIP*, October 2008.

[3] G. Chantas, N. Galatsanos, R. Molina, and A. K. Katsaggelos. Variational bayesian image restoration with a spatially adaptive product of total variation image priors. *IEEE Trans. Image Process.*, 19(2):351–362, Feb. 2010.

[4] G. Chantas, N. P. Galatsanos, A. Likas, and M. Saunders. Variational bayesian image restoration based on a product of t-distributions image prior. *IEEE Trans. Image Process.*, 17(10):1795–1805, October 2008.

[5] S. Farsiu. *MDSP resolution enhancement software*. University of California at Santa Cruz, 2004.

[6] S. Farsiu, M. D. Robinson, M. Elad, and P. Milanfar. Fast and robust multiframe super resolution. *IEEE Trans. Image Process.*, 13(10):1327–1344, Oct. 2004.

[7] R. Hardie, K. Barnard, and E. Armstrong. Joint MAP registration and high-resolution image estimation using a sequence of undersampled images. *IEEE Trans. Image Process.*, 6(12):1621–1633, 1997.

[8] Y. He, K. H. Yap, L. Chen, and L. P. Chau. A nonlinear least square technique for simultaneous image registration and super-resolution. *IEEE Trans. Image Process.*, 16(11):2830–2841, 2007.

[9] G. Hinton. Training products of experts by minimizing contrastive divergence. *Neural Computation*, 14(8):1771–1800, 2002.

[10] A. K. Katsaggelos, R. Molina, and J. Mateos. *Super Resolution of Images and Video*. Morgan and Claypool, 2007.

[11] B. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *Proceedings of Imaging Understanding Workshop*, pages 121–130, 1981.

[12] P. Milanfar. *Super-Resolution Imaging*. Digital Imaging and Computer Vision. Taylor&Francis/CRC Press (Ed.), 2010.

[13] R. Molina, M. Vega, J. Abad, and A. K. Katsaggelos. Parameter estimation in bayesian high-resolution image reconstruction with multisensors. *IEEE Trans. Image Process.*, 12(12):1655–1667, Dec. 2003.

[14] L. C. Pickup, D. P. Capel, S. J. Roberts, and A. Zisserman. Bayesian image super-resolution, continued. In *Advances in Neural Information Processing Systems*, Dec. 2006.

[15] L. C. Pickup, D. P. Capel, S. J. Roberts, and A. Zisserman. Bayesian methods for image super-resolution. *The Computer Journal*, 2007.

[16] S. Roth and M. J. Black. Fields of experts: A framework for learning image priors. In *In CVPR*, pages 860–867, 2005.

[17] J. Starck, M. Elad, and D. Donoho. Image decomposition via the combination of sparse representation and a variational approach. *IEEE Trans. Image Process.*, 14(10):1570–1582, 2005.

[18] D. Sun and 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):2743–2751, 2007.

[19] M. Tipping and C. Bishop. Bayesian image super-resolution. In S. Thrun, S. Becker, and K. Obermayer, editors, *Advances in Neural Information Processing Systems 15*, pages 1279–1286, Cambridge, MA, 2003. MIT Press.

[20] S. Villena, M. Vega, R. Molina, and A. K. Katsaggelos. Bayesian super-resolution image reconstruction using an l1 prior. In *6th ISPA Proc.*, pages 152–157, 2009.

[21] A. Zomet, A. Rav-Acha, and S. Peleg. Robust super-resolution. In *IEEE CVPR*, pages 645–650, 2001.