IMAGE PRIOR COMBINATION IN SUPER-RESOLUTION IMAGE RECONSTRUCTION

Salvador Villena[(a)], Miguel Vega[(a)], Rafael Molina[(b)], Aggelos K. Katsaggelos[(c)]

[(a)] Dept. de Lenguajes y Sistemas
Informáticos
Univ. de Granada
18071 Granada, Spain
Email: {svillena,mvega}@ugr.es

[(b)] Dept. Ciencias de la Computación
e Inteligencia Artificial
Univ. de Granada
18071 Granada, Spain
Email: rms@decsai.ugr.es

[(c)] Dept. of Electrical Engineering
and Computer Science
Northwestern Univ.
Evanston, Illinois 60208-3118
Email: aggk@eecs.northwestern.edu

ABSTRACT

In this paper a new combination of image priors is introduced and applied to *Super Resolution (SR)* image reconstruction. A sparse image prior based on the $\ell 1$ norms of 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-Leiback-(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 [7–9] 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 [12], [14] 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 [12] and [14] 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 develop a variational Bayesian methodology for the combination of the sparse prior, proposed in [15], based on the $\ell 1$ norm of the horizontal and vertical differences between image pixel values, and a non-sparse SAR prior. The novel method is applied to SR image reconstruction.

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, which is modeled as

$$\mathbf{y}_k = \mathbf{A}\mathbf{H}_k\mathbf{D}_k\mathbf{x} + \mathbf{n}_k = \mathbf{B}_k\mathbf{x} + \mathbf{n}_k, \qquad (1)$$

where $\mathbf{A}$ is the $N \times PN$ downsampling matrix, $\mathbf{H}_k$ is the $PN \times PN$ blurring matrix, $\mathbf{D}_k$ is the $PN \times PN$ warping matrix and $\mathbf{n}_k$ is the $N \times 1$ acquisition noise. In this work, we assume that the blurring $\mathbf{H}_k$ matrices as well as the warping matrices $\mathbf{D}_k$ are known. The effects of downsampling, blurring, and warping can be combined into a single $N \times PN$ system matrix $\mathbf{B}_k$. 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 $\mathbf{x}$.

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

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

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

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

As we have already explained in the introduction, in this paper for simplicity we are combining only a sparse prior, the $\ell 1$ model [15], and a non-sparse one, the simultaneous autoregression (SAR) model [10]. Note that the idea of combining sparse and non-sparse models has also been proposed in other contexts, see for instance [13]. The $\ell 1$ prior, which is very effective in preserving edges while

imposing smoothness is defined by

$$p_1(\mathbf{x}|\alpha_1) = \frac{1}{Z(\alpha_1)} \times$$
$$\exp\left\{-\sum_{i=1}^{PN}\left[\alpha_1^h \parallel \Delta_i^h(\mathbf{x}) \parallel_1 + \alpha_1^v \parallel \Delta_i^v(\mathbf{x}) \parallel_1\right]\right\}, \quad (4)$$

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

$$Z(\alpha_1) \propto (\alpha_1^h \alpha_1^v)^{-\frac{PN}{4}}. \quad (5)$$

The use in Eq. (4) 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

$$p_2(\mathbf{x}|\alpha_2) \propto \alpha_2^{\frac{PN}{2}} \exp\left\{-\frac{\alpha_2}{2} \parallel \mathbf{C}\mathbf{x}\parallel^2\right\}, \quad (6)$$

where $\mathbf{C}$ 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

$$p(\mathbf{x}|\alpha_1, \alpha_2) = \frac{1}{Z(\alpha_1, \alpha_2)}$$
$$\exp\left\{-\sum_{i=1}^{PN}\left[\alpha_1^h \parallel \Delta_i^h(\mathbf{x}) \parallel_1 + \alpha_1^v \parallel \Delta_i^v(\mathbf{x}) \parallel_1\right] - \frac{\alpha_2}{2} \parallel \mathbf{C}\mathbf{x}\parallel^2\right\}, \quad (7)$$

but since there is no known approximation to the partition function $Z(\alpha_1, \alpha_2)$, the estimation of the parameters would be impossible with this prior model $p(\mathbf{x}|\alpha_1, \alpha_2)$ (see however [6] in the context of model learning).

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,

$$p(\omega) = \Gamma(\omega|a_\omega^o, b_\omega^o) = \frac{(b_\omega^o)^{a_\omega^o}}{\Gamma(a_\omega^o)} \omega^{a_\omega^o - 1} \exp\left[-b_\omega^o \omega\right], \quad (8)$$

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) and (8), with the two different prior models we obtain the joint probability distributions

$$p_i(\mathbf{y}, \Omega, \alpha_i) = p(\mathbf{y}|\mathbf{x}, \boldsymbol{\beta}) \prod_{k=1}^{L} p(\beta_k) p_i(\mathbf{x}|\alpha_i) p(\alpha_i), \quad (9)$$

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

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 $p(\Theta \mid \mathbf{y})$. We propose here to approximate this distribution by the distribution minimizing the following linear convex combination of KL divergence measures

$$\hat{q}(\Theta) = \underset{q(\Theta)}{\operatorname{argmin}} \sum_{i=1}^{2} \lambda_i C_{KL}(q(\Omega)q(\alpha_i) \parallel p_i(\Omega, \alpha_i|\mathbf{y})) \quad (10)$$

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

$$q(\Omega) = q(\mathbf{x}) \prod_{k=1}^{L} q(\beta_k), \quad q(\Theta) = q(\Omega) \prod_{i=1}^{2} q(\alpha_i), \quad (11)$$

and the KL divergences are given by

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

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

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

expression (10) can be rewritten in the more compact form as

$$\hat{q}(\Theta) = \underset{q(\Theta)}{\operatorname{argmin}} \int q(\Theta)$$
$$\log\left(\frac{q(\Omega)}{p(\mathbf{y}|\mathbf{x}, \{\beta_k\})\prod_{k=1}^{L} p(\beta_k)} \prod_{i=1}^{2}\left[\frac{q(\alpha_i)}{p_i(\mathbf{x}|\alpha_i)p(\alpha_i)}\right]^{\lambda_i}\right) d\Theta. \quad (14)$$

Unfortunately, we can not directly tackle the minimization of (14) because of the $\ell 1$ image prior $p_1(\mathbf{x}|\alpha_1)$ of Eq. (4). 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 (9) which makes the minimization of (14) tractable. Let us first consider the following functional $M(\alpha_1, \mathbf{x}, \mathbf{w})$, 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 $\mathbf{u}_i^v$, $i = 1, \ldots, PN$,

$$M(\alpha_1, \mathbf{x}, \mathbf{w}) \propto (\alpha_1^h \alpha_1^v)^{\frac{PN}{4}} \times$$
$$\exp\left\{-\sum_{i=1}^{PN}\left[\alpha_1^h \frac{(\Delta_i^h(\mathbf{x}))^2 + \mathbf{u}_i^h}{2\sqrt{\mathbf{u}_i^h}} + \alpha_1^v \frac{(\Delta_i^v(\mathbf{x}))^2 + \mathbf{u}_i^v}{2\sqrt{\mathbf{u}_i^v}}\right]\right\}. \quad (15)$$

It can be shown (details can be found in [1]) that the functional $M(\alpha_1, \mathbf{x}, \mathbf{w})$ is a lower bound of the image prior $p_1(\mathbf{x}|\alpha_1)$, that is,

$$p_1(\mathbf{x}|\alpha_1) \geq M(\alpha_1, \mathbf{x}, \mathbf{w}). \quad (16)$$

This lower bound can be used to find a lower bound for the joint distribution, for $i = 1$, in (9), that is,

$$p_1(\mathbf{y}, \Omega, \alpha_1) \geq p(\mathbf{y}|\mathbf{x}, \boldsymbol{\beta}) \prod_{k=1}^{L} p(\beta_k) M(\alpha_1, \mathbf{x}, \mathbf{w}) p(\alpha_1)$$
$$= F(\Omega, \alpha_1, \mathbf{w}, \mathbf{y}), \quad (17)$$

which results in an upper bound of the KL distance as

$$C_{KL}(q(\Omega, \alpha_1) \parallel p_1(\mathbf{y}, \Omega, \alpha_1)) \leq C_{KL}(q(\Omega, \alpha_1) \parallel F(\Omega, \alpha_1, \mathbf{w}, \mathbf{y})). \quad (18)$$

The minimization of $C_{KL}(q(\Omega, \alpha_1) \parallel p_1(\mathbf{y}, \Omega, \alpha_1))$ can then be replaced by the minimization of its upper bound (18), 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 (18) is quadratic and therefore it can be evaluated analytically.

Before we proceed to calculate the posterior approximation, we first observe that to calculate $q(\alpha_i), i = 1, 2$ we only have to look at

the only divergence where that distribution is present. So we can write

$$q(\alpha_1) = \text{const} \times \exp\left(\langle \log F(\boldsymbol{\Omega}, \alpha_1, \mathbf{w}, \mathbf{y})\rangle_{\Omega}\right), \quad (19)$$

$$q(\alpha_2) = \text{const} \times \exp\left(\langle \log p_2(\mathbf{y}, \boldsymbol{\Omega}, \alpha_2)\rangle_{\Omega}\right), \quad (20)$$

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

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

$$q(\xi) = \text{const} \times \exp\left(\left\langle \log\left[p(\mathbf{y}|\mathbf{x}, \boldsymbol{\beta}) \prod_{k=1}^{L} p(\beta_k)\right.\right.\right.$$

$$\left.\left.\left.\times [M(\alpha_1, \mathbf{x}, \mathbf{w})\, p(\alpha_1)]^{\lambda_1} [p_2(\mathbf{x}|\alpha_2)p(\alpha_2)]^{1-\lambda_1}\right]\right\rangle_{\Theta_\xi}\right), \quad (21)$$

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. (21), $q(\mathbf{x})$ is the multivariate Gaussian

$$q(\mathbf{x}) \propto \exp\left(-\frac{1}{2}\lambda_1\left\{<\alpha_1^h>\sum_i \frac{(\Delta_i^h(\mathbf{x}))^2 + u_i^h}{\sqrt{u_i^h}}\right.\right.$$

$$\left.+<\alpha_1^v>\sum_i \frac{(\Delta_i^v(\mathbf{x}))^2 + u_i^v}{\sqrt{u_i^v}}\right\} - \frac{(1-\lambda_1)}{2}<\alpha_2>\|\mathbf{C}\mathbf{x}\|^2$$

$$\left.-\frac{1}{2}\sum_k <\beta_k>\|\mathbf{y}_k - \mathbf{A}\mathbf{H}_k\mathbf{D}_k\mathbf{x}\|^2\right). \quad (22)$$

Furthermore, the elements of $\mathbf{u}^d$, for $d \in \{h, v\}$, in Eq. (15) are equal to

$$u_i^d = E_{\mathbf{x}}[(\Delta_i^d(\mathbf{x}))^2]. \quad (23)$$

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

$$q(\alpha_1^d) \propto \alpha_1^{d \frac{PN}{4}-1+a^o_{\alpha_1^d}} \exp\left[-\alpha_1^d\left(b^o_{\alpha_1^d} + \sum_i \sqrt{u_i^d}\right)\right], \quad (24)$$

$$q(\alpha_2) \propto \alpha_2^{PN/2-1+a^o_{\alpha_2}} \exp\left[-\alpha_2\left(b^o_{\alpha_2} + \frac{E_{\mathbf{x}}\left[\|\mathbf{C}\mathbf{x}\|^2\right]}{2}\right)\right], \quad (25)$$

$$q(\beta_k) \propto \beta_k^{N/2-1+a^o_{\beta}} \exp\left[-\beta_k\left(b^o_{\beta} + \frac{E_{\mathbf{x}}\left[\|\mathbf{y}_k-\mathbf{B}_k\mathbf{x}\|^2\right]}{2}\right)\right] \quad (26)$$

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. (22).
   2. Compute spatial adaptivity vector $\mathbf{u}^d$, for $d \in \{h, v\}$, using Eq. (23).
   3. Estimate the distributions of the hyperparameters α_1, α_2 and $\{\beta_k\}$ using Eqs. (24), (25) and (26).

---

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 as $\text{PSNR} = 10\log_{10}\frac{NP}{\|\hat{\mathbf{x}}-\mathbf{x}\|^2}$, 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* [11], which is an oversmooth estimate of the HR image obtained using the LR images as $\mathbf{x}_a = \mathbf{S}^{-1}\sum_{k=1}^{L}\mathbf{B}_k^T\mathbf{y}_k$, 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 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. (22), our prior model becomes a *SAR*. On the other hand, when we set $\lambda_1 = 1$ in Eq. (22), our models coincides with the Variational Super-Resolution method in [15] (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 [16] (denoted by *ZMT*), which is based on backprojection with median filtering, 3) the robust SR method in [5] (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*.

We generated sets of 5 synthetic LR images from 80×80 fragments of an HR disk image and also of the Lena image, 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 $(0°, 3°, -3°, 5°, -5°)$. 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 Tables 1 and 2. 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 disk image and in Fig. (2) 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.

Fig. (3)(a) 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 $\lambda_1 = 0.85$, for the disk image with SNR=15 dB. Fig. (3)(b) shows this same variation for the SNR=15 dB Lena image, with its maximum value at $\lambda_1 = 0.85$. Both cases are examples of non-degenerate combinations of divergences providing better reconstructions than degenerate ones. Finally, table 3 shows required CPU time, as well as the number of iterations, by our implementation of the different methods, on a Xeon 5460 3.16 GHz processor, when processing the SNR=20dB Lena image.
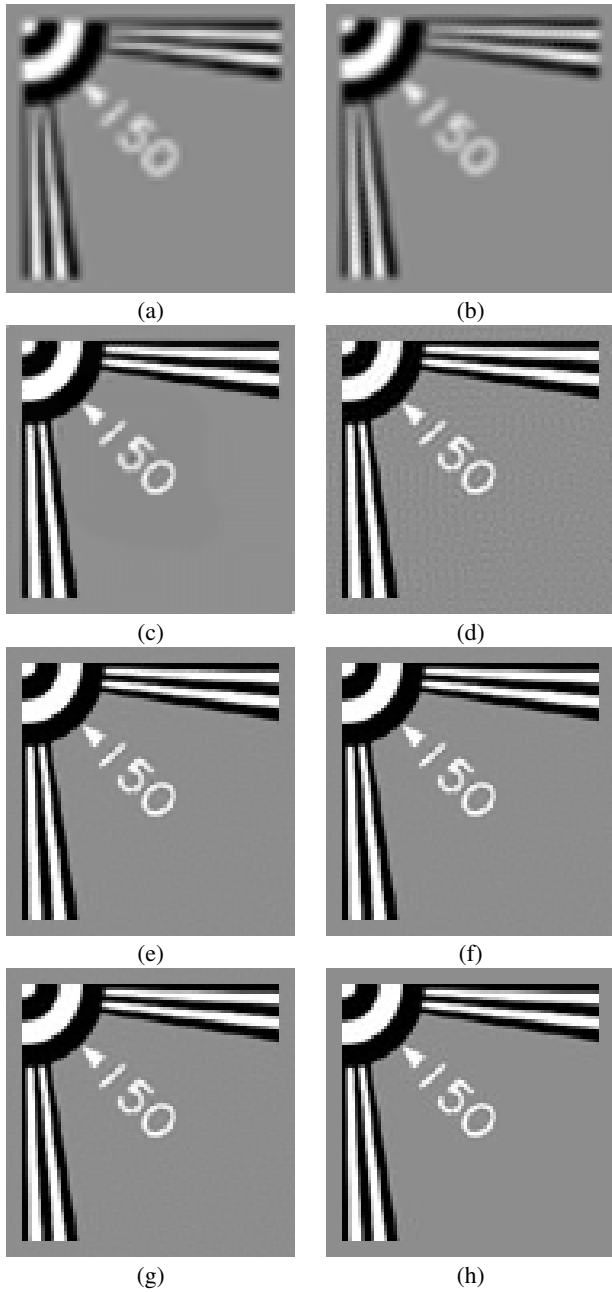
Figure 1: Example estimated HR disk images from different SR methods for the case when SNR=15dB. Results of (a) Bicubic interpolation (PSNR $= 17.61$ dB), (b) *ZMT* (PSNR $= 19.24$ dB), (c) *RSR* (PSNR $= 34.05$ dB), the proposed methods: using (d) *SAR* (PSNR $= 35.98$ dB), (e) *VSR* (PSNR $= 41.03$ dB), (f) *VSRℓ1* (PSNR $= 41.69$ dB), (g) *ALG1*(0.85) (PSNR $= 42.33$ dB) and (h) the original image.
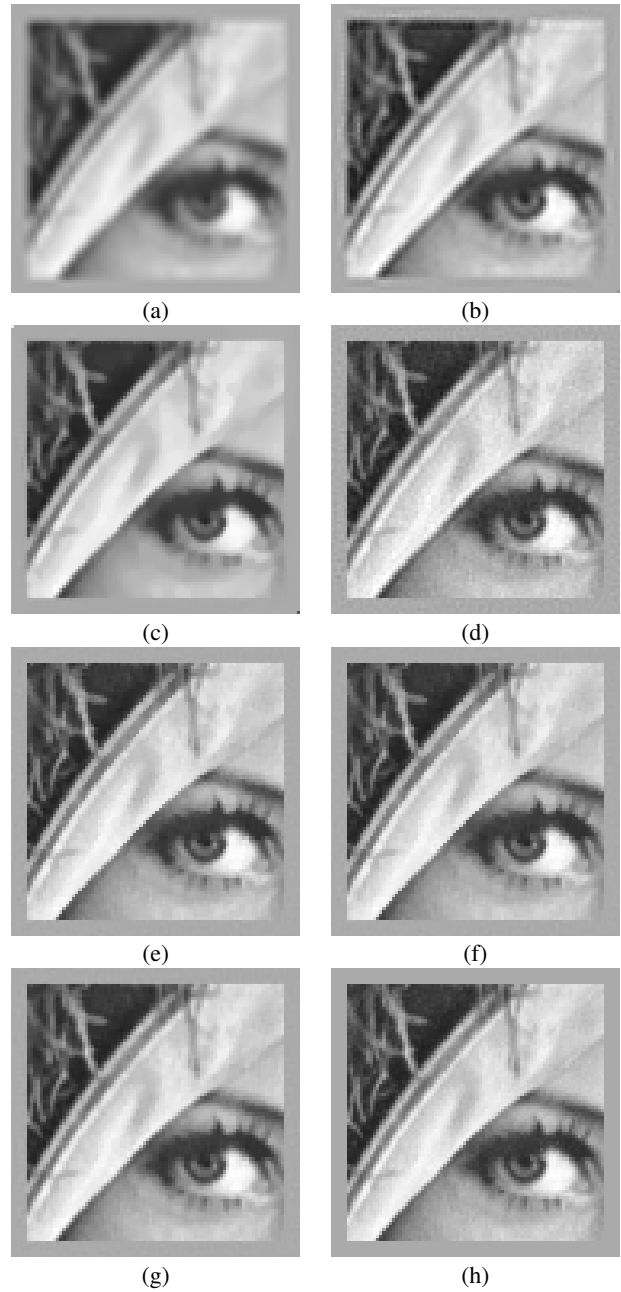
Figure 2: Example estimated HR Lena images from different SR methods for the case when SNR=15dB. Results of (a) Bicubic interpolation (PSNR $= 24.27$ dB), (b) *ZMT* (PSNR $= 30.51$ dB), (c) *RSR* (PSNR $= 33.55$ dB), the proposed methods: using (d) *SAR* (PSNR $= 35.80$ dB), (e) *VSR* (PSNR $= 36.59$ dB), (f) *VSRℓ1* (PSNR $= 36.44$ dB), (g) *ALG1*(0.85) (PSNR $= 37.88$ dB) and (h) the original image.

Table 1: 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 ± 0.00	17.61 ± 0.00	17.61 ± 0.00
ZMT	19.19 ± 0.00	19.22 ± 0.00	19.24 ± 0.00
RSR	34.31 ± 0.012	34.11 ± 0.004	34.10 ± 0.039
SAR	31.67 ± 0.005	36.01 ± 0.002	40.23 ± 0.005
VSR	36.41 ± 0.001	40.88 ± 0.012	46.02 ± 0.014
VSR$\ell 1$	37.54 ± 0.00	41.60 ± 0.006	46.49 ± 0.001
ALG1(λ_1)	37.70 ± 0.00	42.25 ± 0.003	46.95 ± 0.00
λ_1	0.95	0.85	0.85

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

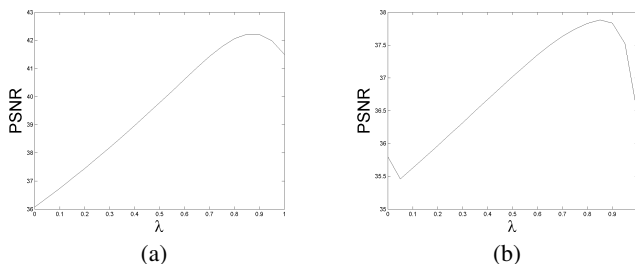| SNR	10dB	15dB	20dB
Bicubic	24.26 ± 0.000	24.27 ± 0.000	24.27 ± 0.000
ZMT	30.33 ± 0.005	30.57 ± 0.002	30.46 ± 0.000
RSR	33.39 ± 0.010	33.55 ± 0.003	33.53 ± 0.002
SAR	32.42 ± 0.0005	35.76 ± 0.010	39.76 ± 0.003
VSR	34.00 ± 0.001	36.63 ± 0.002	40.41 ± 0.023
VSR$\ell 1$	34.05 ± 0.005	36.46 ± 0.001	40.16 ± 0.023
ALG1(λ_1)	34.47 ± 0.003	37.86 ± 0.001	41.36 ± 0.005
λ_1	0.9	0.85	0.8

(a) (b)

Figure 3: PSNR values obtained with $ALG1(\lambda_1)$, as a function of λ_1, (a) for the SNR=15 dB disk image, and (b) for the SNR=15 dB Lena image.

Table 3: CPU time, in seconds, and number of iterations needed by the different SR algorithms for the SNR=20dB Lena image.

	ZMT	RSR	SAR	VSR	VSR$\ell 1$	ALG1
Time	14	17	172	153	165	142
N Iter.	300	300	66	58	63	55

6. CONCLUSIONS

In this paper a new combination of image priors has been introduced, and applied to SR image reconstruction. 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. 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.

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 Inter. Conf. on Image Process. 2008*, 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, M. D. Robinson, M. Elad, and P. Milanfar. Fast and robust multiframe super resolution. *IEEE Trans. Image Process.*, 13(10):1327–1344, Oct. 2004.

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

[7] A. K. Katsaggelos and R. Molina, editors. *Super resolution (special issue)*. The Computer Journal, 52, 395-396, 2009.

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

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

[10] 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.

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

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

[13] 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.

[14] 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.

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

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