Using the Kullback-Leibler Divergence to Combine Image Priors in Super-Resolution Image Reconstruction

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

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

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

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

Abstract—This paper is devoted to the combination of image priors in Super Resolution (SR) image reconstruction. Taking into account that each combination of a given observation model and a prior model 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 minimizes a linear convex combination of the Kullback-Leibler divergences associated with each posterior distribution. We find this distribution in closed form and also relate the proposed approach to other prior combination methods in the literature. The estimated HR images are compared with images provided by other SR reconstruction methods.

Index Terms—Super resolution, combination of priors, variational methods, parameter estimation, Bayesian methods.

I. INTRODUCTION

Image SR refers to the process of obtaining an HR image from a set of degraded Low Resolution (LR) images. The basic principle in SR is that changes in the LR images caused by the blur and the (camera or scene) motion provide additional information that can be utilized to reconstruct the HR image from the set of LR observations.

SR image reconstruction is still an open and widely investigated area (see, for instance, [1]–[3]) where, as in many recovery problems, the selection of priors is crucial to obtain a good reconstruction. While in image restoration there have been several recent attempts to combine image priors (see [4] and references therein), no such attempts have been made in the SR literature.

In this paper, we develop a novel variational Bayesian methodology to combine prior models in SR image reconstruction. While the methodology can be developed more generally we present it here for simplicity for the combination of a sparse and a non-sparse image prior. We also show that this methodology applied to image restoration produces, as a special case, the model developed in [4].

The rest of this paper is organized as follows. Section II provides the mathematical model for the LR image acquisition process. We provide the description of the hierarchical Bayesian framework modeling the unknowns in Section III. The inference procedure which develops the proposed method is presented in Section IV. We demonstrate the effectiveness of the proposed approach with experimental results in Section V. Finally, section VI concludes the paper.

II. 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

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

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 $\mathbf{x}$.

III. 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

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

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

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

As we have already explained in the introduction, in this paper for simplicity we assume that we want to combine only a sparse prior, the TV model [5], and a non-spare one, the simultaneous autoregression (SAR) model [6]. Note that the idea of combining sparse and non-sparse models has also been proposed in other contexts, see for instance [7]. The TV prior, which is very effective in preserving edges while imposing smoothness is defined by

$$p_1(\mathbf{x}|\alpha_1) \propto c\,\alpha_1^{PN/2} \exp\left[-\frac{1}{2}\alpha_1 \mathrm{TV}(\mathbf{x})\right], \tag{4}$$

where c is a constant and

$$\text{TV}(\mathbf{x}) = \sum_{i=1}^{PN} \sqrt{(\Delta_i^h(\mathbf{x}))^2 + (\Delta_i^v(\mathbf{x}))^2}. \quad (5)$$

The operators $\Delta_i^h(\mathbf{x})$ and $\Delta_i^v(\mathbf{x})$ correspond, respectively, to the horizontal and vertical first order differences at pixel i.

We also consider the SAR prior, defined as

$$\text{p}_2(\mathbf{x}|\alpha_2) \propto \alpha_2^{\frac{PN}{2}} \exp\left\{-\frac{\alpha_2}{2} \|\mathbf{C}\mathbf{x}\|^2\right\}, \quad (6)$$

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

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

$$\text{p}(\mathbf{x}|\alpha_1, \alpha_2) = \frac{1}{Z(\alpha_1, \alpha_2)} \exp\left\{-\alpha_1 \text{TV}(\mathbf{x}) - \frac{\alpha_2}{2} \|\mathbf{C}\mathbf{x}\|^2\right\}, \quad (7)$$

but since there is no known approximation to the partition function $Z(\alpha_1, \alpha_2)$, the estimation of these parameters would be impossible if we use the prior model $\text{p}(\mathbf{x}|\alpha_1, \alpha_2)$ (see however [8] 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,

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Unfortunately, we can not directly tackle the minimization of (14) because of the TV image prior $\text{p}_1(\mathbf{x}|\alpha_1)$ of Eq. (4). This difficulty is overcome in this paper by resorting to the majorization-minimization (MM) approach (see [9]).

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 $\text{M}(\alpha_1, \mathbf{x}, \mathbf{w})$, where $\mathbf{w} \in (R^+)^{PN}$ with components w_i, $i = 1, \dots, PN$,

$$\text{M}(\alpha_1, \mathbf{x}, \mathbf{w}) \propto \alpha_1^{PN/2} \exp\left[-\frac{\alpha_1}{2} \sum_i \frac{(\Delta_i^h(\mathbf{x}))^2 + (\Delta_i^v(\mathbf{x}))^2 + w_i}{\sqrt{w_i}}\right]. \quad (15)$$

It can be shown (details can be found in [9]) that the functional $\text{M}(\alpha_1, \mathbf{x}, \mathbf{w})$ is a lower bound of the image prior $\text{p}_1(\mathbf{x}|\alpha_1)$, that is,

$$\text{p}_1(\mathbf{x}|\alpha_1) \geq \text{M}(\alpha_1, \mathbf{x}, \mathbf{w}). \quad (16)$$

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

$$\text{p}_1(\mathbf{y}, \Omega, \alpha_1) \geq \text{p}(\mathbf{y}|\mathbf{x}, \boldsymbol{\beta}) \prod_{k=1}^{L} \text{p}(\beta_k) \text{M}(\alpha_1, \mathbf{x}, \mathbf{w}) \text{p}(\alpha_1)$$
$$= \text{F}(\Omega, \alpha_1, \mathbf{w}, \mathbf{y}), \quad (17)$$

which results in an upper bound of the KL distance as

$$C_{KL}(\text{q}(\Omega, \alpha_1) \parallel \text{p}_1(\mathbf{y}, \Omega, \alpha_1)) \leq C_{KL}(\text{q}(\Omega, \alpha_1) \parallel \text{F}(\Omega, \alpha_1, \mathbf{w}, \mathbf{y})). \quad (18)$$

The minimization of $C_{KL}(\text{q}(\Omega, \alpha_1) \parallel \text{p}_1(\mathbf{y}, \Omega, \alpha_1))$ 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 $\text{q}(\alpha_i)$, $i = 1, 2$ we only have to look at the only divergence where that distribution is present. So we can write

$$\text{q}(\alpha_1) = \text{const} \times \exp\left(\langle \log F(\Omega, \alpha_1, \mathbf{w}, \mathbf{y}) \rangle_\Omega\right), \quad (19)$$
$$\text{q}(\alpha_2) = \text{const} \times \exp\left(\langle \log \text{p}_2(\mathbf{y}, \Omega, \alpha_2) \rangle_\Omega\right), \quad (20)$$

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

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

$$\text{q}(\xi) = \text{const} \times \exp\left(\left\langle \log\left[\text{p}(\mathbf{y}|\mathbf{x}, \boldsymbol{\beta}) \prod_{k=1}^{L} \text{p}(\beta_k)\right.\right.\right.$$
$$\left.\left.\left.\times \left[\text{M}(\alpha_1, \mathbf{x}, \mathbf{w})\text{p}(\alpha_1)\right]^{\lambda_1} \left[\text{p}_2(\mathbf{x}|\alpha_2)\text{p}(\alpha_2)\right]^{1-\lambda_1}\right]\right\rangle_{\Theta_\xi}\right), \quad (21)$$

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

$$q(\mathbf{x}) \propto \exp\left(-\frac{1}{2}<\alpha_1>\lambda_1 \sum_i \frac{(\Delta_i^h(\mathbf{x}))^2 + (\Delta_i^v(\mathbf{x}))^2 + w_i}{\sqrt{w_i}}\right.$$
$$\left.-\frac{(1-\lambda_1)}{2}<\alpha_2>\|\mathbf{Cx}\|^2 - \frac{1}{2}\sum_k <\beta_k>\| \mathbf{y}_k - \mathbf{AH}_k\mathbf{D}_k\mathbf{x} \|^2\right). \quad (22)$$

Furthermore, the elements of $\mathbf{w}$ in Eq. (15) are equal to

$$w_i = \mathrm{E}_\mathbf{x}[(\Delta_i^h(\mathbf{x}))^2 + (\Delta_i^v(\mathbf{x}))^2]. \quad (23)$$

In the last step of the algorithm, the distributions $q(\alpha_1)$, $q(\alpha_2)$ and $q(\beta_k)$ are the Gamma distributions, given by

$$q(\alpha_1) \propto \alpha_1^{PN/2-1+a_{\alpha_1}^o} \exp\left[-\alpha_1(b_{\alpha_1}^o + \sum_i \sqrt{w_i})\right], \quad (24)$$

$$q(\alpha_2) \propto \alpha_2^{PN/2-1+a_{\alpha_2}^o} \exp\left[-\alpha_2\left(b_{\alpha_2}^o + \frac{\mathrm{E}_\mathbf{x}\left[\|\mathbf{Cx}\|^2\right]}{2}\right)\right], \quad (25)$$

$$q(\beta_k) \propto \beta_k^{N/2-1+a_\beta^o}\exp\left[-\beta_k(b_\beta^o + \frac{\mathrm{E}_\mathbf{x}\left[\| \mathbf{y}_k - \mathbf{B}_k\mathbf{x} \|^2\right]}{2})\right] \quad (26)$$

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 $\mathbf{w}$ using Eq. (23).
    3. Estimate the distributions of the hyperparameters α_1, α_2 and $\{\beta_k\}$ using Eqs. (24), (25) and (26).

It is interesting to note that when we combine M prior models the inference procedure proposed in [4] is a particular case of the one developed here assuming that $\lambda_i = 1/M$, $\forall i$.

V. 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 $\mathrm{PSNR} = 10\log_{10}\frac{NP}{\|\hat{\mathbf{x}}-\mathbf{x}\|^2}$, 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* [10], 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.

SNR	5 dB	15 dB	25 dB
Bicubic	17.60 ± 0.000	17.61 ± 0.000	17.61 ± 0.000
ZMT	20.84 ± 5.77	19.25 ± 0.0003	19.24 ± 0.000
RSR	29.19 ± 0.004	34.14 ± 0.02	34.01 ± 0.002
SAR	27.64 ± 0.0002	35.99 ± 0.0002	45.71 ± 0.001
VSR	30.86 ± 0.002	41.06 ± 0.001	51.52 ± 0.001
ALG1(λ_1)	30.95 ± 0.002	41.64 ± 0.007	51.67 ± 0.001
λ_1	0.95	0.9	0.9

SNR	5 dB	15 dB	25 dB
Bicubic	24.143 ± 0.001	24.27 ± 0.000	24.27 ± 0.000
ZMT	27.56 ± 1.561	30.47 ± 0.238	30.63 ± 0.000
RSR	29.75 ± 0.001	33.51 ± 0.008	33.63 ± 0.004
SAR	30.28 ± 0.002	35.72 ± 0.004	44.15 ± 0.031
VSR	30.956 ± 0.004	36.56 ± 0.002	45.44 ± 0.012
ALG1(λ_1)	31.09 ± 0.004	37.81 ± 0.004	45.82 ± 0.012
λ_1	0.9	0.85	0.85

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 [5] (denoted by *VSR*). The optimal value of λ_1, for our Algorithm 1, is found here experimentally, and the obtained reconstruction will be denoted by $ALG1(\lambda_1)$. In this section, we evaluate the performance of the proposed algorithm $ALG1(\lambda_1)$ in comparison with the following methods: 1) Bicubic interpolation, 2) the robust SR method in [11] (denoted by *ZMT*), which is based on backprojection with median filtering, 3) the robust SR method in [12] (denoted by *RSR*), which is based on bilateral TV priors, 4) *SAR* and 5) *VSR*.

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 5 dB, 15 dB, and 25 dB.

We conducted simulations with 10 different noise realizations at each SNR level, and the average PSNR and standard deviations of these experiments are shown in Tables I and II. 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. It is clear that *VSR* 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 $ALG1(\lambda_1)$ when λ_1 changes from $\lambda_1 = 0$ (SAR) to $\lambda_1 = 1$ (VSR), reaching its maximum value at $\lambda_1 = 0.9$, for the disk image with SNR=15 dB, and Fig. (3) 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.

VI. CONCLUSIONS

We have proposed a new methodology to combine image priors in SR image reconstruction. The new methodology is based on

(a)   (b)

(c)   (d)
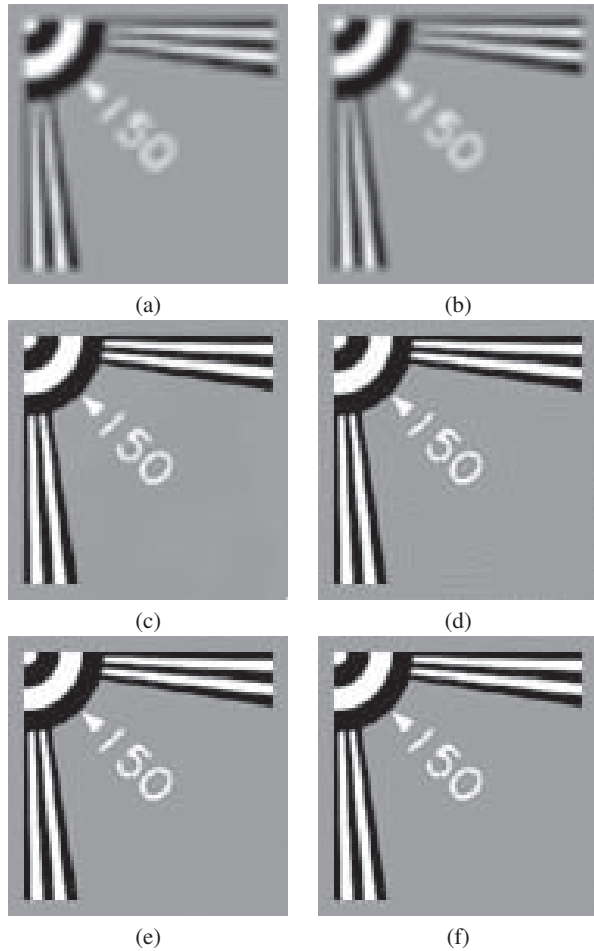
(e)   (f)

Fig. 1. Example estimated HR disk images from different SR methods for the case when SNR=15 dB. Results of (a) Bicubic interpolation (PSNR = 17.61 dB), (b) *ZMT* (PSNR = 19.23 dB), (c) *RSR* (PSNR = 34.32 dB), the proposed methods: using (d) *SAR* (PSNR = 36.15 dB), (e) *VSR* (PSNR = 40.95 dB) and (f) *ALG1*(0.9) (PSNR = 41.61 dB).
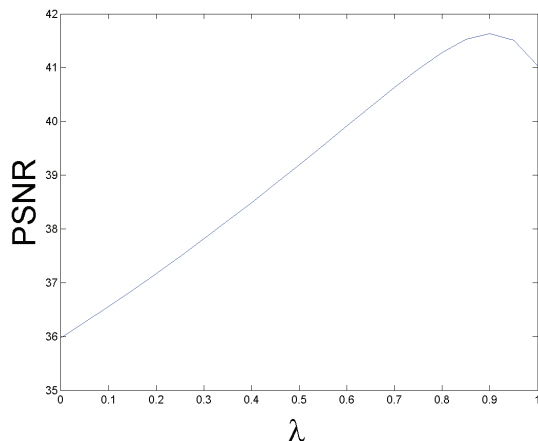

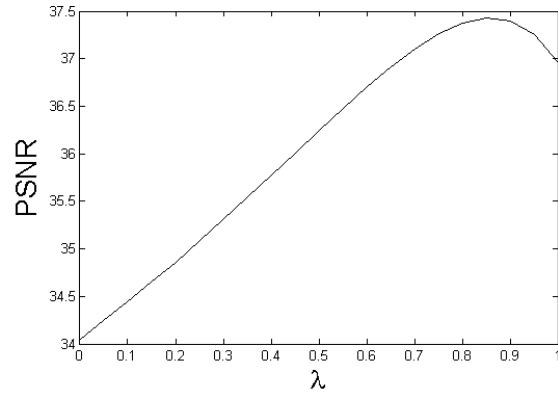Fig. 2. PSNR values obtained with $ALG1(\lambda_1)$, as a function of λ_1, for the SNR=15 dB disk image.

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

finding the distribution on the HR image given the observations that minimizes a linear convex combination of the Kullback-Leibler divergences associated with each pair of observation and prior models. We have found this distribution in closed form. We have also related the proposed methodology to other prior combination methods in the literature. The estimated HR images compare favorably with images provided by other SR reconstruction methods. Their superiority over the HR images obtained when each prior is used independently has also been established experimentally. Future work will address the estimation of the weights assigned to each Kullback-Leibler divergence in the convex combination.

REFERENCES

[1] A. K. Katsaggelos, R. Molina, and J. Mateos, *Super Resolution of Images and Video*. Morgan and Claypool, 2007.
[2] A. K. Katsaggelos and R. Molina, Eds., *Super resolution (special issue)*. The Computer Journal, 52, 395-396, 2009.
[3] P. Milanfar, *Super-Resolution Imaging*, ser. Digital Imaging and Computer Vision. Taylor&Francis/CRC Press (Ed.), 2010.
[4] G. Chantas, N. Galatsanos, R. Molina, and A. Katsaggelos, "Variational bayesian image restoration with a spatially adaptive product of total variation image priors," *IEEE Trans. Image Process.*, vol. 19, no. 2, pp. 351–362, Feb. 2010.
[5] 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.
[6] R. Molina, M. Vega, J. Abad, and A. Katsaggelos, "Parameter estimation in bayesian high-resolution image reconstruction with multisensors," *IEEE Trans. Image Process.*, vol. 12, no. 12, pp. 1655–1667, Dec. 2003.
[7] J. Starck, M. Elad, and D. Donoho, "Image decomposition via the combination of sparse representation and a variational approach," *IEEE Trans. Image Process.*, vol. 14, no. 10, pp. 1570–1582, 2005.
[8] G. Hinton, "Training products of experts by minimizing contrastive divergence," *Neural Computation*, vol. 14, no. 8, pp. 1771–1800, 2002.
[9] S. D. Babacan, R. Molina, and A. Katsaggelos, "Parameter estimation in TV image restoration using variational distribution approximation," *IEEE Trans. Image Process.*, no. 3, pp. 326–339, March 2008.
[10] L. C. Pickup, D. P. Capel, S. J. Roberts, and A. Zisserman, "Bayesian methods for image super-resolution," *The Computer Journal*, 2007.
[11] A. Zomet, A. Rav-Acha, and S. Peleg, "Robust super-resolution," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2001)*, 2001, pp. 645–650.
[12] S. Farsiu, M. D. Robinson, M. Elad, and P. Milanfar, "Fast and robust multiframe super resolution," *IEEE Trans. Image Process.*, vol. 13, no. 10, pp. 1327–1344, Oct. 2004.