Resolution Enhancement of Video Sequences with Simultaneous Estimation of the Regularization Parameter

Hu He, Lisimachos P. Kondi

Dept. of Electrical Engineering, SUNY at Buffalo, Buffalo, NY 14260

ABSTRACT

In this paper, we propose a technique for the estimation of the regularization parameter for image resolution enhancement (super-resolution) based on the assumptions that it should be a function of the regularized noise power of the data and that its choice should yield a convex functional whose minimization would give the desired high-resolution image. The regularization parameter acts adaptively to determine the trade-off between fidelity to the received data and prior information about the image. Experimental results are presented and conclusions are drawn.

Keywords: Resolution enhancement, super-resolution, MAP estimation, image restoration, regularization

1. INTRODUCTION

In many imaging systems, the resolution of the detector array of the camera is not sufficiently high for a particular application. Furthermore, the capturing process introduces additive noise and the point spread function of the lens and the effects of the finite size of the photo-detectors further degrade the acquired video frames. The goal of resolution enhancement is to estimate a high-resolution image from a sequence of low-resolution images while also compensating for the above-mentioned degradations.

Resolution enhancement using multiple frames is possible when there exists subpixel motion between the captured frames. Thus, each of the frames provides a unique look into the scene. An example scenario is the case of a camera that is mounted on an aircraft and is imaging objects in the far field. The vibrations of the aircraft will generally provide the necessary motion between the focal plane array and the scene, thus yielding frames with subpixel motion between them and minimal occlusion effects.

In this paper, we extend our previous results [1] by proposing a technique for the estimation of the regularization parameter. The rest of the paper is organized as follows. In section 2, a Maximum A Posteriori (MAP) formulation is presented for the super-resolution problem and it is shown that, for specific choice of prior model, the MAP cost function is equivalent to a Tikhonov regularization cost function. In section 3, we rewrite the cost function in multi-channel form to establish the relationship between the overall regularization parameter and the individual parameters for each channel. We then develop our technique for the estimation of the regularization parameter. In section 4, experimental results are presented. Finally, in section 5, conclusions are drawn.

2. REGULARIZED COST FUNCTION OF RESOLUTION ENHANCEMENT

The problem of video super-resolution is an active research area. We next outline a few of the approaches that have appeared in the literature. Among the earliest efforts in the field is the work by Tsai and Huang [2]. Their method operates in the frequency domain and capitalizes on the shifting property of the Fourier Transform, the aliasing relationship between the Continuous Fourier Transform (CFD) and the Discrete Fourier Transform (DFT), and the fact that the original scene is assumed to be band limited. The above properties are used to construct a system of equations relating the aliased DFT coefficients of the observed images to samples of the CFT of the unknown high-resolution image. The system of equations is solved, yielding an estimate of the DFT coefficients of the original high-resolution image, which can then be obtained using inverse DFT. This technique was further improved by Tekalp et. al. in [3] by taking into account a Linear Shift Invariant (LSI) blur Point Spread Function (PSF) and using a least squares approach to solving the system of equations. The big advantage of the frequency domain methods is their low computational complexity. However, these methods are applicable only to global motion and a priori information about the high-resolution image cannot be exploited.

Most of the other resolution enhancement techniques that have appeared in the literature operate in the spatial domain. The most computationally efficient techniques involve interpolation of non-uniformly spaced samples. This requires the computation of the optical flow between the acquired low-resolution frames are combined in order to create a high-resolution frame. Interpolation techniques are used to estimate pixels in the high-resolution frame that did not correspond to pixels in one of the acquired frames.

Finally, image restoration techniques are used to compensate for the blurring introduced by the imaging device. A method based on this idea is the Temporal Accumulation of Registered Image Data (TARID) [4], developed by the Naval Research Laboratory (NRL).

Another method that has appeared in the literature is the iterated backprojection method [5]. In this method, the estimate of the high-resolution image is updated by backprojecting the error between motion-compensated, blurred and subsampled versions of the current estimate of the high-resolution image and the observed low-resolution images, using an appropriate backprojection operator.

Another proposed method is the Projection Onto Convex Sets (POCS) [6], [7]. In this method, the space of high-resolution images is intersected with a set of convex constraint sets representing desirable image characteristics, such as positivity, bounded energy, fidelity to data, smoothness, etc.

Another class of resolution enhancement algorithms is based on stochastic techniques. Methods in this class include Maximum Likelihood (ML) [8] and Maximum A Posteriori (MAP) approaches [9], [10], [11], [12]. MAP estimation with an edge preserving Huber-Markov random field image prior is studied in [9], [10], [11]. MAP based resolution enhancement with simultaneous estimation of registration parameters (motion between frames) has been proposed in [1], [12]. The MAP methods lead to solving a regularized cost function. The regularization parameter of the cost function plays a very important role in the reconstruction of high-resolution image, while its selection is a kind of art. The L-curve method chooses the "L-corner", the point with maximum curvature on the L-curve, as the one corresponding to regularization parameter [13]. Iterative adaptive algorithms with automatically updated regularization parameter have been proposed for image restoration [14], [15]. However, little research has been done for the super-resolution scenario. The objective of this paper is to extend these results to super-resolution.

In the following, we use the same formulation and notation as in [12]. The image degradation process can be modeled by a linear blur, motion, subsampling by pixel averaging and an additive Gaussian noise process. In the following, we order all vectors lexicographically. We assume that *p* low-resolution frames are observed, each of size $N_1 \times N_2$. The desired high-resolution image $\mathbf{z} = \left[z_1, z_2, \cdots, z_N\right]^T$ is of size $N = L_1 N_1 L_2 N_2$, where L_1 and L_2 represent the down-sampling factors in the horizontal and

vertical directions respectively. Let the *k*th low-resolution frame be denoted as $\mathbf{y}_k = \left[y_{k,1}, y_{k,2}, \cdots y_{k,M}\right]^T$ for $k = 1,2,\cdots p$ and where $M = N_1 N_2$. The full set of *p* observed low-resolution images can be denoted as

$$\mathbf{y} = \left[\mathbf{y}_1^T, \mathbf{y}_2^T, \cdots, \mathbf{y}_p^T\right]^T = \left[y_1, y_2, \cdots, y_{pM}\right]^T. \tag{1}$$

The observed low-resolution frames are related to the high-resolution image through the following model:

$$y_{k,m} = \sum_{r=1}^{N} w_{k,m,r}(\mathbf{s}_k) z_r + \eta_{k,m}, \tag{2}$$

for $m = 1,2,\cdots M$ and $k = 1,2,\cdots p$. The weight $w_{k,m,r}(\mathbf{s}_k)$ represents the "contribution" of the *r*th high-resolution pixel to the *m*th low-resolution observed pixel of the *k*th frame, which implements blur, motion and pixel averaging. The vector $\mathbf{s}_k = \left[s_{k,1}, s_{k,2}, \cdots s_{k,K}\right]^T$ contains the *K* registration parameters for frame *k*, representing translational shift, rotation, affine transformation parameters, or other motion parameters. This motion is measured in reference to a fixed high-resolution grid. The term $\eta_{k,m}$ represents additive noise samples that are assumed to be independent and identically distributed (i.i.d.) Gaussian noise samples with variance σ_η^2. We can rewrite Equation (2) in matrix notation

$$\mathbf{y} = \mathbf{W}_\mathbf{s}\mathbf{z} + \mathbf{n}, \tag{3}$$

where matrix $\mathbf{W}_\mathbf{s} = \left[\mathbf{W}_{\mathbf{s},1}^T, \mathbf{W}_{\mathbf{s},2}^T, \cdots, \mathbf{W}_{\mathbf{s},p}^T\right]^T$ contains the values $w_{k,m,r}$ and $\mathbf{n} = \left[\eta_1, \eta_2, \cdots \eta_{pM}\right]^T$. The multivariate p.d.f of **y** given **z** and **s** is

$$P_r(\mathbf{y}|\mathbf{z},\mathbf{s}) = \frac{1}{(2\pi)^{\frac{pM}{2}} \sigma_\eta^{pM}} \exp\left\{-\frac{1}{2\sigma_\eta^2}(\mathbf{y} - \mathbf{W}_\mathbf{s}\mathbf{z})^T(\mathbf{y} - \mathbf{W}_\mathbf{s}\mathbf{z})\right\}. \tag{4}$$

We can form a MAP estimate of the high-resolution image **z** and the registration parameters **s** simultaneously, given the observed **y**. The estimates can be computed as

$$\hat{\mathbf{z}}, \hat{\mathbf{s}} = \arg\max_{\mathbf{z},\mathbf{s}} P_r(\mathbf{z}, \mathbf{s}|\mathbf{y}). \tag{5}$$

Using Bayes' rule, the above equation can be expressed as

$$\hat{\mathbf{z}}, \hat{\mathbf{s}} = \arg\max_{\mathbf{z},\mathbf{s}} \frac{P_r(\mathbf{y}|\mathbf{z}, \mathbf{s}) P_r(\mathbf{z}, \mathbf{s})}{P_r(\mathbf{y})}. \tag{6}$$

Clearly, the denominator of the above equation is not a function of $\mathbf{z}$ or $\mathbf{s}$. If we further assume $\mathbf{z}$ and $\mathbf{s}$ are statistically independent and all possible vectors $\mathbf{s}$ are equally probable, we have

$$\hat{\mathbf{z}}, \hat{\mathbf{s}} = \arg\max_{\mathbf{z},\mathbf{s}} P_r(\mathbf{y}|\mathbf{z}, \mathbf{s}) P_r(\mathbf{z}). \tag{7}$$

It is very important to choose an appropriate model for the p.d.f of the desired image **z**. As in [12], we choose Gauss-Markov random field (GMRF) as the image prior, with density of the form:

$$P_r(\mathbf{z}) = \frac{1}{(2\pi)^{\frac{N}{2}} |\mathbf{C}|^{\frac{1}{2}}} \exp\left\{ -\frac{1}{2} \mathbf{z}^T \mathbf{C}^{-1} \mathbf{z} \right\}, \tag{8}$$

where matrix **C** is the $N \times N$ covariance matrix of **z**. For a specific choice of the covariance matrix **C**, the above equation can be written as

$$P_r(\mathbf{z}) = \frac{1}{(2\pi)^{\frac{N}{2}} |\mathbf{C}|^{\frac{1}{2}}} \exp\left\{ -\frac{1}{2\lambda} \sum_{i=1}^{N} \left(\sum_{j=1}^{N} d_{i,j} z_j \right)^2 \right\}, \tag{9}$$

where $d_i = \left[d_{i,1}, d_{i,2}, \cdots d_{i,N} \right]^T$ is the coefficient vector and λ is called temperature parameter. The above equation results if we assume that the elements $C_{i,j}^{-1}$ of the inverse of **C** satisfy

$$C_{i,j}^{-1} = \frac{1}{\lambda} \sum_{r=1}^{N} d_{r,i} d_{r,j}. \tag{10}$$

$d_{i,j}$ can be chosen as the 2-D Laplacian kernel

$$d_{i,j} = \begin{cases} 1 & \text{for } i = j \\ -1/4 & \text{for } j : z_j \text{ is a cardinal neighbor of } z_j \end{cases}. \tag{11}$$

Following the same procedure as in [12], we can reach the following regularized cost function to minimize

$$L(\mathbf{z},\mathbf{s}) = \frac{1}{2\sigma_\eta^{\,2}}\left(\mathbf{y}-\mathbf{W}_\mathbf{s}\mathbf{z}\right)^T\left(\mathbf{y}-\mathbf{W}_\mathbf{s}\mathbf{z}\right)+\frac{1}{2}\mathbf{z}^T\mathbf{C}^{-1}\mathbf{z}\,. \tag{12}$$

Or, equivalently,

$$\begin{aligned} L(\mathbf{z},\mathbf{s}) &= \sum_{m=1}^{pM}\left(y_m - \sum_{r=1}^{N} w_{m,r}(\mathbf{s}) z_r\right)^2 + \frac{\sigma_\eta^{\,2}}{\lambda}\sum_{i=1}^{N}\left(\sum_{j=1}^{N} d_{i,j} z_j\right)^2 \\ &= \left\|\mathbf{y}-\mathbf{W}_\mathbf{s}\mathbf{z}\right\|^2 + \frac{\sigma_\eta^{\,2}}{\lambda}\left\|\mathbf{D}\mathbf{z}\right\|^2 \\ &= \left\|\mathbf{y}-\mathbf{W}_\mathbf{s}\mathbf{z}\right\|^2 + \alpha\left\|\mathbf{D}\mathbf{z}\right\|^2 \end{aligned}, \tag{13}$$

where $w_{m,r}$ is the "contribution" of z_r to y_m, for $m = 1,2,\cdots, pM$ and $r = 1,2,\cdots N$, and α is the regularization parameter defined as

$$\alpha = \frac{\sigma_\eta^{\,2}}{\lambda}, \tag{14}$$

and **D** is the $N \times N$ matrix representing the 2-D Laplacian kernel, which is a high-pass filter.

The above cost function can be minimized using the coordinate-descent method [12]. This iterative method starts with an initial estimate of **z** obtained using interpolation from a low-resolution frame. Then, for a fixed **z**, the cost function is minimized with respect to **s**, using Equation (15)

$$\hat{\mathbf{s}}_k^n = \arg\min_{\mathbf{s_k}}\left\{\sum_{m=1}^{M}\left(y_{k,m} - \sum_{r=1}^{N} w_{k,m,r}(\mathbf{s}_k)\hat{z}_r^n\right)^2\right\}. \tag{15}$$

Thus, the motion of each frame is estimated. *n* is the iteration number starting from 0. Then, for fixed **s**, a new estimate for **z** is obtained (Equation (16))

$$\hat{z}_r^{n+1} = \hat{z}_r^n - \varepsilon g_r(\hat{\mathbf{z}}^n, \hat{\mathbf{s}}^n). \tag{16}$$

The gradient $g_r(\hat{\mathbf{z}}^n, \hat{\mathbf{s}}^n)$ can be obtained from

$$g_r(\mathbf{z},\mathbf{s}) = \frac{\partial L(\mathbf{z},\mathbf{s})}{\partial z_r} = 2\left\{\sum_{m=1}^{pM} w_{m,r}(\mathbf{s})\left(\sum_{j=1}^{N} w_{m,j}(\mathbf{s}) z_j - y_m\right) + \alpha \sum_{i=1}^{N} d_{i,r}\left(\sum_{j=1}^{N} d_{i,j} z_j\right)\right\}. \tag{17}$$

The choice of step size ε is addressed in the paper later.

3. ESTIMATION OF REGULARIZATION PARAMETERS

It can be seen that the cost function in Equation (13) is a Tikhonov regularization cost function. Thus, for the specific choice of prior model $P_r(\mathbf{z})$ considered here, the MAP formulation is equivalent to a Tikhonov regularization formulation. Equation (13) has two terms: a term representing the fidelity of the solution to the received data (residual norm $\|\mathbf{y} - \mathbf{W}_s\mathbf{z}\|^2$) and a term representing a priori information about the high-resolution image (smoothness norm $\|\mathbf{D}\mathbf{z}\|^2$). The latter involves a high-pass filter and thus dictates that the solution be smooth by penalizing discontinuities. The relative weighting of the two terms is determined by a regularization parameter α, which is the ratio of the power of noise σ_η^2 over the temperature parameter λ. In the most general case, we have no prior information for both σ_η^2 and λ. In this case, the regularization parameter can be explicitly expressed as a function of the original image [14]. We rewrite the regularized cost function as

$$L(\mathbf{z},\mathbf{s}) = \|\mathbf{y} - \mathbf{W}_s\mathbf{z}\|^2 + \alpha(\mathbf{z})\|\mathbf{D}\mathbf{z}\|^2. \tag{18}$$

Furthermore, we can rewrite the cost function as the sum of individual smoothing functionals for each of the *p* low-resolution images as:

$$L(\mathbf{z},\mathbf{s}) = \sum_{k=1}^{p}\left\{\|\mathbf{y}_k - \mathbf{W}_{s,k}\mathbf{z}\|^2 + \alpha_k(\mathbf{z})\|\mathbf{D}\mathbf{z}\|^2\right\}. \tag{19}$$

We drop the subscript k from $\mathbf{D}_k$ in the above equation because $\mathbf{D}_k = \mathbf{D}$, i.e., the same high-pass filter (Laplacian kernel) is used for all low-resolution image $k = 1,2,\cdots p$. Then we can define the individual functional for each low-resolution image (channel) as:

$$L_k(\alpha_k(\mathbf{z}), \mathbf{z}, \mathbf{s}) = \left\|\mathbf{y}_k - \mathbf{W}_{\mathbf{s},k}\mathbf{z}\right\|^2 + \alpha_k(\mathbf{z})\left\|\mathbf{D}\mathbf{z}\right\|^2, \quad k = 1,2,\cdots p. \tag{20}$$

Following the same procedure as in [14], we impose the following requirements for each $\alpha_k(\mathbf{z})$: it should be a function of the regularized noise power of the data and its choice should yield a convex functional whose minimization would give the high-resolution image. Then we reach the same iteration expressions as in equations (15), (16) and (17), except for that α is replaced with $\alpha(\mathbf{z}) = \sum_{k=1}^{p} \alpha_k(\mathbf{z})$.

The imposed properties on $\alpha_k(\mathbf{z})$ require a linear function between $\alpha_k(\mathbf{z})$ and each term of the cost function:

$$\alpha_k(\mathbf{z}) = f\left(L_k(\alpha_k(\mathbf{z}), \mathbf{z})\right) = \gamma_k \left\{\left\|\mathbf{y}_k - \mathbf{W}_{\mathbf{s},k}\mathbf{z}\right\|^2 + \alpha_k(\mathbf{z})\left\|\mathbf{D}\mathbf{z}\right\|^2\right\}. \tag{21}$$

Thus, the choice of regularization parameter for the multichannel regularization functional is given by

$$\alpha_k(\mathbf{z}) = \frac{\left\|\mathbf{y}_k - \mathbf{W}_{\mathbf{s},k}\mathbf{z}\right\|^2}{\frac{1}{\gamma_k} - \left\|\mathbf{D}\mathbf{z}\right\|^2}. \tag{22}$$

Also, following the same procedure for convergence requirement as in [14], we get

$$\frac{1}{\gamma_k} > \frac{\varepsilon p\left\|\mathbf{y}_k - \mathbf{W}_{\mathbf{s},k}\mathbf{z}\right\|^2 \phi_{\max}(\mathbf{D}^T\mathbf{D})}{2 - \varepsilon p \phi_{\max}(\mathbf{W}_{\mathbf{s},k}^T \mathbf{W}_{\mathbf{s},k})} + \left\|\mathbf{D}\mathbf{z}\right\|^2 \tag{23}$$

where $\phi_{\max}(\cdot)$ stands for the maximum eigenvalue of a matrix.

For the image resolution enhancement, the $N \times N$ degradation matrix $\mathbf{W}_{\mathbf{s},k}$ for channel $k = 1, 2, \cdots, p$ can be expressed as the multiplication of subsampling matrix $\mathbf{S}$, motion matrix $\mathbf{M}_k$ and blur matrix $\mathbf{B}_k$, with size $N_1 N_2 \times N$, $N \times N$, and $N \times N$ respectively.

$$\mathbf{W}_{\mathbf{s},k} = \mathbf{S}\mathbf{M}_k\mathbf{B}_k . \tag{24}$$

Therefore,

$$\phi_{\max}(\mathbf{W}_{\mathbf{s},k}^T \mathbf{W}_{\mathbf{s},k}) = \phi_{\max}(\mathbf{B}_k^T \mathbf{M}_k^T \mathbf{S}^T \mathbf{S}\mathbf{M}_k\mathbf{B}_k) . \tag{25}$$

Now, for subsampling by pixel averaging, we can easily verify

$$\mathbf{S}^T\mathbf{S} = \frac{1}{(L_1 L_2)^2}\mathbf{I} , \tag{26}$$

where $\mathbf{I}$ is the identity matrix with size $N \times N$.

Since no information is lost or added due to motion operation $\mathbf{M}_k$, the element of $\mathbf{M}_k$ are “1”s and “0”s, with each column and each row having only a single “1”. For such kind of matrix $\mathbf{M}_k$, we can easily verify that

$$\mathbf{M}_k^T\mathbf{M}_k = \mathbf{I} . \tag{27}$$

For a PSF generated from Gaussian blur, we can assume that the impulse response coefficients are normalized to add to 1, which is equivalent to

$$\phi_{\max}(\mathbf{B}_k^T\mathbf{B}_k) = 1 . \tag{28}$$

By substituting equations (26), (27) , and (28) into (25), we have

$$\phi_{\max}(\mathbf{W}_{\mathbf{s},k}^T \mathbf{W}_{\mathbf{s},k}) = \frac{1}{(L_1 L_2)^2}\phi_{\max}(\mathbf{B}_k^T\mathbf{B}_k) = \frac{1}{(L_1 L_2)^2} . \tag{29}$$

Therefore, inequality (23) becomes

$$\frac{1}{\gamma_k} > \frac{\varepsilon p \phi_{\max}(\mathbf{D}^T\mathbf{D})}{2-\left(\varepsilon p/\left(L_1L_2\right)^2\right)}\left\|\mathbf{y}_k - \mathbf{W}_{\mathbf{s},k}\mathbf{z}\right\|^2 + \left\|\mathbf{D}\mathbf{z}\right\|^2 \tag{30}$$

Now, we can select step size ε to make

$$\frac{\varepsilon p \phi_{\max}(\mathbf{D}^T\mathbf{D})}{2-\left(\varepsilon p/\left(L_1L_2\right)^2\right)} = 1. \tag{31}$$

That is,

$$\varepsilon = \frac{2}{p}\left(\frac{\left(L_1L_2\right)^2}{\left(L_1L_2\right)^2\phi_{\max}(\mathbf{D}^T\mathbf{D})+1}\right). \tag{32}$$

Then, inequality (30) becomes

$$\frac{1}{\gamma_k} > \left\|\mathbf{y}_k - \mathbf{W}_{\mathbf{s},k}\mathbf{z}\right\|^2 + \left\|\mathbf{D}\mathbf{z}\right\|^2. \tag{33}$$

Now, $\left\|\mathbf{y}_k\right\|^2 \geq \left\|\mathbf{y}_k - \mathbf{W}_{\mathbf{s},k}\mathbf{z}\right\|^2$, since the low-resolution image is assumed to have more energy than the additive noise, and $\left\|\mathbf{y}_k\right\|^2 \approx \frac{\left\|\mathbf{z}\right\|^2}{L_1L_2} > \left\|\mathbf{D}\mathbf{z}\right\|^2$ for small subsampling ratio $L_1 = L_2 = 4$, since $\mathbf{z}$ is assumed to have much less energy at high frequencies than at low frequencies and each low-resolution image $\mathbf{y}_k$ has $1/L_1L_2$ of the energy of $\mathbf{z}$ for noiseless cases. We list two commonly used examples ("Cameraman" and "Lena") in Tables 4, 5 in the Appendix with subsampling ratio $L_1 = L_2 = 4$ as an example to show the validity of the above inequalities. From the two ratios in the tables (Ratio 1 and Ratio 2), we can see that the inequalities are satisfied for all cases.

For a subsampling ratio $L_1 = L_2 = 4$, as used in this paper, we can safely say that the choice of

$$\frac{1}{\gamma_k} = 2\left\|\mathbf{y}_k\right\|^2 \tag{34}$$

satisfies the condition for convergence and also provides a positive $\alpha_k(\mathbf{z})$

$$\alpha_k(\mathbf{z}) = \frac{\left\|\mathbf{y}_k - \mathbf{W}_{\mathbf{s},k}\mathbf{z}\right\|^2}{2\left\|\mathbf{y}_k\right\|^2 - \left\|\mathbf{D}\mathbf{z}\right\|^2}. \tag{35}$$

We can see from the experimental results that the choice of $1/\gamma_k$ in Equation (34) not only provides a fixed, simple and tight choice for inequality (33), but also results in good reconstructions. During the iterations, the regularization parameter $\alpha_k(\mathbf{z})$ is adaptively updated according to the current estimate of high-resolution image **z**.

4. EXPERIMENTAL RESULTS

A number of experiments were conducted, some of which are presented here. To test the performance of our algorithm, we first use the 256x256 "Cameraman" and "Lena" test images for a synthetic test. Four cases, Case I-IV, as listed in Table 1, are tested. The first frame is selected as reference frame and bilinear interpolation of the first frame is chosen as the first estimate of high-resolution image **z**. Algorithm is carried out for 20 iterations or while convergence is reached when $\left\|\hat{\mathbf{z}}^{n+1} - \hat{\mathbf{z}}^{n}\right\|^2 / \left\|\hat{\mathbf{z}}^{n}\right\|^2 < 10^{-6}$.

In the above four cases, global shift is implemented as the motion degradation. Therefore $\mathbf{s}_k = \left[s_{k,1}, s_{k,2}\right]^T$, i.e., *K*=2 registration parameters for frame *k*.

To compare with the proposed algorithm, an exhaustive search over the parameter space was conducted, each time with one fixed value of α. We then find the α corresponding to the maximum value of

$$PSNR_\alpha = 10 \times \log_{10} \frac{255^2 \times N}{\left\|\mathbf{z} - \hat{\mathbf{z}}_\alpha\right\|^2}, \tag{36}$$

and denote it as α_{fixed}. The plots of $PSNR_\alpha$ versus α for "Cameraman" and "Lena" are shown in Fig. 1 and 3 respectively. The PSNR of the reconstructed image for "Cameraman" and "Lena" using the three methods (Bilinear, Method with α_{fixed}, Proposed Method) are listed in Tables 2 and 3 respectively.

The reference frame, bilinear interpolation of the reference frame and the reconstructed images from Method with α_{fixed} and Proposed Method of "Cameraman" in Case I are shown in Fig. 2(a)-(d), respectively. In Fig. 2(e) and 2(f), the absolute convergence curve (the plot of $\left\|\hat{\mathbf{z}}^{n}-\mathbf{z}\right\|^{2}\Big/\left\|\mathbf{z}\right\|^{2}$ versus iteration number *n*) and relative convergence curve (the plot of $\left\|\hat{\mathbf{z}}^{n+1}-\hat{\mathbf{z}}^{n}\right\|^{2}\Big/\left\|\hat{\mathbf{z}}^{n}\right\|^{2}$ versus iteration number *n*) are shown for the Proposed Method.

The reference frame, bilinear interpolation of the reference frame and the reconstructed images from Method with α_{fixed} and Proposed Method of "Lena" in Case I are shown in Fig. 4(a)-(d), respectively. In Fig. 4(e) and 4(f), the absolute convergence curve and relative convergence curve are shown for Proposed Method.

For all the four cases in Tables 2 and 3, the Proposed Method gives the highest PSNR and best visual quality for both "Cameraman" and "Lena" (2~3 dB better than bilinear interpolation, and up to 0.25dB better than Method with α_{fixed}). For real data, the original high-resolution image is not available, thus $PSNR_{\alpha}$ in Equation (36) can not be evaluated. We will show next that Proposed Method provides a good regularization parameter without exhaustive search or trial and error method.

Next, we used real data of an infrared video sequence provided to us by the Naval Research Laboratory, Washington, D.C., to test the proposed algorithm. 20 frames of low-resolution image with size 128x128 pixels were used. Up-sample ratio is L_1=L_2=4. We assumed a Gaussian point spread function for the lens and estimated its variance at 1.7 using trial and error. Bilinear interpolation of the first frame was chosen as the first estimate of high-resolution image **z**. Joint MAP registration technique was applied to estimate the motion, with the high-resolution estimate partitioned into macro blocks (Equation (15)). Three Point Search (TSS) with search window ±7 and Sum of Squared Errors (SSE) criteria was applied to decide the motion vector between the 16x16 macro block in the high-resolution estimate and the corresponding 4x4 macro block in the low-resolution image. When reconstructing the high-resolution image, these motion vectors

(registration parameters) were used to compensate the motion. We assumed that convergence was reached when $\left\|\hat{\mathbf{z}}^{n+1}-\hat{\mathbf{z}}^{n}\right\|^2 / \left\|\hat{\mathbf{z}}^{n}\right\|^2 < 10^{-6}$.

There were totally $(512\times512)\div(16\times16)=1024$ macro blocks in the high-resolution image, and each macro block had two translational shift parameters. Therefore $\mathbf{s}_k = \left[s_{k,1}, s_{k,2}, \cdots s_{k,2048}\right]^T$, i.e., $K = 2\times1024 = 2048$ registration parameters for frame *k*.

The first frame of the low video sequence, bilinear interpolation of the first frame, Proposed Method, the relative convergence curve for Proposed Method (compared with the previous estimate of high-resolution image "truck") are shown in Fig. 5(a)-(d). From the relative convergence curves, we can see the Proposed Method converges very fast.

The visually best reconstructed image using a fixed valued regularization parameter is obtained with $\alpha = 0.1$, shown in Fig. 5(e) via trial and error, with the help of human interaction. Although it provides a good result, the computation cost is huge and involves human's subjective influence. Reconstructed images using two fixed valued regularization parameters $\alpha = 10^3$ and $\alpha = 10^{-3}$ are shown in Fig. 5(f) and Fig. 5(g). As expected, image in Fig. 5(f) is "over-smoothened" and the image in Fig. 5(g) is "noise-amplified", which means arbitrarily selected and fixed valued regularization parameter may lead to bad results. On the contrary, reconstructed high-resolution image using the proposed method with simultaneous estimation of regularization parameter provides the best visual effect and much less computation cost efficiently.

5. CONCLUSIONS

We have proposed a technique for the estimation of the regularization parameter for digital image resolution enhancement. Our experimental results demonstrate the performance of the proposed algorithm. Experimental results using both synthetic and real data are presented. For the synthetic results, in addition to a subjective evaluation, we also present objective PSNR results. In all the cases considered, the proposed algorithm gives a better reconstruction than results obtained using an optimal fixed-valued choice of the regularization parameter, obtained using exhaustive search.

6. REFERENCES

1. L.P. Kondi, D. Scribner, J. Schuler, "A Comparison of Digital Image Resolution Enhancement Techniques", *SPIE AeroSense*, Orlando, FL, April 2002.
2. R. Tsai and T. Huang, "Multiframe image restoration and registration," in *Advances in Computer Vision and Image Processing*, vol. 1, pp. 317-339, JAI Press Inc., 1984
3. A. M. Tekalp, M. K. Ozkan, and M.I. Sezan, "High-resolution image reconstruction from low-resolution image sequences and space-varying image restoration," in *Proceeding of the International Conference on Acoustics, Speech and Signal Processing*, vol. III, (San Francisco, CA), pp. 169-172, 1992
4. J. M. Schuler, J. G. Howard, P. R. Warren, and D. A. Scribner, "TARID-based image super-resolution," in *Proceeding of SPIE AeroSense*, vol. 4719, (Orlando, Florida), pp. 247-254, 2002.
5. M. Irani and S. Peleg, "Motion analysis for image enhancement: Resolution, occlusion, and transparency," *Journal of Visual Communications and Image Representation*, vol. 4. pp. 324-335, Dec. 1993.
6. A. J. Patti, M.I. Sezan, and A. M. Tekalp, "Super-resolution video reconstruction with arbitrary sampling lattices and nonzero apperature time," *IEEE Transactions on Image Processing*, vol. 6. pp. 1064-1076, Aug. 1997.
7. P. E. Erem, M.I. Sezan, and A. M. Tekalp, "Robust, object-based high-resolution image reconstruction from low-resolution video," *IEEE Transactions on Image Processing*, vol. 6. pp. 1446-1451, Oct. 1997.
8. B. C. Tom and A. K. Katsaggelos, "Reconstruction of a high-resolution image from multiple degraded mis-registered low-resolution images," in *Proceeding of the Conference on Visual Communications and Image Processing*, vol. 2308, pp.971-981. Sept. 1994
9. R. R. Schultz and R. L. Stevenson, "Extraction of high-resolution frames from video sequences," *IEEE Trans. Image Processing*, vol. 5, pp. 996–1011, June 1996.
10. R. R. Schultz and R. L. Stevenson, "Extraction of high-resolution frames from video sequences," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, Detroit, MI, May 1995.

11. R. R. Schultz and R. L. Stevenson, “A Bayesian approach to image expansion for improved definition,” *IEEE Trans. Image Processing*, vol. 3, pp. 233–242, May 1994.
12. R. C. Hardie, K. J. Barnard, “Joint MAP Registration and High-Resolution Image Estimation Using a Sequence of Undersampled Images,” *IEEE Trans. Image Processing*, Vol. 6, No. 12, pp. 621–1633, December 1997.
13. P. C. Hansen, “Rank-Deficient and Discrete Ill-Posed Problems”, SIAM 1997.
14. M. G. Kang, A. K. Katsaggelos, "Simultaneous Multichannel Image Restoration and Estimation of the Regularization parameters", *IEEE Trans. Image Processing*, May 1997, pp. 774-778.
15. M. G. Kang, A. K. Katsaggelos, “General Choice of the Regularization Functional in Regularized Image Restoration,” *IEEE Trans. Image Processing*, Vol. 4, No. 12, pp. 594–602, May 1995.
16. B. C. Tom, A. K. Katsaggelos, “Resolution Enhancement of Monochrome and Color Video Using Motion Compensation,” *IEEE Trans. Image Processing*, Vol. 10, No. 2, pp. 278–287, February 2001.
17. E. Kaltenbacher and R. C. Hardie, “High-resolution infrared image reconstruction using multiple low-resolution aliased frames,” in *Proc. IEEE Nat. Aerospace Electronics Conf.*, Dayton, OH, May 1996.
18. N. P. Galatsanos, A.K. Katsaggelos, R.T. Chin, and A. D. Hillery, “Least squares restoration of multichannel images”, *IEEE Trans. Acoust. Speech, Signal Processing*, Vol. 39, no. 10, pp. 2222-2236, Oct., 1991.
19. R. R. Schultz and R. L. Stevenson, “Improved Definition Video Frame Enhancement,” *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, vol. 4, pp. 2169–2171, May 1995.
20. H. Stark and P. Oskoui, “High-resolution image recovery from imageplane arrays, using convex projections,” *J. Opt. Soc. Amer. A*, vol. 6,pp. 1715–1726, 1989.
21. A. M. Tekalp, Digital Video Processing. *Englewood Cliffs, NJ: Prentice-Hall*, 1995.

7. APPENDIX: JUSTIFICATION OF INEQUALITY (33)

We define the following two ratios for the convenience to test inequality (33).

Ratio 1: $\min_{k=1,2,\cdots,p} \left\{ \left\| \mathbf{y}_k \right\|^2 \Big/ \left\| \mathbf{y}_k - \mathbf{W}_{\mathbf{s},k} \mathbf{z} \right\|^2 \right\}$

Ratio 2: $\min_{k=1,2,\cdots,p} \left\{ \left\| \mathbf{y}_k \right\|^2 \Big/ \left\| \mathbf{Dz} \right\|^2 \right\}$

256x256 “Cameraman” and “Lena” are used as the two original high-resolution images. The results are summarized in Tables 4, 5.

8. BIOGRAPHICS

Hu He received the M.S. in the Department of Electrical Engineering of State University of New York at Buffalo in 2002. Currently, He is a Ph.D. candidate in the same university. His research interests include image and video resolution enhancement. He has been a student member of IEEE since 2001 till now.

Lisimachos P. Kondi received the Diploma degree in electrical engineering from the Aristotle University of Thessaloniki, Greece, in 1994 and the M.S. and Ph.D. degrees in electrical and computer engineering from Northwestern University, Evanston, IL, USA, in 1996 and 1999, respectively. During the 1999-2000 academic year, he was a post-doctoral research associate at Northwestern University. Since August 2000, he is an Assistant Professor of electrical engineering at the State University of New York at Buffalo, USA. During the summer of 2001, he was a U.S. Navy/ASEE Summer Faculty Fellow at the U.S. Naval Research Laboratory, Washington, DC, USA. His current research interests include image and video processing and compression, wireless communications and wireless video transmission, image restoration and super-resolution, and shape coding.

Table 1. Four cases of synthetic test for “Cameraman” and “Lena”

	Variance of Gaussian blur	L_1, L_2: Subsampling ratio	*p*: Number of frames of low-resolution images	$\mathbf{s}_k^T = [s_{k,1}, s_{k,2}]$: Global shift of **z** to obtain low-resolution image *k*. $\mathbf{s}_k^T$ belongs to the set generated from the Cartesian product below	σ_η^2: variance of AWGN noise
Case I	1.7	4	16	$\{0, 1, 2, 3\} \times \{0, 1, 2, 3\}$	1
Case II	1.7	4	16	$\{0, 1, 2, 3\} \times \{0, 1, 2, 3\}$	20
Case III	1.7	4	16	$\{0, 1, 2, 3\} \times \{0, 1, 2, 3\}$	65.025
Case IV	1.7	4	16	$\{0, 1, 2, 3\} \times \{0, 1, 2, 3\}$	100

Table 2. Results of “Cameraman” using the three methods

PSNR (dB)	Bilinear	Method with α_{fixed}	Proposed Method
Case I	21.2597	23.9752	24.1319
Case II	21.1919	23.6352	23.8105
Case III	21.0088	23.1422	23.1929
Case IV	20.9008	22.8038	22.8881

Table 3. Results of “Lena” using the three methods

PSNR (dB)	Bilinear	Method with α_{fixed}	Proposed Method
Case I	23.2106	26.7533	27.0913
Case II	23.1044	26.2251	26.4797
Case III	22.8536	25.4049	25.4085
Case IV	22.6685	24.8298	24.8895

Table 4. Justification of (33) for “Cameraman”

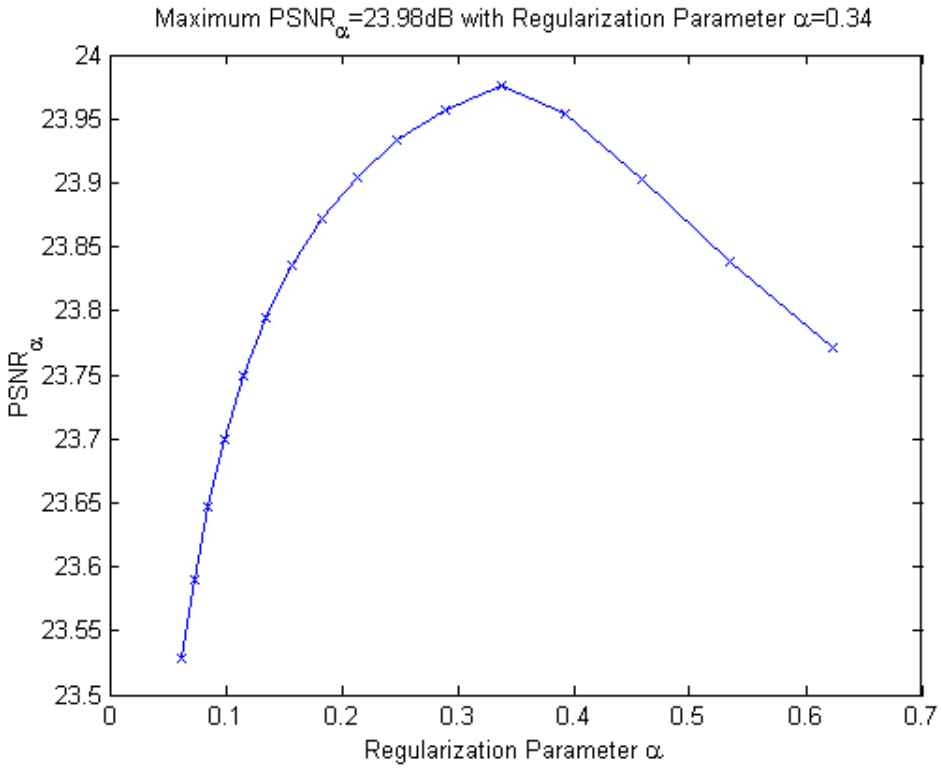
	Case I	**Case II**	**Case III**	**Case IV**
$\min\left\{\left\|\mathbf{y}_k\right\|^2\right\}^*$	6.8784e+007	6.8843e+007	6.9009e+007	6.9082e+007
$\frac{\left\|\mathbf{z}\right\|^2}{L_1 L_2}$	7.3654e+007	7.3654e+007	7.3654e+007	7.3654e+007
$\left\|\mathbf{y}_k - \mathbf{W}_{\mathbf{s},k}\mathbf{z}\right\|^2$	4.096e+003	8.192e+003	2.6634e+005	4.096e+005
$\left\|\mathbf{Dz}\right\|^2$	1.0509e+007	1.0509e+007	1.0509e+007	1.0509e+007
Ratio 1	1.6793e+004	840.3708	259.0986	168.6582
Ratio 2	6.5450	6.5507	6.5665	6.5735

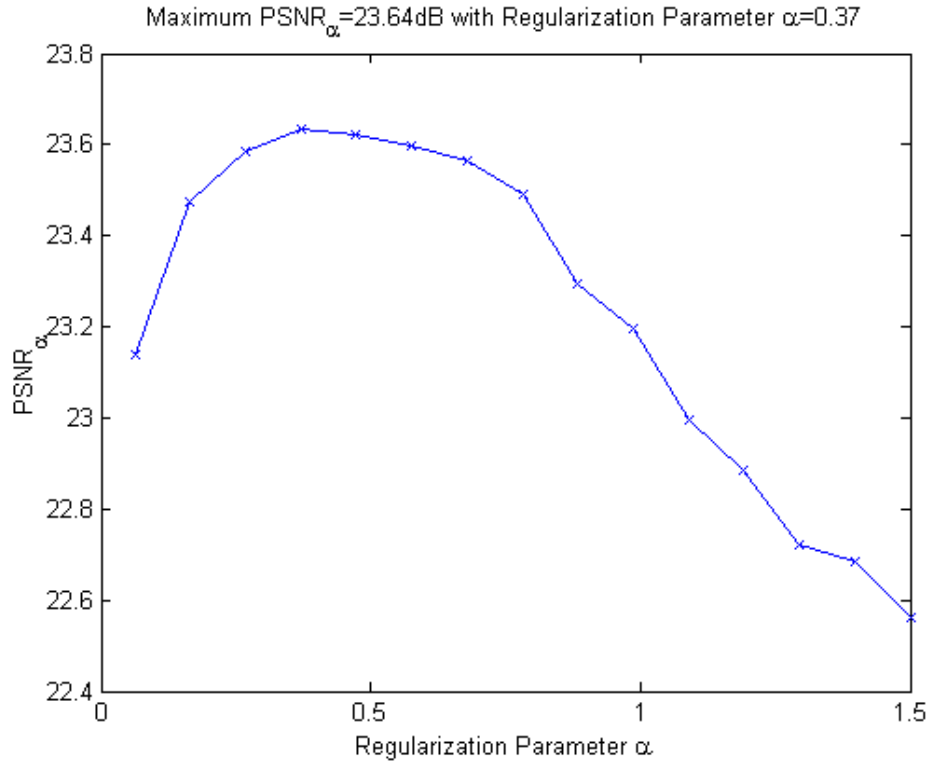
* The minimum value is calculated based on the low-resolution images used in the synthetic simulations.

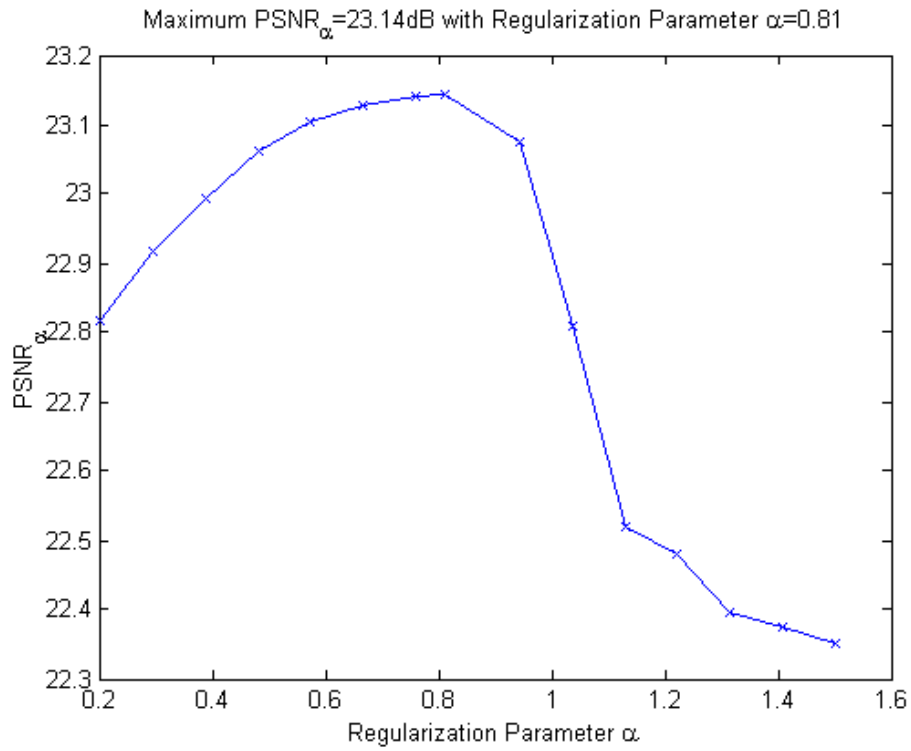
Table 5. Justification of (33) for “Lena”

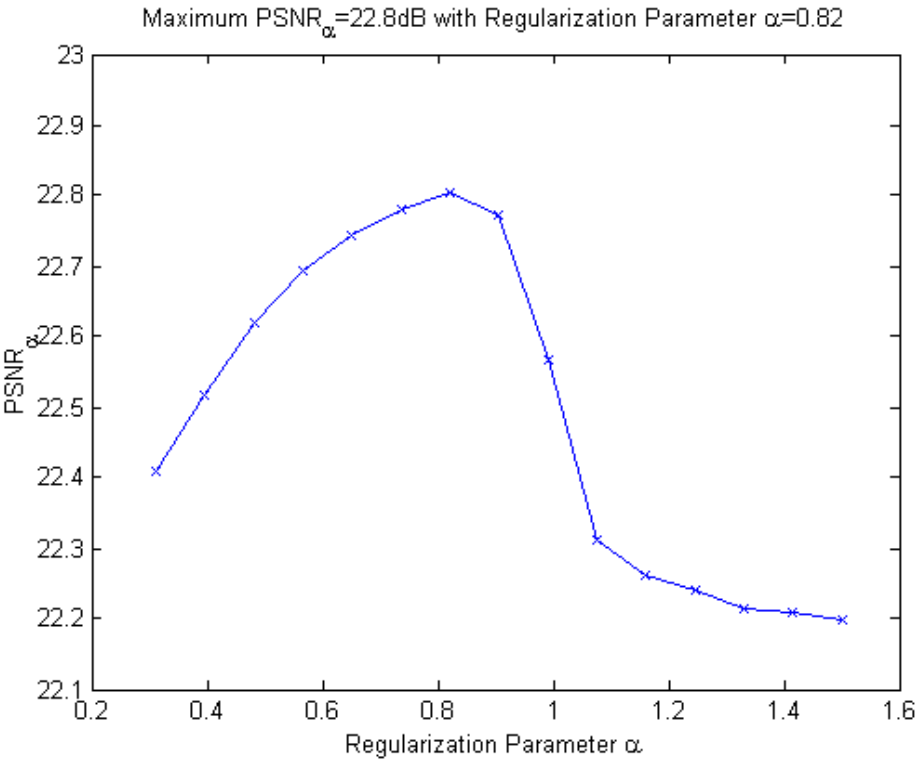
	Case I	**Case II**	**Case III**	**Case IV**
$\min\left\{\left\|\mathbf{y}_k\right\|^2\right\}^*$	8.9106e+007	8.9063e+007	8.9487e+007	8.9531e+007
$\frac{\left\|\mathbf{z}\right\|^2}{L_1 L_2}$	9.3498e+007	9.3498e+007	9.3498e+007	9.3498e+007
$\left\|\mathbf{y}_k - \mathbf{W}_{\mathbf{s},k}\mathbf{z}\right\|^2$	4.096e+003	8.192e+003	2.6634e+005	4.096e+005
$\left\|\mathbf{D}\mathbf{z}\right\|^2$	7.0386e+006	7.0386e+006	7.0386e+006	7.0386e+006
Ratio 1	2.1755e+004	1.0872e+003	335.9853	218.5827
Ratio 2	12.6598	12.6535	12.7138	12.7201

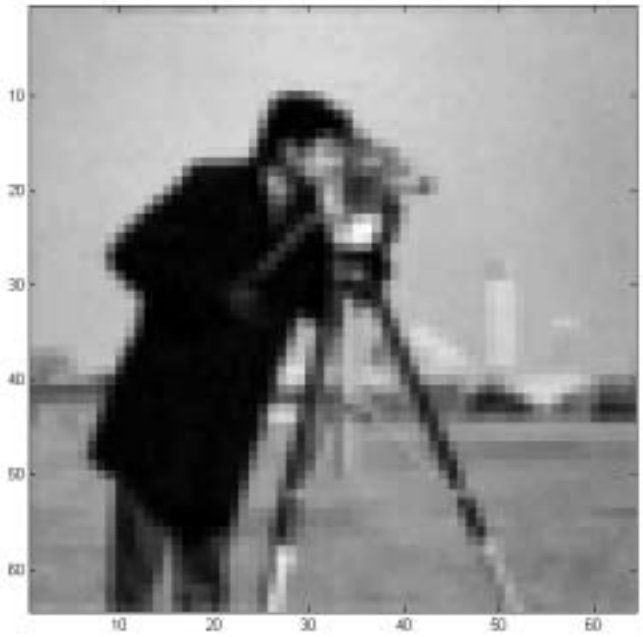
* The minimum value is calculated based on the low-resolution images used in the synthetic simulations.


Maximum PSNR$_\alpha$=23.98dB with Regularization Parameter α=0.34
PSNR$_\alpha$
Regularization Parameter α
24
23.95
23.9
23.85
23.8
23.75
23.7
23.65
23.6
23.55
23.5
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7


Maximum PSNR$_\alpha$=23.64dB with Regularization Parameter α=0.37
PSNR$_\alpha$
Regularization Parameter α
23.8
23.6
23.4
23.2
23
22.8
22.6
22.4
0
0.5
1
1.5

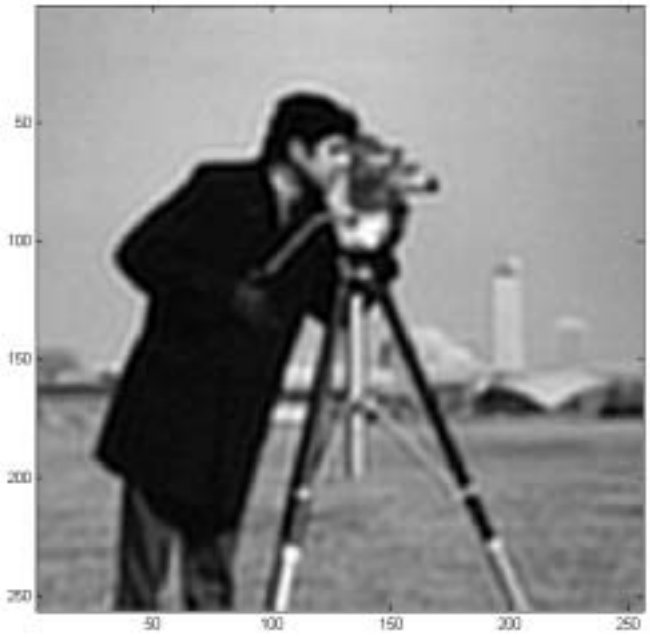

Maximum PSNR$_\alpha$=23.14dB with Regularization Parameter α=0.81
PSNR$_\alpha$
Regularization Parameter α
23.2
23.1
23
22.9
22.8
22.7
22.6
22.5
22.4
22.3
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6

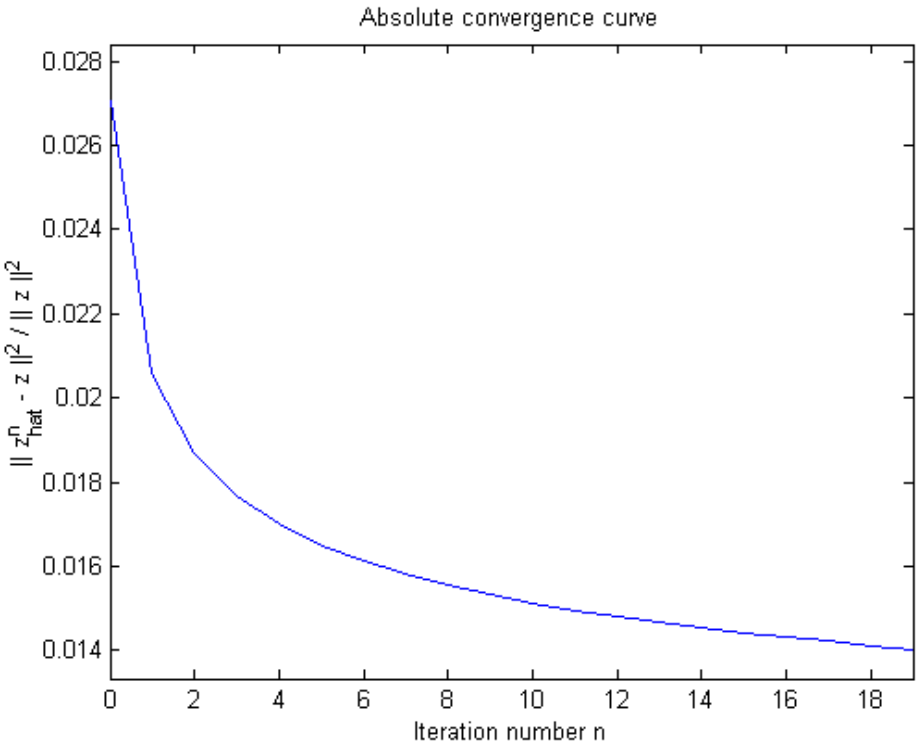

Maximum PSNR$_\alpha$=22.8dB with Regularization Parameter α=0.82
PSNR$_\alpha$
Regularization Parameter α
23
22.9
22.8
22.7
22.6
22.5
22.4
22.3
22.2
22.1
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6

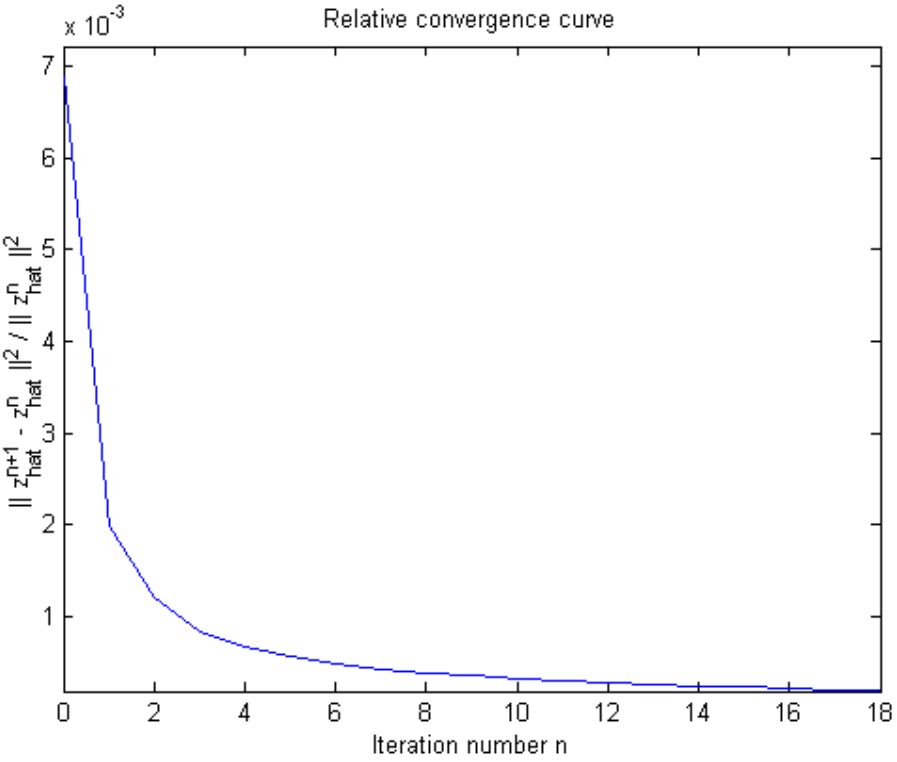

10
20
30
40
50
60
10
20
30
40
50
60

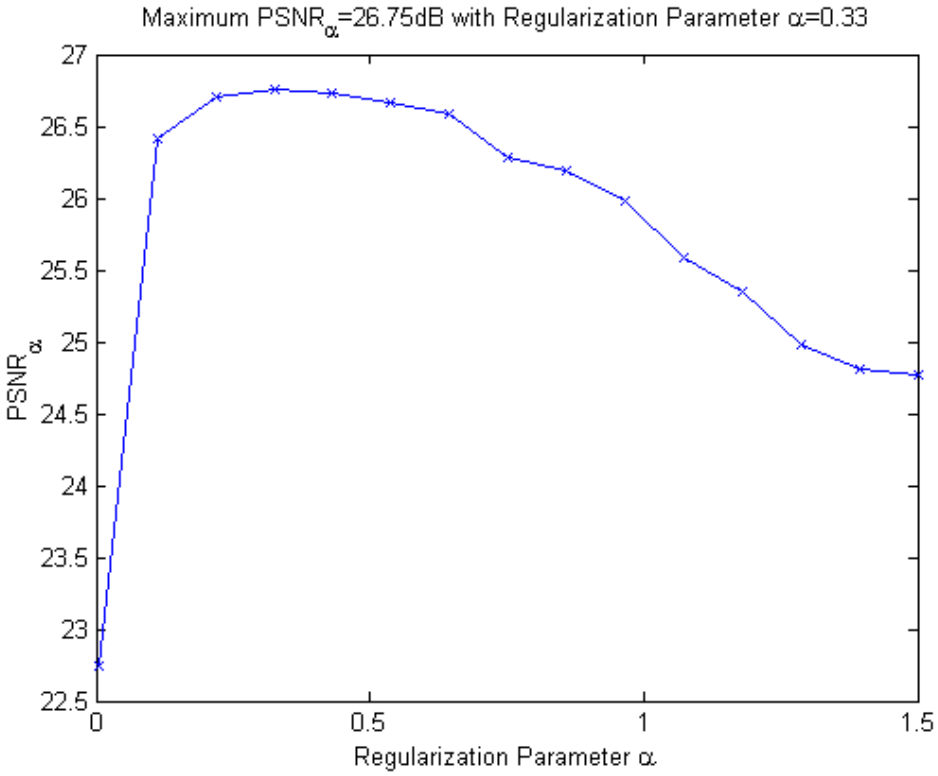

50
100
150
200
250
50
100
150
200
250

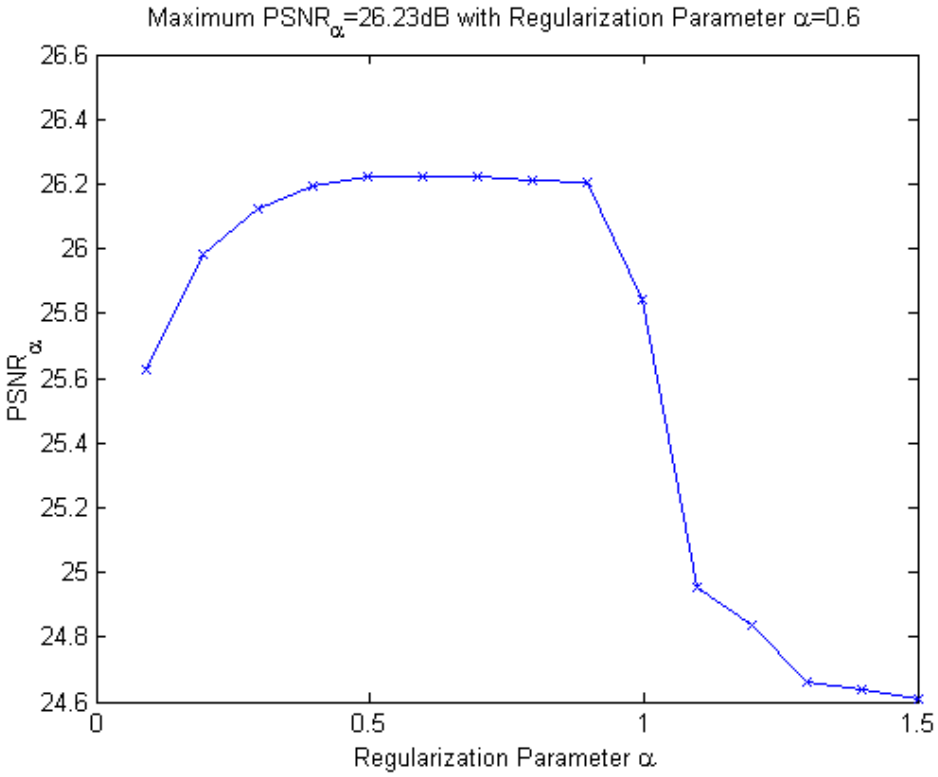

50
100
150
200
250
50
100
150
200
250

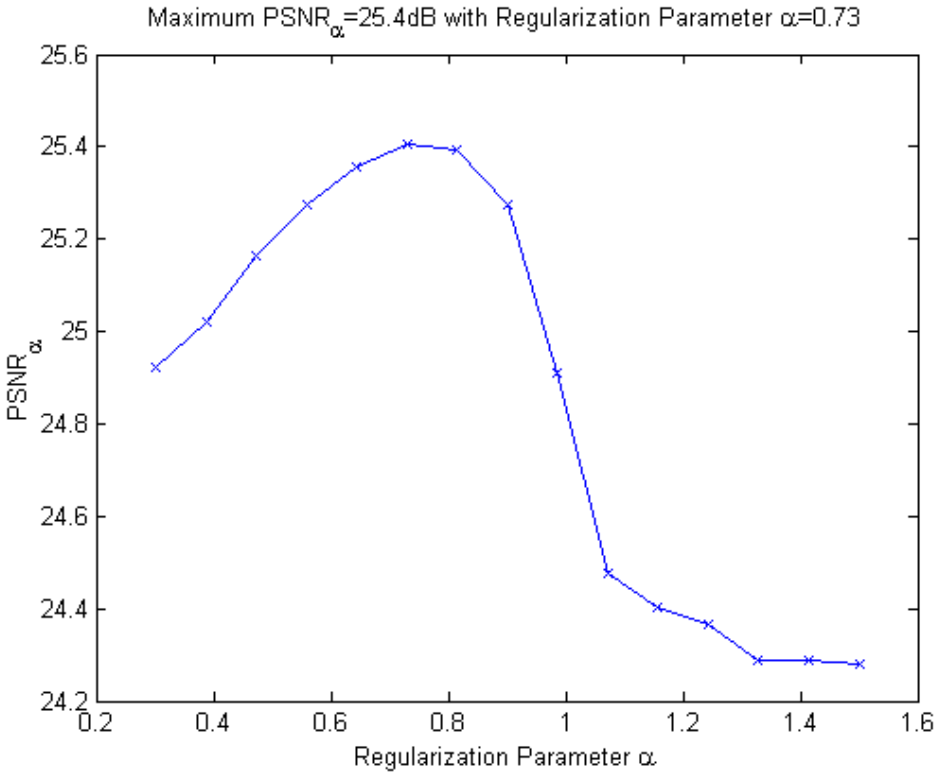

50
100
150
200
250
50
100
150
200
250

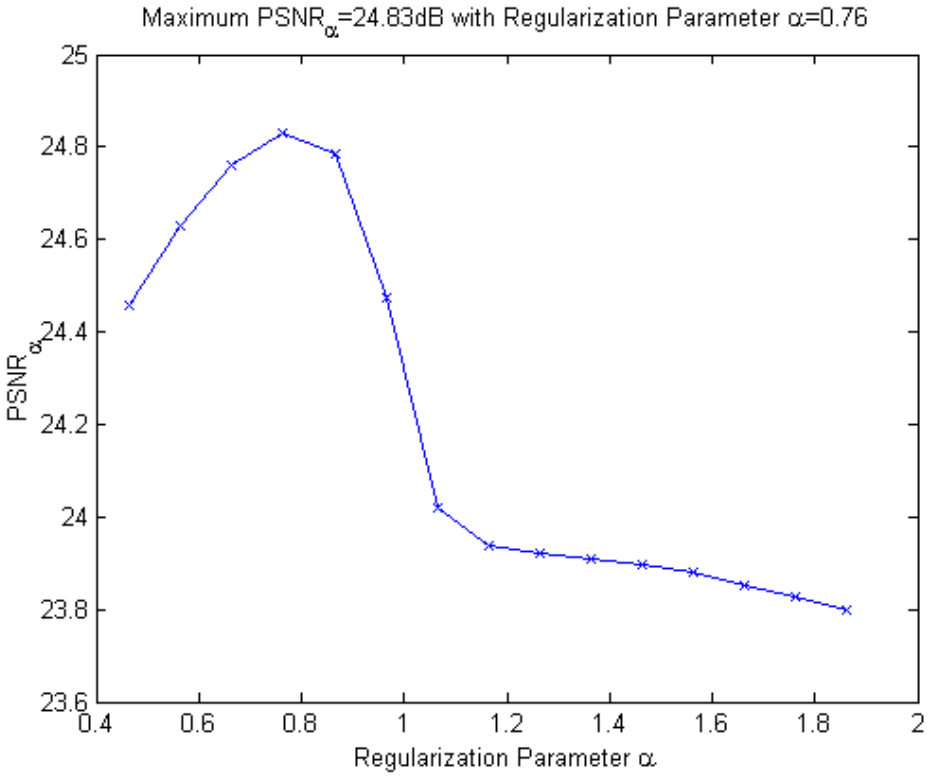

Absolute convergence curve
0.028
0.026
0.024
0.022
0.02
0.018
0.016
0.014
|| z^n_hat - z ||^2 / || z ||^2
0
2
4
6
8
10
12
14
16
18
Iteration number n


Relative convergence curve
x 10^-3
7
6
5
4
3
2
1
|| z^n+1_hat - z^n_hat ||^2 / || z^n_hat ||^2
0
2
4
6
8
10
12
14
16
18
Iteration number n


Maximum PSNR$_\alpha$=26.75dB with Regularization Parameter α=0.33
PSNR$_\alpha$
Regularization Parameter α
27
26.5
26
25.5
25
24.5
24
23.5
23
22.5
0
0.5
1
1.5


Maximum PSNR$_\alpha$=26.23dB with Regularization Parameter α=0.6
PSNR$_\alpha$
Regularization Parameter α
26.6
26.4
26.2
26
25.8
25.6
25.4
25.2
25
24.8
24.6
0
0.5
1
1.5


Maximum PSNR$_\alpha$=25.4dB with Regularization Parameter α=0.73
PSNR$_\alpha$
Regularization Parameter α
25.6
25.4
25.2
25
24.8
24.6
24.4
24.2
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6

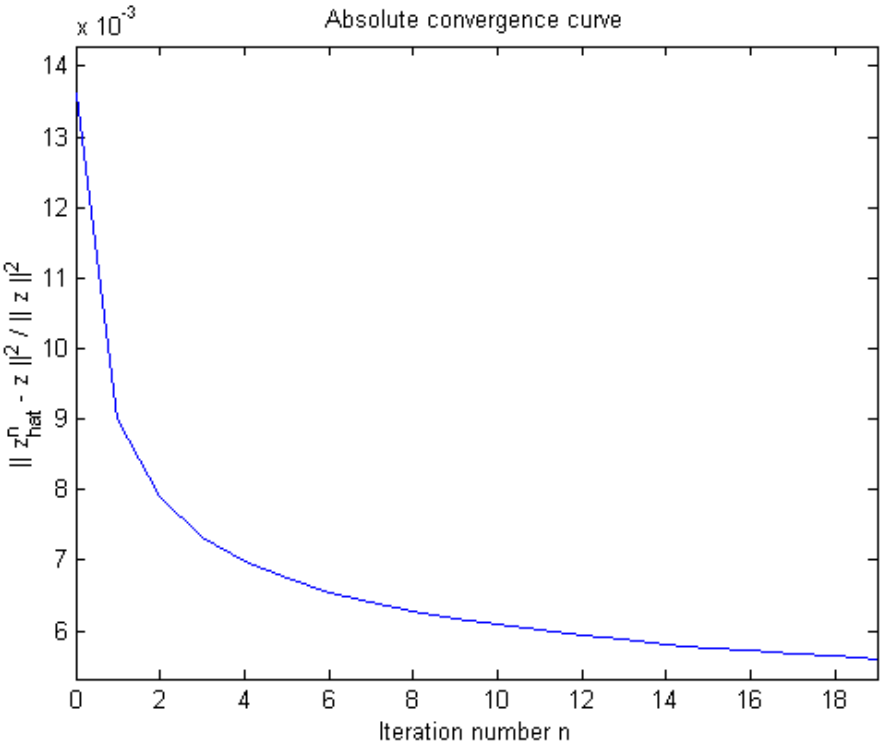

Maximum PSNR$_\alpha$=24.83dB with Regularization Parameter α=0.76
PSNR$_\alpha$
Regularization Parameter α
25
24.8
24.6
24.4
24.2
24
23.8
23.6
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
2

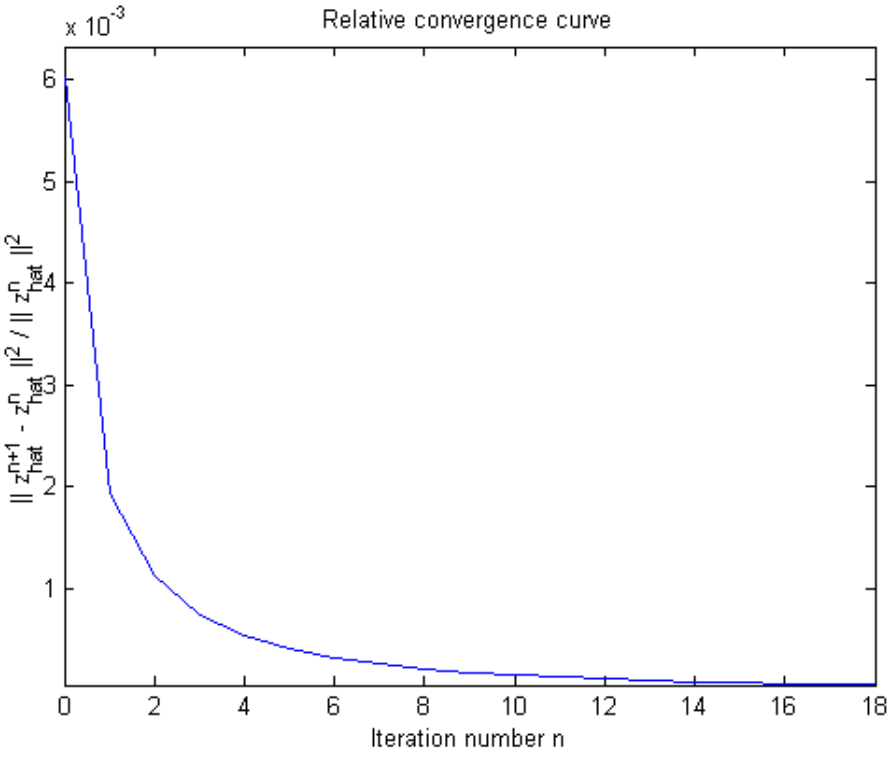

10
20
30
40
50
60
10
20
30
40
50
60

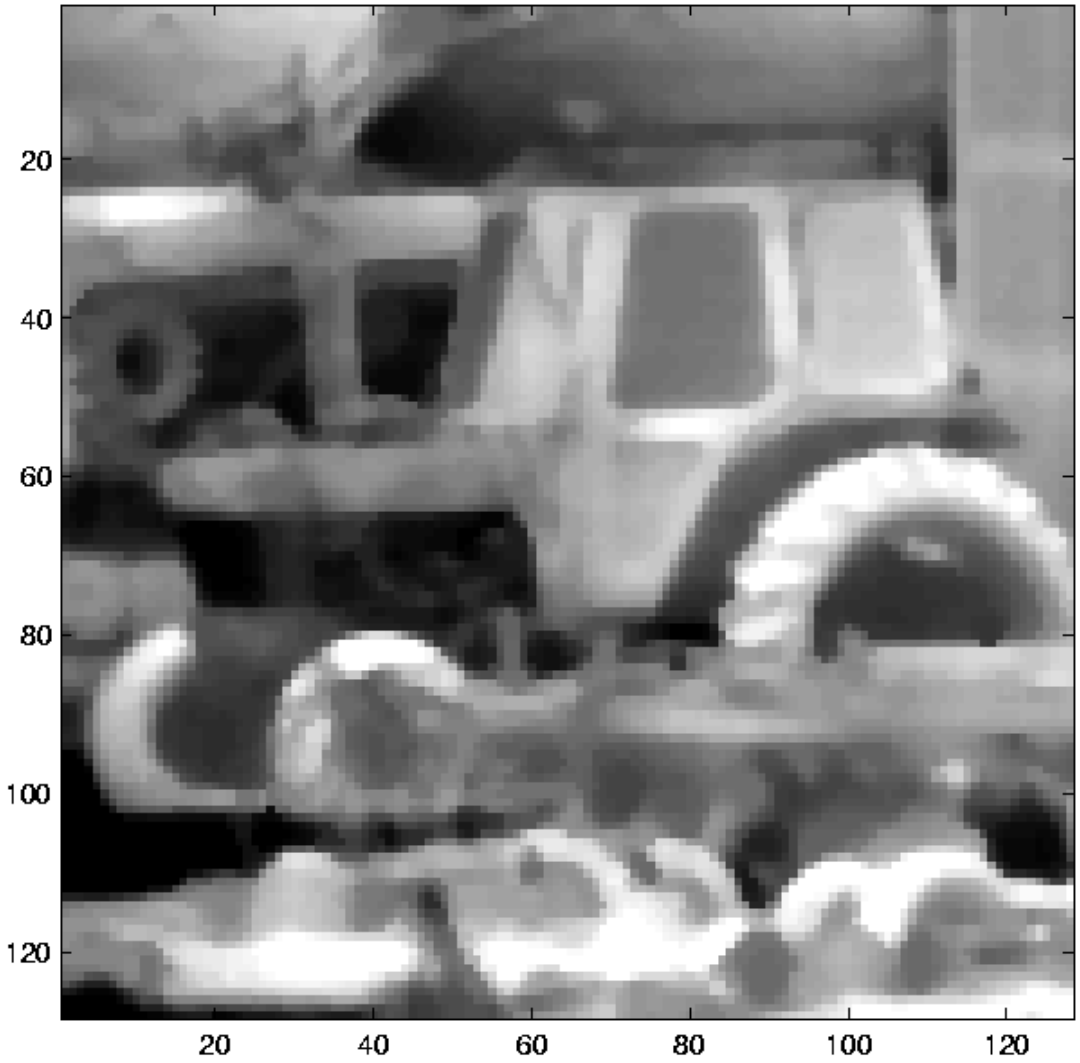

50
100
150
200
250
50
100
150
200
250

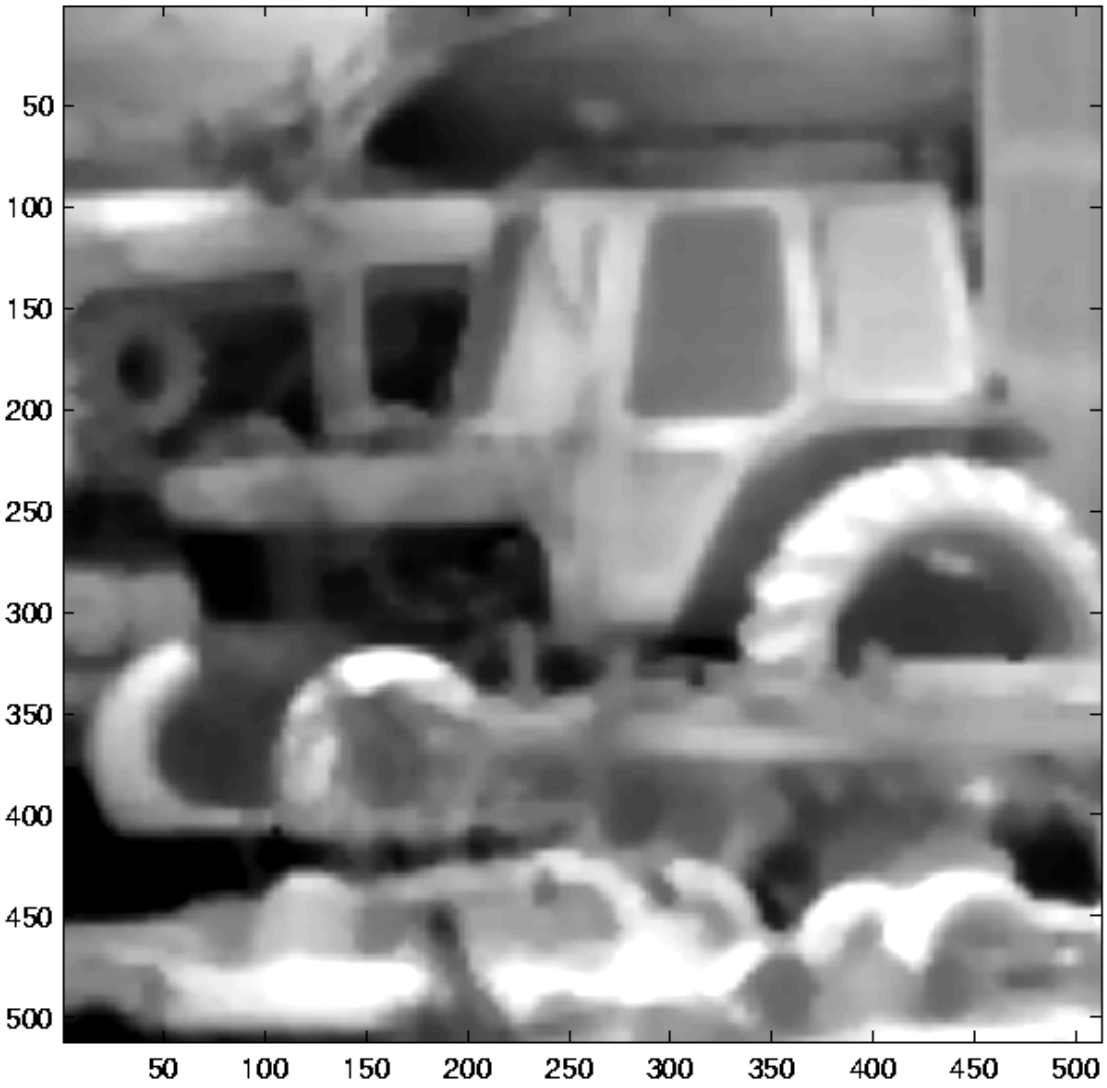

50
100
150
200
250
50
100
150
200
250

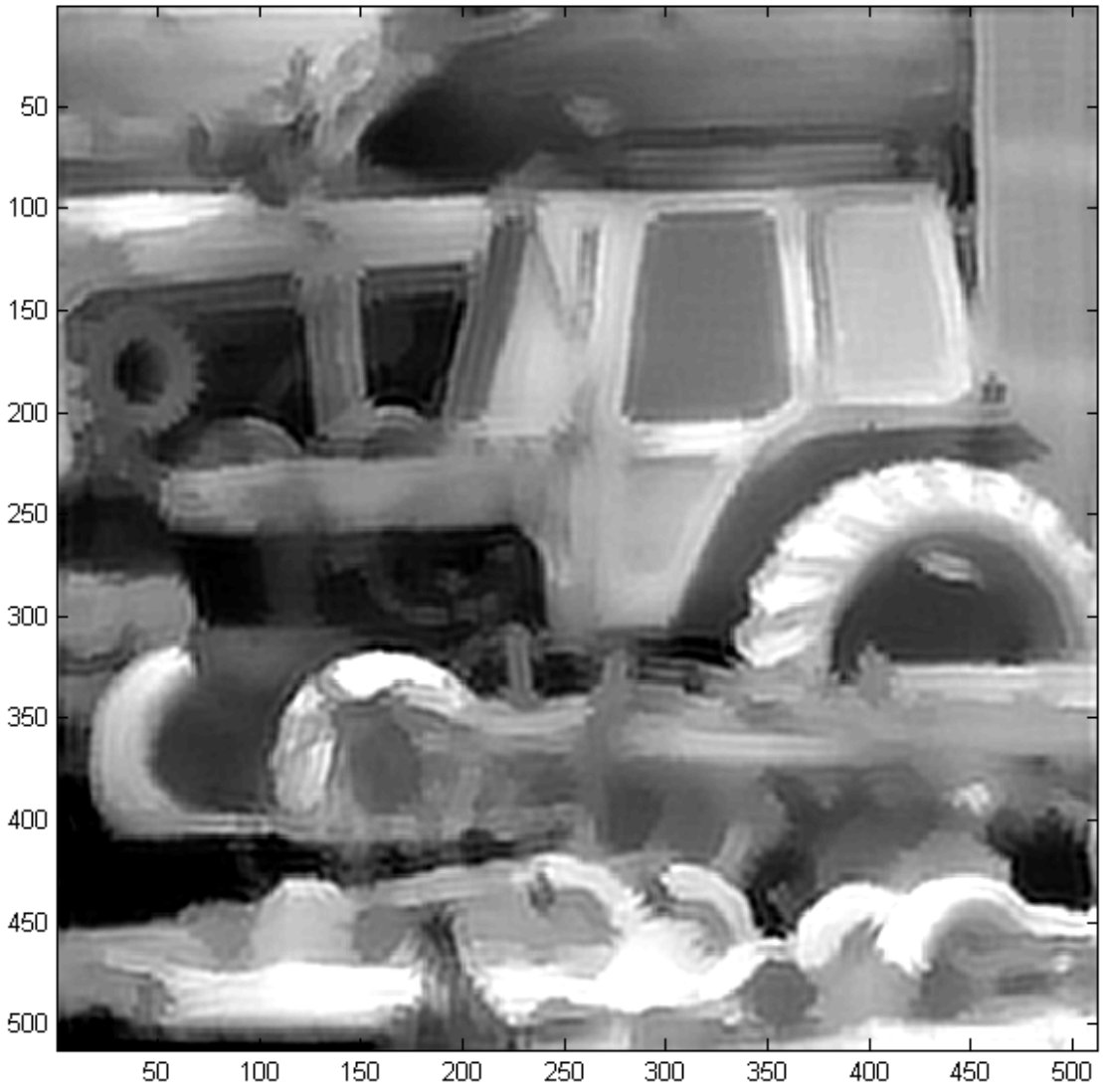

50
100
150
200
250
50
100
150
200
250

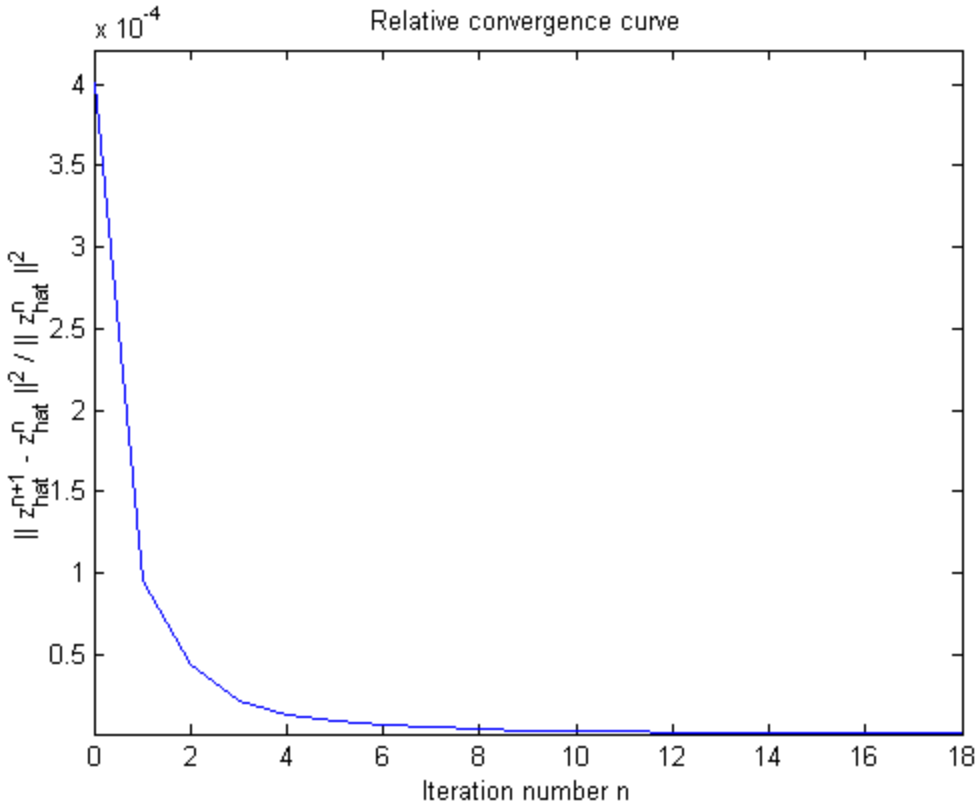

Absolute convergence curve
x 10^-3
14
13
12
11
10
9
8
7
6
0
2
4
6
8
10
12
14
16
18
Iteration number n
|| z^n_hat - z ||^2 / || z ||^2

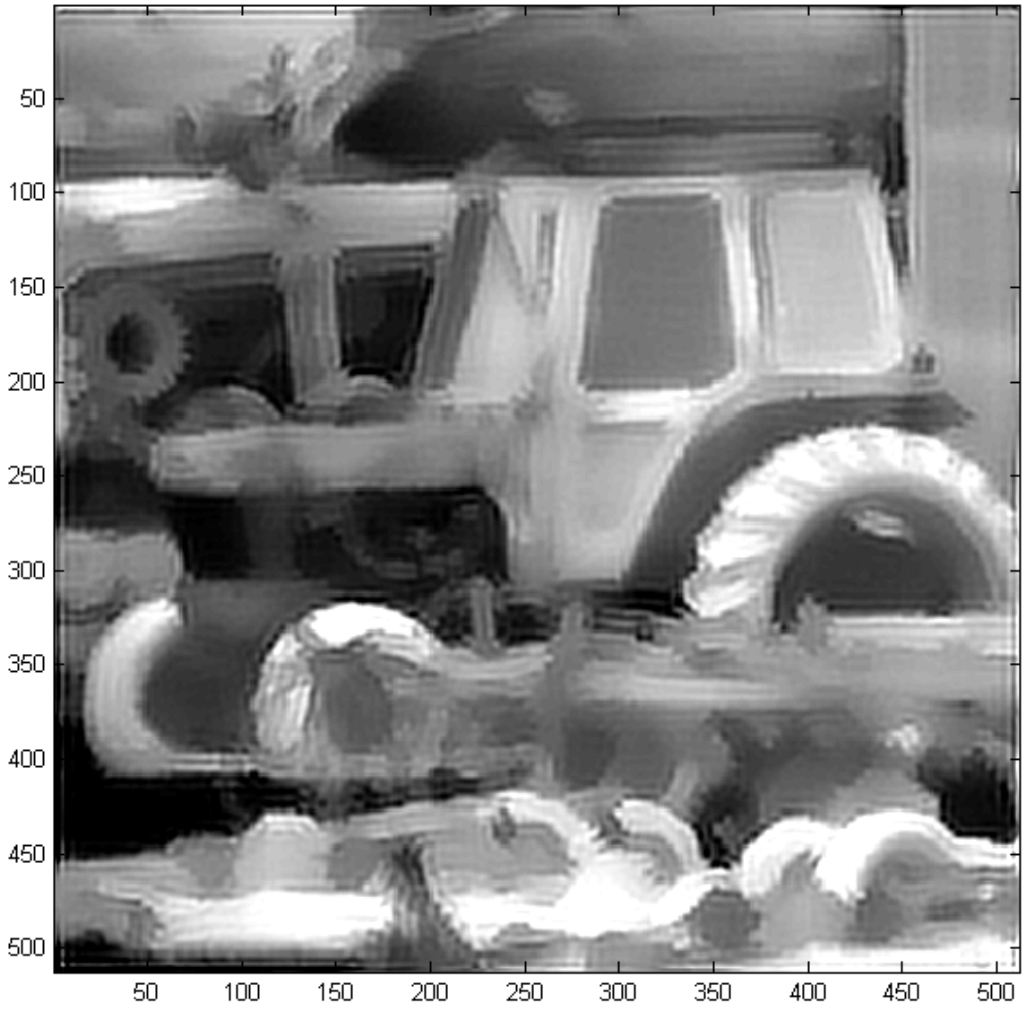

Relative convergence curve
x 10^-3
6
5
4
3
2
1
0
2
4
6
8
10
12
14
16
18
Iteration number n
|| z^n+1_hat - z^n_hat ||^2 / || z^n_hat ||^2

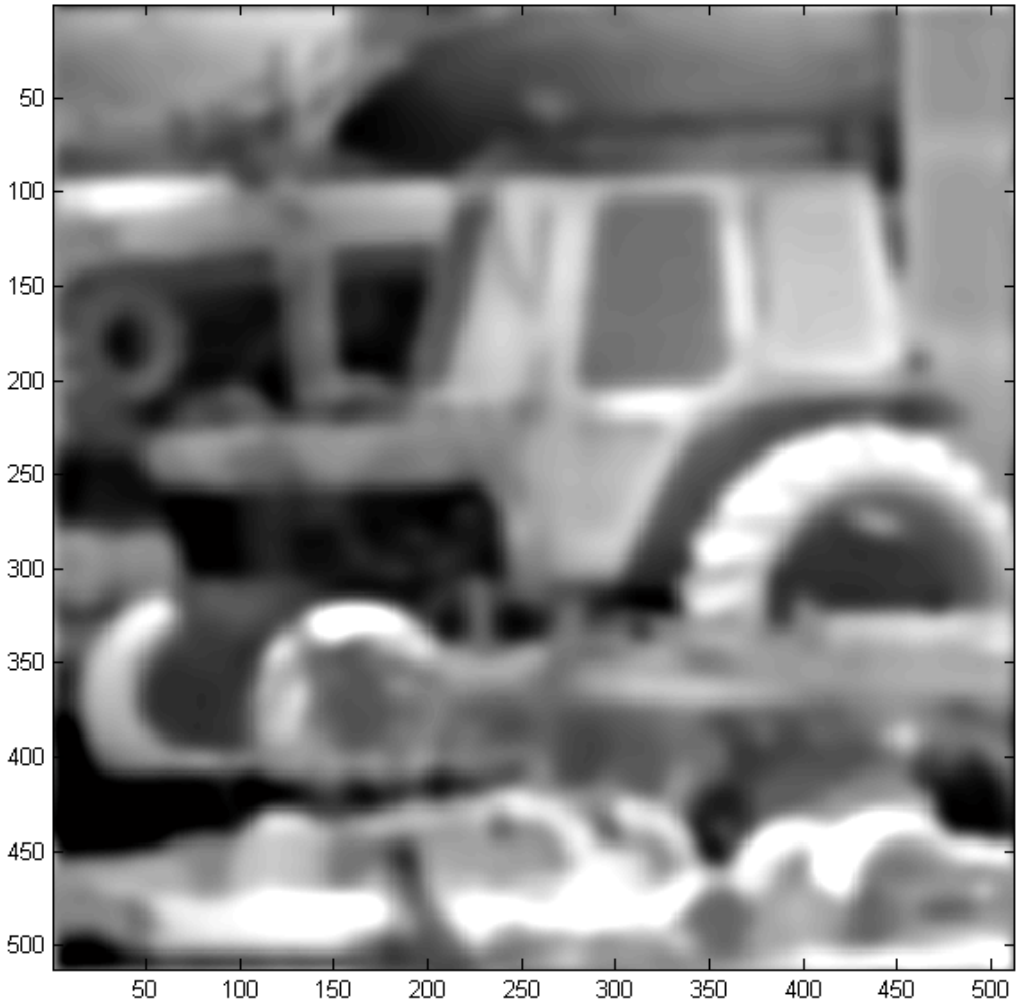

20
40
60
80
100
120
20
40
60
80
100
120

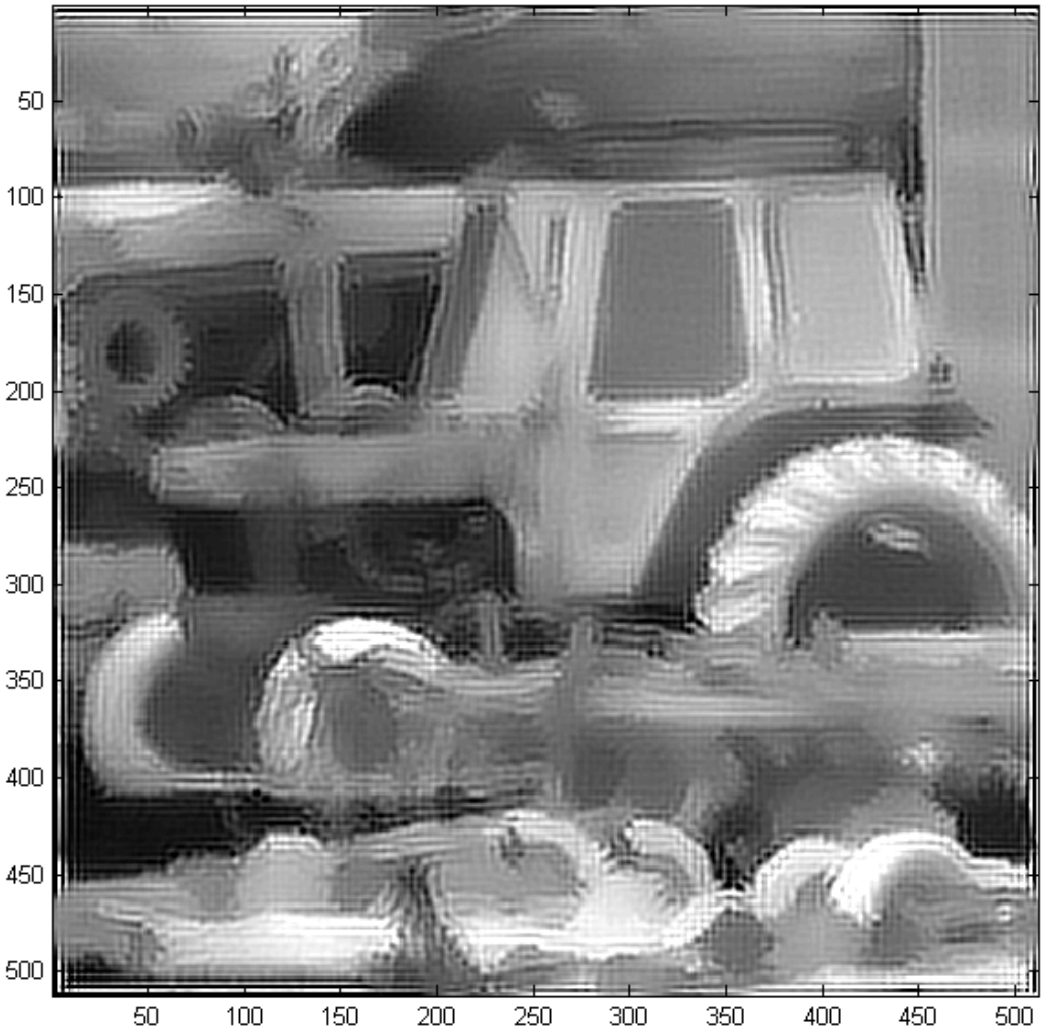

50
100
150
200
250
300
350
400
450
500
50
100
150
200
250
300
350
400
450
500


50
100
150
200
250
300
350
400
450
500
50
100
150
200
250
300
350
400
450
500


Relative convergence curve
x 10^-4
4
3.5
3
2.5
2
1.5
1
0.5
0
2
4
6
8
10
12
14
16
18
Iteration number n
|| z^n+1_hat - z^n_hat ||^2 / || z^n_hat ||^2


50
100
150
200
250
300
350
400
450
500
50
100
150
200
250
300
350
400
450
500


50
100
150
200
250
300
350
400
450
500
50
100
150
200
250
300
350
400
450
500


50
100
150
200
250
300
350
400
450
500
50
100
150
200
250
300
350
400
450
500

10. LIST OF FIGURES

Fig. 5(a). Original image (first frame of “truck” sequence)

Fig. 5(b). Bilinear interpolation of first frame of “truck” sequence

Fig. 5(c). Reconstructed high-resolution “truck” image using MAP with simultaneous estimation of regularization parameter

Fig. 5(d). Relative convergence curve of constructed high-resolution “truck” image using MAP with simultaneous estimation of regularization parameter

Fig. 5(e). Reconstructed high-resolution “truck” image using joint MAP with fixed regularization parameter $\alpha = 0.1$

Fig. 5(f). Reconstructed high-resolution “truck” image using joint MAP with fixed regularization parameter $\alpha = 10^{3}$

Fig. 5(g). Reconstructed high-resolution “truck” image using joint MAP with fixed regularization parameter $\alpha = 10^{-3}$