Parameter Estimation in Bayesian High-Resolution Image Reconstruction With Multisensors

Rafael Molina, *Member, IEEE*, Miguel Vega, Javier Abad, and Aggelos K. Katsaggelos, *Fellow, IEEE*

***Abstract*—In this paper, we consider the estimation of the unknown parameters for the problem of reconstructing a high-resolution image from multiple undersampled, shifted, degraded frames with subpixel displacement errors. We derive mathematical expressions for the iterative calculation of the maximum likelihood estimate of the unknown parameters given the low resolution observed images. These iterative procedures require the manipulation of block-semi circulant (BSC) matrices, that is block matrices with circulant blocks. We show how these BSC matrices can be easily manipulated in order to calculate the unknown parameters. Finally the proposed method is tested on real and synthetic images.**

***Index Terms*—Bayesian methods, high-resolution image reconstruction, parameter estimation.**

I. Introduction

HIGH-RESOLUTION images can, in some cases, be obtained directly from high precision optics and charge coupled devices (CCDs). However, due to hardware and cost limitations, imaging systems often provide us with only multiple low resolution images and in addition, there is a lower limit as to how small each CCD can be, due to the presence of shot noise [1] and the fact that the associated signal to noise ratio (SNR) is proportional to the size of the detector [2].

Low resolution images are common in many imaging applications, such as remote sensing, surveillance and astronomy. For example, the optical imaging camera on the Hubble Space Telescope (HST), the Wide Field Planetary Camera 2 (WFPC2) is composed of four separated 800×800 pixel CCD cameras, one of which, the planetary camera (PC), has an image scale of $0.046''$/pixel, while the other three arranged in a chevron around the PC have a scale of $0.0996''$/pixel (see [3] for details).

Over the last two decades research has been devoted to the problem of reconstructing a high-resolution image from multiple undersampled, shifted, degraded frames with subpixel displacement errors. Since the early work by Tsai and Huang [4], researchers, primarily within the engineering community, have focused on formulating the high resolution problem as a reconstruction (see [5] for a comprehensive review) or a recognition one (see [6]–[8], see also [9]). The astronomical community has also being working on the high resolution problem and has made available the Drizzle method to reconstruct high resolution images (see [3]). However, as reported in [5], not much work has been devoted to the efficient calculation of the reconstruction or to the estimation of the associated parameters.

Bose and Boo [10] use a block semi-circulant (BSC) matrix decomposition in order to calculate the *maximum a posteriori* (MAP) reconstruction, Chan *et al.* ([11]–[13]) and Nguyen ([14]–[16]) use preconditioning, wavelets, as well as BSC matrix decomposition. The efficient calculation of the MAP reconstruction is also addressed by Ng *et al.* ([17], [18]) and Elad and Hel-Or [19].

To our knowledge only the works by Bose *et al.* [20], Nguyen [15], [16], [21], [22] and to some extend [13] and [23]–[25] address the problem of parameter estimation. Furthermore, in those works the same parameter is assumed for all the low resolution images, although in the case of [20] the proposed method can be extended to different parameter for low resolution images (see [26]).

In this paper we use the general framework for frequency domain multi-channel signal processing developed by Katsaggelos *et al.* in [27] and Banham *et al.* in [28] (a formulation that was also obtained later by Bose and Boo [10] for the high resolution problem) to tackle the parameter estimation in high resolution problems. With the use of BSC matrices we show that all the matrix calculations involved in the parameter maximum likelihood estimation can be performed in the Fourier domain. The proposed approach can be used to assign the same parameter to all low resolution images or make them image dependent. We also show that the results are extensions of maximum likelihood estimation for single channel restoration problems.

The rest of the paper is organized as follows. The problem formulation is described in Section II. The high resolution prior image model and the process to obtain the low resolution images from the high resolution one is described in Section III. The application of the Bayesian paradigm to calculate the maximum a posteriori high resolution image and to estimate the parameters is described in Section IV. Experimental results are described in Section V. Finally, Section VI concludes the paper. The paper also contains two appendices on the application of EM algorithms to high resolution problems and the matrix manipulations needed to calculate the MAP estimate and the parameters.

II. Problem Formulation

Consider a sensor array with $L_1 \times L_2$ sensors (a sensor is, for instance, a CCD camera), where each sensor has $N_1 \times N_2$

Manuscript received October 15, 2002; revised April 10, 2003. This work was supported by the "Comisión Nacional de Ciencia y Tecnología" under Contract TIC2000-1275. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Mario A. T. Figueiredo.

R. Molina and J. Abad are with the Departamento de Ciencias de la Computación e I.A. Universidad de Granada, 18071 Granada, Spain (e-mail: rms@decsai.ugr.es).

M. Vega is with the Departamento de Lenguajes y Sistemas Informáticos. Universidad de Granada, 18071 Granada, Spain.

A. K. Katsaggelos is with the Department of Electrical and Computer Engineering, Northwestern University, Evanston, IL 60208-3118 USA.

Digital Object Identifier 10.1109/TIP.2003.818117

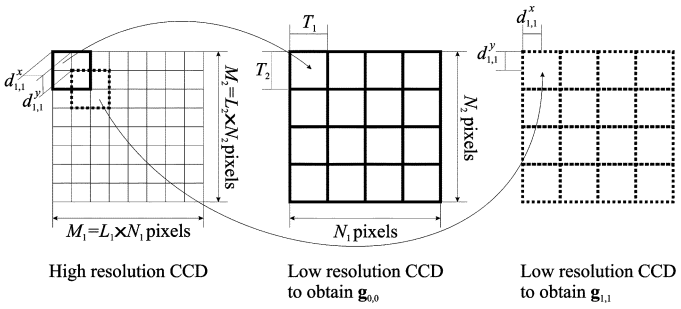


Fig. 1. Correspondence between high and low resolution pixels.

pixels and the size of each sensing element is $T_1 \times T_2$. Our aim is to reconstruct an $M_1 \times M_2$ high resolution image, where $M_1 = L_1 \times N_1$ and $M_2 = L_2 \times N_2$, from $L_1 \times L_2$ low-resolution observed images. Fig. 1 shows a visual description of the problem formulation for $L_1 = L_2 = 2$ and $N_1 = N_2 = 4$.

Note that in order for our goal make sense we need to assume that the original high-resolution scene is bandlimited to wavenumbers $L_1/(2T_1)$ and $L_2/(2T_2)$ along the horizontal and vertical directions, respectively. In this case the original $T_1N_1 \times T_2N_2$ high-resolution image can be reconstructed with $M_1 = L_1 \times N_1$ and $M_2 = L_2 \times N_2$ samples along the horizontal and vertical directions, respectively (see [29]).

To maintain the aspect ratio of the reconstructed image we consider the case where $L_1 = L_2 = L$, for simplicity we also assume that L is an even number. Each observed undersampled image is a shifted, downsampled version of the high-resolution image.

In the ideal case, the low resolution sensors are shifted with respect to each other by a value proportional to $(T_1/L)\times(T_2/L)$ (note that if the sensors are shifted by values proportional to $T_1 \times T_2$ the high-resolution image reconstruction problem becomes singular). However, in practice there can be small perturbations around those ideal locations (see [30] for a formulation without perturbations). Thus, the horizontal and vertical displacements $d^x_{l1,l2}$ and $d^y_{l1,l2}$ of the $[l_1, l_2] - th$ sensor with respect to the $[0, 0] - th$ reference sensor are given by (see Fig. 1)

$$\begin{aligned} d^x_{l1,l2} &= \frac{T_1}{L}\left(l_1 + \epsilon^x_{l1,l2}\right) \text{ and} \\ d^y_{l1,l2} &= \frac{T_2}{L}\left(l_2 + \epsilon^y_{l1,l2}\right), \\ l_1, l_2 &= 0, \ldots, L-1 \end{aligned} \tag{1}$$

where $\epsilon^x_{l1,l2}$ and $\epsilon^y_{l1,l2}$ denote respectively the normalized horizontal and vertical displacement errors. We assume that $|\epsilon^x_{l1,l2}| < 1/2$ and $|\epsilon^y_{l1,l2}| < 1/2$ with $\epsilon^x_{0,0} = \epsilon^y_{0,0} = 0$. The normalized horizontal and vertical displacement may be assumed to be known (see [10] and [18] for details). An approach where the displacements are assumed unknown and are estimated simultaneously with the high-resolution image is presented in [23]–[25].

III. Image and Degradation Models

Let $\mathbf{f}$ be the $(M_1 \times M_2) \times 1$ high resolution image and $\mathbf{g}_{l1,l2}$ the $(N_1 \times N_2) \times 1$ observed low resolution image from the $(l1, l2) - th$ sensor, $(l1, l2) \in \{0, \ldots L-1\}^2$. Our goal is to reconstruct $\mathbf{f}$ from $\{\mathbf{g}_{l1,l2} | (l1, l2) \in \{0, \ldots L-1\}^2\}$ using the Bayesian paradigm.

The first step with this paradigm is the definition of a prior distribution, a probability distribution over high resolution images $\mathbf{f}$, $p(\mathbf{f}|\alpha)$. It is here where we incorporate information on the expected structure of $\mathbf{f}$. It is also necessary to specify $p(\mathbf{g}_{l1,l2}|\mathbf{f}, \beta_{l1,l2})$ the probability distribution of the observed low resolution image $\mathbf{g}_{l1,l2}$ if $\mathbf{f}$ were the 'true' high resolution image. These image and high to low resolution degradation models depend on the unknown parameters α and $\beta_{l1,l2}$, $(l1, l2) \in \{0, \ldots L-1\}^2$ that have to be estimated.

In order to apply the Bayesian paradigm to this problem we define next our image and high to low degradation models.

A. Image Model

Our prior knowledge about the smoothness of the original high resolution image makes it possible to model the distribution of $\mathbf{f}$ by a simultaneous autoregression (SAR) [31]. Thus,

$$p(\mathbf{f}|\alpha) \propto \exp\left\{-\frac{1}{2}\alpha \mathbf{f}^t (I - \phi\mathbf{N})^t (I - \phi\mathbf{N})\mathbf{f}\right\} \tag{2}$$

where the entries of the matrix $\mathbf{N}$, N_{ij}, are equal to 1 if cells i and j are spatial neighbors (pixels at distance one) and zero otherwise and ϕ is just less than 0.25 in order to model smoothness and result in a positive definite matrix $(I - \phi\mathbf{N})$.

This model is characterized by

$$(I - \phi\mathbf{N})\mathbf{f} = \epsilon \tag{3}$$

where $\epsilon \sim \mathcal{N}(0, \alpha^{-1}I)$. Note that the parameter α measures the smoothness of the high resolution image $\mathbf{f}$.

Assuming a toroidal edge correction, the eigenvalues of the matrix $(I - \phi\mathbf{N})$ are $\lambda_{ij} = 1 - 2\phi(\cos(2\pi i/M_1) + \cos(2\pi j/M_2))$, $i = 1, 2, \ldots, M_1$, $j = 1, 2, \ldots, M_2$. Then the SAR distribution is given by

$$p(\mathbf{f}|\alpha) = \frac{1}{Z_{prior}(\alpha)} \exp\left\{-\frac{1}{2}\alpha \|\mathbf{C}\mathbf{f}\|^2\right\} \tag{4}$$

where $Z_{prior}(\alpha) = (\prod_{i,j} \lambda^2_{ij})^{-1/2}(2\pi/\alpha)^{(M_1 \times M_2)/2}$, $\mathbf{C} = I - \phi\mathbf{N}$.

From the regularization point of view, the SAR model imposes conditions on the second differences of the image.

B. Model for Obtaining the Low-Resolution Observed Images

The process to obtain the observed low resolution image by the $(l1, l2) - th$ sensor, $\mathbf{g}_{l1,l2}$, from $\mathbf{f}$ can be modeled as follows (see Fig. 1 for the correspondence between the high and low resolution image pixels). First, $\mathbf{f}^{l1,l2}$ is obtained. This image represents a blurred version of the original high-resolution image, according to

$$\mathbf{f}^{l1,l2} = \mathbf{H}_{l1,l2}\mathbf{f} \tag{5}$$

where $\mathbf{H}_{l1,l2}$ is an $(M_1 \times M_2) \times (M_1 \times M_2)$ matrix and may have different forms. In [10], [18], $\mathbf{H}_{l1,l2}$ is associated to the blurring function

$$h_{l1,l2}(x, y) = h^1_{l1}(x) h^2_{l2}(y) \tag{6}$$

with

$$h_l^i(u) = \begin{cases} \frac{1}{L}\left(\frac{1}{2} - \epsilon_l^i\right) & u = -\frac{L}{2} \\ \frac{1}{L} & |u| < \frac{L}{2} \\ \frac{1}{L}\left(\frac{1}{2} + \epsilon_l^i\right) & u = \frac{L}{2} \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

where $l \in \{0, \dots L-1\}$ and $i = 1, 2$. In [32] h_l^i has the form

$$h_l^i(u) = \begin{cases} \frac{1}{L} & u = -(L-1), \dots, 0 \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

note that in this case, $h_{l1}^1 = h_{l2}^2, \forall i$, $\epsilon_l^i = 0$, the normalized horizontal and vertical displacement errors in (1) satisfy $\epsilon_{l1,l2}^x = \epsilon_{l1,l2}^y = 0$, and $\mathbf{H}_{l1,l2} = \mathbf{H}, \forall l1, l2 = 0, \dots, L-1$.

Let $\mathbf{D}_{l1}$ and $\mathbf{D}_{l2}$ now be the 1-D downsampling matrices defined by

$$\mathbf{D}_{l1} = \mathbf{I}_{N_1} \otimes \mathbf{e}_l^t \tag{9}$$

$$\mathbf{D}_{l2} = \mathbf{I}_{N_2} \otimes \mathbf{e}_l^t \tag{10}$$

where $\mathbf{I}_{N_i}$ is the $N_i \times N_i$ identity matrix, $\mathbf{e}_l$ is the $L \times 1$ unit vector whose nonzero element is in the $l-th$ position, $\otimes$ denotes the Kronecker product operator and t the transpose operator.

Then for each sensor the discrete low-resolution observed image $\mathbf{g}_{l1,l2}$ can be written as

$$\mathbf{g}_{l1,l2} = \mathbf{D}_{l1,l2}\mathbf{H}_{l1,l2}\mathbf{f} + \mathbf{v}_{l1,l2} \tag{11}$$

where $\mathbf{H}_{l1,l2}$ has been defined in (5)

$$\mathbf{D}_{l1,l2} = \mathbf{D}_{l1} \otimes \mathbf{D}_{l2} \tag{12}$$

denotes the $(N_1 \times N_2) \times (M_1 \times M_2)$ 2D downsampling matrix and $\mathbf{v}_{l1,l2}$ is modeled as independent white noise with variance $\beta_{l1,l2}^{-1}$.

If $\mathbf{W}_{l1,l2}$ denotes the $(N_1 \times N_2) \times (M_1 \times M_2)$ matrix

$$\mathbf{W}_{l1,l2} = \mathbf{D}_{l1,l2}\mathbf{H}_{l1,l2} \tag{13}$$

then we have

$$p(\mathbf{g}_{l1,l2}|\mathbf{f}, \beta_{l1,l2}) \propto \frac{1}{Z(\beta_{l1,l2})} \exp\left[-\frac{\beta_{l1,l2}}{2}||\mathbf{g}_{l1,l2} - \mathbf{W}_{l1,l2}\mathbf{f}||^2\right] \tag{14}$$

where $Z(\beta_{l1,l2}) = (2\pi/\beta_{l1,l2})^{(N_1 \times N_2)/2}$. We denote by $\mathbf{g}$ the sum of the upsampled low-resolution images, that is,

$$\mathbf{g} = \sum_{u=0}^{L-1}\sum_{v=0}^{L-1} \mathbf{D}_{u,v}^t \mathbf{g}_{u,v}. \tag{15}$$

Then

$$p(\mathbf{g}|\mathbf{f}, \underline{\beta}) \propto \frac{1}{Z_{noise}(\underline{\beta})} \times \exp\left[-\frac{1}{2}\sum_{l1=0}^{L-1}\sum_{l2=0}^{L-1} \beta_{l1,l2}||\mathbf{g}_{l1,l2} - \mathbf{W}_{l1,l2}\mathbf{f}||^2\right] \tag{16}$$

where $\underline{\beta} = (\beta_{l1,l2}|(l1, l2) \in \{0, \dots, L-1\}^2)$ and $Z_{noise}(\underline{\beta}) = \prod_{l1=0}^{L-1}\prod_{l2=0}^{L-1} Z(\beta_{l1,l2})$.

From now on, given an $(M_1 \times M_2) \times 1$ column vector $\mathbf{u}$, we will denote by $\mathbf{u}_{l1,l2}$ the $(N_1 \times N_2) \times 1$ column vector given by

$$\mathbf{u}_{l1,l2} = \mathbf{D}_{l1,l2}\mathbf{u}. \tag{17}$$

IV. Bayesian Analysis

The steps we follow in this paper to estimate the parameters, α and $\underline{\beta}$, and the original image are as follows:

1) Step I: Estimation of the Parameters: $\hat{\alpha}$ and $\hat{\underline{\beta}} = (\hat{\beta}_{l1,l2}|(l1, l2) \in \{0, \dots, L-1\}^2)$ are first selected as

$$\hat{\alpha}, \hat{\underline{\beta}} = \arg\max_{\alpha, \underline{\beta}} \mathcal{L}_{\mathbf{g}}(\alpha, \underline{\beta}) = \arg\max_{\alpha, \underline{\beta}} \log p(\mathbf{g}|\alpha, \underline{\beta}) \tag{18}$$

where

$$p(\mathbf{g}|\alpha, \underline{\beta}) = \int_{\mathbf{f}} p(\mathbf{f}|\alpha) p(\mathbf{g}|\mathbf{f}, \underline{\beta}) d\mathbf{f}. \tag{19}$$

2) Step II: Estimation of the Original Image: Once the parameters have been estimated, the estimation of the original image, $\mathbf{f}_{(\hat{\alpha}, \hat{\underline{\beta}})}$, is selected as the image satisfying

$$\mathbf{f}_{(\hat{\alpha}, \hat{\underline{\beta}})} = \arg\min_{\mathbf{f}} \hat{\alpha}||\mathbf{C}\mathbf{f}||^2 + \sum_{l1=0}^{L-1}\sum_{l2=0}^{L-1} \hat{\beta}_{l1,l2}||\mathbf{g}_{l1,l2} - \mathbf{W}_{l1,l2}\mathbf{f}||^2 \tag{20}$$

which produces

$$\mathbf{f}_{(\hat{\alpha}, \hat{\underline{\beta}})} = \mathbf{Q}(\hat{\alpha}, \hat{\underline{\beta}})^{-1} \sum_{l1=0}^{L-1}\sum_{l2=0}^{L-1} \hat{\beta}_{l1,l2}\mathbf{W}_{l1,l2}^t \mathbf{g}_{l1,l2} \tag{21}$$

where

$$\mathbf{Q}(\hat{\alpha}, \hat{\underline{\beta}}) = \hat{\alpha}\mathbf{C}^t\mathbf{C} + \sum_{l1=0}^{L-1}\sum_{l2=0}^{L-1} \hat{\beta}_{l1,l2}\mathbf{W}_{l1,l2}^t\mathbf{W}_{l1,l2}. \tag{22}$$

Note that we are using maximum likelihood for estimating the parameter and *maximum a posteriori* (MAP) for estimating $\mathbf{f}$. Furthermore, although steps I and II are separated, the iterative scheme to be proposed performs both estimations simultaneously.

The estimation process we are using could be performed within the so called hierarchical Bayesian approach (see [33]) by including priors on the unknown parameter $\hat{\alpha}$ and vector $\hat{\underline{\beta}}$. However, the possibility of incorporating additional knowledge on them by means of gamma or other distributions will not be discussed here (see [33] and [34]).

Let us examine the estimation process in detail. Fixing α and $\underline{\beta}$ and expanding the function

$$M(\mathbf{f}, \mathbf{g}|\alpha, \underline{\beta}) = \alpha||\mathbf{C}\mathbf{f}||^2 + \sum_{l1=0}^{L-1}\sum_{l2=0}^{L-1} \beta_{l1,l2}||\mathbf{g}_{l1,l2} - \mathbf{W}_{l1,l2}\mathbf{f}||^2 \tag{23}$$

around $\mathbf{f}_{(\alpha, \underline{\beta})}$, we have

$$M(\mathbf{f}, \mathbf{g}|\alpha, \underline{\beta}) = M\left(\mathbf{f}_{(\alpha, \underline{\beta})}, \mathbf{g}|\alpha, \underline{\beta}\right) + \frac{1}{2}\left(\mathbf{f} - \mathbf{f}_{(\alpha, \underline{\beta})}\right)^t \mathbf{Q}(\alpha, \underline{\beta})\left(\mathbf{f} - \mathbf{f}_{(\alpha, \underline{\beta})}\right). \tag{24}$$

Therefore

$$p(\mathbf{g}|\alpha,\underline{\beta}) = \frac{1}{Z_{prior}(\alpha)Z_{noise}(\underline{\beta})}\exp\left\{-\frac{1}{2}M\left(\mathbf{f}_{(\alpha,\underline{\beta})},\mathbf{g}|\alpha,\underline{\beta}\right)\right\} \times \int_{\mathbf{f}} \exp\{\ldots\text{the same}\ldots\}\,d\mathbf{f} = \text{the same}. \quad (25)$$

We then have

$$2\mathcal{L}_{\mathbf{g}}(\alpha,\underline{\beta}) = -\alpha\left\|\mathbf{C}\mathbf{f}_{(\alpha,\underline{\beta})}\right\|^2 - \sum_{l1=0}^{L-1}\sum_{l2=0}^{L-1}\beta_{l1,l2} \times \left\|\mathbf{g}_{l1,l2} - \mathbf{W}_{l1,l2}\mathbf{f}_{(\alpha,\underline{\beta})}\right\|^2 - \log\left|\mathbf{Q}(\alpha,\underline{\beta})\right| - 2\log Z_{prior}(\alpha) - 2\log Z_{noise}(\underline{\beta}) + \text{const}.$$

We now differentiate $-2\mathcal{L}_{\mathbf{g}}(\alpha,\underline{\beta})$ with respect to α and $\underline{\beta}$ so as to find the conditions which are satisfied at the maxima. We have

$$\left\|\mathbf{C}\mathbf{f}_{(\alpha,\underline{\beta})}\right\|^2 + \text{trace}\left[\mathbf{Q}(\alpha,\underline{\beta})^{-1}\mathbf{C}^t\mathbf{C}\right] = \frac{M_1\times M_2}{\alpha} \quad (26)$$

$$\left\|\mathbf{g}_{l1,l2} - \mathbf{W}_{l1,l2}\mathbf{f}_{(\alpha,\underline{\beta})}\right\|^2 + \text{trace}\left[\mathbf{Q}(\alpha,\underline{\beta})^{-1}\mathbf{W}_{l1,l2}^t\mathbf{W}_{l1,l2}\right] = \frac{N_1\times N_2}{\beta_{l1,l2}}, \quad l1,l2 = 0,\ldots,L-1. \quad (27)$$

The following algorithm is proposed for the simultaneous estimation of the parameters and the high resolution image.

Algorithm 1

1) Choose α^0 and $\underline{\beta}^0$.
2) Compute $\mathbf{f}_{(\alpha^0,\underline{\beta}^0)}$ using (21) with $\hat{\alpha} = \alpha^0$ and $\hat{\underline{\beta}} = \underline{\beta}^0$.
3) For $k = 1,2,\ldots$
   a) Calculate α^k and $\underline{\beta}^k$ by substituting α^{k-1} and $\underline{\beta}^{k-1}$ in the left hand side of (26) and (27).
   b) Compute $\mathbf{f}_{(\alpha^k,\underline{\beta}^k)}$ using (21) with $\hat{\alpha} = \alpha^k$ and $\hat{\underline{\beta}} = \underline{\beta}^k$.
4) Go to 3 until $\|\mathbf{f}_{(\alpha^k,\underline{\beta}^k)} - \mathbf{f}_{(\alpha^{k-1},\underline{\beta}^{k-1})}\|^2 / \|\mathbf{f}_{(\alpha^{k-1},\underline{\beta}^{k-1})}\|^2$ is less than a prescribed bound.

A number of comments are now made regarding these equations.

A) If the same parameter is used for some low resolution observations, (26) and (27) become easier to solve. In particular if all the noise variances are assumed to be the same, that is $\beta_{l1,l2} = \beta, \forall l1, l2$, (27) becomes

$$\sum_{l1=0}^{L-1}\sum_{l2=0}^{L-1}\left(\left\|\mathbf{g}_{l1,l2} - \mathbf{W}_{l1,l2}\mathbf{f}_{(\alpha,\underline{\beta})}\right\|^2 + \text{trace}\left[\mathbf{Q}(\alpha,\underline{\beta})^{-1}\mathbf{W}_{l1,l2}^t\mathbf{W}_{l1,l2}\right]\right) = \frac{M_1\times M_2}{\beta}. \quad (28)$$

Fig. 2. Original 256 × 256 high resolution image.

Note that β refers here to a value and should not be confused with $\underline{\beta}$ that refers to a vector.

B) Since

$$\alpha\text{trace}\left[\mathbf{Q}(\alpha,\underline{\beta})^{-1}\mathbf{C}^t\mathbf{C}\right] + \sum_{l1=0}^{L-1}\sum_{l2=0}^{L-1}\beta_{l1,l2} \times\text{trace}\left[\mathbf{Q}(\alpha,\underline{\beta})^{-1}\mathbf{W}_{l1,l2}^t\mathbf{W}_{l1,l2}\right] = \text{trace}\left[\mathbf{I}_{M_1\times M_2}\right] \quad (29)$$

by using (26) and (27) we have that

$$\alpha\left\|\mathbf{C}\mathbf{f}_{(\alpha,\underline{\beta})}\right\|^2 + \alpha\text{trace}\left[\mathbf{Q}(\alpha,\underline{\beta})^{-1}\mathbf{C}^t\mathbf{C}\right] = M_1\times M_2 \quad (30)$$

and

$$\sum_{l1=0}^{L-1}\sum_{l2=0}^{L-1}\beta_{l1,l2}\left\|\mathbf{g}_{l1,l2} - \mathbf{W}_{l1,l2}\mathbf{f}_{(\alpha,\underline{\beta})}\right\|^2 + \sum_{l1=0}^{L-1}\sum_{l2=0}^{L-1} \times\beta_{l1,l2}\text{trace}\left[\mathbf{Q}(\alpha,\underline{\beta})^{-1}\mathbf{W}_{l1,l2}^t\mathbf{W}_{l1,l2}\right] = M_1\times M_2. \quad (31)$$

So we see that the maximum likelihood estimate (mle), satisfies,

$$\alpha\left\|\mathbf{C}\mathbf{f}_{(\alpha,\underline{\beta})}\right\|^2 + \sum_{l1=0}^{L-1}\sum_{l2=0}^{L-1}\beta_{l1,l2} \times\left\|\mathbf{g}_{l1,l2} - \mathbf{W}_{l1,l2}\mathbf{f}_{(\alpha,\underline{\beta})}\right\|^2 = M_1\times M_2 \quad (32)$$

which means that a fraction of the observations are used to calculate the misfit to the prior model ($\alpha\|\mathbf{C}\mathbf{f}_{(\alpha,\underline{\beta})}\|^2$), and another proportion is used for the misfit to the high to low resolution process ($\beta_{l1,l2}\|\mathbf{g}_{l1,l2} - \mathbf{W}_{l1,l2}\mathbf{f}_{(\alpha,\underline{\beta})}\|^2, l1,l2 = 0,\ldots,L-1$).

C) Algorithm 1 is, in fact an EM-algorithm [35] with complete data $\mathcal{X}^t = (\mathbf{f}^t, \mathbf{g}^t)$ and incomplete data $\mathcal{Y} = \mathbf{g} = [\mathbf{0}\,\mathbf{I}]^t\mathcal{X}$. Steps 3a and 3b iteratively increase $\mathcal{L}_{\mathbf{g}}(\alpha,\underline{\beta})$ (see Appendix I for details).

D) We note that in order to find the MAP estimate we need to invert $\mathbf{Q}(\alpha,\underline{\beta})$ and in order to estimate the parameters we have to calculate for $l1,l2 = 0,\ldots,L-1$ the values of $\text{trace}[\mathbf{Q}(\alpha,\underline{\beta})^{-1}\mathbf{C}^t\mathbf{C}]$ and $\text{trace}[\mathbf{Q}(\alpha,\underline{\beta})^{-1}\mathbf{W}_{l1,l2}^t\mathbf{W}_{l1,l2}]$. In Appendix II it is shown how these calculations can be performed in an efficient way in the frequency domain by utilizing properties of SBC matrices [27].

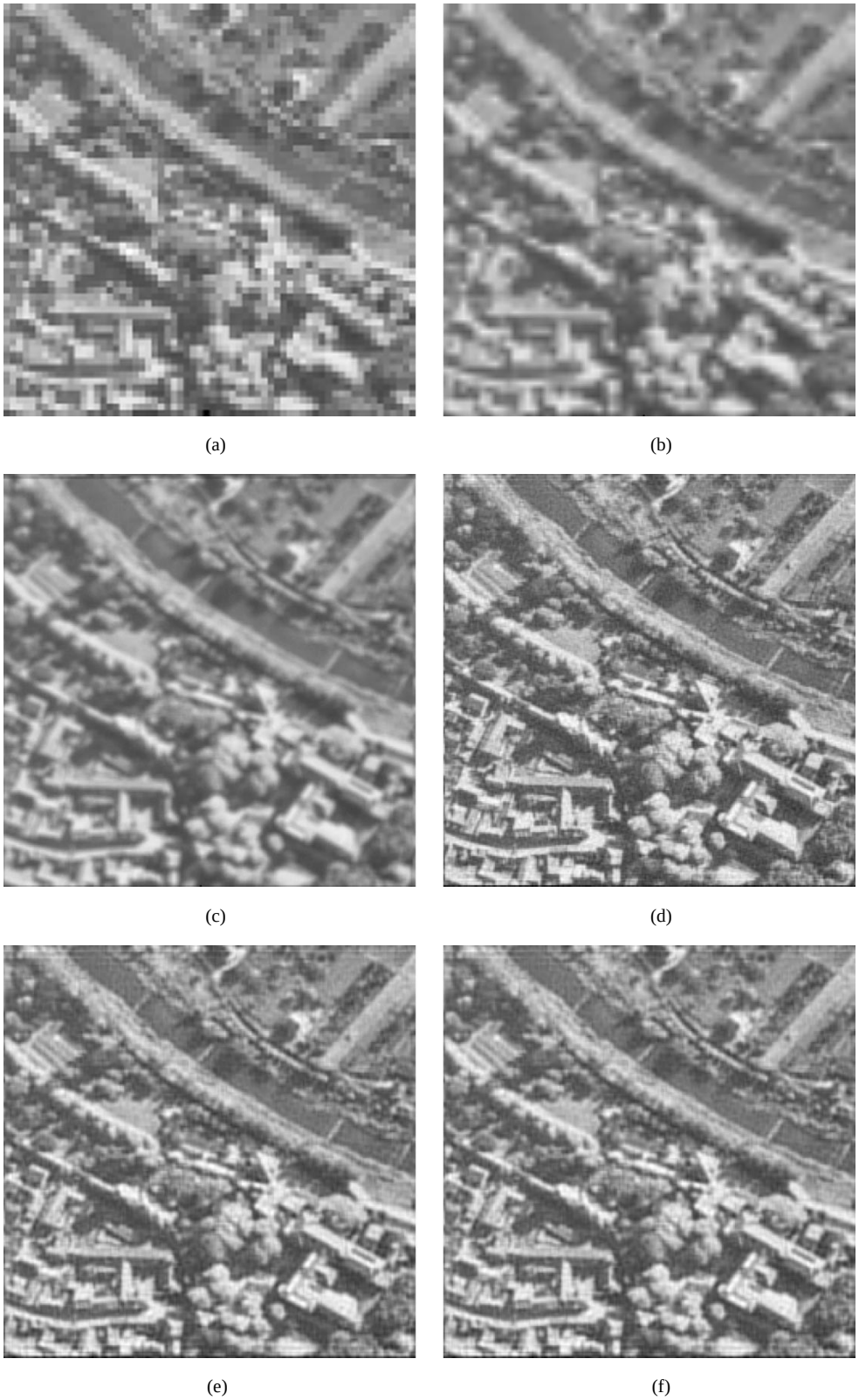


Fig. 3. Experiment 1, 30 dB case. (a) Zero-order hold for $\mathbf{g}_{0,0}$, (b) best bilinear interpolated image, (c) initial high resolution image, (d) estimated high resolution image with the proposed method, (e) estimated high resolution image with GCV, and (f) estimated high resolution image with L-curve.

E) Finally, we also note that a very interesting low to high resolution problem occurs when $\mathcal{I} < L \times L$ low resolution observations are available. We are currently working on this problem, and preliminary results can be found in [36] and [37].

V. Experimental Results

A number of simulations have been performed with the proposed algorithm over a set of images. We present results with

TABLE I
SUMMARY OF RESULTS FOR THE THREE DIFFERENT LOW RESOLUTION IMAGE SETS. EACH COLUMN SHOWS STATISTICS FOR THE TEN SIMULATIONS

	161.29 ≈ 10dB	14.44 ≈ 20dB	1.44 ≈ 30dB
Bilinear interpolation mean Δ_{SNR}	0.54 dB	0.24 dB	0.21 dB
Proposed algorithm, 1 noise parameter, mean Δ_{SNR}	6.17 dB	7.96 dB	10.70 dB
Proposed algorithm, 16 noise parameters, mean Δ_{SNR}	6.17 dB	7.95 dB	10.70 dB
GCV, mean Δ_{SNR}	3.67 dB	6.29 dB	8.01 dB
L-curve, mean Δ_{SNR}	2.24 dB	6.09 dB	8.01 dB
Proposed algorithm, 1 noise parameter, standard deviation	0.03	0.02	0.03
Proposed algorithm, 16 noise parameters, standard deviation	0.03	0.02	0.03
Proposed algorithm, 1 noise parameter, No. of iterations	4	13	12
Proposed algorithm, 16 noise parameters, No. of iterations	4	13	13

two images evaluating the performance of the proposed method under different noise conditions and comparing it with other existing approaches for estimating the parameters.

For all the experiments, the criterion $\|\mathbf{f}_{(\alpha^k,\underline{\beta}^k)} - \mathbf{f}_{(\alpha^{k-1},\underline{\beta}^{k-1})}\|^2/\|\mathbf{f}_{(\alpha^{k-1},\underline{\beta}^{k-1})}\|^2 < 10^{-6}$ was used for terminating the iteration. We set $\mathbf{f}^0 = \mathbf{g}$ in all our experiments, where $\mathbf{g}$ has been defined in (15).

The performance of the restoration algorithms was evaluated by measuring the signal to noise ratio (SNR) improvement denoted by Δ_{SNR} and defined by

$$\Delta_{SNR} = 10\log_{10}\frac{\|\mathbf{f}-\mathbf{g}\|^2}{\|\mathbf{f}-\hat{\mathbf{f}}\|^2}, \tag{33}$$

where $\mathbf{f}$ and $\hat{\mathbf{f}}$ are the original and estimated high resolution images, respectively.

We compared our proposed algorithm with the Generalized Cross-Validation (GCV) [16] and L-curve [20] methods. Both methods estimate the high resolution image using

$$\hat{\mathbf{f}}_\lambda = \arg\min_{\mathbf{f}}\left\{\lambda\|\mathbf{Cf}\|^2 + \sum_{l1=0}^{L-1}\sum_{l2=0}^{L-1}\|\mathbf{g}_{l1,l2} - \mathbf{W}_{l1,l2}\mathbf{f}\|^2\right\} \tag{34}$$

where λ is selected using the GCV [16] or L-curve [20] methods, the respective values of λ will be denoted by λ_{GCV} and $\lambda_{L-curve}$. See [38] for a description of GCV and other methods for choosing the regularization parameter in image restoration problems.

Note that the L-curve and GCV methods estimate a single parameter. This corresponds to the low to high resolution problem where the same noise variance is assumed for all observed low resolution images (see (11)). Note also that selecting the high resolution image according to (34) is the same as finding the solution of (20) with $\alpha = \lambda$ and $\beta_{l1,l2} = \beta,\ \forall l1, l2$.

A first experiment was devoted to test the performance of the proposed method under different noise conditions on the low resolution images. The 256×256 image, $\mathbf{f}$, shown in Fig. 2 was blurred using the blurring function defined in (8) with $L = 4$, which produced the image $\mathbf{u} = \mathbf{Hf}$. Then $\mathbf{u}$ was downsampled with $L_1 = L_2 = 4$ obtaining the sixteen 64×64 low resolution images

$$u_{l1,l2}(x,y) = u(L_1x + l1, L_2y + l2)$$
$$x,y = 0,\ldots,\frac{256}{4} - 1, \quad l1,l2 = 0,\ldots,3. \tag{35}$$

Gaussian noise with the same variance, that is, $\beta_{l1,l2} = \beta$, $\forall l1, l2$ was added to each low resolution image to obtain three sets of sixteen low resolution images. The noise variance for each set of images was set to 161.29, 14.44 and 1.44, respectively, thus obtaining an SNR of approximately 10 dB, 20 dB and 30 dB in the low resolution images of each image set.

Fig. 3 shows reconstructions for the 30 dB SNR case. Fig. 3(a) depicts the zero-order hold upsampled image $\mathbf{g}_{0,0}$ (see (11)). Each low resolution image in each set was bilinearly interpolated to obtain a 256×256 image. The best one in terms of Δ_{SNR} is shown in Fig. 3(b). The starting image, $\mathbf{f}^0$, is shown in Fig. 3(c). Fig. 3(d) depicts the estimated high resolution image by the proposed algorithm when sixteen noise parameters are considered. The high resolution image obtained by the proposed method with the same noise parameter for the sixteen low resolution observations, see (28), is almost identical and it is not shown here. Figs. 3(e) and 3(f) show the images obtained by estimating the high resolution image using the parameter values selected by the GCV and L-curve methods, respectively. From

TABLE II
STATISTICS OF THE ESTIMATED NOISE VARIANCES AND REGULARIZATION PARAMETER FOR THE TEN SIMULATIONS OF THE THREE LOW RESOLUTION IMAGE SETS WITH ONE NOISE PARAMETER

Noise variance β	161.29 ≈ 10dB	14.44 ≈ 20dB	1.44 ≈ 30dB
Proposed algorithm, mean $\hat{\beta^{-1}}$	160.87	14.55	1.42
Proposed algorithm $\lambda = \alpha/\beta$	0.74	0.07	0.007
λ_{GCV}	0.14	0.03	0.010
$\lambda_{L-curve}$	0.06	0.02	0.010
Proposed algorithm, standard deviation of the estimates	1.1	0.12	0.02

these figures, it is clear that the proposed method provides the visually best reconstructions.

In order to validate the proposed parameter estimation algorithm on a number of simulations, ten realizations of the noise were generated for each noise level. Table I shows the mean values of Δ_{SNR} for the different methods under consideration. From this table it is clear that the proposed method improves the SNR even in the case of severe noise although higher improvements are obtained as the noise decreases. The improvement in SNR obtained by the proposed algorithm is greater than the ones obtained by the GCV and L-curve procedures, resulting in less noisy results as previously commented. In Table I we have included the Δ_{SNR} obtained by the proposed method when using only one and sixteen different noise parameters. The results are almost identical in both cases, so validating the estimation process. Table I also shows the standard deviation of the ten Δ_{SNR} obtained by our proposed method with one and sixteen noise parameters and the number of iterations needed.

The estimated image model parameters and their standard deviations, in brackets, for the three sets of images with one noise parameter were equal to $\alpha^{-1} = 217(1.0)$, $\alpha^{-1} = 192.3(1.4)$ and $\alpha^{-1} = 198.5(2.1)$, respectively. The corresponding estimated mean noise parameters for the low resolution images are presented in Table II together with their standard deviations. Examining these tables we conclude that the proposed method produces accurate estimates of the low resolution noise variances. The results of the estimation of sixteen noise parameters, one for each low resolution image, were also very close to their real values. In order to compare the proposed method with the GCV and L-curve approaches we have included in Table II the equivalent value of λ obtained by our method when estimating only one noise parameter. We can see from this table that both GCV and L-curve obtain a smaller regularization parameter and therefore noisier reconstructions.

Fig. 4 shows the evolution of the Δ_{SNR} for the three low resolution image sets. Note that most of the improvement was obtained in the first few iterations. Each iteration took 15.5 seconds on a Pentium IV 1700.

A second experiment was performed to test the proposed algorithm when different noise variances are used on the low resolution observations. The original image in Fig. 2 was blurred and downsampled as in the previous experiment and Gaussian noise was added to each low resolution image to obtain a set

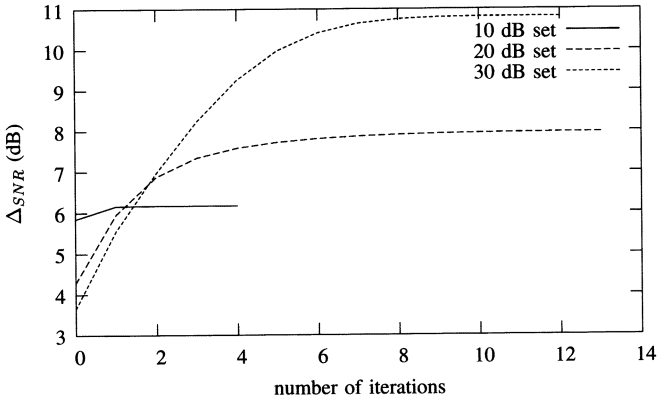


Fig. 4. SNR improvement evolution with the number of iterations for the three low resolution image sets.

TABLE III
SNR AND NOISE VARIANCES FOR LOW RESOLUTION IMAGES OF THE SECOND EXPERIMENT

$SNR/\beta^{-1}_{l1,l2}$	0	1	2	3
0	10dB / 161.29	20dB / 14.44	30dB / 1.44	10dB / 161.29
1	20dB / 14.44	30dB / 1.44	10dB / 161.29	30dB / 1.44
2	30dB / 1.44	20dB / 14.44	30dB / 1.44	20dB / 14.44
3	20dB / 14.44	30dB / 1.44	10dB / 161.29	20dB / 14.44

of sixteen low resolution images with different noise characteristics, with SNRs of 10 dB, 20 dB or 30 dB randomly selected (see Table III). Again, we generated ten realizations of the noise. Fig. 5(a) depicts the zero-order hold upsampled image $\mathbf{g}_{0,0}$. The bilinear interpolation is shown in Fig. 5(b). Fig. 5(c) and 5(d) depict the results obtained by the proposed algorithm estimating one and sixteen noise parameters, respectively, and Fig. 5(e) and 5(f) show the results with the GCV and L-curve estimation procedures, respectively. From these images it is clear that the proposed method outperforms all other reported methods, for both cases of one and sixteen noise parameters estimation. The best visual result is obtained when one noise parameter is estimated for each low resolution image.

The Δ_{SNR} obtained by the proposed algorithm estimating one and sixteen noise parameters and, also, by the GCV and L-curve methods are shown in Table IV. We can see that even when the proposed algorithm estimates only one noise parameter the results are better than those obtained by the GCV and

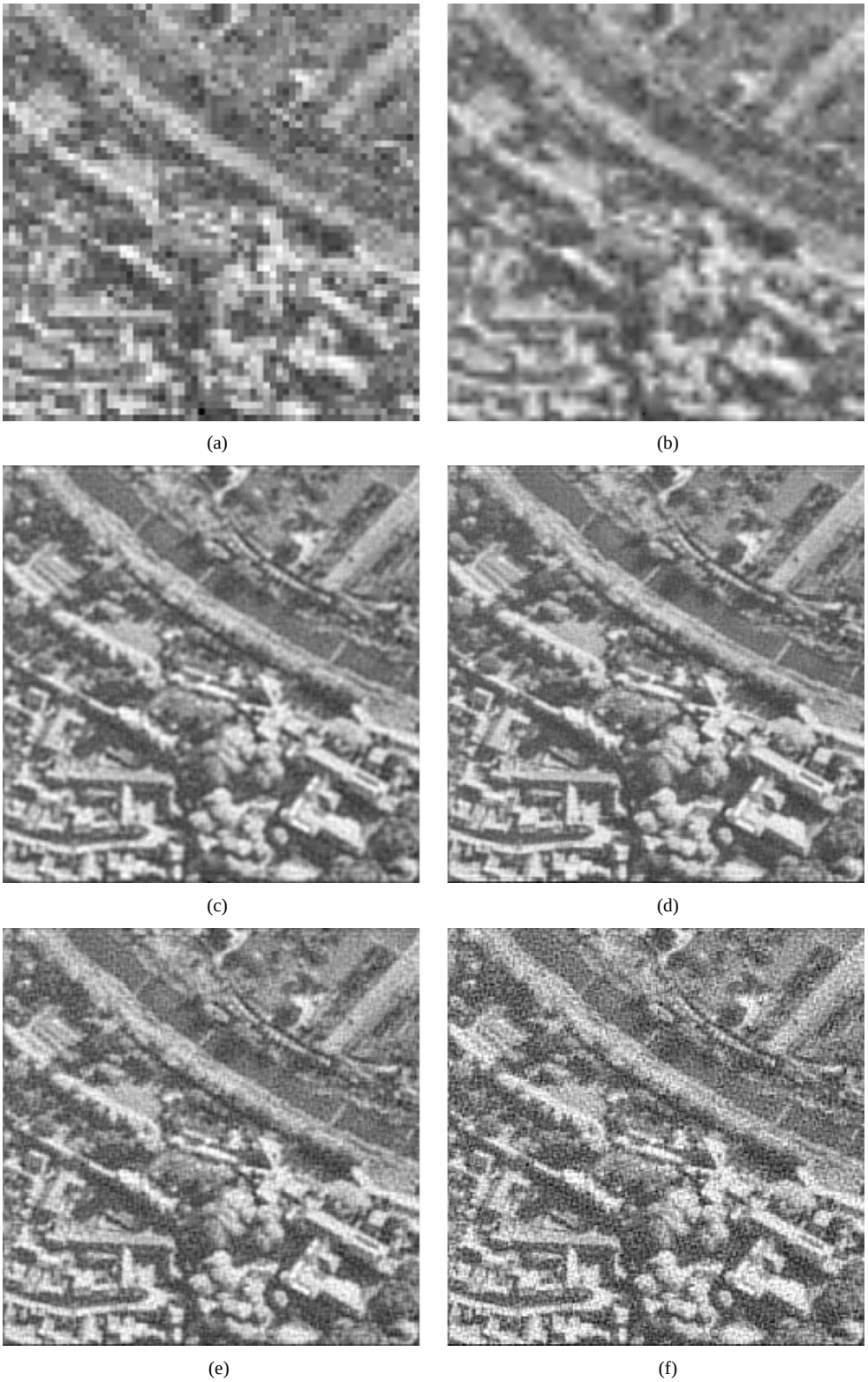
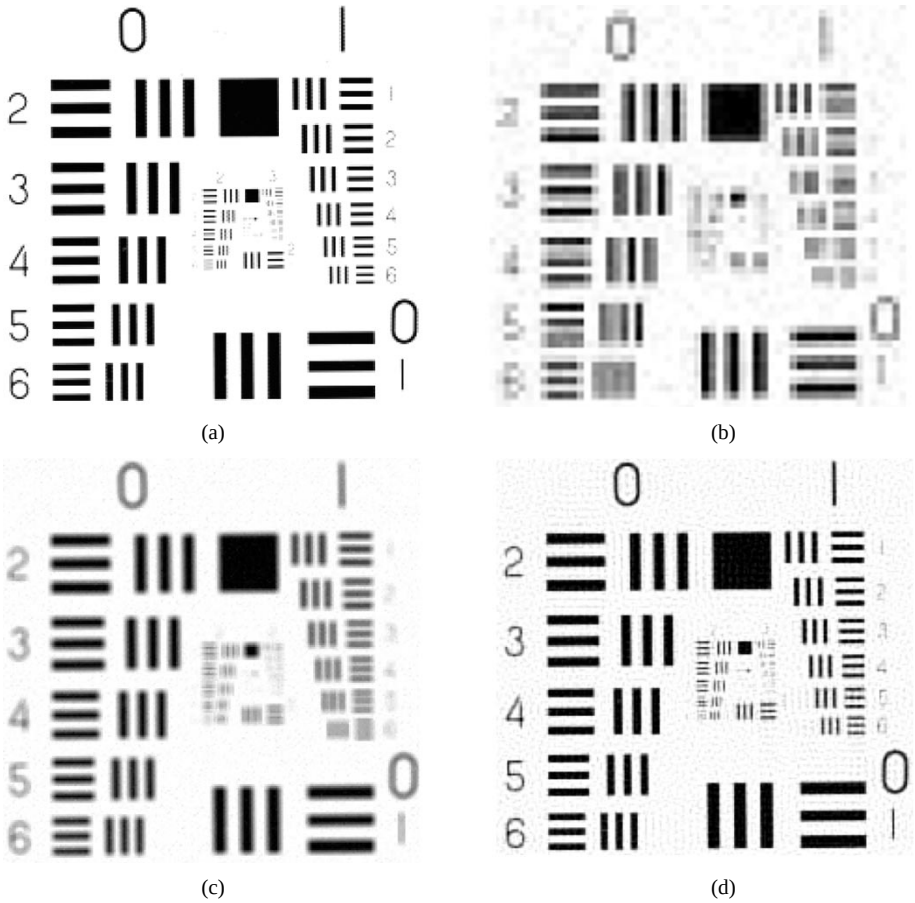


Fig. 5. (a) Zero-order hold for $\mathbf{g}_{0,0}$, (b) best bilinear interpolated image, (c) estimated high resolution image (one noise parameter), (d) estimated high resolution image (sixteen noise parameters), (e) estimated high resolution image with GCV, and (f) estimated high resolution image with L-curve.

L-curve methods. The improvement in SNR obtained by the proposed algorithm with sixteen noise parameters clearly represents the best result. The number of iterations required by the proposed algorithm to satisfy the convergence criterion are also shown in this table. Table IV also shows the estimated values of λ by the proposed method when estimating one noise parameter, and the GCV and L-curve. From these figures it is clear that both GCV and L-curve methods obtain significantly lower Δ_{SNR} than the proposed method.

The mean value of the image model parameter estimated by the proposed algorithm is $\alpha^{-1} = 182.8$, with standard deviation 3.1, when estimating one noise parameter, and



(a) (b) (c) (d)

Fig. 6. (a) Original high resolution image, (b) zero order hold of low resolution observed image $\mathbf{g}_{0,0}$, (c) Initial high resolution image, and (d) estimated high resolution image.

TABLE IV
Δ_{SNR} FOR THE SECOND EXPERIMENT

Method	Δ_{SNR}
Bilinear interpolation	0.80 dB
Proposed algorithm, 1 noise parameter. 12 iterations, $\lambda = 0.26$	6.86 dB
Proposed algorithm, 16 noise parameters 14 iterations	9.09 dB
GCV ($\lambda_{GCV} = 0.016$)	5.95 dB
L-curve ($\lambda_{L-curve} = 0.06$)	2.91 dB

$\alpha^{-1} = 203.9$, with standard deviation 1.0, when estimating sixteen noise parameters. Table V shows the mean noise variance parameters, and their corresponding standard deviations in brackets, estimated by the proposed algorithm. This table shows that accurate estimates are obtained by the proposed model although small variance values are slightly overestimated when their corresponding low resolution image is close in terms of shifts to other low resolution images with larger noise variances. A value of $\beta^{-1} = 46.86$, with standard deviation 0.61, was obtained when estimating one noise parameter. This value is close to the mean of the noise variances in the low resolution images.

TABLE V
MEANS OF THE ESTIMATED NOISE PARAMETERS FOR THE SECOND EXPERIMENT, SEE TABLE III. IN BRACKETS THEIR STANDARD DEVIATIONS FOR THE TEN REALIZATIONS

$\hat{\beta}^{-1}_{l1,l2}$	0	1	2	3
0	162.04 (3.66)	14.23 (0.41)	1.33 (0.09)	159.40 (1.61)
1	14.18 (0.31)	2.89 (0.12)	159.90 (3.95)	1.44 (0.04)
2	1.67 (0.06)	13.60 (0.33)	1.72 (0.08)	13.72 (0.53)
3	13.96 (0.33)	1.53 (0.04)	160.44 (3.24)	13.86 (0.23)

In a third experiment we also tested the proposed method on low resolution images with different noise variances. First, the set of low resolution images was obtained by blurring and downsampling the 256 × 256 original image depicted in Fig. 6(a) following the same procedure as in the previous experiments. Then, to each 64 × 64 low resolution image, Gaussian noise was added to obtain at random degraded images with SNR 20, 30, or 40 dB. The noise variances utilized are shown in Table VI. Fig. 6(b) depicts zero-order hold of the observed low resolution image $\mathbf{g}_{00}$.

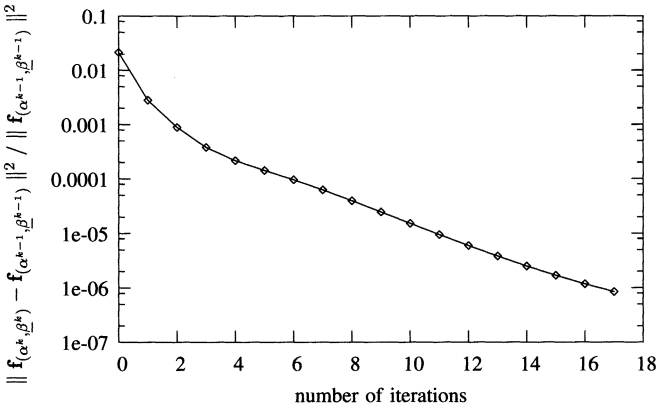
We ran the proposed method starting with the initial image (Fig. 6(c)) obtaining at convergence the estimated high resolution image shown in Fig. 6(d). For the resulting image the SNR improvement was 11.63 dB and the estimated image model parameter $\alpha^{-1} = 683.9$. The estimated variances of the degrading noise are shown in Table VII.

TABLE VI
SNR AND NOISE VARIANCES FOR LOW RESOLUTION IMAGES IN FIG. (6b)

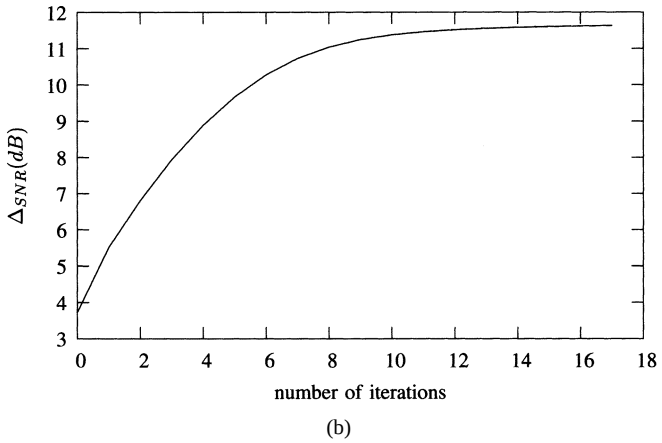
$SNR/\beta_{l1,l2}^{-1}$	0	1	2	3
0	20dB / 38.94	30dB / 4.00	40dB / 0.39	20dB / 40.63
1	30dB / 3.90	20dB / 38.72	40dB / 0.39	30dB / 3.86
2	40dB / 0.39	20dB / 41.03	30dB / 3.88	40dB / 0.40
3	20dB / 39.73	40dB / 0.39	30dB / 3.93	20dB / 39.20

TABLE VII
ESTIMATED NOISE VARIANCES FOR HIGH RESOLUTION IMAGE IN FIG. (6d)

$\hat{\beta}_{l1,l2}^{-1}$	0	1	2	3
0	37.70	3.37	0.80	38.75
1	3.59	36.13	1.12	2.73
2	0.95	37.54	3.27	1.76
3	37.36	1.27	3.02	37.28



(a)



(b)

Fig. 7. (a) Convergence criterion versus iterations plot and (b) SNR improvement versus iterations plot for the image shown in Fig. 6(d).

By comparing the real and estimated degradation model parameters (see Tables VI and VII, respectively) we conclude that the proposed model produces good estimates for all parameters although, again, small variance values are overestimated when their corresponding low resolution images are close in terms of shifts to other low resolution images with larger noise variances.

Fig. 7(a) depicts $\|\mathbf{f}_{(\alpha^k,\underline{\beta}^k)} - \mathbf{f}_{(\alpha^{k-1},\underline{\beta}^{k-1})}\|^2 / \|\mathbf{f}_{(\alpha^{k-1},\underline{\beta}^{k-1})}\|^2$ versus number of iterations. Note that the vertical axis is plotted in logarithm scale. From this plot it is clear that the method converges fast, needing only a few iterations to obtain a good estimation of the image, validating, this way, the theoretical results.

Fig. 7(b) shows the signal to noise ratio improvement, Δ_{SNR}, versus the number of iterations. It is clear that Δ_{SNR} increases monotonically and that most of the improvement is obtained in the first four iterations.

VI. CONCLUSIONS

A new method to estimate the unknown parameters in a high resolution image reconstruction problem has been proposed. Using BSC matrices we have shown that all the matrix calculations involved in the parameter maximum likelihood estimation can be performed in the Fourier domain. The approach followed can be used to assign the same parameter to all low resolution image parameters or to make them image dependent. We have also shown that the results are extensions of maximum likelihood estimation for single channel restoration problems. The proposed method has been validated experimentally.

APPENDIX I
EM ALGORITHM APPLIED TO THE RESOLUTION PROBLEM

In this section, we show how the EM-algorithm can be used to estimate the unknown parameters in the low to high reconstruction problem under consideration. Let $\Psi^t = (\alpha, \underline{\beta})$ be the vector containing the unknown parameters. Note that in our high resolution problem we are dealing with an exponential family since we have

$$p_\Psi(\mathbf{f},\mathbf{g}) = b(\mathbf{f},\mathbf{g})\frac{\exp\left[\Psi^t c(\mathbf{f},\mathbf{g})\right]}{a(\Psi)} \tag{36}$$

where

$$c^t(\mathbf{f},\mathbf{g}) = \left(-\frac{1}{2}\|\mathbf{Cf}\|^2, \left(-\frac{1}{2}\|\mathbf{g}_{l1,l2} - \mathbf{W}_{l1,l2}\mathbf{f}\|^2\right)\right.$$
$$\left.|l1,l2 = 0,\ldots,L-1)\right. \tag{37}$$
$$a(\Psi) = Z_{prior}(\alpha)Z_{noise}(\underline{\beta}) \tag{38}$$

and $b(\mathbf{f},\mathbf{g})$ is a scalar function depending only on the complete data $(\mathbf{f},\mathbf{g})$.

We note that the expectation of the sufficient statistic $c(\mathbf{f},\mathbf{g})$ is given by

$$E_\Psi\left[c(\mathbf{f},\mathbf{g})\right] = \frac{\partial}{\partial\Psi}\log a(\Psi) \tag{39}$$

and so we have

$$E_\Psi\left[\|\mathbf{Cf}\|^2\right] = \frac{M_1 \times M_2}{\alpha} \tag{40}$$

and

$$E_\Psi\left[\|\mathbf{g}_{l1,l2} - \mathbf{W}_{l1,l2}\mathbf{f}\|^2\right] = \frac{N_1 \times N_2}{\beta_{l1,l2}}. \tag{41}$$

Furthermore

$$E_\Psi\left[\|\mathbf{Cf}\|^2|\mathbf{g}\right] = \left\|\mathbf{Cf}_{(\alpha,\underline{\beta})}\right\|^2 + \text{trace}\left[\mathbf{Q}(\alpha,\underline{\beta})^{-1}\mathbf{C}^t\mathbf{C}\right] \tag{42}$$

and

$$E_\Psi\left[\|\mathbf{g}_{l1,l2}-\mathbf{W}_{l1,l2}\mathbf{f}\|^2|\mathbf{g}\right]=\left\|\mathbf{g}_{l1,l2}-\mathbf{W}_{l1,l2}\mathbf{f}_{(\alpha,\underline{\beta})}\right\|^2 +\text{trace}\left[\mathbf{Q}(\alpha,\underline{\beta})^{-1}\mathbf{W}_{l1,l2}^t\mathbf{W}_{l1,l2}\right]. \quad (43)$$

The EM-algorithm for this family requires the maximization with respect to Ψ (see [39])

$$Q(\Psi,\Psi^k)=\Psi^t E_{\Psi^k}\left[c(\mathbf{f},\mathbf{g})|\mathbf{g}\right]-\log a(\Psi) \quad (44)$$

whose unique solution is provided by steps 3a and 3b of algorithm 1.

Appendix II
Calculating the Map and Parameters

In this appendix we show how with known $(\alpha,\underline{\beta})$ to efficiently calculate the corresponding MAP, $\mathbf{f}_{(\alpha,\underline{\beta})}$, that is, how to obtain the solution of

$$\left(\alpha\mathbf{C}^t\mathbf{C}+\sum_{l1=0}^{L-1}\sum_{l2=0}^{L-1}\beta_{l1,l2}\mathbf{W}_{l1,l2}^t\mathbf{W}_{l1,l2}\right)\mathbf{f} =\sum_{l1=0}^{L-1}\sum_{l2=0}^{L-1}\beta_{l1,l2}\mathbf{W}_{l1,l2}^t\mathbf{g}_{l1,l2}. \quad (45)$$

Let us consider the convolution process given by

$$p(x,y)=\sum_{m_1=0}^{M_1-1}\sum_{m_2=0}^{M_2-1}h(x-m_1,y-m_2)q(m_1,m_2). \quad (46)$$

By lexicographically ordering the signals in (46) $\mathbf{p}$ and $\mathbf{q}$ we have

$$\mathbf{p}=\mathbf{H}\mathbf{q} \quad (47)$$

where $\mathbf{p}$ and $\mathbf{q}$ are $M_1\times M_2$ vectors and $\mathbf{H}$ is a convolution matrix of size $(M_1M_2)\times(M_1M_2)$. We will assume that $\mathbf{H}$ is a circular convolution operator (see however [18] for the issues when using circulant approximations in high-resolution problems).

We examine how to calculate $\mathbf{D}_{l1,l2}\mathbf{p}$ as the sum of $L\times L$ convolutions. Each convolution will involve a circulant convolution matrix of size $(N_1N_2)\times(N_1N_2)$ and a vector of size N_1N_2.

Proposition 1: Let $\mathbf{p}=\mathbf{H}\mathbf{q}$ be the circular convolution process defined in (47) and $\mathbf{D}_{l1,l2}$ the downsampling matrix defined in (12), then

$$\mathbf{D}_{l1,l2}\mathbf{p}=\sum_{u=0}^{L-1}\sum_{v=0}^{L-1}\mathbf{H}_{l1,l2}^{u,v}\mathbf{D}_{u,v}\mathbf{q} \quad (48)$$

where $\mathbf{H}_{l1,l2}^{u,v}$ is the $N_1N_2\times N_1N_2$ circular convolution matrix defined by

$$h_{l1,l2}^{u,v}(x,y)=h\left(xL+(l1-u),yL+(l2-v)\right),\quad x=0,\dots,N_1,\ y=0,\dots,N_2. \quad (49)$$

Proof: First, note that for an $M_1M_2\times 1$ column vector, $\mathbf{r}$, $\mathbf{D}_{l1,l2}\mathbf{r}=\mathbf{r}_{l1,l2}$ is the $N_1N_2\times 1$ column vector $\mathbf{r}_{l1,l2}$ with lexicographically ordered components $r(n_1L+l1,n_2L+l2)$, $n_1=0,\dots,N_1-1$, $n_2=0,\dots,N_2-1$.

If in (46) we write $m=aL+u$, $n=bL+v$ with $a=0,\dots,N_1-1$, $b=0,\dots,N_2-1$ and $u,v=0,\dots,L-1$ we have

$$\begin{aligned}p(n_1L+l1,n_2L+l2)&=\sum_{u=0}^{L-1}\sum_{v=0}^{L-1}\left(\sum_{a=0}^{N_1-1}\sum_{b=0}^{N_2-1}h\left(n_1L+l1-(aL+u),\right.\right.\\&\qquad\left.\left.n_2L+l2-(bL+v)\right)q(aL+u,bL+v)\right)\\&=\sum_{u=0}^{L-1}\sum_{v=0}^{L-1}\left(\sum_{a=0}^{N_1-1}\sum_{b=0}^{N_2-1}h\left((n_1-a)L+(l1-u),\right.\right.\\&\qquad\left.\left.(n_2-b)L+(l2-v)\right)q(aL+u,bL+v)\right)\end{aligned}$$

and so

$$p_{l1,l2}(n_1,n_2)=\sum_{u=0}^{L-1}\sum_{v=0}^{L-1}\left(\sum_{a=0}^{N_1-1}\sum_{b=0}^{N_2-1}h_{l1,l2}^{u,v}(n_1-a,n_2-b)q_{l_1,l_2}(a,b)\right) \quad (50)$$

or

$$\mathbf{p}_{l1,l2}=\mathbf{D}_{l1,l2}\mathbf{p}=\sum_{u=0}^{L-1}\sum_{v=0}^{L-1}\mathbf{H}_{l1,l2}^{u,v}\mathbf{D}_{u,v}\mathbf{q} =\sum_{u=0}^{L-1}\sum_{v=0}^{L-1}\mathbf{H}_{l1,l2}^{u,v}\mathbf{q}_{u,v}. \quad (51)$$

■

It is interesting to note that $H_{l1,l2}^{u,v}=H_{l1+k1,l2+k2}^{u+k1,v+k2}$, $\forall\, k1,k2$. Let $\mathbf{D}$ be the matrix defined by

$$\mathbf{D}^t=\left(\mathbf{D}_{00}^t\dots\mathbf{D}_{0L-1}^t\mathbf{D}_{10}^t\dots\mathbf{D}_{1L-1}^t\dots\mathbf{D}_{L-1L-1}^t\right). \quad (52)$$

Note that $\mathbf{D}$ is a permutation matrix, see [40], and therefore $\mathbf{D}^t\mathbf{D}=\mathbf{D}\mathbf{D}^t=I$. Since $\mathbf{D}_{l1l2}\mathbf{H}\mathbf{q}=\mathbf{D}_{l1l2}\mathbf{H}\mathbf{D}^t\mathbf{D}\mathbf{q}$ and $(\mathbf{D}\mathbf{q})^t=(\mathbf{q}_{(u,v)}^t|u,v=0,\dots,L-1)$, we have from the above proposition [see (48)].

Property 1:

$$\mathbf{D}_{l1l2}\mathbf{H}\mathbf{D}^t=\left(\mathbf{H}_{l1,l2}^{0,0}\ \dots\ \mathbf{H}_{l1,l2}^{0,L-1}\ \mathbf{H}_{l1,l2}^{1,0}\ \dots\ \mathbf{H}_{l1,l2}^{1,L-1}\ \dots\ \mathbf{H}_{l1,l2}^{L-1,L-1}\right) \quad (53)$$

and so we have the following:

Property 2: See (54) at the top of the next page.

Furthermore, from property 1 we have the following.

Property 3:

$$\mathbf{D}\mathbf{H}^t\mathbf{D}_{l1l2}^t=\begin{pmatrix}\mathbf{H}_{l1,l2}^{0,0\ t}\\ \vdots\\ \mathbf{H}_{l1,l2}^{0,L-1\,t}\\ \mathbf{H}_{l1,l2}^{1,0\ t}\\ \vdots\\ \mathbf{H}_{l1,l2}^{1,L-1\,t}\\ \vdots\\ \mathbf{H}_{l1,l2}^{L-1,L-1\,t}\end{pmatrix}. \quad (55)$$

$$\mathbf{DHD}^t = \begin{pmatrix} \mathbf{H}_{0,0}^{0,0} & \dots & \mathbf{H}_{0,0}^{0,L-1} & \mathbf{H}_{0,0}^{1,0} & \dots & \mathbf{H}_{0,0}^{1,L-1} & \dots & \mathbf{H}_{0,0}^{L-1,L-1} \\ \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots & \\ \mathbf{H}_{0,L-1}^{0,0} & \dots & \mathbf{H}_{0,L-1}^{0,L-1} & \mathbf{H}_{0,L-1}^{1,0} & \dots & \mathbf{H}_{0,L-1}^{1,L-1} & \dots & \mathbf{H}_{0,L-1}^{L-1,L-1} \\ \mathbf{H}_{1,0}^{0,0} & \dots & \mathbf{H}_{1,0}^{0,L-1} & \mathbf{H}_{1,0}^{1,0} & \dots & \mathbf{H}_{1,0}^{1,L-1} & \dots & \mathbf{H}_{1,0}^{0,L-1} \\ \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots & \\ \mathbf{H}_{1,L-1}^{0,0} & \dots & \mathbf{H}_{1,L-1}^{0,L-1} & \mathbf{H}_{1,L-1}^{1,0} & \dots & \mathbf{H}_{1,L-1}^{1,L-1} & \dots & \mathbf{H}_{1,L-1}^{L-1,L-1} \\ \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots & \\ \mathbf{H}_{L-1,L-1}^{0,0} & \dots & \mathbf{H}_{L-1,L-1}^{0,L-1} & \mathbf{H}_{L-1,L-1}^{1,0} & \dots & \mathbf{H}_{L-1,L-1}^{1,L-1} & \dots & \mathbf{H}_{L-1,L-1}^{L-1,L-1} \end{pmatrix} \tag{54}$$

Let us now proceed to solve (45). First, we rewrite this equation as

$$\mathbf{D}\left(\alpha\mathbf{C}^t\mathbf{C} + \sum_{l1=0}^{L-1}\sum_{l2=0}^{L-1}\beta_{l1,l2}\mathbf{W}_{l1,l2}^t\mathbf{W}_{l1,l2}\right)\mathbf{D}^t\mathbf{D}\mathbf{f} = \sum_{l1=0}^{L-1}\sum_{l2=0}^{L-1}\beta_{l1,l2}\mathbf{D}\mathbf{W}_{l1,l2}^t\mathbf{g}_{l1,l2}. \tag{56}$$

Then, we use property 2 on $\mathbf{DC}^t\mathbf{CD}^t$, properties 1 and 3 on each term of the form $\mathbf{DW}_{l1,l2}^t\mathbf{W}_{l1,l2}\mathbf{D}^t = (\mathbf{DH}_{l1,l2}^t\mathbf{D}_{l1,l2}^t)(\mathbf{D}_{l1,l2}\mathbf{H}_{l1,l2}\mathbf{D}^t)$ and property 3 on each term of the form $\mathbf{DW}_{l1,l2}^t = \mathbf{DH}_{l1,l2}^t\mathbf{D}_{l1,l2}^t$ to obtain that (56) can be written as

$$\sum_{u=0}^{L-1}\sum_{v=0}^{L-1}\mathbf{A}_{l1,l2}^{u,v}\mathbf{f}_{u,v} = \sum_{u=0}^{L-1}\sum_{v=0}^{L-1}\mathbf{B}_{l1,l2}^{u,v}\mathbf{g}_{u,v} \quad l1,l2=0,\dots,L-1 \tag{57}$$

where $\mathbf{A}_{l1,l2}^{u,v}$ and $\mathbf{B}_{l1,l2}^{u,v}$ are $N_1N_2 \times N_1N_2$ circular convolution matrices.

In order to solve the system in (57) we only need to apply Fourier transform to each equation of the form

$$\sum_{u=0}^{L-1}\sum_{v=0}^{L-1}\mathbf{A}_{l1,l2}^{u,v}\mathbf{f}_{u,v} = \sum_{j=0}^{L-1}\sum_{k=0}^{L-1}\mathbf{B}_{l1,l2}^{u,v}\mathbf{g}_{u,v} \tag{58}$$

in that system, since in the Fourier domain $\mathbf{A}_{l1,l2}^{u,v}$ and $\mathbf{B}_{l1,l2}$ become diagonal matrices. Then, our problem becomes the solution of N_1N_2 systems of equations each one of size $L \times L$.

It is interesting to note that a similar decomposition approach to the one used here is followed in [28] and [10] to solve restoration and high-resolution problems.

Finally in order to estimate the parameters we have to calculate $\mathrm{trace}[\mathbf{Q}(\alpha,\underline{\beta})^{-1}\mathbf{W}_{l1,l2}^t\mathbf{W}_{l1,l2}]$ and $\mathrm{trace}[\mathbf{Q}(\alpha,\underline{\beta})^{-1}\mathbf{C}^t\mathbf{C}]$. To do so we only need to take into account that

$$\begin{aligned}&\mathrm{trace}\left[\mathbf{Q}(\alpha,\underline{\beta})^{-1}\mathbf{C}^t\mathbf{C}\right] \\ &\quad = \mathrm{trace}\left[\mathbf{D}\mathbf{Q}(\alpha,\underline{\beta})^{-1}\mathbf{D}^t\mathbf{D}\mathbf{C}^t\mathbf{C}\mathbf{D}^t\right], \\ &\mathrm{trace}\left[\mathbf{Q}(\alpha,\underline{\beta})^{-1}\mathbf{W}_{l1,l2}^t\mathbf{W}_{l1,l2}\right] \\ &\quad = \mathrm{trace}\left[\mathbf{D}\mathbf{Q}(\alpha,\underline{\beta})^{-1}\mathbf{D}^t\mathbf{D}\mathbf{W}_{l1,l2}^t\mathbf{W}_{l1,l2}\mathbf{D}^t\right]\end{aligned}$$

and then apply Fourier transforms as we did to calculate the MAP.

REFERENCES

[1] K. Aizawa, T. Komatsu, and T. Saito, "A scheme for acquiring very high resolution images using multiple cameras," in *Proc. IEEE Conf. Audio, Speech, Signal Processing*, vol. 3, 1992, pp. 289–292.

[2] H. Stark and P. Oskoui, "High-resolution image recovery from image-plane arrays, using convex projections," *J. Opt. Soc. A*, vol. 6, no. 11, pp. 1715–1726, 1989.

[3] A. S. Frutcher and R. N. Hook, "Drizzle: a method for the linear reconstruction of undersampled images," *Pub. Astron. Soc. Pacific*, vol. 114, pp. 144–152, 2002.

[4] R. Y. Tsai and T. S. Huang, "Multiframe image restoration and registration," in *Advances in Computer Vision and Image Processing*, R. Y. Tsai and T. S. Huang, Eds. New York: JAI, 1984, vol. 1, pp. 317–339.

[5] S. Borman and R. Stevenson, "Spatial resolution enhancement of low-resolution image sequences. A comprehensive review with directions for future research," in *Laboratory for Image and Signal Analysis*. Notre Dame, IN: Tech. Rep., Univ. Notre Dame, 1998.

[6] S. Baker and T. Kanade, "Limits on super-resolution and how to break them," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, vol. 2, 2000, pp. 372–379.

[7] ——, "Super-resolution: reconstruction or recognition?," in *IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing*, 2001.

[8] W. T. Freeman, E. C. Pasztor, and O. T. Carmichael, "Learning low-level vision," *Int. J. Comput. Vis.*, vol. 40, no. 1, pp. 25–47, 2000.

[9] D. P. Capel and A. Zisserman, "Super-resolution from multiple views using learnt image models," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, vol. 2, 2001, pp. 627–634.

[10] N. K. Bose and K. J. Boo, "High-resolution image reconstruction with multisensors," *Int. J. Imag. Syst. Technol.*, vol. 9, pp. 141–163, 1998.

[11] R. H. Chan, T. F. Chan, M. K. Ng, W. C. Tang, and C. K. Wong, "Preconditioned iterative methods for high-resolution image reconstruction with multisensors," *Proc. SPIE*, vol. 3461, pp. 348–357, 1998.

[12] R. H. Chan, T. F. Chan, L. Shen, and S. Zuowei, "A wavelet method for high-resolution image reconstruction with displacement errors," in *Proc. Int. Symp. Intelligent Multimedia, Video and Speech Processing*, 2001, pp. 24–27.

[13] R. H. Chan, T. F. Chan, L. X. Shen, and Z. W. Shen, "Wavelet Algorithms for High-Resolution Image Reconstruction," Tech. Rep., Dept. Math., Chinese Univ. Hong Kong, 2001.

[14] N. Nguyen and P. Milanfar, "A wavelet-based interpolation-restoration method for superresolution," *Circuits, Syst., Signal Process.*, vol. 19, pp. 321–338, 2000.

[15] N. Nguyen, "Numerical Algorithms for Superresolution," Ph.D. dissertation, Stanford Univ., Stanford, CA, 2001.

[16] N. Nguyen, P. Milanfar, and G. Golub, "A computationally efficient superresolution image reconstruction algorithm," *IEEE Trans. Image Processing*, vol. 10, pp. 573–583, 2001.

[17] M. K. Ng, R. H. Chan, and T. F. Chan, "Cosine transform preconditioners for high resolution image reconstruction," *Linear Algebra Applicat.*, vol. 316, pp. 89–104, 2000.

[18] M. K. Ng and A. M. Yip, "A fast MAP algorithm for high-resolution image reconstruction with multisensors," *Multidimen. Syst. Signal Process.*, vol. 12, pp. 143–164, 2001.

[19] M. Elad and Y. Hel-Or, "A fast super-resolution reconstruction algorithm for pure translational motion and common space invariant blur," *IEEE Trans. Image Processing*, vol. 10, pp. 1187–1193, 2001.

[20] N. K. Bose, S. Lertrattanapanich, and J. Koo, "Advances in superresolution using L-curve," in *Proc. IEEE Int. Symp. Circuits Systems*, vol. 2, 2001, pp. 433–436.

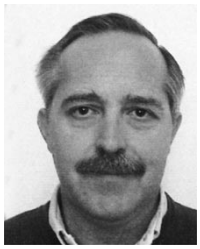
[21] N. Nguyen, P. Milanfar, and G. Golub, "Blind superresolution with generalized cross-validation using Gauss-type quadrature rules," in *Proc. 33rd Asilomar Conf. Signals, Systems, Computers*, vol. 2, 1999, pp. 1257–1261.

[22] N. Nguyen, P. Milanfar, and G. H. Golub, "Efficient generalized cross-validation with applications to parametric image restoration and resolution enhancement," *IEEE Trans. Image Processing*, vol. 10, pp. 1299–1308, 2001.
[23] B. C. Tom, N. P. Galatsanos, and A. K. Katsaggelos, "Reconstruction of a high resolution image from multiple low resolution images," in *Super-Resolution Imaging*, S. Chaudhuri, Ed. Norwell, MA: Kluwer, 2001, ch. 4, pp. 73–105.
[24] B. C. Tom, A. K. Katsaggelos, and N. P. Galatsanos, "Reconstruction of a high-resolution image from registration and restoration of low-resolution images," in *Proc. IEEE Int. Conf. Image Processing*, vol. 3, 1994, pp. 553–557.
[25] B. C. Tom and A. K. Katsaggelos, "Reconstruction of a high-resolution image by simultaneous registration, restoration, and interpolation of low-resolution images," in *Proc. IEEE Int. Conf. Image Processing*, vol. 2, 1995, pp. 539–542.
[26] M. Belge, M. E. Kilmer, and E. L. Miller, "Simultaneous multiple regularization parameter selection by means of the L-hypersurface with applications to linear inverse problems posed in the wavelet domain," *Proc. SPIE*, 1998.
[27] A. K. Katsaggelos, K. T. Lay, and N. P. Galatsanos, "A general framework for frequency domain multi-channel signal processing," *IEEE Trans. Image Processing*, vol. 2, pp. 417–420, 1993.
[28] M. R. Banham, N. P. Galatsanos, H. L. Gonzalez, and A. K. Katsaggelos, "Multichannel restoration of single channel images using a wavelet-based subband decomposition," *IEEE Trans. Image Processing*, vol. 3, pp. 821–833, 1994.
[29] A. K. Jain, *Fundamentals of Digital Image Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1989.
[30] R. R. Schultz and R. L. Stevenson, "Extraction of high resolution frames from video sequences," *IEEE Trans. Image Processing*, vol. 5, pp. 996–1011, 1996.
[31] B. D. Ripley, *Spatial Statistics*. New York: Wiley, 1981.
[32] D. Rajan and S. Chaudhuri, "An MRF-based approach to generation of super-resolution images from blurred observations," *J. Math. Imag. Vis.*, vol. 16, pp. 5–153, 2002.
[33] R. Molina, A. K. Katsaggelos, and J. Mateos, "Bayesian and regularization methods for hyperparameter estimation in image restoration," *IEEE Trans. Image Processing*, vol. 8, pp. 231–246, 1999.
[34] N. P. Galatsanos, V. Z. Mesarovic, R. Molina, A. K. Katsaggelos, and J. Mateos, "Hyperparameter estimation in image restoration problems with partially-known blurs," *Opt. Eng.*, vol. 41, pp. 1845–1854, 2002.
[35] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data," *J. R. Statist. Soc. B*, vol. 39, pp. 1–38, 1972.
[36] J. Mateos, R. Molina, and A. K. Katsaggelos, "Bayesian high resolution image reconstruction with incomplete multisensor low resolution systems," in *Proc. IEEE Int. Conf. Acoustic, Speech, Signal Processing*, 2003.
[37] J. Mateos, M. Vega, R. Molina, and A. K. Katsaggelos, "Bayesian image estimation from an incomplete set of blurred, undersampled low resolution images," in *Proc. 1st Iberian Conf. Pattern Recognition and Image Analysis (IbPRIA2003)*, 2003.
[38] N. P. Galatsanos and A. K. Katsaggelos, "Methods for choosing the regularization parameter and estimating the noise variance in image restoration and their relation," *IEEE Trans. Image Processing*, vol. 1, pp. 322–336, 1992.
[39] G. J. McLachlan and T. Krishnan, *The EM Algorithm and Extensions*. New York: Wiley, 1997.
[40] G. H. Golub and C. F. Van Loan, *Matrix Computations*. Baltimore, MD: Johns Hopkins Univ. Press, 1991.

Rafael Molina (M'88) was born in 1957. He received the degree in mathematics (statistics) in 1979 and the Ph.D. degree in optimal design in linear models in 1983.

He became Professor of computer science and artificial intelligence at the University of Granada, Granada, Spain, in 2000. His areas of research interest are image restoration (applications to astronomy and medicine), parameter estimation in image restoration, low to high image and video, and blind deconvolution.

Dr. Molina is a member of SPIE, the Royal Statistical Society, and the Asociación Española de Reconocimiento de Formas y Análisis de Imágenes (AERFAI).



Miguel Vega was born 1956 in Spain. He received the Bachelor Physics degree from the Universidad de Granada, Granada, Spain, in 1979 and the Ph.D. degree from the Universidad de Granada in 1984.

He was a Staff Member (1984–1987) and Director (1989–1992) of the Computing Center facility of the Universidad de Granada. He has been a Lecturer since 1987 in the E.T.S. Ing. Informática of the Universidad de Granada (Departemento de Lenguajes y Sistemas Informáticos). He teaches software engineering. His research focuses on image processing (multichannel and superresolution image reconstruction). He has collaborated at several projects from the Spanish Research Council.

Javier Abad was born in Granada, Spain, in 1968. He received his degree in computer science from the University of Granada in 1991, and recently received the Ph. D. degree in wavelet image restoration with applications to super-resolution problems.

He has been Assistant Professor since 1991 at the Department of Computer Science and Artificial Intelligence of the University of Granada. His research interests are image and video processing, including multichannel image restoration, image and video recovery and compression, and super-resolution from (compressed) stills and video sequences.

Aggelos K. Katsaggelos (F'98) received the Diploma degree in electrical and mechanical engineering from the Aristotelian University of Thessaloniki, Greece, in 1979 and the M.S. and Ph.D. degrees both in electrical engineering from the Georgia Institute of Technology, Atlanta, in 1981 and 1985, respectively.

In 1985, he joined the Department of Electrical and Computer Engineering at Northwestern University, Evanston, IL, where he is currently Professor, holding the Ameritech Chair of Information Technology. He is also the Director of the Motorola Center for Communications. During the 1986–1987 academic year, he was an Assistant Professor at Department of Electrical Engineering and Computer Science, Polytechnic University, Brooklyn, NY.

Dr. Katsaggelos is a Fellow of the IEEE, an Ameritech Fellow, a member of the Associate Staff, Department of Medicine, at Evanston Hospital, and a member of SPIE. He is a member of the Publication Board of the IEEE PROCEEDINGS, the IEEE Technical Committees on Visual Signal Processing and Communications, and Multimedia Signal Processing, Editorial Board Member of Academic Press, Marcel Dekker: Signal Processing Series, *Applied Signal Processing*, and *Computer Journal*. He has served as editor-in-chief of the *IEEE Signal Processing Magazine* (1997–2002), a member of the Publication Boards of the IEEE Signal Processing Society, the IEEE TAB Magazine Committee, an Associate editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING (1990–1992), an area editor for the journal *Graphical Models and Image Processing* (1992–1995), a member of the Steering Committees of the IEEE TRANSACTIONS ON IMAGE PROCESSING (1992–1997) and the IEEE TRANSACTIONS ON MEDICAL IMAGING (1990–1999), a member of the IEEE Technical Committee on Image and Multidimensional Signal Processing (1992–1998), and a member of the Board of Governors of the IEEE Signal Processing Society (1999–2001). He is the editor of *Digital Image Restoration* (New York: Springer-Verlag, Heidelberg, 1991), coauthor of *Rate Distortion Based Video Compression* (Norwell, MA: Kluwer, 1997), and coeditor of *Recovery Techniques for Image and Video Compression and Transmission* (Norwell, MA: Kluwer, 1998). He is the coinventor of eight international patents, and the recipient of the IEEE Third Millennium Medal (2000), the IEEE Signal Processing Society Meritorious Service Award (2001), and an IEEE Signal Processing Society Best Paper Award (2001).