Bayesian Multichannel Image Restoration Using Compound Gauss-Markov Random Fields

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

***Abstract*—In this paper, we develop a multichannel image restoration algorithm using Compound Gauss Markov Random Fields (CGMRF) models. The line process in the CGMRF will allow the channels to share important information regarding the objects present in the scene. In order to estimate the underlying multichannel image two new iterative algorithms are presented, whose convergence is established. They can be considered as extensions of the classical simulated annealing and iterative conditional methods. Experimental results with color images demonstrate the effectiveness of the proposed approaches.**

***Index Terms*—Compound Gauss-Markov random fields, iterative conditional mode, multichannel image restoration, simulated annealing.**

I. Introduction

In many applications the images to be processed have a multichannel nature; that is, there are several image planes available called channels with redundant, as well as complementary information. The different channels may correspond, for instance, to different frequencies, different sensors, or different time frames. Application areas where multichannel images are used include, among others, multi-spectral satellite remote sensing, multi-sensor robot guidance, medical image analysis, electron microscopy image analysis, and color image processing.

The goal of multichannel image restoration is to obtain an estimate of the source multichannel image from its blurred and noisy observation, exploiting the known complementarity of the different channels. Information that is not readily available from the observed multichannel image is recovered by the restoration process.

Processing multichannel images is much more complicated than processing single channel (grayscale) images, due to, among other reasons, the higher dimensionality of the problem. However, relevant information present in a given channel can be taken into account when processing the others, providing therefore more accurate restoration results. In order to effectively deconvolve the observed image, this joint processing of the different channels is particularly essential when cross-channel blurring is present. However, even in the absence of cross-channel blurring, more precise restorations can be obtained by combining information from the various channels [1]. Consider, for instance, the case of an object having a very low intensity in one of the channels; knowledge of its presence in the other channels may prevent us from smoothing its boundary, which may produce texture errors and color bleeding.

Previous works have approached the restoration of noisy and blurred multichannel images by using Wiener filtering [1], [2], set theoretic and constrained least squares [3]–[6], Bayesian methods [7]–[11] and Total Variation methods [12]. See [13] for an overview of multichannel restoration techniques. See also [14] and [15] for new restoration methods based on highly redundant multiscale transforms.

In order to effectively deconvolve the image taking into account the information from the other channels, it is necessary to discern what is relevant and what is spurious information coming from the rest of the channels. A way to deal with this problem is to enforce the similarity between intensity values of corresponding pixels in different channels using so called spectral, or cross-channel, regularization. In 3-D Laplacian regularization this cross-channel regularization operator is combined with spatial, within channel, regularization [5]. These techniques give results which range from a clear improvement of multichannel over single-channel processing, if the different channel images are highly correlated, to a worsening if they are uncorrelated [3]. Another significant approach is weighted 3-D Laplacian regularization [3] where the influence of each channel is weighted according to the value of its norm or its optical flux [7]. The weighted approach yields better results than the 3-D Laplacian regularization approach if the channels are highly correlated; if the bands, however, do not have similar values, which is the case with most real color images, color bleeding and flux leaking is observed in the resulting restoration. Therefore, techniques adapted to the properties of the different image region should be included in order to prevent these problems from occurring.

The correct treatment of object edges, whose variability across channels and from object to object is to a large extent responsible for the nonstationary nature of cross-channel correlations [8], has been shown to be necessary for the success of multichannel restoration approaches. Experiments have been reported where nonadaptive multichannel techniques yield worse results than single-channel ones as a consequence of object boundary misalignments among different channels [8].

Manuscript received November 26, 2001; revised May 8, 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. Mark S. Drew.

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

A. K. Katsaggelos is with the Department of Electrical and Computer Engineering, Northwestern University, Evanston, IL 60208-3118 USA (e-mail: aggk@ece.nwu.edu).

M. Vega is with the Departamento de Lenguajes y Sistemas Informáticos, Universidad de Granada, 18071 Granada, Spain (e-mail: mvega@ugr.es).

Digital Object Identifier 10.1109/TIP.2003.818015

Since the correlation between the different channels may be small, in this paper we propose a solution to the multichannel image restoration problem from the point of view of similarity among the region edges in the different channels. Hence, similarity between pixel values in different channels will not be enforced but edge preserving spatial regularization will be applied within channel regions. Similarity between different channel edges is introduced by favoring the presence of an edge in a channel if the same edge is present in the other channels, thus detecting similar regions in all image channels and successfully preventing color bleeding in image edges while preserving the channel local characteristics.

In this paper we will use the results in [9], where we apply Compound Gauss Markov Random Fields (CGMRF) based edge preserving regularization to the restoration of severely blurred high range images, and extend them to the multichannel processing problem, advancing along the lines of our preliminary results [10], [11].

This paper is organized as follows. In Section II, notation, the proposed model for the image and line processes, and the noise model, are introduced. In Section III, algorithms to obtain the MAP estimate without blurring using stochastic and deterministic approaches are described. Problems with such methods in the presence of blurring are addressed in Section IV. In Section V we present a modified version of the stochastic algorithm and its corresponding deterministic one to restore multichannel images when blurring is present. In Section VI, the performance of the proposed algorithms is presented and compared with other methods. Finally, Section VII concludes the paper.

II. Notation and Models

A. Bayesian Model

We will distinguish between f, the 'true' or source image which would be observed under ideal conditions (i.e., no noise and no distortions produced by blurring and instrumental effects), and g, the observed image. Let us assume that there are L channels, each one of size $M \times N$, represented by

$$f = \begin{pmatrix} f^1 \\ f^2 \\ \vdots \\ f^L \end{pmatrix} \qquad \text{and } g = \begin{pmatrix} g^1 \\ g^2 \\ \vdots \\ g^L \end{pmatrix}$$

where each one of the $M \times N$ vectors $g^c, f^c, c = 1, \ldots, L$, results from the lexicographic ordering of the two-dimensional signal in each channel. We will denote by f_i^c the intensity of the original channel c at the pixel location i on the lattice. The convention applies equally to the observed image g. Our aim will be to reconstruct f from g. Bayesian methods start with a *prior distribution*, a probability distribution over images f where we incorporate information on the expected structure within an image. When using a CGMRF as prior distribution, we also introduce a line process, l, that, intuitively, acts as an activator or inhibitor of the relation between two neighboring pixels depending on whether or not the pixels are separated by an edge. In the Bayesian framework it is also necessary to specify $\mathrm{p}(g \,|\, f, l)$, the probability distribution of observed images g if f were the 'true' image and l the line process. This distribution models how the observed image has been obtained from the 'true' one. The Bayesian paradigm dictates that inference about the true f and l should be based on $\mathrm{p}(f, l \,|\, g)$ given by

$$\mathrm{p}(f, l \,|\, g) = \frac{\mathrm{p}(g \,|\, f, l)\mathrm{p}(f, l)}{\mathrm{p}(g)} \propto \mathrm{p}(g \,|\, f, l)\mathrm{p}(f, l). \qquad (1)$$

Maximization of (1) with respect to f and l yields

$$\hat{f}, \hat{l} = \arg\max_{f, l} \mathrm{p}(f, l \,|\, g) \qquad (2)$$

the maximum *a posteriori* (MAP) estimate. Let us now examine the degradation and prior models.

B. Degradation Model

The model to obtain g from f is given by

$$g = Hf + n$$

where $n = (n^1\ n^2 \cdots n^L)$, is a Gaussian noise vector of independent components with $n_i^c \sim N(0, \sigma_{\mathrm{n}}^{2^c})$. Note that each n^c is an $M \times N$ column vector. H is the $[L \times (M \times N)] \times [L \times (M \times N)]$ multichannel blurring matrix of the form

$$H = \begin{pmatrix} \lambda^{11}H^{11} & \lambda^{12}H^{12} & \cdots & \lambda^{1L}H^{1L} \\ \lambda^{21}H^{21} & \lambda^{22}H^{22} & \cdots & \lambda^{2L}H^{2L} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda^{L1}H^{L1} & \lambda^{L2}H^{L2} & \cdots & \lambda^{LL}H^{LL} \end{pmatrix} \qquad (3)$$

whose components, $H^{cc'}$, are $(M \times N) \times (M \times N)$ matrices defining systematic blurs (assumed to be known) and $\lambda^{cc'}$ denotes the contribution of each channel c' to the blurring of a given channel c. Note, that $\forall c, c', H^{cc'}$ is a "classical" energy preserving linear space-invariant blurring function used in single channel restoration problems. Note also, that for H to be an energy preserving multichannel blurring matrix, $\lambda^{cc'} \geq 0, \forall c, c'$, and $\sum_{c'=1}^{L} \lambda^{cc'} = 1, \forall c$.

With the above formulation, each component of g, g_i^c, is obtained as

$$g_i^c = (Hf)_i^c + n_i^c = \sum_{c'=1}^{L} \lambda^{cc'} \sum_j h_{(i-j)}^{cc'} f_j^{c'} + n_i^c$$

where $h_i^{cc'}$ are the coefficients defining the blurring function represented in $H^{cc'}$.

We therefore have that the probability of the observed image g if f were the 'true' image and l the 'true' line process is

$$\mathrm{p}(g \,|\, f, l) \propto \exp\left\{-\frac{1}{2}(g - Hf)^t\, D_n\, (g - Hf)\right\} \qquad (4)$$

(note that $\mathrm{p}(g \,|\, f, l) = \mathrm{p}(g \,|\, f)$) where D_n, is an $[L \times (M \times N)] \times [L \times (M \times N)]$ block-diagonal matrix of the form

$$D_n = \begin{pmatrix} D_n^1 & 0 & \cdots & 0 \\ 0 & D_n^2 & \ddots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & D_n^L \end{pmatrix} \qquad (5)$$

where, for each $c \in \{1, \ldots, L\}$, D_n^c is an $(M \times N) \times (M \times N)$ diagonal matrix of the form $\beta^c I$ with $\beta^c = 1/\sigma_{\mathrm{n}}^{2^c}$.

Let us now briefly comment on the structure of the matrix H. In [8] examples are reported for the case $H^{ii} = H^0$ with $\lambda^{ii} = 1, \lambda^{ij} = 0, i \neq j$. This matrix H corresponds, for instance, to motion blur. As pointed out in [8] $\lambda^{ij} = 0, i \neq j$ with different blurring matrices H^{ii} in each band corresponds to chromatic aberrations sometimes present in camera lenses (see also [16]). The case with $\lambda^{ij} \neq 0 \, \forall i, j$ is encountered in remote sensing of multispectral satellite images. Less than ideal spectral characteristics of detectors produce leakage between adjacent bands, resulting in between-channel degradation [17]. Blurring matrices with a great deal of inter-channel blur also appear in the restoration of color images taken with a single CCD sensor [18]. In this case, however, the observed multichannel (color) images are also subsampled, something that is not addressed by our problem formulation.

An equally important issue is the modeling of the noise process. In all the reported work in the literature on multichannel image restoration that we are aware of, the noise between channels is assumed to be independent and Gaussian independent noise within each channel. In other words, the model of (4) is used. Additional noise models are also captured by (4), for example, if the noise is independent between channels and within channel independent Gaussian with signal dependent variance $\epsilon^c (Hf)_i^c$, for each channel c and each i pixel, then (4) is still valid with each diagonal term of D_n^c in (5) given by $1/(\epsilon^c (Hf)_i^c)$.

For applications with independent Poisson noise within each channel (4) needs to be rewritten and the algorithms in [19] need to be modified to deal with the multichannel problem when using the prior model proposed next.

C. Prior Model

The use of a CGMRF as prior was first presented in [20] using an Ising model to represent the upper level and a line process to model the abrupt transitions. The model was extended to continuous range images in [21] (see also [22] and [23]).

Let us first describe the prior model for one channel, i.e., $f = f^1$, and without any edges. Our prior knowledge about the smoothness of the object luminosity distribution makes it possible to model the distribution of f by a CAR (Conditional Auto-Regressive) model (see [24]). Thus,

$$\mathrm{p}(f^1) \propto \exp\left\{-\frac{1}{2}\alpha^1 f^{1T}(I - \phi N) f^1\right\}$$

where $N_{ij} = 1$ if cells i and j are spatial neighbors (pixels at distance one), zero otherwise and ϕ just less than 0.25. The parameter α^1 measures the smoothness of the "true" image.

The above prior distribution can be written as

$$-\log \mathrm{p}(f^1) = \mathrm{const} + \frac{1}{2}\alpha^1 \sum_i f_i^1 \left(f_i^1 - \phi (N f^1)_i\right).$$

Then, if we assume a "toroidal correction" and $i{:}\ +1, i{:}\ +2, i{:}\ +3, i{:}\ +4$ denote the four pixels around pixel i as described in Fig. 1 (note that if $i = (u, v)$ they correspond to $(u+1, v), (u, v+1), (u-1, v)$, and $(u, v-1)$, respectively), we have

$$-\log \mathrm{p}(f^1) = \mathrm{const} + \frac{1}{2}\alpha^1 \sum_i \Big[\phi \left(f_i^1 - f_{i:+1}^1\right)^2 + \phi \left(f_i^1 - f_{i:+2}^1\right)^2 + (1 - 4\phi) f_i^{1\,2}\Big].$$

We now introduce a line process by defining the function $l_{[i,j]}^1$ as taking the value zero if pixels i and j are not separated by an active line and one otherwise (see Fig. 1). Then we would have, in the single channel case

$$\begin{aligned}-\log \mathrm{p}(f^1, l^1) = \mathrm{const} + \frac{1}{2}\alpha^1 & \times \sum_i \Big[\phi \left(f_i^1 - f_{i:+1}^1\right)^2 \left(1 - l_{[i,i:+1]}^1\right) \\ & + \phi \left(f_i^1 - f_{i:+2}^1\right)^2 \left(1 - l_{[i,i:+2]}^1\right) \\ & + \tau^1 l_{[i,i:+1]}^1 + \tau^1 l_{[i,i:+2]}^1 + (1 - 4\phi) f_i^{1\,2}\Big].\end{aligned}$$

The introduction of an active line element in the position $[i, j]$ is penalized by the term $\tau^1 l_{[i,j]}^1$ since otherwise the above equation would obtain its minimum value by setting all line elements equal to one. The intuitive interpretation of this line process is simple; it acts as an activator or inhibitor of the relation between two neighbor pixels depending on whether or not the pixels are separated by an edge. The straightforward extension of the above prior model to the multichannel case, that is, $f = (f^1, \ldots, f^L)^t$ and $l = (l^1, \ldots, l^L)^t$, is to write

$$\begin{aligned}-\log \mathrm{p}(f, l) = \mathrm{const} + \sum_{c=1}^{L} \Bigg[\frac{1}{2}\alpha^c & \times \sum_i \Big[\phi \left(f_i^c - f_{i:+1}^c\right)^2 \left(1 - l_{[i,i:+1]}^c\right) \\ & + \tau^c l_{[i,i:+1]}^c + \phi \left(f_i^c - f_{i:+2}^c\right)^2 \left(1 - l_{[i,i:+2]}^c\right) \\ & + \tau^c l_{[i,i:+2]}^c + (1 - 4\phi) f_i^{c^2}\Big]\Bigg].\end{aligned}$$

We want, however, to increase the probability of a new active line element in the position $[i, j]$ in a given channel if the other channels have a line in the same position (see Fig. 2 for a three channels–RGB–example).

To do so we include in the prior model the term $\epsilon^{cc'}(l_{[i,i:+1]}^c l_{[i,i:+1]}^{c'} + l_{[i,i:+2]}^c l_{[i,i:+2]}^{c'})$ with $c, c' = 1, \ldots, L, c \neq$

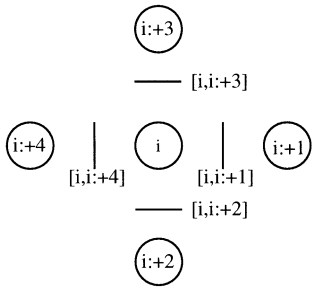


Fig. 1. Image and line sites.

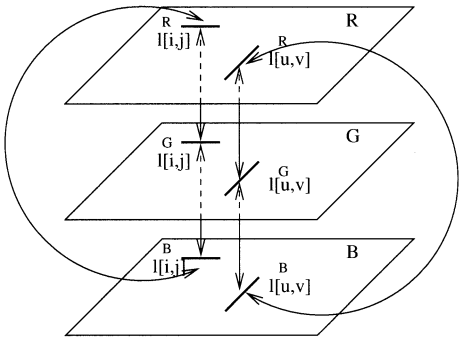


Fig. 2. Interactions between line processes at different channels.

c' and $\epsilon^{cc'} > 0$ and finally write the proposed prior multichannel model as

$$
\begin{aligned}
-\log \mathrm{p}(f,l) = \text{const} \\
&+ \sum_{c=1}^{L} \left[\frac{1}{2}\alpha^c \times \sum_i \left[\phi \left(f_i^c - f_{i:+1}^c \right)^2 \left(1 - l_{[i,i:+1]}^c \right) \right. \right. \\
&+ \tau^c l_{[i,i:+1]}^c + \phi \left(f_i^c - f_{i:+2}^c \right)^2 \left(1 - l_{[i,i:+2]}^c \right) \\
&+ \left. \tau^c l_{[i,i:+2]}^c + (1-4\phi) f_i^{c^2} \right. \\
&- \left. \sum_{\substack{c'=1 \\ c' \neq c}}^{L} \epsilon^{cc'} \left(l_{[i,i:+1]}^c l_{[i,i:+1]}^{c'} + l_{[i,i:+2]}^c l_{[i,i:+2]}^{c'} \right) \right] \Bigg] .
\end{aligned}
\tag{6}
$$

The introduction of these cross terms will help to recognize the same objects in all the channels even if they do not have similar intensities. Note that if $\epsilon^{cc'} = 0, \forall c, c'$ no prior edge information will be passed between channels. Furthermore, if $\epsilon^{cc'}$ is very large the restoration will tend to have the same edge map in all channels.

It could be argued that by the introduction of these cross-channel terms in the line process, the prior may end up creating nonexisting regions in some channels. Consider, for instance, an 8-bit per channel RGB image consisting of an object with intensity 128 in the R and B channels and a blue background, also of intensity 128. In this case the prior may create the object in the blue band when, in fact, if we look only in terms of intensity values in that band, the object does not exist. This creates no problems, however, since the introduction of a line will only force smoothing with fewer than four neighbors. On the other hand, for this same example, if the intensity of the background changes from the value of 128 to 100, the prior will allow the recognition of the object, despite the small intensity difference between object and background in the blue channel.

III. MAP Estimation Using Stochastic and Deterministic Relaxation

In order to find the MAP estimate for problems where the prior model explicitly favors the presence of discontinuities, stochastic algorithms such as simulated annealing (see [20] and also [9]), deterministic methods (see [23], [25] and also [26]) and mixed-annealing algorithms, where stochastic steps alternate with deterministic ones (see [27] and [28]), have been proposed.

The classical simulated annealing (SA) algorithm, for our multichannel MAP estimation problem defined in (2), uses the probability distribution, derived from (4) and (6)

$$
\begin{aligned}
&\mathrm{p}_T(f,l \,|\, g) \\
&= \frac{1}{Z_T} \exp \left\{ -\frac{1}{T} \left[\frac{1}{2} (g - Hf)^t \, D_n \, (g - Hf) \right. \right. \\
&+ \sum_{c=1}^{L} \frac{1}{2} \alpha^c \sum_i \left[\phi \left(f_i^c - f_{i:+1}^c \right)^2 \left(1 - l_{[i,i:+1]}^c \right) \right. \\
&+ \tau^c l_{[i,i:+1]}^c + \phi \left(f_i^c - f_{i:+2}^c \right)^2 \left(1 - l_{[i,i:+2]}^c \right) \\
&+ \tau^c l_{[i,i:+2]}^c + (1-4\phi) f_i^{c^2} \\
&- \left. \left. \sum_{\substack{c'=1 \\ c' \neq c}}^{L} \epsilon^{cc'} \left(l_{[i,i:+1]}^c l_{[i,i:+1]}^{c'} + l_{[i,i:+2]}^c l_{[i,i:+2]}^{c'} \right) \right] \right] \Bigg\}
\end{aligned}
\tag{7}
$$

where T is the temperature and Z_T a normalization constant.

The algorithm simulates the conditional *a posteriori* density function of $l_{[i,j]}^c$, given the rest of l, f and g and the conditional *a posteriori* density function of f_i^c given the rest of f, l and g.

To simulate $\mathrm{p}_T(l_{[i,j]}^c \,|\, l_{[m,n]}^c \colon [m,n] \neq [i,j], l^{c'} \colon c' \neq c, f, g)$, using a raster-scanning scheme we have

$$
\begin{aligned}
&\mathrm{p}_T \left(l_{[i,j]}^c = 0 \,|\, l_{[m,n]}^c : [m,n] \neq [i,j], l^{c'} : c' \neq c, f, g \right) \\
&\quad \propto \exp \left[-\frac{1}{T} \frac{\alpha_c \phi}{2} \left(f_i^c - f_j^c \right)^2 \right],
\end{aligned}
\tag{8}
$$

$$
\begin{aligned}
&\mathrm{p}_T \left(l_{[i,j]}^c = 1 \,|\, l_{[m,n]}^c \colon [m,n] \neq [i,j], l^{c'} \colon c' \neq c, f, g \right) \\
&\quad \propto \exp \left[-\frac{1}{T} \frac{\alpha_c}{2} \left(\tau^c - \sum_{c' \neq c} \epsilon^{cc'} l_{[i,j]}^{c'} \right) \right].
\end{aligned}
\tag{9}
$$

To simulate the conditional *a posteriori* density function of f_i^c given the rest of f, l and g, using a raster-scanning scheme we have

$$
\mathrm{p}_T \left(f_i^c \,|\, f_j^c \colon \forall j \neq i, f^{c'} \colon \forall c' \neq c, l, g \right) \sim \mathcal{N} \left(\mu_i^c, T \sigma_i^{c^2} \right)
\tag{10}
$$

where

$$
\begin{aligned}
\mu_i^c = \gamma_i^c \phi \sum_{j \mathrm{nhbr} i} \frac{f_j^c \left(1 - l_{[i,j]}^c \right)}{n n_i^c(l)} \\
+ (1 - \gamma_i^c) \left(\frac{[H^t \, D_n \, (g - Hf)]_i^c}{\mathrm{s^c}} + f_i^c \right)
\end{aligned}
\tag{11}
$$

and

$$
\sigma_i^{c^2} = \frac{1}{\alpha^c n n_i^c(l) + \mathrm{s^c}}
\tag{12}
$$

where the suffix "$j\,\mathrm{nhbr}\,i$" denotes the four neighbor pixels at distance one from pixel i (see Fig. 1), $nn_i^c(l) = \phi \sum_{j\mathrm{nhbr}i}(1 - l^c_{[i,j]}) + (1 - 4\phi)$

$$\gamma_i^c = \frac{\alpha^c nn_i^c(l)}{\alpha^c nn_i^c(l) + \mathrm{s^c}}, \tag{13}$$

and

$$\mathrm{s^c} = \sum_{c'=1}^{L} \beta^{c'} \lambda^{c'c^2} \sum_i h_i^{c'c^2}. \tag{14}$$

Then the sequential SA to find the MAP estimate, when no blurring is present ($H = I$), proceeds as described in [21] and [29] (see also [9]).

Instead of using a stochastic approach, we can use a deterministic method to search for a local maximum. An advantage of the deterministic method is that its convergence is much faster than that of the stochastic approach, since instead of simulating the distributions, the mode from the corresponding conditional distribution is chosen, hence called Iterative Conditional Mode (ICM). Note that with this approach, 0 is selected for the line process if the value of (8) is greater than that of (9), and 1 otherwise, and for f_i^c the mean, i.e., the value given by (11) is selected. A disadvantage of the method is that a local optimum is obtained. Another disadvantage that we have observed experimentally is that it tends to produce either very few active lines or excessive clusters of them in the reconstruction. This method can be seen as a particular case of SA where the temperature is always set to zero.

IV. Instability of the SA and ICM Solutions

Unfortunately, due to the presence of blurring the convergence of SA has not been established for this problem. Let us examine intuitively and formally why there may exist convergence problems with the SA and ICM algorithms.

Following [9], let us assume for simplicity that there is no line process and examine the iterative procedure where we update the whole image at the same time. For this simplified case, we have

$$f^{(k)} = \Gamma(\phi N) f^{(k-1)} - (I - \Gamma) \times \left[S^{-1} H^t D_n H f^{(k-1)} - f^{(k-1)}\right] + (I - \Gamma) S^{-1} H^t D_n g \tag{15}$$

where k is the iteration number, understood as a sweep of the whole image. Γ is a block diagonal matrix of the form

$$\Gamma = \begin{pmatrix} \Gamma^1 & 0 & \cdots & 0 \\ 0 & \Gamma^2 & \ddots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & \Gamma^L \end{pmatrix} \tag{16}$$

where each Γ^c is a diagonal matrix with diagonal entries given by γ_i^c as defined in (13) with $nn_i^c = 1, \forall i, c$. S is a block diagonal matrix of the form

$$S = \begin{pmatrix} S^1 & 0 & \cdots & 0 \\ 0 & S^2 & \ddots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & S^L \end{pmatrix} \tag{17}$$

where each S^c is a diagonal matrix of the form $\mathrm{s^c} I$ with $\mathrm{s^c}$ defined in (14). Finally, N is a block diagonal matrix of the form

$$N = \begin{pmatrix} N^1 & 0 & \cdots & 0 \\ 0 & N^2 & \ddots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & N^L \end{pmatrix} \tag{18}$$

where $\forall c, N_{ij}^c = 1$ if cells i and j are spatial neighbors (pixels at distance one) and zero otherwise.

Note that (15) can be written as

$$f^k = A f^{(k-1)} + \mathrm{const} \tag{19}$$

where

$$A = [I - \Gamma(I - \phi N) - (I - \Gamma) S^{-1} H^t D_n H]. \tag{20}$$

For this algorithm to converge A must be a contraction mapping, that is $\| A \| < 1$. However this may not be the case. For instance, if the image suffers from severe blurring then the elements of S are close to zero and the matrix $S^{-1} H^t D_n H$ has eigenvalues greater than one. Furthermore, if the image has a high dynamic range it is natural to assume that α^c is small and thus, $(I - \Gamma)[S^{-1} H^t D_n H]$ has eigenvalues greater than one, which could imply A having eigenvalues less than minus one. Therefore, this iterative method may not converge. It is important to note that, when there is no blurring, $H = S = I$ and therefore $A = [I - \Gamma(I - \phi N) - (I - \Gamma) I]$, which is a contraction mapping.

To establish the convergence of the SA algorithm, we need to show that A is a contraction (Lemma 1 in [29]). Since, for the blurring case, A is not a contraction, convergence cannot be established for the multichannel deconvolution problem under consideration.

Let us then now modify A in order to make it contractive. Writing the iterative method in (15) as a gradient descent method we have

$$\begin{aligned} f^{(k)} &= f^{(k-1)} - \Gamma(I - \phi N) f^{(k-1)} - (I - \Gamma) \\ &\quad \times \left[S^{-1} H^t D_n H f^{(k-1)}\right] + (I - \Gamma) S^{-1} H^t D_n g \\ &= f^{(k-1)} - \Omega \Big\{ [D_p (I - \phi N) + H^t D_n H] f^{(k-1)} \\ &\quad - H^t D_n g \Big\} \end{aligned} \tag{21}$$

where D_p is a block diagonal matrix of the form

$$D_p = \begin{pmatrix} D_p^1 & 0 & \cdots & 0 \\ 0 & D_p^2 & \ddots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & D_p^L \end{pmatrix} \tag{22}$$

where each D_p^c is a diagonal matrix of the form $\alpha^c I$ and

$$\Omega = (I - \Gamma) S^{-1}$$

that is, Ω is a block diagonal matrix of the form

$$\Omega = \begin{pmatrix} \Omega^1 & 0 & \cdots & 0 \\ 0 & \Omega^2 & \ddots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & \Omega^L \end{pmatrix} \tag{23}$$

with diagonal block $\Omega^c = (1)/\alpha^c + \mathrm{s}^c I$.

We note that for the iterative gradient descent method defined in (21) to converge the step size Ω has to be small enough. However, using the step size in (23) our modified iterative scheme is not guaranteed to converge. In Appendix A we show that if we use a modified step size Ψ given by

$$\Psi = \frac{1}{\max_c(\alpha^c + \sum_{c'} \beta^{c'} \lambda^{c'c})} I \tag{24}$$

instead of Ω, the gradient descent algorithm in (21) will converge.

The use of Ψ in (21) instead of Ω results in the modified iteration

$$\begin{aligned} f^{(k)} &= f^{(k-1)} - \Psi \Big\{ [D_p(I - \phi N) \\ &\quad + H^t D_n H] f^{(k-1)} - H^t D_n g \Big\} \\ &= (I - \Lambda) f^{(k-1)} + \Lambda \Big\{ f^{(k-1)} \\ &\quad - \Omega \Big\{ [D_p(I - \phi N) + H^t D_n H] f^{(k-1)} - H^t D_n g \Big\} \Big\} \end{aligned} \tag{25}$$

where $\Lambda = \Psi \Omega^{-1}$. All entries of Λ are now less than one. The interpretation of the iterative method in (25) is that the whole image is updated by using a convex combination of the old value of the image and the conditional mean described in (11). We will use these ideas in the following section to propose our modified multichannel simulated annealing algorithm.

V. Proposed Modified Multichannel Simulated Annealing Algorithm

Having studied the problem of simulating the conditional distribution when blurring but no line process is present, let us now examine how to replace the mean of the conditional distribution in order to have a contraction when a CGMRF is used.

In [9] modified versions of the classical SA [21] and ICM [30] algorithms were proposed, to deal with the presence of blurring in single channel image restoration. The convergence of these modified algorithms (referred to as MSA and MICM) was established. These methods were extended to the multichannel restoration case in [10] and [11], for block diagonal blurring matrices (intra-channel degradations). In the following we extend such methods to the general case described by (3).

We start by rewriting the conditional mean in (11) as

$$\mu_i^c = f_i^c - \frac{1}{\alpha^c n n_i^c(l) + \mathrm{s}^c} \left[\alpha^c \phi \sum_{j \mathrm{nhbr} i} \left(f_i^c - f_j^c \right) \left(1 - l_{[i,j]}^c \right) + \alpha^c (1 - 4\phi) f_i^c + \left[H^t D_n (g - Hf) \right]_i^c \right].$$

Again, as explained in Section IV, the resulting iterative restoration algorithm is not guaranteed to converge since the mapping resulting from the right hand side of the above equation is not contractive.

In order to have a contraction we replace this conditional mean by

$$\underline{\mu}_i^c = (1 - \varepsilon_i^c) f_i^c + \varepsilon_i^c \mu_i^c \tag{26}$$

with

$$\varepsilon_i^c = \frac{\alpha^c n n_i^c(l) + \mathrm{s}^c}{\max(\alpha^c + \sum_{c'} \beta^{c'} \lambda^{c'c})}. \tag{27}$$

In the modified simulated annealing procedure we propose, we sample the whole image at the same time, each component according to the distribution

$$\mathcal{N}\left(\underline{\mu}_i^c, T \hat{\sigma}_i^{c^2} \right) \tag{28}$$

where

$$\hat{\sigma}_i^{c^2} = \left(1 - (1 - \varepsilon_i^c)^2 \right) \sigma_i^{c^2}. \tag{29}$$

The reason for using this modified variance is clear if we take into account that, if

$$X \sim \mathcal{N}(m, \sigma^2)$$

and

$$Y \,|\, X \sim \mathcal{N}\left(\lambda X + (1 - \lambda) m, (1 - \lambda^2) \sigma^2 \right)$$

where $0 < \lambda < 1$, then

$$Y \sim \mathcal{N}(m, \sigma^2).$$

The resulting proposed Multichannel Modified Simulated Annealing algorithm (MMSA) is as follows.

Algorithm 1 (MMSA Procedure):

1) Set $t = 0$ and assign an initial configuration denoted as $f^{(-1)}, l^{(-1)}$ and initial temperature $T^{(0)} = 1$.
2) The evolution $l^{(t-1)} \rightarrow l^{(t)}$ of the line process can be obtained by sampling the next point of the line process from the raster-scanning scheme based on the conditional probability mass function defined in (8) and (9) and keeping the rest of $l^{(t-1)}$ unchanged.
3) Set $t = t + 1$. Go back to step 2 until a complete sweep of the field l is finished.
4) The evolution $f^{(t-1)} \rightarrow f^{(t)}$ of the image f can be obtained by sampling the next value of the whole image based on the conditional probability mass function given in (28).
5) Go to step 2 until $t > t_f$, where t_f is a specified integer.

Two points are worth mentioning now. In the proposed algorithm we have now a contraction and so Lemma 1 in [29] is satisfied (see Appendix B). Unfortunately we have to update the whole image at the same time and in addition we are no longer sampling from the conditional distribution of f_i^c given the rest of f, l and g. In other words, we do not have a stationary distribution. This will prevent us from using Lemma 3 in [20] that shows that starting from the uniform distribution over all

global maxima the process of simulating the conditional distribution of f_i^c given the rest of f, l and g when the temperature decreases converges to the above uniform distribution over all global maxima. In Appendix B we also prove the following theorem that guarantees the convergence of the MMSA to a local MAP estimate, even in the presence of intra and inter-channel blurring.

Theorem 1: If the following conditions are satisfied:

1) $|\phi| < 0.25$;
2) $T(t) \to 0$ as $t \to \infty$; such that
3) $T(t) \geq C_T / \log(1 + k(t))$;

then for any starting configuration $f^{(-1)}, l^{(-1)}$, we have

$$\mathrm{p}\left(f^{(t)}, l^{(t)} \,\middle|\, f^{(-1)}, l^{(-1)}, g\right) \to \mathrm{p}_0(f, l) \quad \text{as } t \to \infty$$

where $\mathrm{p}_0(f, l)$ is a probability distribution over local MAP solutions, C_T is a constant and $k(t)$ is the sweep iteration number at time t.

We notice that if the method converges to a configuration $(\bar{f}, \bar{l})$, then

$$\bar{f} = \arg\max_f p(f|\bar{l}, g),$$

and

$$\bar{l} = \arg\max_l p(l|\bar{f}, g).$$

We conjecture that the proposed MMSA method converges to a distribution over global maxima. However, the difficulty of using synchronous models prevent us from proving this result (see [31]).

A modified multichannel ICM procedure (MMICM) can be obtained by selecting in steps 2 and 4 of Algorithm 1 the mode of the corresponding transition probabilities.

Finally, we make a comment on the block diagonal blurring case. In [10] and [11], for block diagonal blurring matrices, that is, $\lambda^{c'c} = 0$ if $c \neq c'$ and $\lambda^{cc} = 1$, we used

$$\underline{\varepsilon}_i^c = \frac{\alpha^c nn_i^c(l) + \mathrm{s}^c}{\alpha^c + \beta^c}$$

instead of

$$\varepsilon_i^c = \frac{\alpha^c nn_i^c(l) + \mathrm{s}^c}{\max(\alpha^c + \beta^c)}$$

defined in (27). Using $\underline{\varepsilon}_i^c$, assigns a larger weight to μ_i^c in (26), equivalent to replacing Ψ in (24) by a block diagonal matrix $\mathcal{X}$ with diagonal block $\mathcal{X}^c = (1/(\alpha^c + \beta^c))I$.

It is shown at the end of Appendices A and B that when H is a block diagonal blurring matrix (the case studied in [10] and [11]), $\mathcal{X}$ can be used instead of Ψ.

VI. Experimental Results

Experiments were carried out with RGB color images in order to evaluate the performance of the proposed methods and compare them with other existing ones. Although visual inspection of the restored images is a very important quality measure for the restoration problem, in order to get quantitative image quality comparisons, the peak signal-to-noise ratio (PSNR) of each channel is used. Given two image channels f^c and g^c of size $M \times N$ and values in the range [0, 255], the PSNR of channel c is defined as

$$\mathrm{PSNR} = 10 \log_{10} \left[\frac{M \times N \times 255^2}{\| g^c - f^c \|^2} \right].$$

The mean of the PSNR values of all channels is used as a figure of merit for the color image.

Results obtained with the 256×256 Lena image are reported next. In order to test the proposed methods, the original image was blurred with a color bleeding multichannel point spread function, H, given by

$$H = \begin{pmatrix} 0.7\,H^0 & 0.2\,H^0 & 0.1\,H^0 \\ 0.1\,H^0 & 0.7\,H^0 & 0.2\,H^0 \\ 0.2\,H^0 & 0.1\,H^0 & 0.7\,H^0 \end{pmatrix} \tag{30}$$

where H^0 defines an out-of-focus blurring with radius 5 (see [32] for an analytical expression of this blurring). Gaussian noise of variance $\sigma_{\mathrm{n}}^{2^c} = 16$ was added to each channel, c, obtaining the degraded image shown in Fig. 3(a) with a PSNR of 18.593 dB, 20.286 dB, and 21.146 dB for the R, G, and B channels, respectively, resulting in a mean PSNR of 20.008 dB. Note that only the 192×192 central part of the image is shown.

The MMSA and MMICM algorithms were run on this multichannel image for 4000 and 400 iterations, respectively (see comments on the number of iterations later). The parameters for the prior model were chosen to be $\alpha^c = 1/150, \tau^c = 260$ and $\epsilon^{cc'} = 105$, for all c and c'. The restored images by the MMSA and MMICM algorithms are shown in Fig. 3(b) and 3(c), respectively. Again, for a better visual inspection, only the 192×192 central part of the images is shown. The corresponding PSNR values are summarized in Table I. These results show that both the MMSA and the MMICM algorithms produce very good restorations both in terms of visual quality and PSNR values; the mean PSNR value for the MMSA algorithm is slightly better. Fig. 4 depicts the line processes for the R, G, and B bands. A black pixel at location i in this figure means that either $l_{[i,i:+1]}$ or $l_{[i,i:+2]}$ are equal to 1. Although both methods detect a large number of lines (edges) present in the image, the MMICM method shows fewer lines [see Fig. 4(b)]. The MMSA method, on the other hand, gives us more complete contours [Fig. 4(a)] and details are better resolved. There are some gaps in the line process and some of the diagonal edges show a staircase effect. Better results are expected if 8 neighbors instead of 4 are used or, in general, more intra-channel terms are added to the line process energy function.

The effect of the cross term, $\epsilon^{cc'}(l^c_{[i,i:+1]} l^{c'}_{[i,i:+1]} + l^c_{[i,i:+2]} l^{c'}_{[i,i:+2]})$, introduced in (6) is considered next. Fig. 4(c) shows the line process obtained by the MMSA algorithm when the parameter $\epsilon^{cc'}$ in (9) is set to zero for all c, c'. Note that since the similarity between the different line processes is not taken into account, fewer edges are now detected and some features on the face of Lena are lost in most of the channels. Color bleeding and artifacts may now be observed in the restored image [Fig. 3(d)]. Fig. 3(e) shows an enlarged part of Fig. 3(d). Note that the border of the hat and the white of the eye show color bleeding. This effect is clearly corrected

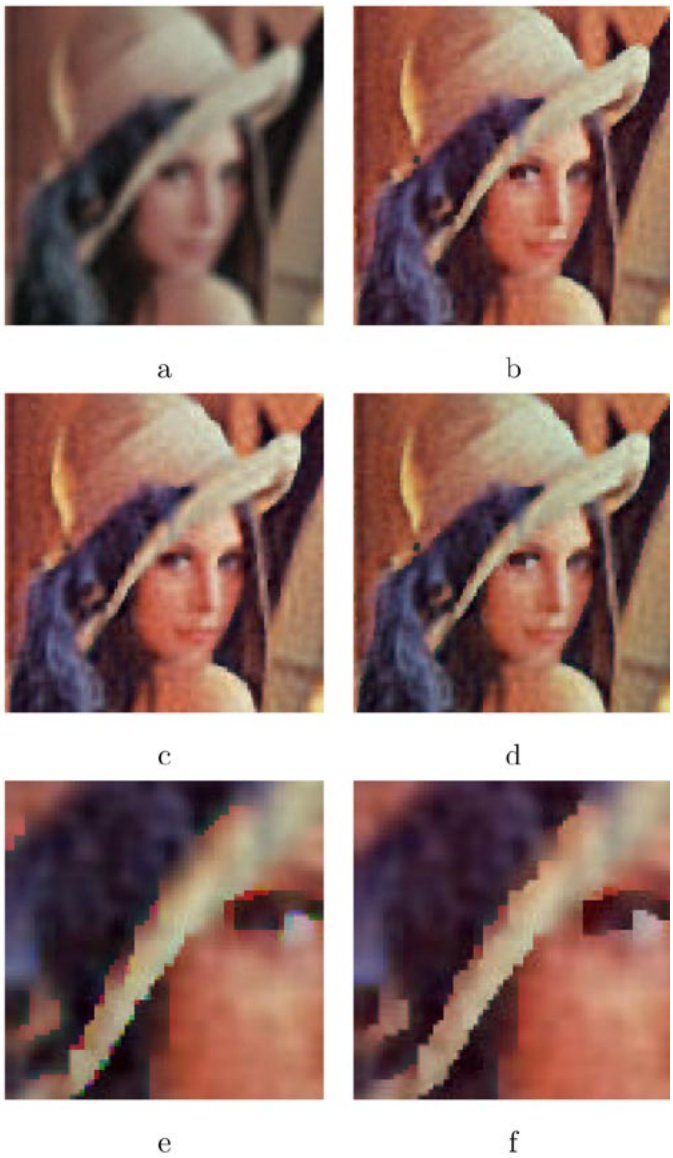


Fig. 3. (a) Degraded image. (b) Restoration with the MMSA method. (c) Restoration with the MMICM method. (d) Restoration with the MMSA method with $\epsilon^{cc'} = 0, \forall c, c'$. (e) An enlarged part of (d). (f) An enlarged part of (b).

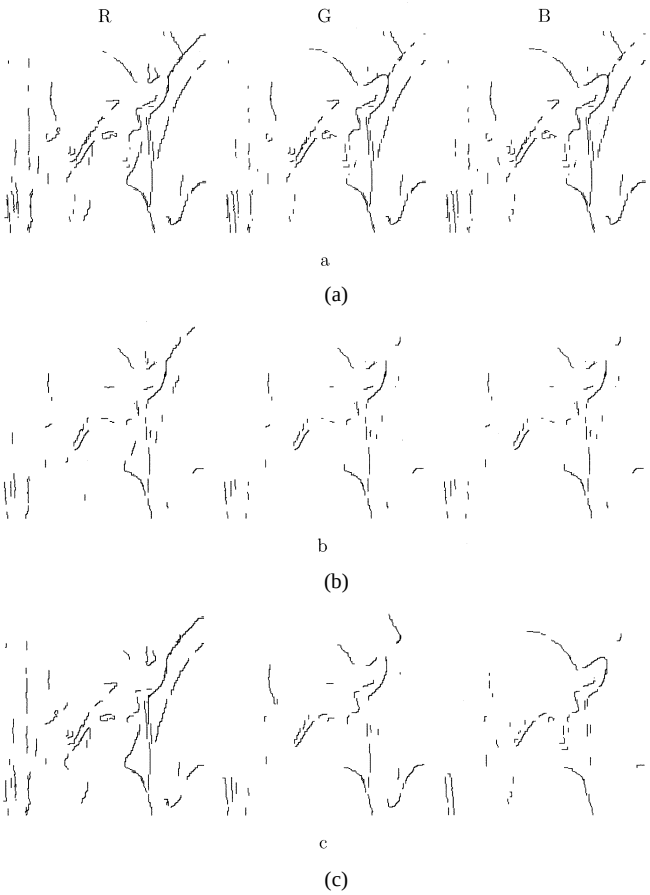


Fig. 4. (a) Line process for the MMSA restoration shown in Fig. 3(b). (b) Line process for the MMICM restoration shown in Fig. 3(c). (c) Line process for the MMSA restoration shown in Fig. 3(d).

TABLE I
PSNR FOR THE DIFFERENT RESTORATIONS OF LENA SHOWN IN FIG. 3

$PSNR$ (dB)	red	green	blue	mean
Observed	18.593	20.286	21.146	20.008
MMSA	24.194	24.133	24.835	24.387
MMICM	24.154	24.075	24.843	24.357
MMSA ($\epsilon^{cc'} = 0,\ \forall c, c'$)	24.018	23.945	24.750	24.237

by introducing the cross terms as shown in Fig. 3(b) and its enlargement shown in Fig. 3(f) while obtaining, at the same time, sharp edges. The effect of the line process is also clear by examining the PSNR values shown in Table I. As can be seen in it, the incorporation of an inter-channel line process increases the PSNR value. It is mentioned here that clearly the introduction of line processes affects primarily edge pixels. Due to this, although visual quality of the restored image is considerably increased, the resulting PSNR increase is small, since only a small number of pixels is affected. Fig. 5 shows the number of line elements detected by the MMSA method on the three channels of the Lena image when the parameter $\epsilon^{cc'}$ varies from 0 to 150 and shows the sensitivity of the model to changes in the parameter. For $\epsilon^{cc'} = 150$ the line processes obtained for the three channels are very similar.

The total computation time for the MMSA and MMICM methods on images of size 256×256 pixels is 5700 s (for 4000 iterations) and 500 s (for 400 iterations), respectively, on an AMD K7 1000 processor. These numbers imply that a single iteration of the algorithms takes 1.425 s and 1.25 s for the MMSA and MMICM methods, respectively. Note that the small difference in the computation time of one iteration between the two methods is due to the fact that most of the time is spent in convolving images, which is done by both methods. This clearly favors the deterministic MMICM algorithm, since typically it needs a considerably smaller number of iterations for producing acceptable results. Although we ran the MMICM algorithm for 400 iterations in our experiments, similar results are obtained with a much smaller number of iterations since after about 100 iterations the method has practically converged. This is shown in Fig. 6, where the norm of the difference between two consecutive iterations divided by the norm of the image in this

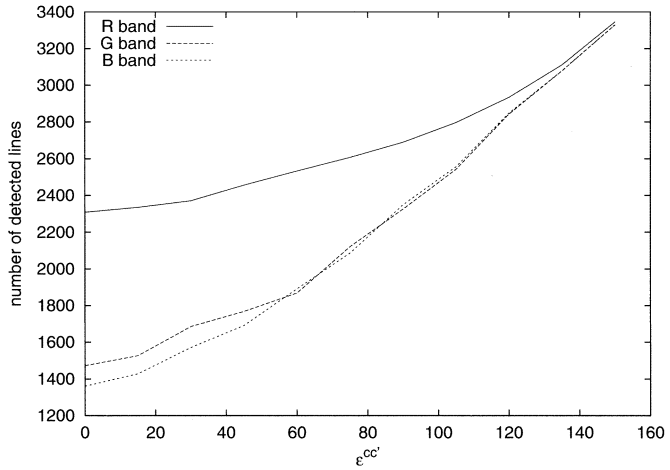


Fig. 5. Number of line elements detected by the MMICM method.

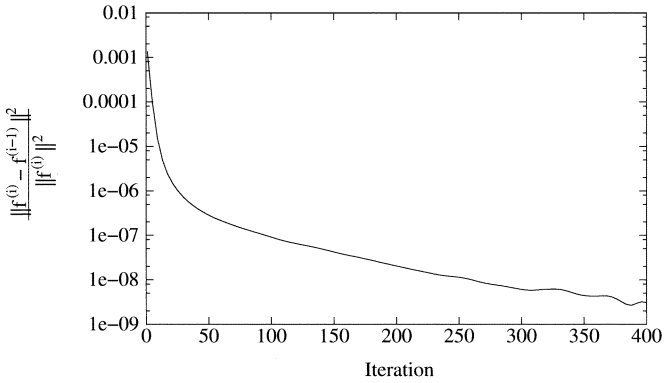


Fig. 6. Convergence of the MMICM algorithm (Y axis is plotted in logarithmic scale).

iteration is used to establish convergence (note that the vertical axis is plotted in logarithmic scale). For images of size 512×512 and 1024×1024, computation times per iteration are 7.6 and 36.1 s, respectively, for the MMSA algorithm, and 7.1 and 34.7 s, respectively for the MMICM algorithm. The algorithm's complexity is $O(n \log n)$ since we compute part of the algorithm in the Fourier domain, using FFT.

The proposed algorithms were also compared with other existing methods, such as the one proposed by Guo *et al.* [5], Galatsanos *et al.* [3] and Molina and Mateos [7]. These methods are based on a 3-D regularization operator that enforces similarity between the pixel values in the different image channels. Although the method proposed in [5] is based on the 3-D Laplacian we can easily find its Conditional Autoregressive (CAR) version, given by

$$\begin{aligned} \mathrm{p}(f) \propto \exp\left[-\sum_{c=1}^{L-1}\sum_{c'=c+1}^{L}\frac{\alpha^{cc'}}{2}\|f^c - f^{c'}\|^2\right] \\ \times \prod_{c=1}^{L}\exp\left[-\frac{\alpha^c}{2}\sum_i\left[\phi\left(f_i^c - f_{i:+1}^c\right)^2\right.\right. \\ \left.\left. + \phi\left(f_i^c - f_{i:+2}^c\right)^2 + (1-4\phi)\, f_i^{c^2}\right]\right], \end{aligned} \quad (31)$$

where the parameters $\alpha^{cc'}$ control the similarity between channels while the parameters α^c control the smoothness within channels. The method proposed in [3] takes into account the norm of each channel, that is, f^c is replaced by $f^c/\|f^c\|$ in the CAR model given by (31). Finally, the method proposed in [7] tries to preserve the flux within each channel by replacing f^c in (31) by $f^c/\sum_i f_i^c$.

Results of the comparison of these methods with the proposed MMSA and MMICM algorithms are presented again for the Lena image. Unfortunately, some of these methods do not allow for cross channel blurring so we used a diagonal point spread function, i.e., each channel was blurred independently of the other channels. The degradation used for each channel was an out-of-focus blur with radius 5 and additive Gaussian noise of variance $\sigma_{\mathrm{n}}^{2^c} = 16$. The resulting degraded image is shown in Fig. 7(a). Note that, although the whole image was processed, Fig. 7 depicts only the 192, $\times$ 192 pixel central part of the images.

All methods were run with the parameters that empirically gave the best PSNR value. Resulting restored images are shown in Figs. 7(b)–7(d) by the methods presented in [5], [3] and [7], respectively. The MMICM and MMSA methods were run for 4000 and 400 iterations, respectively. The parameters for the prior model were chosen to be $\alpha^c = 1/200, \tau^c = 200$ and $\epsilon^{cc'} = 80$, for all c and c'. The restored images by the MMSA and MMICM methods are shown in Fig. 7(e) and 7(f), respectively. For a better visual inspection, Fig. 8 presents a magnified view of Fig. 7. The corresponding PSNR values for all methods are summarized in Table II. It can be observed that in all cases the proposed methods produce better image quality and higher PSNR. It is clear that edges are sharper and there is no ringing in the restorations provided by the proposed methods. It is also important to note that the model in [7] and the ones proposed in this paper are the only ones that preserve the flux within each channel.

Finally, the proposed method was tested on a real astronomical image. Although the original image size was 512×512 pixels, only a small 192×192 region of interest is presented in Fig. 9. The observed image, depicted in Fig. 9(a), corresponds to three different images of the same object taken at different wavelength. The image was scaled to the range $[0, 255]$ for printing since the range of the original bands was extremely low.

Since the images were taken at different times no cross-channel blurring is present. The blurring functions, h^{cc}, can be approximated by

$$h_i^{cc} \propto (1 + i^2/r^2)^{-\delta}, \quad c \in \{R, G, B\}.$$

We found $\delta \sim 3$ and $r \sim 3.4$ pixels in all the channels.

In order to compare the effects of the line process parameters, the MMSA algorithm was run on this multichannel image for 4000 iterations. The noise parameters were chosen to be 11.20, 1.75 and 1.75 for the R, G and B bands, respectively. The parameters for the prior model were chosen to be $\alpha^R = 1/30, \alpha^G = 2$ and $\alpha^B = 4$. Three test were performed using different values for the parameters that control the line process, τ^c and $\epsilon^{cc'}$. In the first test, shown in Fig. 9(b), $\tau^c = 10\,000$ and $\epsilon^{cc'} = 0.0$, for all c and c' were selected, that is, no line process is considered. The second test, see Fig. 9(c), allows for the line process but does not take into account the relations between the bands ($\tau^c = 0.4$ and $\epsilon^{cc'} = 0.0$, for all c and c'). A magnified view of a part of this figure is depicted in Fig. 9(e). Finally, Fig. 9(d)

Fig. 7. (a) Degraded image. (b) Restoration with the method proposed in [5]. (c) Restoration with the method proposed in [3]. (d) Restoration with the method proposed in [7]. (e) Restoration with the proposed MMSA method. (f) Restoration with the proposed MMICM method.

Fig. 8. Magnified view of Fig. 7: (a) Degraded image. (b) Restoration with the method proposed in [5]. (c) Restoration with the method proposed in [3]. (d) Restoration with the method proposed in [7]. (e) Restoration with the proposed MMSA method. (f) Restoration with the proposed MMICM method.

(magnified in Fig. 9(f)) shows the results when considering the full model by using $\tau^c = 0.4$ and $\epsilon^{cc'} = 0.2$, for all c and c'.

By comparing the resulting images in Figs. 9(b)–9(d), it is clear that, although improvement is clear in all of them, restoration with the line process produces crisp edges and better definition of the high luminosity objects in the image. Note that, using the cross terms, the color bleeding, that appears in Fig. 9(c), is highly reduced or eliminated.

VII. Conclusion

In this paper, we have used CGMRF prior models in multichannel image restoration problems in order to exchange information among channels. This exchange of information is carried out through the line process with the objective of preserving similar regions in the channels. This approach is different from the classical ones where the exchange of information is performed with the use of the flux of each channel. In order to find the MAP estimates for these problems we have proposed two new methods that can be considered as extensions of the

TABLE II
PSNR Comparisons of Different Restoration Algorithms for the Lena Image.

$PSNR$ (dB)	red	green	blue	mean
Observed	21.079	22.229	22.690	22.000
Guo *et al.*	23.317	23.220	24.334	23.624
Galatsanos *et al.*	23.718	23.555	24.567	23.947
Molina and Mateos	23.791	23.629	24.588	24.003
MMSA	24.692	24.480	24.931	24.701
MMICM	24.554	24.321	24.896	24.590

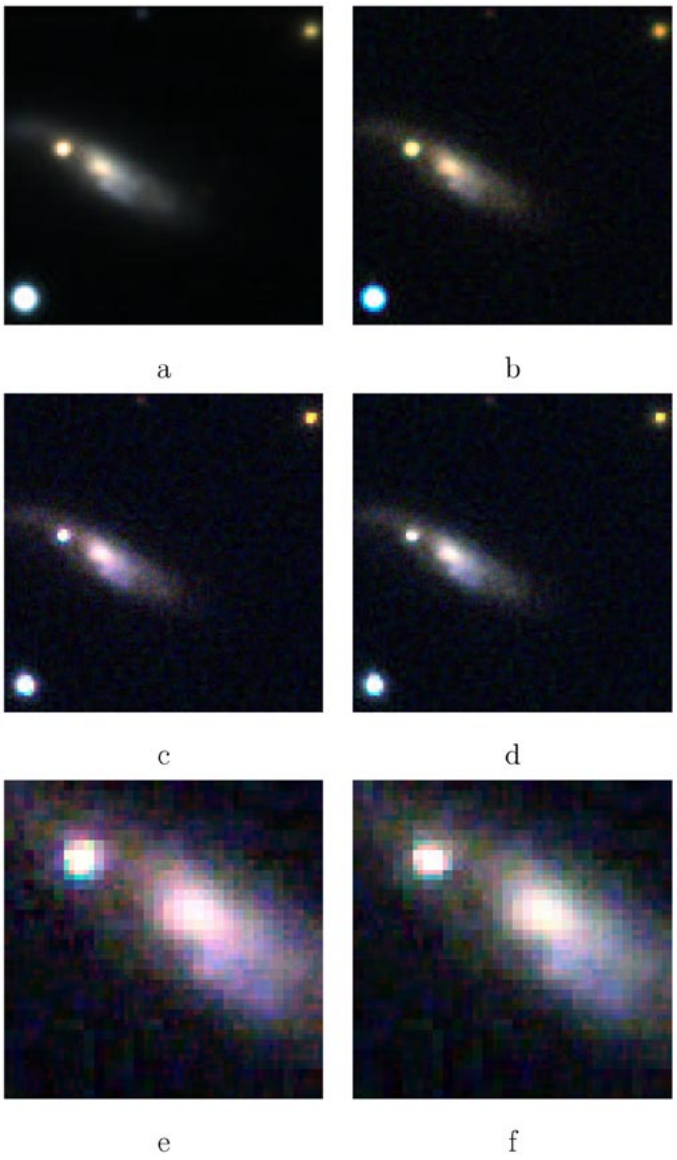


Fig. 9. (a) Degraded image. (b) Restoration with the MMSA method with $\tau^c = 0, \forall c, \epsilon^{cc'} = 0, \forall c, c'$. (c) Restoration with the MMSA method with $\epsilon^{cc'} = 0, \forall c, c'$. (d) Restoration with the MMSA method. (e) An enlarged part of (c). (f) An enlarged part of (d).

classical ICM and SA procedures. The convergence of the algorithms has been theoretically established. The experimental results verify the derived theoretical results and show improvement on the restoration of multichannel images based both on visual and PSNR criteria.

APPENDIX

VIII. CONVERGENCE OF THE SIMPLIFIED PROBLEM

Theorem 2: Let D_p, D_n, H and Ψ be the matrices defined in (22), (5), (3) and (24), respectively. Then the iterative method

$$f^{(k)} = f^{(k-1)} - \Psi[D_p(I-\phi N) + H^t D_n H] f^{(k-1)} + \Psi H^t D_n g$$

is a contraction.

Proof: From the above equation we have

$$f^{(k)} - f^{(k-1)} = [I - \Psi[D_p(I-\phi N) + H^t D_n H]\left(f^{(k-1)} - f^{(k-2)}\right).$$

Hence, we have to study the properties of the matrix

$$\hat{A} = [I - \Psi[D_p(I-\phi N) + H^t D_n H]$$

and show that

$$-I \leq \hat{A} \leq I.$$

It is clear that $\hat{A} \leq I$ since $[D_p(I-\phi N) + H^t D_n H] \geq 0$. To prove that

$$-I \leq I - \Psi[D_p(I-\phi N) + H^t D_n H] \tag{32}$$

or

$$\Psi[D_p(I-\phi N) + H^t D_n H] \leq 2I \tag{33}$$

we use the fact that

$$\begin{aligned} f^t H^t D_n H f &= \sum_{c=1}^{L} \beta^c \left(\sum_{c'=1}^{L} \lambda^{cc'} H^{cc'} f^{c'}\right)^t \\ &\quad \times \left(\sum_{c'=1}^{L} \lambda^{cc'} H^{cc'} f^{c'}\right) \\ &\leq \sum_{c=1}^{L} \beta^c \left(\sum_{c'=1}^{L} \lambda^{cc'} \|H^{cc'} f^{c'}\|^2\right) \\ &\leq \sum_{c=1}^{L} \beta^c \left(\sum_{c'=1}^{L} \lambda^{cc'} \|f^{c'}\|^2\right) \\ &= \sum_{c'=1}^{L} \|f^{c'}\|^2 \left(\sum_{c=1}^{L} \beta^c \lambda^{cc'}\right) \end{aligned}$$

where we have taken into account that the matrix $H^{cc'}$ is a blurring matrix and that, $\forall f^c$ vectors and $\lambda^{cc'} \geq 0, \sum_{c'} \lambda^{cc'} = 1$

$$\|\sum_{c'=1}^{L} \lambda^{cc'} f^{c'}\|^2 \leq \sum_{c'=1}^{L} \lambda^{cc'} \|f^{c'}\|^2$$

due to the convexity of the function x^2 and $\lambda^{cc'} \leq 1, \forall c'$.

So we will have

$$\begin{aligned} & f^t D_p(I-\phi N) f + f^t H^t D_n H f \\ &\leq \sum_{c=1}^{L} 2\alpha^c \|f^c\|^2 \\ &\quad + \sum_{c=1}^{L} \|f^c\|^2 \left(\sum_{c'=1}^{L} \beta^{c'} \lambda^{c'c}\right) \\ &= \sum_{c=1}^{L} \|f^c\|^2 \left(2\alpha^c + \sum_{c'=1}^{L} \beta^{c'} \lambda^{c'c}\right). \end{aligned}$$

Now, if we choose

$$\psi = \frac{1}{\max_c (\alpha^c + \sum_{c'} \beta^{c'} \lambda^{c'c})}$$

the iterative procedure

$$f^{(k)} = f^{(k-1)} - \frac{1}{\max_c (\alpha^c + \sum_{c'} \beta^{c'} \lambda^{c'c})} \times \left\{[D_p(I-\phi N) + H^t D_n H] f^{(k-1)} - H^t D_n g\right\}$$

converges. ■

Note that if H is a block diagonal blurring matrix, that is, $\lambda^{c'c}=0$ if $c\neq c'$ and $\lambda^{cc}=1, \psi I$ can be replaced by a block diagonal matrix $\mathcal{X}$ with diagonal block $\mathcal{X}^c=(1/(\alpha^c+\beta^c))I$.

IX. Convergence of the MMSA Procedure

In this section we shall examine the convergence of the MMSA algorithm (Algorithm 1). It is important to make clear that in this new iterative procedure we simulate the whole multichannel image at the same time using (28) and to simulate $l_{[i,j]}$ we keep using (8) and (9).

Step 4 in algorithm 1 can be written as

$$f^{(t)}=M(l)f^{(t-1)}+Qg+\Theta \tag{34}$$

where Θ is a random column vector with independent components $\mathcal{N}(0,\hat{\sigma}_i^{c^2})$, where $\hat{\sigma}_i^{c^2}$ has been defined in (29).

Following the same steps as in [9] (the steps used in [9] are the same as the ones used in [29]) to prove Theorem 1 we only need to show the following Lemma.

Lemma 1: If $|\phi|<0.25$, then $\forall l$, the matrix $M(l)$ defined in (34) is a contraction.

Proof: First we note that from (34)

$$\begin{aligned}f_i^{c(t)}=f_i^{c(t-1)}-\psi\Bigg[&\alpha^c\phi\sum_{j\text{nhbr}i}\left(f_i^{c(t-1)}-f_j^{c(t-1)}\right)\\&\times\left(1-l_{[i,j]}^c\right)+\alpha^c(1-4\phi)f_i^{c(t-1)}\Bigg]\\&+\psi\left[H^t D_n\left(g-Hf^{(t-1)}\right)\right]_i^c\end{aligned}$$

with $\psi=1/\max(\alpha^c+\sum_{c'}\beta^{c'}\lambda^{c'c})$.

So, $M(l)$ is symmetric and for any column vector $x\in R^{L\times M\times N}$ we have

$$\begin{aligned}&x^tM(l)x\\&=\sum_c\sum_i x_i^{c2}-\psi\Bigg[\sum_c\alpha_c(1-4\phi)\sum_i x_i^{c2}\\&\quad+\sum_c\alpha_c\sum_i\Big(\phi\left(x_i^c-x_{i:+1}^c\right)^2\left(1-l_{[i,i:+1]}^c\right)\\&\quad+\phi\left(x_i^c-x_{i:+2}^c\right)^2\left(1-l_{[i,i:+2]}^c\right)\Big)+x^tH^tD_nHx\Bigg].\end{aligned}$$

We now examine the value of $x^tA(l)x$, where

$$\begin{aligned}x^tA(l)x=\sum_c\alpha_c\sum_i\Big(&\phi\left(x_i^c-x_{i:+1}^c\right)^2\left(1-l_{[i,i:+1]}^c\right)\\&+\phi\left(x_i^c-x_{i:+2}^c\right)^2\left(1-l_{[i,i:+2]}^c\right)+(1-4\phi)x_i^{c2}\Big).\end{aligned}$$

If $-0.25<\phi\leq 0$ we have

$$\begin{aligned}\sum_c 2\alpha_c\|x^c\|^2&>x^tA(l)x\\&\geq\sum_c\alpha_c\sum_i\Big(\phi\left(x_i^c-x_{i:+1}^c\right)^2\\&\quad+\phi\left(x_i^c-x_{i:+2}^c\right)^2+(1-4\phi)x_i^{c2}\Big)\\&=\sum_c\alpha_c x^{ct}(I-\phi N)x^c>0.\end{aligned}$$

If $0<\phi<0.25$ we have

$$\begin{aligned}0&<x^tA(l)x\\&\leq\sum_c\alpha_c\sum_i\Big(\phi\left(x_i^c-x_{i:+1}^c\right)^2\\&\quad+\phi\left(x_i^c-x_{i:+2}^c\right)^2+(1-4\phi)x_i^{c2}\Big)\\&=\sum_c\alpha_c x^{ct}(I-\phi N)x^c\\&<\sum_c 2\alpha_c\|x^c\|^2.\end{aligned}$$

So, if $|\phi|<0.25$

$$x^tM(l)x\leq\|x\|^2-\psi x^tA(l)x<\|x\|^2$$

and, since from Theorem 2, $x^tH^tD_nHx<\sum_c\|x^c\|^2\sum_{c'}\beta^{c'}\lambda^{c'c}$, we have

$$\begin{aligned}x^tM(l)x&\geq\sum_c\left(\|x^c\|^2-\psi\|x^c\|^2\left(2\alpha_c+\sum_{c'}\beta^{c'}\lambda^{c'c}\right)\right)\\&>-\|x\|^2.\end{aligned}$$

Thus, if $|\phi|<0.25$

$$-I<M(l)<I$$

which proves that $M(l)$ is a contraction matrix. ■

Note that, as was the case in Appendix A, if H is a block diagonal blurring matrix, that is, $\lambda^{c'c}=0$ if $c\neq c'$ and $\lambda^{cc}=1, \psi I$ can be replaced by a block diagonal matrix $\mathcal{X}$ with diagonal block $\mathcal{X}^c=(1/(\alpha^c+\beta^c))I$.

References

[1] N. P. Galatsanos and R. T. Chin, "Digital restoration of multichannel images," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, no. 3, pp. 415–421, 1989.

[2] B. R. Hunt and O. Kubler, "Karhunen-Loeve multispectral image restoration. Part 1: Theory," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-32, no. 3, pp. 592–600, 1984.

[3] N. P. Galatsanos, A. K. Katsaggelos, R. T. Chin, and A. D. Hillery, "Least squares restoration of multichannel images," *IEEE Trans. Signal Processing*, vol. 39, pp. 2222–2236, Oct. 1991.

[4] M. G. Kang and A. K. Katsaggelos, "Simultaneous multichannel image restoration and estimation of the regularization parameters," *IEEE Trans. Image Processing*, vol. 6, no. 5, pp. 774–778, 1997.

[5] Y. P. Guo, H. P. Lee, and C. L. Teo, "Multichannel image restoration using an iterative algorithm in space domain," *Image Vis. Comput.*, vol. 14, no. 6, pp. 389–400, 1996.

[6] M. G. Kang, "Generalized multichannel image deconvolution approach and its applications," *Opt. Eng.*, vol. 37, no. 11, pp. 2953–2964, 1998.

[7] R. Molina and J. Mateos, "Multichannel image restoration in Astronomy," *Vistas Astron.*, vol. 41, no. 3, pp. 373–379, 1997.

[8] R. R. Schultz and R. L. Stevenson, "Stochastic modeling and estimation of multispectral image data," *IEEE Trans. Image Processing*, vol. 4, no. 8, pp. 1109–1119, 1995.

[9] R. Molina, A. K. Katsaggelos, J. Mateos, A. Hermoso, and C. A. Segall, "Restoration of severely blurred high range images using stochastic and deterministic relaxation algorithms in compound Gauss Markov random fields," *Pattern Recognit.*, vol. 33, no. 3, pp. 557–571, 2000.

[10] R. Molina, J. Mateos, and A. K. Katsaggelos, "Multichannel image restoration using compound Gauss-Markov random fields," in *Proc. Int. Conf. on Acoustics, Speech, and Signal Processing, (ICASSP 2000)*, vol. 1, Vancouver, BC, Canada, 2000, pp. 141–144.

[11] J. Mateos, R. Molina, and A. K. Katsaggelos, "Color image restoration using compound Gauss-Markov random fields," in *Proc. X Eur. Signal Processing Conf. (EUSIPCO'2000)*, vol. III, Tampere, Finland, 2000, pp. 1341–1344.

[12] P. V. Blomgren, "Total Variation Methods for Restoration of Vector Valued Images," Ph.D. dissertation, Univ. California, 1998.
[13] F. Sroubek and J. Flusser, "An overview of multichannel image restoration techniques," in *Week of Doctoral Students 1999*, J. Safrànkovà, Ed., 1999, pp. 580–585.
[14] E. J. Candes and F. Gou, "New multiscale transforms, minimum total variation synthesis: Applications to edge-preserving image reconstruction," *Signal Process.*, vol. 82, no. 11, pp. 1519–1534, 2002.
[15] J. Portilla and E. P. Simoncelli, "Image restoration using Gaussian scale mixtures in the wavelet domain," in *Proc. Int. Conf. in Image Processing (ICIP 2003)*, 2003.
[16] J. Bescos, I. Glaser, and A. A. Sawchuk, "Restoration of color images degraded by chromatic aberrations," *Appl. Opt.*, vol. 19, no. 22, pp. 3869–3876, 1980.
[17] N. P. Galatsanos and R. T. Chin, "Restoration of color images by multichannel Kalman filtering," *IEEE Trans. Signal Processing*, vol. 39, no. 10, pp. 2237–2252, 1991.
[18] X. Zhang, Y. Obuchi, T. Kambe, N. N. Kubo, and I. Suzuki, "Color imaging for digital cameras with a single CCD sensor," in *Proc. 26th Annu. Conf. IEEE Industrial Electronics Society, IECON 2000*, vol. 3, 2000, pp. 2007–2012.
[19] A. Lípez, R. Molina, A. K. Katsaggelos, and J. Mateos, "SPECT image reconstruction using compound models," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 16, no. 3, pp. 317–330, 2002.
[20] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 9, no. 6, pp. 721–742, 1984.
[21] F. C. Jeng, "Compound Gauss-Markov Random Fields for Image Estimation and Restoration," Ph.D. dissertation, Rensselaer Polytechnic Inst., Troy, NY, 1988.
[22] R. Chellapa, T. Simchony, and Z. Lichtenstein, "Image estimation using 2D noncausal Gauss-Markov random field models," *Digital Image Restor.*, 1991. vol. 23 of *Springer Series in Information Science*.
[23] A. Blake and A. Zisserman, *Visual Reconstruction*. Cambridge, MA: MIT Press, 1987.
[24] B. D. Ripley, *Spatial Statistics*. New York: Wiley, 1981.
[25] P. Charbonnier, L. Blanc-Féraud, G. Aubert, and M. Barlaud, "Deterministic edge-preserving regularization in computed imaging," *IEEE Trans. Image Processing*, vol. 6, no. 2, pp. 298–311, 1997.
[26] R. H. Chan and K. P. Ng, "Conjugate gradient method for Toeplitz systems," *SIAM Rev.*, vol. 38, no. 3, pp. 427–482, 1996.
[27] J. L. Marroquin, *Probabilistic Solution of Inverse Problems*. Cambridge, MA: MIT-AI 860, Artif. Intell. Lab., MIT, 1985.
[28] L. Bedini, M. del Corso, and A. Tonazzini, "A preconditioning technique for edge preserving image restoration," in *Proc. Int. Conf. Information Intelligence and Systems (ISNLS-99)*, 1999, pp. 519–526.
[29] F. C. Jeng and J. W. Woods, "Simulated annealing in compound Gaussian random fields," *IEEE Trans. Inform. Theory*, vol. 36, no. 1, pp. 94–107, 1990.
[30] J. Besag, "On the statistical analysis of dirty pictures," *J. R. Statist. Soc. B*, vol. 48, pp. 259–302, 1986.
[31] L. Younes, "Synchronous Random Fields and Image Restoration," Ecole Normale Supérieure de Cachan, Tech. Rep., CMLA, 1998.
[32] M. R. Banham and A. K. Katsaggelos, "Digital image restoration," *IEEE Signal Processing Mag.*, vol. 14, no. 2, pp. 24–41, 1997.

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).

Javier Mateos was born in Granada, Spain, in 1968. He received the degree in computer science from the University of Granada in 1991. He received the Ph.D. degree in computer science from the same university in July 1998.

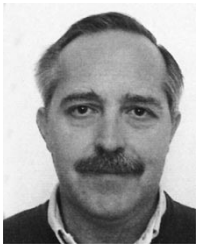
He has been assistant professor at the Department of Computer Science and Artificial Intelligence of the University of Granada from October 1992 to March 2001, when he became a permanent Associate Professor. He is conducting his research on image and video processing, including image restoration, image and video recovery and compression and super-resolution from (compressed) stills and video sequences.

Dr. Mateos is a member of the AERFAI (Asociación Española de Reconocimento de Formas y Análisis de Imágenes) and IAPR (International Association for Pattern Recognition).

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).



Miguel Vega was born in 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 (Department 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.