Stochastic Methods for Joint Regisration, Restoration, and Interpolation of Multiple Under Sampled Images

N.A. Woods, N.P. Galatsanos and A.K. Katsaggelos

12 – 2004

Preprint, no 12 – 9 / 2004

Department of Computer Science
University of Ioannina
45110 Ioannina, Greece

Stochastic Methods for Joint Registration, Restoration, and Interpolation of Multiple Under Sampled Images

by

Nathan A. Woods[1], Nikolas P. Galatsanos[2], and Aggelos K. Katsaggelos[3]*

[1]Binary Machines, Inc.
1320 Tower Rd.
Schaumburg, IL 60173
nathan@BinaryMachinesInc.com

[2]Department of Computer Science
University of Ioannina
PO Box 1186 – GR 45110
Ioannina, Greece
galatsanos@cs.uoi.gr

[3]Department of Electrical and Computer Engineering
Northwestern University
2145 Sheridan Rd.
Evanston, IL 60208
aggk@ece.northwestern.edu

EDICS 2-REST

Abstract

Using a stochastic framework, we propose two algorithms for the problem of obtaining a single high-resolution image from multiple noisy, blurred, and under-sampled images. The first algorithm is based on a Bayesian formulation that is implemented via the expectation maximization (E-M) algorithm, in which the desired high-resolution image is considered to be a hidden variable and is marginalized. The second algorithm is based on a maximum a posteriori (MAP) formulation. In both of our formulations, the registration, noise, and image statistics are unknown parameters. All unknown parameters and the high-resolution image are estimated jointly based on the available observations. We present an efficient implementation of these algorithms in the frequency domain that allows their application to large images. Simulations are presented that test and compare the proposed algorithms.

* Corresponding author.

I. Introduction

The problem of obtaining a high-resolution image from a sequence of aliased low-resolution images, also known as the super-resolution problem, has been a very active area of research over the last several years. Tsai and Huang [1] were the first to demonstrate that unique information in a sequence of translated and aliased images can be exploited to produce an enhanced resolution image. In their early work, the low-resolution images were neither blurred nor noisy, and they assumed that the shifts between the low-resolution stills were known. Kim *et al.* extended this work to include additive noise [2] and blur [3] by formulating the problem as a restoration problem and solving it using a weighted least-squares algorithm. However, estimation of the translational shifts among the low-resolution frames, also referred to as the image registration problem, was not addressed.

In [4], Irani and Peleg suggested a solution to the super-resolution problem based on iterative back-projection adapted from computer-aided tomography. In this technique, an initial guess of the high-resolution image is projected through the observation model, and the high-resolution image is updated according to the error in the obtained low-resolution projections.

Another approach proposed by Stark and Oskoui [5] used a projection onto convex set (POCS) algorithm to reconstruct the high-resolution image. This method was later extended by Tekalp *et al.* [6] to include noise. Patti and Tekalp [7] extended the formulation again to account for time-varying motion blur and various video sampling patterns. Block matching or phase correlation was suggested in [6] and [7] as a means of estimating the registration parameters. Elad and Feuer [8] proposed a hybrid *maximum a posteriori* (MAP)/POCS super-resolution algorithm that combines the ML solution with non-ellipsoidal constraints, assuming that the blur (which may be time and/or space variant), motion parameters, and noise statistics are known.

Stevenson and Shultz [9] presented a Bayesian MAP formulation to the super-resolution problem with an edge-preserving image prior. They addressed the more difficult problem of independent object motion in a video sequence, as opposed to the simple cases of global displacement or rotation. Hierarchical block matching was used to estimate the displacement vector field. To obtain sub-pixel displacement vector estimates, the resolution of the low-resolution images was first increased using the single frame algorithm in [10]. The combination of estimating the motion parameters using a single-frame resolution enhancing technique and block matching (which compares blocks between two frames) was sub-optimal in the sense that it did not exploit all of the observed frames when estimating the motion parameters for an image. Later, Borman and Stevenson [11] extended the same approach to include spatio-temporal priors to improve the robustness of the reconstruction to errors in the motion estimation.

In [12], Hardie *et al.* presented a solution to the super-resolution problem wherein the registration and restoration was performed jointly using a MAP formulation. The registration parameters were estimated by finding the parameters that minimized the squared error of the projection of the high-resolution estimate through the observation model and the observations. Since this minimization was intractable, the authors suggested a somewhat *ad-hoc* pyramidal search strategy to numerically find the registration parameters.

In [13], [14], and [15], Tom *et al.* formulated the super-resolution problem as a restoration problem. In particular, in [15] a solution was presented utilizing the E-M algorithm, in which the registration parameters and noise variance were estimated jointly. Our proposed Bayesian solution, is also implemented using the E-M algorithm, and was originally inspired by the approach in [15]. However, our proposed image model differs significantly from the model employed in [15] (leading to a different solution), and our formulation admits an efficient solution in the frequency domain that bypasses the need to introduce the approximations adopted by Tom *et al.* in their solution. More specifically, in [15], the high-resolution image was first decimated, then blurred and shifted. We believe that our imaging model is more realistic, with blurring and shifting preceding decimation. Also, due to their formulation, the authors in [15] chose to truncate the shift and blur operators for practical run-time considerations, which reduced the quality of the reconstructions.

More recently, Lee and Kang [18] proposed a regularized least-squares solution solved via gradient decent. The regularization parameter was calculated from the data using a functional that accounts for registration parameter error under the assumption that the registration error manifests itself as additive white Gaussian noise in the observations. Their technique treats the registration parameters as fixed during restoration, requiring a separate registration step during pre-processing.

Nguyen *et al.* [16] presented a solution based on Tikhonov regularization of a least-squares formulation. The regularization parameter was obtained using generalized cross-validation. They solved the system using the conjugate gradient method, preconditioning the system to accelerate convergence. Registration parameters were estimated during a pre-processing step under the assumption of smooth translational motion using the simple Taylor series approximation proposed in [4]. For recent surveys of super-resolution techniques, the reader is referred to [19], [20], and [21].

In this paper, we solve the super-resolution problem in which the registration parameters, noise, and image statistics are unknown. We restrict our model to the special case of global translational motion among low resolution frames, common space-invariant blur, additive white noise, and common integer decimation. In our solution, all unknown parameters are estimated jointly along with the restored high-resolution image using all of the available data. We present two solutions derived within a stochastic framework. The first solution is based on a Bayesian formulation in which the desired high-resolution image is modeled as a latent, or hidden, variable and is marginalized over the image prior. The second

solution is based on the popular maximum *a posteriori* (MAP) formulation in which the desired high-resolution image is taken as the mode of the posterior density function. Both of our proposed solutions can be efficiently implemented in the frequency domain and therefore can be applied to large images.

One of the goals of this paper is to compare the performance of these two stochastic approaches and to discuss their relative merits. By comparing both approaches on equal footing using the same prior image model, we attempt to determine whether or not the Bayesian treatment of the unknown high-resolution image provides an estimation advantage. Another goal is to present a case for the joint estimation of the high-resolution image and the unknown model parameters. This is in contrast to traditional approaches, which, if they consider the parameter estimation problem at all, consider the estimation of only one or two model parameters (for example, the registration parameters or the noise variance). Such approaches treat the remaining model parameters as known quantities during restoration and therefore assume that they have been estimated independently from the high-resolution image during a pre-processing step.

The rest of the paper is organized as follows. The imaging model is described in Section II. In Section III, we present the stochastic estimation framework used in the rest of the paper. In Section IV, we derive the Bayesian solution to the super-resolution problem, based on the E-M algorithm, in the spatial domain. The efficient calculation of the E-M iterations in the frequency domain is described in Section V, supplemented by detailed calculations in Appendix A. In Section VI, we derive our MAP solution to the super-resolution problem. Experimental results comparing the approaches are presented in Section VII, and Section VIII concludes the paper.

II. Image Model

In this section, we present a generative, stochastic imaging model. We first approximate the underlying, or actual, scene as a discrete, high-resolution image of size $M_H \times N_H$. Let $\boldsymbol{x}$ be an $M_H N_H \times 1$ vector containing the discrete intensity values, arranged lexicographically, of the underlying scene. We model the intensity values as samples from a Gaussian random process given by,

$$p\left(\mathbf{x} \mid \alpha^2\right) \propto \left(\alpha^2\right)^{\frac{M_H N_H - 1}{2}} \exp\left(-\alpha^2 \frac{\mathbf{x}^T \mathbf{Q}^T \mathbf{Q} \mathbf{x}}{2}\right) \tag{1}$$

where $\mathbf{Q}$ is the Laplacian operator matrix, α^2 an unknown covariance parameter and $(\,)^T$ denotes matrix transposition. For this model the covariance matrix is given by $\mathbf{\Lambda}_x = E\left[\mathbf{x}\mathbf{x}^T\right] = (\alpha^2 \mathbf{Q}^T \mathbf{Q})^{-1}$. This image model is termed simultaneously autoregressive (SAR) and has been used successfully in image restoration [26], [27]. The observations, $\mathbf{y}_i$, are related to the underlying image $\mathbf{x}$ using the linear model

$$\mathbf{y}_i = \mathbf{B}_i \mathbf{x} + \mathbf{n}_i \quad \text{for } i = 0,1,\ldots,P-1 \tag{2}$$

where $\mathbf{y}_i$ $i = 0, 1, \ldots, P-1$, is an $M_L N_L$ x 1 vector containing the discrete intensity values of the i^{th} observed image arranged lexicographically, $\mathbf{B}_i$ is a linear degradation operator of size $M_L N_L$ x $M_H N_H$, and $\mathbf{n}_i$ is an $M_L N_L$ x 1 noise vector. The noise is modeled as a realization of a white Gaussian random process with zero mean

$$\mathbf{n}_i \sim N(\mathbf{0}, \Lambda_n) \tag{3}$$

where $\Lambda_n = \sigma^2 \mathbf{I}$, σ^2 is the unknown noise variance, and $\mathbf{I}$ is the identity matrix. We assume that the noise is uncorrelated with the underlying scene.

The linear degradation operator, $\mathbf{B}_i$, blurs, spatially shifts, and decimates $\mathbf{x}$. Therefore, we decompose $\mathbf{B}_i$ as follows,

$$\mathbf{B}_i = \mathbf{D}_i \mathbf{S}(\boldsymbol{\delta}_i) \mathbf{H}_i \tag{4}$$

where $\mathbf{D}_i$ is the i^{th} decimation operator of size $M_L N_L$ x $M_H N_H$, $\mathbf{S}$ is the shift operator of size $M_H N_H$ x $M_H N_H$ parameterized by the i^{th} shift vector $\boldsymbol{\delta}_i$, and $\mathbf{H}_i$ is the i^{th} blur operator also of size $M_H N_H$ x $M_H N_H$. We define the i^{th} shift vector to be a 1 x 2 vector, $\boldsymbol{\delta}_i = \begin{bmatrix} \delta_{i,x} & \delta_{i,y} \end{bmatrix}$, containing the horizontal and vertical shifts, measured in pixels of $\mathbf{x}$, of the i^{th} image relative to the 0^{th} image.

In this paper, we restrict ourselves to integer vertical and horizontal decimation factors, d_y and d_x, given by

$$\frac{M_H}{M_L} = d_y \quad \text{and} \quad \frac{N_H}{N_L} = d_x \tag{5}$$

With this restriction, the decimation matrix $\mathbf{D}_i$ can be obtained from an $M_H N_H$ x $M_H N_H$ identity matrix by copying from the identity matrix to the $\mathbf{D}_i$ matrix only those rows that satisfy both of the following constraints,

$$\begin{aligned} & j \bmod d_x = 0 \\ & \lfloor j / N_H \rfloor \bmod d_y = 0 \quad \text{for } j = 0,1,\ldots,M_H \cdot N_H \end{aligned} \tag{6}$$

where j is the row index of the identity matrix, mod is the integer modulus operator, and $\lfloor x \rfloor$ rounds x to the nearest integer towards minus infinity.

The shift operator, $\mathbf{S}(\boldsymbol{\delta}_i)$, is the Shannon 2-D interpolation operator which is shift invariant [25]. The 2-D impulse response of the shift operator is given by

$$h_{shift}(m,n;\boldsymbol{\delta}_i)=\frac{\sin\left(\pi(m-\delta_{i,y})\right)}{\pi(m-\delta_{i,y})}\frac{\sin\left(\pi(n-\delta_{i,x})\right)}{\pi(n-\delta_{i,x})} \tag{7}$$

for $m=0,1,\ldots,M_H-1$, and $n=0,1,\ldots N_H-1$. In this paper, we shall assume that $\mathbf{S}(\boldsymbol{\delta}_i)$ and $\mathbf{H}$ can be approximated as doubly circulant matrices, which commute since they have the same eigenvectors [17].

Note that our image model does not account for global rotation since, as we describe later, in order to improve the computational efficiency of our approach we employ the Discrete Fourier Transform (DFT) to diagonalize circulant matrices. Since rotation is spatially variant, the DFT provides no computational advantage here. However, it may be possible to estimate global rotation between observed frames using our approaches in the spatial domain, although at great additional computational cost. We defer to future research the investigation of computationally efficient methods to handle global rotation.

Now, let us represent the entire sequence of observed images as a single PM_LN_L x 1 vector, given by

$$\mathbf{y}=\begin{bmatrix}\mathbf{y}_0^T & \mathbf{y}_1^T & \cdots & \mathbf{y}_{P-1}^T\end{bmatrix}^T \tag{8}$$

Likewise, the noise vector and degradation matrix can be written as

$$\mathbf{n}=\begin{bmatrix}\mathbf{n}_0^T & \mathbf{n}_1^T & \cdots & \mathbf{n}_{P-1}^T\end{bmatrix}^T \tag{9}$$

$$\mathbf{B}=\text{blockdiag}\left(\mathbf{B}_i\right) \quad i=0,1,\ldots,P-1 \tag{10}$$

where blockdiag( ) returns a block diagonal matrix from the supplied component matrices. Written compactly, the model is now

$$\mathbf{y}=\mathbf{B}\mathbf{x}+\mathbf{n} \tag{11}$$

We further restrict the model to the case where all of the observations are degraded by the same blur and the same decimation, so that $\mathbf{D}_i=\mathbf{D}_j$ and $\mathbf{H}_i=\mathbf{H}_j$, $\forall i,j$. With this restriction, substituting (4) into (10) and (11) yields

$$\mathbf{y}=\mathbf{D}\mathbf{S}\mathbf{H}\mathbf{x}+\mathbf{n} \tag{12}$$

where $\mathbf{D}$ is the PM_LN_L x M_HN_H block diagonal decimation matrix,

$$\mathbf{D}=\text{blockdiag}\left(\mathbf{D}_i\right) \quad i=0,1,\ldots,P-1 \tag{13}$$

$\mathbf{S}$ is a PM_HN_H x M_HN_H matrix containing the shift operators stacked on top of one another,

$$\mathbf{S} = \begin{bmatrix} \mathbf{S}(\boldsymbol{\delta}_0) \\ \mathbf{S}(\boldsymbol{\delta}_1) \\ \vdots \\ \mathbf{S}(\boldsymbol{\delta}_{P-1}) \end{bmatrix} \tag{14}$$

and **H** is a common spatially invariant $M_H N_H$ x $M_H N_H$ blur matrix..

Equation (12) is the spatial-domain version of the model. It relates a discrete, high-resolution image to P low-resolution blurred, shifted, and noisy images. In addition, each observed image, $\mathbf{y}_i$, can be degraded by aliasing due to decimation. In this paper, we produce an estimate of **x** from the observation vector **y,** given the decimation factors d_x and d_y, the blur matrix **H**, and the Laplacian matrix **Q**. All other model parameters are considered unknown, including the covariance parameter α, the noise variance, σ^2, and the P-1 shift vectors $\boldsymbol{\delta}_i$.

Estimating **x** from **y** requires the inversion of **B,** which is a well-known ill-posed problem. To ameliorate the effects of the noise and the ill-posedness of the problem, we regularize the solution by choosing a particular **Q** that determines (to within a scalar) the covariance in our statistical model of the image. In other words, we regularize the solution by exploiting prior knowledge of the second order statistics of **x**.

III. Stochastic Frameworks for Super-resolution

In this section, we present two stochastic estimation frameworks, the Bayesian and the maximum a posteriori (MAP) frameworks, and discuss their relative merits.

A. Bayesian Framework

Perhaps the most popular approach for parameter estimation is maximum likelihood. In this approach, the most probable parameters that gave rise to the data are estimated. That is, the parameters of our image model are estimated by

$$\left(\alpha^2, \sigma^2, \boldsymbol{\delta}\right)_{ML} = \arg\max_{\alpha^2, \sigma^2, \boldsymbol{\delta}} p\left(\mathbf{y} \mid \alpha^2, \sigma^2, \boldsymbol{\delta}\right) \tag{15}$$

However, in many cases the direct assessment, in closed form, of the likelihood function that captures the statistical relation between the unknown parameters and the observations, $p\left(\mathbf{y} \mid \alpha^2, \sigma^2, \boldsymbol{\delta}\right)$, is either difficult or impossible.

The computation of this likelihood is greatly facilitated by the introduction of "hidden" variables that are subsequently integrated out, or marginalized. These variables act as links that connect the observations to the unknown parameters via Bayes' law. The choice of hidden variables is problem dependent. However, in the super-resolution problem the choice is clear; the high-resolution image, **x**, is the appropriate hidden

variable. With this choice, both $p(\mathbf{y}|\mathbf{x},\sigma^2,\boldsymbol{\delta})$ and $p(\mathbf{x}|\alpha^2)$ are known and the marginal likelihood is obtained from the integral

$$p(\mathbf{y}|\alpha^2,\sigma^2,\boldsymbol{\delta}) = \int p(\mathbf{y},\mathbf{x}|\alpha^2,\sigma^2,\boldsymbol{\delta})d\mathbf{x} = \int p(\mathbf{y}|\mathbf{x},\sigma^2,\boldsymbol{\delta})p(\mathbf{x}|\alpha^2)d\mathbf{x}. \quad (16)$$

Despite the introduction of the hidden variables, in many cases the above integral is either impossible to compute in closed form or yields a marginal likelihood that is complicated so that its direct maximization is difficult. In such cases, iterative algorithms can be used to maximize either the marginal likelihood or a lower bound of it. Algorithms that do the latter are the Expectation Maximization (E-M) algorithms, and algorithms that do the former are the so-called variational E-M algorithms (see, for example [31], [32], and [33]). Both of these algorithms are guaranteed to converge—to a local maximum of the marginal likelihood in the case of the E-M, and to a local maximum of a strict lower bound on the marginal likelihood in the case of the variational E-M.

Once the marginal likelihood and the ML estimates of the parameters are obtained, the posterior with respect to the hidden variables can be computed according to

$$p(\mathbf{x}|\mathbf{y},\alpha^2,\sigma^2,\boldsymbol{\delta}) = \frac{p(\mathbf{y}|\mathbf{x},\sigma^2,\boldsymbol{\delta})p(\mathbf{x}|\alpha^2)}{p(\mathbf{y}|\alpha^2,\sigma^2,\boldsymbol{\delta})} \quad (17)$$

If the posterior is available, its mean can be used as an estimate of the hidden variables. One pleasing side effect of the standard E-M algorithm and the variational E-M algorithm is that the mean of the hidden variables is calculated during the E-step of these algorithms [22],[33].

More specifically, using the model in Section II, we have

$$p(\mathbf{y}|\mathbf{x},\sigma^2,\boldsymbol{\delta}) = \prod_{i=0}^{P-1} p(\mathbf{y}_i|\mathbf{x},\sigma^2,\boldsymbol{\delta}_i) \text{ with } p(\mathbf{y}_i|\mathbf{x},\sigma^2,\boldsymbol{\delta}_i) \propto (\sigma^2)^{-\frac{M_L N_L}{2}} \exp\left(-\frac{1}{2\sigma^2}\|\mathbf{y}_i - \mathbf{B}_i\mathbf{x}\|^2\right) \quad (18)$$

Thus, the Bayesian integral in (16) is

$$p(\mathbf{y}|\alpha^2,\sigma^2,\boldsymbol{\delta}) \propto \alpha^{M_H N_H - 1}\sigma^{-PM_L N_L} \int \exp\left(-\frac{1}{2\sigma^2}\sum_{i=0}^{P-1}\|\mathbf{y}_i - \mathbf{B}_i\mathbf{x}\|^2 - \frac{1}{2}\alpha^2\mathbf{x}^T\mathbf{Q}^T\mathbf{Q}\mathbf{x}\right)d\mathbf{x} \quad (19)$$

Completing the square in the exponential of (19) and using the multidimensional Gaussian integral one can easily show that

$$p(\mathbf{y}|\alpha^2,\sigma^2,\boldsymbol{\delta}) \propto (\alpha^2)^{\frac{M_H N_H - 1}{2}} (\sigma^2)^{-\frac{PM_L N_L}{2}} [\det\mathbf{C}]^{-\frac{1}{2}} \exp\left\{-\frac{1}{2\sigma^2}\left[\mathbf{y}^T\left(\mathbf{I} - \frac{\mathbf{B}\mathbf{C}^{-1}\mathbf{B}^T}{\sigma^2}\right)\mathbf{y}\right]\right\} \quad (20)$$

where $\mathbf{C} = \frac{\mathbf{B}^T\mathbf{B}}{\sigma^2} + \alpha^2\mathbf{Q}^T\mathbf{Q}$, and det( ) denotes the determinant of a matrix. Clearly, the marginal likelihood above is difficult to maximize with respect to the parameters α^2, σ^2, and $\boldsymbol{\delta}$. It is hard to obtain derivatives for these parameters and the parameters α^2 and σ^2 have to be positive. This is the main motivation for resorting to the E-M algorithm in this paper.

B. MAP Framework

The MAP estimates of the parameters are obtained by maximizing

$$\left(\alpha^2,\sigma^2,\boldsymbol{\delta}\right)_{\text{MAP}} = \arg\max_{\alpha^2,\sigma^2,\boldsymbol{\delta}} \prod_{i=0}^{P-1} p\left(\mathbf{y}_i \mid \mathbf{x}^*,\sigma^2,\boldsymbol{\delta}_i\right) p\left(\mathbf{x}^* \mid \alpha^2\right) \tag{21}$$

with

$$\mathbf{x}^* = \arg\max_{\mathbf{x}} p\left(\mathbf{x} \mid \mathbf{y},\alpha^2,\sigma^2,\boldsymbol{\delta}\right) = \arg\max_{\mathbf{x}} \prod_{i=0}^{P-1} p\left(\mathbf{y}_i \mid \mathbf{x},\sigma^2,\boldsymbol{\delta}_i\right) p\left(\mathbf{x} \mid \alpha^2\right) \tag{22}$$

where we have used in (22) the fact that in Bayes' law for $p(\mathbf{x}|\mathbf{y})$ the denominator is independent of $\mathbf{x}$. Clearly, if both α^2, σ^2, $\boldsymbol{\delta}$, and $\mathbf{x}$ have to be estimated, we have to iterate between (21) and (22). However, unlike the E-M, there is no proof in the general MAP case (e.g. for any $p(\mathbf{y},\mathbf{x},\alpha^2,\sigma^2,\boldsymbol{\delta})$) that guarantees convergence of this iteration. Instead, this burden is placed on the user who must prove this on a case-by-case basis. For example, if it can be shown that the joint density function $p\left(\mathbf{y},\mathbf{x},\alpha^2,\sigma^2,\boldsymbol{\delta}\right) = p\left(\mathbf{y} \mid \mathbf{x},\sigma^2,\boldsymbol{\delta}\right) p\left(\mathbf{x} \mid \alpha^2\right)$ is strictly convex with respect to α^2, σ^2, $\boldsymbol{\delta}$, and $\mathbf{x}$, then convergence follows from the properties of cyclic coordinate decent.

C. Comparison of the Bayesian and MAP Frameworks

Comparing the Bayesian and MAP approaches, we see that for the former in order to obtain the estimates of the unknown parameters, α^2, σ^2, and $\boldsymbol{\delta}$, we marginalize the hidden variables, while for the latter we use the mode of the hidden variables. We also see that to estimate the hidden variables, in the Bayesian approach we use the mean of the posterior while in the MAP approach we use the mode of the posterior. In principle, the Bayesian approach offers the advantage of more reliable parameter estimates since all of the information that is known about the hidden variables is incorporated into the estimation process. In other words, the mode of the posterior of the hidden variables may not be representative of the density function[31]. The main potential disadvantage of the Bayesian framework is that computing or even approximating the Bayesian integral in (16) is much more difficult than computing the mode of the posterior in the MAP approach for complex prior models. Finally, we note that despite the fact that both the

Bayesian and MAP methodologies employ Bayes' theorem, there is a clear distinction between them. Thus, the terms Bayesian and MAP estimation are not identical and should not be confused.

IV. Bayesian Super-resolution Using the E-M Algorithm

In this section, we derive a Bayesian algorithm for super-resolution, based on the E-M-algorithm, in the spatial domain. In this paper, we use the maximum-likelihood (ML) criterion to find the best estimate of the parameters of the image model. Direct maximization of the likelihood function is difficult, however, due to the high non-linearity of the likelihood function with respect to those parameters [22]. Instead, an iterative approach called the Expectation-Maximization (E-M) algorithm, originally proposed by Dempster *et al.* [23], is employed to find the ML estimate of the model parameters and **x**.

In the E-M approach, instead of maximizing the likelihood of the observations with respect to the model parameters, one maximizes the expectation of the complete data conditioned on the incomplete data. The names incomplete and complete data suggest a many-to-one mapping from the complete data to the incomplete data. The incomplete data is the set of observations. The user must choose the complete data, denoted herein as **z**, properly. Let us denote by θ the set of the unknown parameters and by $f_z(\mathbf{z};\theta)$ the probability density function (PDF) of the complete data. In the E-step of the E-M algorithm, the conditional expectation of $\log f_z(\mathbf{z};\theta)$, conditioned upon the observed data **y** and the current estimate of the model parameters, is computed. In the M-step, this expectation is maximized. In mathematical form, the E-M is given by the alternate computation of

$$Q\left(\theta;\theta^{(p)}\right) = E\left[\log f_z(\mathbf{z};\theta)\,|\,\mathbf{y};\theta^{(p)}\right] \tag{23}$$

where $E[\mathbf{a}\,|\,\mathbf{b}\,]$ denotes expectation of **a** with respect to **b**, and

$$\theta^{(p+1)} = \arg\left\{\max_{\theta} Q(\theta;\theta^{(p)})\right\} \tag{24}$$

where $\theta^{(p)}$ is the estimate of θ at the p-th iteration. It was shown in [22] and [34] that choosing $\mathbf{z} = \begin{bmatrix}\mathbf{x}\\ \mathbf{y}\end{bmatrix}$ as the complete data permits the simultaneous identification of the system matrix (i.e., **B** in our model) and the image parameters and the restoration of the image. With this choice of complete data the negative of the conditional log likelihood, dropping the constant term, is equal to [22],[34]

$$\begin{aligned} Q'(\theta;\theta^{(p)}) = {} & \log\left|\mathbf{\Lambda}_x\right| + \log\left|\mathbf{\Lambda}_n\right| + \mathrm{tr}\left[\left(\mathbf{\Lambda}_x^{-1} + \mathbf{B}^H\mathbf{\Lambda}_n^{-1}\mathbf{B}\right)\mathbf{\Lambda}_{x|y}^{(p)}\right] \\ & + \boldsymbol{\mu}_{x|y}^{H\,(p)}\left(\mathbf{\Lambda}_x^{-1} + \mathbf{B}^H\mathbf{\Lambda}_n^{-1}\mathbf{B}\right)\boldsymbol{\mu}_{x|y}^{(p)} - 2\,\mathrm{Re}\left\{\mathbf{y}^H\mathbf{\Lambda}_n^{-1}\mathbf{B}\boldsymbol{\mu}_{x|y}^{(p)}\right\} + \mathbf{y}^H\mathbf{\Lambda}_n^{-1}\mathbf{y} \end{aligned} \tag{25}$$

where $(\cdot)^H$ denotes the Hermitian (complex conjugate transpose) of a matrix, $\boldsymbol{\mu}_{x|y} = E[\mathbf{x}|\mathbf{y}]$, and $\boldsymbol{\Lambda}_{x|y} = E[\mathbf{x}\mathbf{x}^T|\mathbf{y}]$. It is well-known for random vectors related by $\mathbf{y}=\mathbf{B}\mathbf{x}+\mathbf{n}$, where $\Lambda_x = E[\mathbf{x}\mathbf{x}^T]$ and $\mathbf{n} \sim N(\mathbf{0}, \Lambda_n)$, that

$$\boldsymbol{\mu}_{x|y} = \boldsymbol{\Lambda}_x \mathbf{B}^H \left(\mathbf{B}\boldsymbol{\Lambda}_x \mathbf{B}^H + \boldsymbol{\Lambda}_n\right)^{-1} \mathbf{y} \tag{26}$$

$$\boldsymbol{\Lambda}_{x|y} = \boldsymbol{\Lambda}_x - \boldsymbol{\Lambda}_x \mathbf{B}^H \left(\mathbf{B}\boldsymbol{\Lambda}_x \mathbf{B}^H + \boldsymbol{\Lambda}_n\right)^{-1} \mathbf{B}\boldsymbol{\Lambda}_x \tag{27}$$

Equations (26), and (27) constitute the E-step of the E-M algorithm in the spatial domain. However, computation of the E-step in the spatial domain is impractical due to the matrix inversion required in the updates of the conditional mean in (26) and conditional covariance in (27). As will be shown in the next section, by transforming the E-M iterations to the frequency domain, the matrix inversion becomes manageable.

V. Efficient Computation of the E-M Iterations in the Frequency Domain

The computation of (26) and (27) requires the inversion of the matrix $\mathbf{B}\boldsymbol{\Lambda}_x\mathbf{B}^H + \boldsymbol{\Lambda}_n$, which for $P = 4$ and $M_L = N_L = 256$ is of size 262,144 x 262,144. Direct inversion of a matrix of this size is impractical. This matrix due to the decimation matrix $\mathbf{D}$ is not doubly circulant; thus, unlike the classical restoration problem, this matrix cannot be diagonalized by the 2-D DFT [24].

Nonetheless, let us define an interlaced observation vector, $\mathbf{Y}'$, in the frequency domain as follows:

$$\mathbf{Y}' = \left[\mathbf{Y}''(0) \quad \mathbf{Y}''(1) \quad \cdots \quad \mathbf{Y}''(M_L N_L - 1)\right]^T \tag{28}$$

where $\mathbf{Y}''(n) = \left[\mathbf{Y}_0(n) \quad \mathbf{Y}_1(n) \quad \cdots \quad \mathbf{Y}_{P-1}(n)\right]^T$ and $\mathbf{Y}_i(n)$ is the n^{th} element of the 2-D DFT of the intensity values of $\mathbf{y}_i$ arranged lexicographically. With this definition, it is shown in Appendix A that the conditional mean and conditional covariance are given in the frequency domain respectively by

$$\mathbf{M}_{x|y} = d_x d_y \mathbf{U}\mathbf{V}^{-1}\mathbf{Y}' \tag{29}$$

$$\mathbf{S}_{x|y} = \mathbf{S}_x - \mathbf{U}\mathbf{V}^{-1}\mathbf{U}^{\mathrm{H}} \tag{30}$$

where $\mathbf{V}$ and $\mathbf{U}$ are defined in Appendix A, and $\mathbf{V}$ is a *block diagonal matrix* composed of P x P submatrices. This decoupling of the system is the key to the frequency domain approach, as clearly, the system in (29) is far easier to compute than the equivalent system in (26). Equations (29) and (30) constitute the E-step of the E-M algorithm in the frequency domain.

We are now ready to compute the M-step of the E-M algorithm in the frequency domain. In the M-step, we must minimize the likelihood function given by (25) with respect to the unknown parameters, α^2, σ^2, and δ. This is conveniently computed in the frequency domain. After significant algebra, the frequency domain equivalent of (25) is given by

$$\begin{aligned} Q''\left(\theta;\theta^{(p)}\right) = {} & pM_L N_L \log\sigma^2 + \\ & \sum_{m=1}^{M_H N_H - 1}\left\{\log\left(\left[\mathbf{S}_x\right]_{m,m}\right) + \frac{\left[\mathbf{S}_{x|y}\right]_{m,m} + 1/M_H N_H \left|\mathbf{M}_{x|y}(m)\right|^2}{\left[\mathbf{S}_x\right]_{m,m}}\right\} + \\ & \frac{1}{d_x d_y \sigma^2}\sum_{l=0}^{d_y-1}\sum_{j=0}^{d_y-1}\sum_{k=0}^{M_L-1}\sum_{n=0}^{d_x-1}\sum_{m=0}^{d_x-1}\sum_{x=0}^{N_L-1}\left\{\frac{1}{M_H N_H}\mathbf{M}_{x|y}(u)\mathbf{M}^*_{x|y}(v) + \right. \\ & \left.\left[\mathbf{H}_D\right]_{u,u}\left[\mathbf{H}^*_D\right]_{v,v}\sum_{i=0}^{P-1}\left[\mathbf{S}_D\left(\boldsymbol{\delta}_i\right)\right]_{u,u}\left[\mathbf{S}^*_D\left(\boldsymbol{\delta}_i\right)\right]_{v,v}\right\} - \\ & \frac{2}{M_H N_H \sigma^2}\mathrm{Re}\left\{\sum_{j=0}^{d_y-1}\sum_{k=0}^{M_L-1}\sum_{m=0}^{d_x-1}\sum_{x=0}^{N_L-1}\mathbf{M}_{x|y}(u)\left[\mathbf{H}_D\right]_{u,u}\sum_{i=0}^{P-1}\left[\mathbf{S}_D\left(\boldsymbol{\delta}_i\right)\right]_{u,u}\cdot\right. \\ & \left.\mathbf{Y}^*_i\left(kN_L + x\right)\right\} + \frac{1}{M_L N_L \sigma^2}\sum_{m=0}^{pM_L N_L}\left|\mathbf{Y}(m)\right|^2 \end{aligned} \tag{31}$$

where $\left[\mathbf{A}\right]_{m,n}$ selects the $(m,n)^{\text{th}}$ element of matrix $\mathbf{A}$; $(\cdot)^*$ denotes complex conjugate; $\mathbf{H}_D$, $\mathbf{S}_D\left(\boldsymbol{\delta}_i\right)$, and $\mathbf{S}_X$ are diagonal matrices and are the frequency domain equivalent of $\mathbf{H}$, $\mathbf{S}\left(\boldsymbol{\delta}_i\right)$, and $\boldsymbol{\Lambda}_x$ as defined in Appendix A in equations (55), (56), and (66), respectively; and for notational convenience we have defined as a function of the summation indices in (31)

$$u = (jM_L + k)N_H + mN_L + x \tag{32}$$

$$v = (lM_L + k)N_H + nN_L + x \tag{33}$$

To minimize (31) with respect to the unknown covariance parameter, α^2, we substitute (66) into (31), differentiate the result with respect to α^2, set it equal to 0, and solve for α^2. The result of this procedure yields the following update equation for α^2:

$$\alpha^{2(p+1)} = \frac{M_H N_H - 1}{\sum_{m=1}^{M_H N_H - 1}\left|\left[\tilde{\mathbf{Q}}\right]_{m,m}\right|^2\left(\left[\mathbf{S}^{(p)}_{x|y}\right]_{m,m} + \frac{1}{M_H N_H\left|\mathbf{M}^{(p)}_{x|y}(m)\right|^2}\right)} \tag{34}$$

where $\tilde{\mathbf{Q}}$ is the diagonal frequency domain version of the matrix $\mathbf{Q}$. Repeating the same procedure with respect to σ^2 yields the following update equation for the noise variance:

$$\sigma^{2(p+1)} = \frac{1}{pM_H N_H} \sum_{l=0}^{d_x-1} \sum_{j=0}^{d_y-1} \sum_{k=0}^{M_L-1} \sum_{n=0}^{d_x-1} \sum_{m=0}^{d_y-1} \sum_{x=0}^{N_L-1} \left\{ \left[\frac{1}{M_H N_H} \mathbf{M}_{x|y}^{(p)}(u) \cdot \mathbf{M}_{x|y}^{*(p)}(v) + \left[\mathbf{S}_{x|y}^{(p)} \right]_{u,v} \right] \left[\mathbf{H}_D \right]_{u,u} \left[\mathbf{H}_D^* \right]_{v,v} \cdot \sum_{i=0}^{P-1} \left[\mathbf{S}_D \left(\boldsymbol{\delta}_i^{(p)} \right) \right]_{u,u} \left[\mathbf{S}_D^* \left(\boldsymbol{\delta}_i^{(p)} \right) \right]_{v,v} \right\} - \frac{2 d_x d_y}{P (M_H N_H)^2 \sigma^2} \cdot \operatorname{Re} \left\{ \sum_{j=0}^{d_y-1} \sum_{k=0}^{M_L-1} \sum_{m=0}^{d_x-1} \sum_{x=0}^{N_L-1} \mathbf{M}_{x|y}^{(p)}(u) \left[\mathbf{H}_D \right]_{u,u} \sum_{i=0}^{P-1} \left[\mathbf{S}_D \left(\boldsymbol{\delta}_i^{(p)} \right) \right]_{u,u} \mathbf{Y}_i^* \left(kN_L + x \right) \right\} + \frac{1}{P} \left(\frac{d_x d_y}{M_H N_H} \right)^2 \sum_{m=0}^{pM_L N_L} \left| \mathbf{Y}(m) \right|^2 \tag{35}$$

If we repeat the procedure for $\boldsymbol{\delta}$, we find that the equation expressing the root of the partial derivative is a transcendental form. Therefore, we must resort to numerical methods to minimize (31) with respect to $\boldsymbol{\delta}$. Gradient decent methods can be employed here since closed-form expressions exist for the first partial derivatives. However, in this paper, we employed direction set methods [29] to find $\boldsymbol{\delta}$.

VI. MAP Super-resolution

Our second proposed approach is based on the joint MAP estimation of the unknown parameters and the high-resolution image. In this section, we derive our MAP algorithm. From Bayes' law, the posterior required for the MAP estimate is

$$p\left(\mathbf{x} \mid \mathbf{y}, \sigma^2, \alpha^2, \boldsymbol{\delta}\right) = \frac{p\left(\mathbf{y} \mid \mathbf{x}, \sigma^2, \delta\right) p\left(\mathbf{x} \mid \alpha^2\right)}{p\left(\mathbf{y}, \sigma^2, \alpha^2, \boldsymbol{\delta}\right)} \tag{36}$$

where we assume $p\left(\mathbf{x} \mid \alpha^2\right)$ is distributed according to (1), and $p\left(\mathbf{y} \mid \mathbf{x}, \sigma^2, \boldsymbol{\delta}\right)$ is distributed according to (18). Thus, to compare the Bayesian and MAP estimation methods on equal footing we employ in our MAP method the same imaging model and the same prior model of $\mathbf{x}$ used in our Bayesian method. The MAP estimate of the high-resolution image and the unknown parameters is given by

$$\left\{ \hat{\mathbf{x}}, \hat{\sigma}^2, \hat{\alpha}^2, \hat{\boldsymbol{\delta}} \right\}_{MAP} = \arg \max_{\mathbf{x}, \sigma^2, \alpha^2, \boldsymbol{\delta}} \left[\log p\left(\mathbf{y} \mid \mathbf{x}, \sigma^2, \boldsymbol{\delta}\right) + \log p\left(\mathbf{x} \mid \alpha^2\right) \right] \tag{37}$$

Substituting (1) and (18) into (37) and maximizing with respect to each unknown separately yields

$$\hat{\mathbf{x}}^{(p+1)} = \left(\mathbf{B}\left(\hat{\boldsymbol{\delta}}^{(p)}\right)^T \mathbf{B}\left(\hat{\boldsymbol{\delta}}^{(p)}\right) + \frac{\hat{\alpha}^{2(p)}}{\hat{\sigma}^{2(p)}} \mathbf{Q}^T \mathbf{Q} \right)^{-1} \mathbf{B}\left(\hat{\boldsymbol{\delta}}^{(p)}\right)^T \mathbf{y} \tag{38}$$

$$\hat{\alpha}^{2(p+1)} = \frac{M_H N_H - 1}{\left\| \mathbf{Q}\hat{\mathbf{x}}^{(p+1)} \right\|^2} \tag{39}$$

$$\hat{\sigma}^{2(p+1)} = \frac{1}{PM_L N_L} \sum_{i=1}^{P} \left\| \mathbf{y}_i - \mathbf{B}_i\left(\hat{\boldsymbol{\delta}}_i^{(p)}\right)\hat{\mathbf{x}}^{(p+1)} \right\|^2 \tag{40}$$

$$\hat{\boldsymbol{\delta}}_i^{(p+1)} = \arg\min_{\boldsymbol{\delta}_i} \left\| \mathbf{y}_i - \mathbf{B}_i\left(\boldsymbol{\delta}_i\right)\hat{\mathbf{x}}^{(p+1)} \right\|^2 \tag{41}$$

where $(\,)^{(p)}$ indicates the value at the p^{th} iteration. Our MAP algorithm requires the iterative computation of (38) through (41) and thus maximizes the joint probability by cyclic coordinate descent. This algorithm is very similar in spirit to the algorithm in [12]. Note that for simulation purposes, the minimization required in (41) was computed numerically using direction set methods, although gradient descent methods can also be employed. We note that the update for **x** is equivalent in our particular MAP and Bayesian formulations. This is a direct result of employing Gaussian statistics in our prior model, since for the Gaussian, the mode and the mean are the same point. (this is obviously not the case for all prior distributions.) The updates for the model parameters, however, are quite different in the MAP case compared to the E-M case. Finally, the MAP solution presented above has an equivalent version in the frequency domain, the derivation of which is straightforward and is omitted here.

VII. Experimental Results

A number of simulations have been performed using the proposed algorithms on a set of images. We present results and evaluate the performance of the proposed methods under different noise conditions. To present a case for the joint estimation of the parameters, we also compare our approaches to a third approach in which the registration parameters are estimated independently from the other parameters and the restored high-resolution image.

In order to measure parameter estimation and image restoration performance, a synthetic set of degraded images were computed from a single high-resolution image. In all of our simulations, we employed the following model for the blurring operator **H**, which we assumed to be known. We decomposed **H** into two circulant matrices, $\mathbf{H} = \mathbf{H}_1\mathbf{H}_2$, where $\mathbf{H}_1$ is a doubly circulant approximation to a 2D linear convolution matrix assembled from the following 2D impulse response

$$h_1\left(m,n;\beta^2\right) = \frac{1}{2\pi\beta^2}\exp\left[-\frac{1}{2\beta^2}\left(m^2+n^2\right)\right] \tag{42}$$

This component of the blur model approximates the combined effects of diffraction, geometric blur, and atmospheric blur. The second component of the blur, $\mathbf{H}_2$, models the integration of photons by the sensor

and is a doubly circulant approximation to a 2D linear convolution matrix assembled from the following 2D kernel

$$h_2(m,n)=\left\{\begin{array}{ll}\dfrac{1}{d_x d_y} & m=0,1,\ldots,d_y-1, n=0,1,\ldots,d_x-1\\ 0 & \text{otherwise}\end{array}\right\} \tag{43}$$

After blurring the high-resolution image by applying **H**, the blurred images were shifted in the spatial domain using a truncated 2-D separable sinc function, where the support of the sinc function was equal to the data support of the high-resolution image. Per simulation trial, the shifting process was repeated with different random shift vectors $\boldsymbol{\delta}_i$ for $i=0,1,\ldots,P-1$, with $\delta_{i,x}$ and $\delta_{i,y}$ samples from uniform distributions $U(-d_x,+d_x)$ and $U(-d_y,+d_y)$, respectively. In all of our experiments, we set $P=d_x d_y$. In order to preserve their aspect ratio, blurred and shifted images were decimated by equal integer decimation factors in the horizontal and vertical direction. Finally, white Gaussian noise with zero mean was added to the blurred and shifted low-resolution images.

We initialized our iterative algorithm using random values. The shift vector estimate was set randomly from the same uniform distributions as the shifts used to produce the degraded images. The noise variance was initialized to a random value drawn from $U[300, 600]$. The covariance parameter was initialized to a random value drawn from $U[0.03, 0.20]$. The regularization matrix **Q** was set to correspond to linear convolution with the following filter mask:

$$\frac{1}{40}\begin{bmatrix}-1 & -4 & -1\\ -4 & 20 & -4\\ -1 & -4 & -1\end{bmatrix} \tag{44}$$

For all experiments, the E-M and MAP iterations were terminated when the criterion

$$\left\|\mathbf{x}_{\left(\alpha^{(p)},\sigma^{(p)},\boldsymbol{\delta}^{(p)}\right)}-\mathbf{x}_{\left(\alpha^{(p-1)},\sigma^{(p-1)},\boldsymbol{\delta}^{(p-1)}\right)}\right\|^2 \Big/ \left\|\mathbf{x}_{\left(\alpha^{(p-1)},\sigma^{(p-1)},\boldsymbol{\delta}^{(p-1)}\right)}\right\|^2 < 10^{-7} \tag{45}$$

was satisfied. This criterion was motivated by the desire to stop both the E-M and MAP iterations at a similar point. Were it not for this constraint, a more natural choice for the E-M algorithm would be to threshold the value of the conditional log likelihood, which is guaranteed to increase monotonically [23]. As a quantitative measure of image restoration performance, we used the peak signal-to-noise ratio (PSNR) given by

$$PSNR = 10\log_{10}\left(\frac{255^2}{\frac{1}{M_H N_H}\|\mathbf{x}-\hat{\mathbf{x}}\|^2}\right) \tag{46}$$

One of the motivations for the development of the proposed joint super-resolution techniques was the generally accepted notion that the quality of reconstruction depends largely on the accuracy of the estimated registration parameters. Our hypothesis was that a technique that estimates all of the unknowns jointly using all of the available data, like our presented techniques, should produce superior registration parameter estimates as compared to other commonly employed techniques that register the images using a separate preprocessing step that operates directly on a pair of noisy, blurred, low-resolution observations.

To test this hypothesis, we compared the performance of our two proposed joint estimation algorithms to a third algorithm. This third algorithm was based on an independent estimation of the registration parameters (i.e., the registration parameters were estimated separately from the restoration.) We chose the sub-pixel registration method in [30], based on a rank-1 approximation of the phase correlation matrix, to estimate the registration parameters directly from the low-resolution observations. The registration parameters estimated using the method in [30] were then supplied to our proposed Bayesian technique, modified to treat the registration parameters obtained from the registration step as known. We denote this approach as Algorithm 1. We denote our MAP approach described in Section VI as Algorithm 2, and our Bayesian approach described in Sections IV and V as Algorithm 3.

For our experiments, we chose the 256 x 256 resolution chart image shown in Fig. 1. Henceforth, we shall refer to this image as the ideal high-resolution image. In our experiments, we set $d_x = d_y = 2$ and created four observations per trial (i.e. $P = 4$).

A number of experiments were performed with varying amounts of noise and blur. The noise variance was set to 315.8, 31.58, 3.158, and 0.3158, corresponding to a SNR of approximately 10, 20, 30, and 40 dB, respectively, relative to the power in the zero mean version of the image. In other words, we have defined SNR in terms of the zero mean version of the high-resolution image as $SNR = 10\log_{10}\left(\|\mathbf{x}-\bar{\mathbf{x}}\|^2 / M_H N_H \sigma^2\right)$ with $\bar{\mathbf{x}}$ equal to the mean of $\mathbf{x}$. This is a convenient measure for simulation, as in our formulation the estimation of the noise variance is performed on images with zero mean. In order to measure and compare the performance of the proposed parameter estimation algorithm, ten realizations of the noise were generated for each noise level. Since the same uniform blur modeled by $\mathbf{H}_2$ was present in all trials, our blur model was parameterized by a single parameter, β^2. We experimented with several different values for β^2 with 40 trials (10 per SNR) for each value of blur variance. We report the results for $\beta^2 = 0.0$ and β^2=0.5 as

representative results and denote the experiments with $\beta^2 = 0.0$ as Experiment 1 and experiments with $\beta^2=0.5$ as Experiment 2.

To measure the quality of registration parameter estimation, we calculated the mean of the magnitude of the registration error averaged over both components, in units of the number of high-resolution pixels, defined as

$$\mu_{|\delta-\hat{\delta}|} = \frac{1}{20P}\sum_{k=1}^{10}\sum_{i=0}^{P-1}\sum_{j\in x,y}\left|\delta_{i,j}^{(k)} - \hat{\delta}_{i,j}^{(k)}\right| \tag{47}$$

where $\delta_{i,j}^{(k)}$ is the true shift and $\hat{\delta}_{i,j}^{(k)}$ is the estimated shift in the k^{th} trial. We also calculated the standard deviation of the magnitude of the registration error, again over both translational components of the shift vector and over all ten realizations of the noise, as follows

$$\sigma_{|\delta-\hat{\delta}|} = \sqrt{\frac{1}{20P-1}\sum_{k=1}^{10}\sum_{i=0}^{P-1}\sum_{j\in x,y}\left(\left|\delta_{i,j}^{(k)} - \hat{\delta}_{i,j}^{(k)}\right| - \mu_{|\delta-\hat{\delta}|}\right)^2} \tag{48}$$

The results for $\beta^2=0.0$ are displayed in Table I, and the results for $\beta^2=0.5$ are displayed in Table II. The table entries are blank for the 10 dB case for Algorithm 2 because the MAP method did not converge in any of the trials for the 10 dB SNR case in either experiment. This highlights one of the key advantages of the Bayesian approach utilizing the E-M algorithm over the MAP approach—convergence of the E-M is guaranteed.

The results show that both of the proposed approaches produce superior registration estimates as compared to independent estimation of motion followed by restoration. This is true even in the very high SNR case of 40 dB where one would expect the chosen separate registration method to perform well. The performance improvement is particularly pronounced at high noise levels. This is despite the best efforts of the authors to improve the quality of the registration estimates using all of the suggestions made in [30] to improve the performance of the algorithm in the presence of noise. In addition, we experimented with the application of mild low-pass pre-filtering (e.g., a uniform 3x3 low-pass filter) before registration. We present the best results we were able to obtain. From these results, we conclude that the proposed joint estimation approach is both more accurate and more robust than the method in [30]. The Bayesian method produces more accurate registration parameter estimates than the MAP approach.

Tables I and II also list the mean and standard deviation of the estimates of the noise variance, σ^2, and covariance parameter, α^2. The results of Algorithm 1 confirm that poor registration estimates hamper the estimation of the other parameters, providing additional justification for a joint estimation procedure. Comparing the Bayesian and MAP methods, we see that the Bayesian approach consistently provides

superior noise variance estimates as compared to our MAP algorithm, as evidenced by the significantly smaller average mean-square error of the Bayesian noise variance estimates.

Finally, because the covariance parameter is a parameter of the image prior, in the ideal case, its value should not change as the noise variance is varied from trial-to-trial. In other words, this parameter is a function of the image signal, not the noise signal. Thus, although we do not know the "correct" value of the covariance parameter, the fact that in the Bayesian approach this parameter does not vary as much with SNR as compared to the MAP approach provides some evidence, albeit indirect, that the Bayesian approach was better able to utilize this prior model to distinguish between the signal and the noise. The superior accuracy of the noise variance estimates in the Bayesian case certainly provides more direct evidence of this. Since both methods employ the same image prior model, we infer that this advantage can be attributed solely to the Bayesian approach.

Table III and Table IV show the PSNR results (averaged over the trials) of Experiment 1 and Experiment 2, respectively. In all cases, the proposed joint estimation approaches produce superior restorations with higher PSNR compared to the restorations produced by the two-step procedure consisting of first, the separate registration estimation step, and second, the restoration. From this, we conclude that the joint estimation approach is superior to the independent estimation approach. Finally, the Bayesian approach provides a small PSNR gain over the MAP approach.

We therefore conclude that the Bayesian approach produces superior parameter estimates and slightly improved restorations compared to the MAP approach. From the fact that the MAP approach did not converge for the 10 dB case in either experiment, we conclude that our Bayesian approach is more robust than our MAP approach, as well.

We display a reconstruction for the 30 dB case in Experiment 1 in Fig. 2 for all three algorithms. For display purposes, we have chosen the trial that resulted in the most visible difference between Algorithm 1 and the other two algorithms. In this trial, the registration method in [30] produced good estimates for two of the shift vectors, but a poor estimate for the remaining shift vectors, resulting in the poor restoration shown. This was not an aberration, however, as we observed this phenomenon repeatedly, particularly at low SNRs. Similarly, Fig. 3 shows the reconstructions produced by the three methods for the 20 dB case in Experiment 2. In this case, we show the best result of Algorithm 1 of the ten trials. In this experiment, the restoration produced by the Bayesian algorithm is sharper than the restoration produced by the MAP algorithm. This can be attributed to the more accurate estimation of the noise variance.

Figure 4 shows the evolution of the PSNR for Algorithm 2 and Algorithm 3 versus the number of iterations for a typical trial in Experiment 1 with the noise level set at 40 dB. We have chosen to display the worst-case (i.e., the trials with SNR of 40 dB required the largest number of iterations). Note that most of the improvement was obtained in the first few iterations. Each iteration took an average of about 75 seconds

and 90 seconds for Algorithm 2 and Algorithm 3, respectively, on an Intel 2.4 GHz Pentium 4 CPU. In both cases, the vast majority of execution time was devoted to the numerical search for the registration parameters within the M-step of the Bayesian method and the calculation of (41) in the MAP method. Figure 5 shows the evolution of the convergence criterion versus the number of iterations for the same experimental trial. Note that the vertical axis is plotted on a logarithmic scale. From these plots, it is clear that both algorithms converge rapidly and that the PSNR increases monotonically.

VIII. Conclusions

We have presented two iterative super-resolution algorithms, one based on a Bayesian formulation and employing the E-M algorithm, and another based on a MAP formulation solved iteratively by cyclic coordinate decent. Our proposed approaches are novel in that the noise variance, regularization, and registration parameters are all treated as unknowns and are estimated jointly using all of the available data. Our numerical experiments show that the proposed joint estimation approaches offer significantly more accurate and robust estimation of the unknown registration parameters, leading to improved estimates of the other unknown model parameters as compared to conventional approaches that employ registration as an independent estimation procedure during pre-processing. Our numerical experiments also show that the Bayesian approach provides measurable advantages over the MAP approach, including improved estimation of the model parameters and improved restorations. Since both methods share the same imaging model (including the image prior), we conclude that the Bayesian approach provides a fundamental estimation advantage compared to the MAP approach for this problem. However, the main disadvantage of the Bayesian framework is that computing or even approximating the Bayesian integral required to compute the marginal likelihood is much more difficult than computing the mode of the posterior in the MAP approach for complex prior models. We plan in the future to explore joint estimation of the blur, as well as efficient handling of global rotational motion.

Acknowledgement

The authors wish to thank Dr. Lisimachos Kondi for providing the image of the resolution chart used in our experiments. The authors also wish to thank the anonymous reviewers whose insightful comments greatly improved the quality of this manuscript.

Appendix A

In this Appendix, we derive the expressions for the E-step of the E-M algorithm in which the conditional mean and conditional covariance of $\mathbf{x}$ are updated in the frequency domain. Let us define the following 2-D DFT matrices:

$$\mathbf{A}_i' = \mathbf{A}_{N_H} \otimes \mathbf{A}_{M_H} \tag{49}$$

$$\mathbf{A}_j'' = \mathbf{A}_{N_L} \otimes \mathbf{A}_{M_L} \tag{50}$$

where $\mathbf{A}_K$ is a 1-D DFT of size K x K and $\otimes$ represents the Kronecker product. For example, $\mathbf{A}_{N_H}$ is a 1-D DFT matrix of size N_H x N_H, and $\mathbf{A}'$ is the 2-D DFT matrix of size $M_H N_H$ x $M_H N_H$. Let us define the block diagonal matrices

$$\mathbf{A} = \text{blockdiag}\left(\mathbf{A}_i'\right) \quad i = 0,1,...,P-1 \tag{51}$$

$$\bar{\mathbf{A}} = \text{blockdiag}\left(\mathbf{A}_j''\right) \quad i = 0,1,...,P-1 \tag{52}$$

From (11) and (12),

$$\mathbf{B} = \mathbf{DSH} \tag{53}$$

Using (49), (51), and (52), let us transform the degradation matrix **B** to the frequency domain as follows:

$$\bar{\mathbf{A}}\mathbf{B}\mathbf{A}'^{-1} = \bar{\mathbf{A}}\mathbf{D}\mathbf{A}^{-1}\mathbf{A}\mathbf{S}\mathbf{A}'^{-1}\mathbf{A}'\mathbf{H}\mathbf{A}'^{-1} \tag{54}$$

The blur matrix **H** is a 2-D convolution matrix and therefore can be written as a doubly circulant matrix. Therefore,

$$\mathbf{A}'\mathbf{H}\mathbf{A}'^{-1} = \mathbf{H}_D \tag{55}$$

where $\mathbf{H}_D$ is a diagonal matrix containing the frequency domain samples of the blur matrix. The shift matrix **S** consists of P 2-D convolution submatrices stacked vertically and therefore can be written as a matrix of P doubly circulant submatrices. Therefore,

$$\mathbf{A}\mathbf{S}\mathbf{A}'^{-1} = \begin{bmatrix} \mathbf{S}_D(\boldsymbol{\delta}_0) \\ \mathbf{S}_D(\boldsymbol{\delta}_1) \\ \vdots \\ \mathbf{S}_D(\boldsymbol{\delta}_{P-1}) \end{bmatrix} = \mathbf{S}_D(\boldsymbol{\delta}) \tag{56}$$

where $\mathbf{S}_D(\boldsymbol{\delta}_i)$ is a complex, diagonal matrix containing the frequency domain samples of the i^{th} shift matrix. Note that since δ_0=[0 0], $\mathbf{S}_D(\delta_0) = \mathbf{I}$. Furthermore, let us define

$$\mathbf{F} = \bar{\mathbf{A}}\mathbf{D}\mathbf{A}^{-1} \tag{57}$$

We call the matrix **F** the "folding" matrix because it describes how the frequency domain version of the high-resolution signal folds over on itself due to the decimation operation. The folding matrix has the following form:

$$\mathbf{F} = \text{blockdiag}\left(\mathbf{F}_i\right) \quad i = 0,1,\ldots,P-1 \tag{58}$$

where

$$\mathbf{F}_i = \frac{1}{d_x d_y}\begin{bmatrix}\mathbf{F}_0' & \mathbf{F}_1' & \cdots & \mathbf{F}_{d_y-1}'\end{bmatrix} \tag{59}$$

$$\mathbf{F}_j' = \text{blockdiag}\left(\mathbf{F}_i''\right) \quad i = 0,1,\ldots,M_L-1 \tag{60}$$

$$\mathbf{F}_k'' = \begin{bmatrix}\mathbf{I}_0 & \mathbf{I}_1 & \cdots & \mathbf{I}_{d_x-1}\end{bmatrix} \tag{61}$$

The identity matrices in (61) are N_H x N_H. Substituting (55), (56), and (57) into (54) yields,

$$\bar{\mathbf{A}}\mathbf{B}\mathbf{A}'^{-1} = \mathbf{F}\mathbf{S}_{\mathrm{D}}(\boldsymbol{\delta})\mathbf{H}_{\mathrm{D}} \tag{62}$$

from which it follows that

$$\mathbf{A}'\mathbf{B}^{\mathrm{H}}\bar{\mathbf{A}}^{-1} = d_x d_y \mathbf{H}_{\mathrm{D}}^{*}\mathbf{S}_{\mathrm{D}}^{*}(\boldsymbol{\delta})\mathbf{F}^{\mathrm{T}} \tag{63}$$

where * denotes the complex conjugate. The frequency-domain versions of the observation vector, noise covariance, and underlying scene covariance are given by

$$\mathbf{Y} = \bar{\mathbf{A}}\mathbf{y} \tag{64}$$

$$\mathbf{S}_n = \bar{\mathbf{A}}\boldsymbol{\Lambda}_n\bar{\mathbf{A}}^{-1} = \sigma^2\mathbf{I} \tag{65}$$

$$\mathbf{S}_x = \mathbf{A}\frac{1}{\alpha^2}\left(\mathbf{Q}^T\mathbf{Q}\right)^{-1}\mathbf{A}^{-1} \Rightarrow \left[\mathbf{S}_x\right]_{m,m} = \frac{1}{\alpha^2}\left|\tilde{\mathbf{Q}}(m)\right|^{-2} \tag{66}$$

where $\tilde{\mathbf{Q}}(m)$ are the eigenvalues of the regularization matrix $\mathbf{Q}$. Note that the matrices on the left-hand side of (65) and (66) are diagonal.

With the above definitions, we are ready to express the E-step of the E-M algorithm in the frequency domain. Starting with the conditional mean,

$$\begin{aligned} M_{x|y} &= \mathbf{A}'\mu_{x|y} = \mathbf{A}'\boldsymbol{\Lambda}_x\mathbf{B}^{\mathrm{H}}\left(\mathbf{B}\boldsymbol{\Lambda}_x\mathbf{B}^{\mathrm{H}} + \boldsymbol{\Lambda}_n\right)^{-1}\mathbf{y} \\ &= \mathbf{A}'\boldsymbol{\Lambda}_x\mathbf{A}'^{-1}\mathbf{A}'\mathbf{B}^{\mathrm{H}}\bar{\mathbf{A}}^{-1}\left(\bar{\mathbf{A}}\mathbf{B}\mathbf{A}'^{-1}\mathbf{A}'\boldsymbol{\Lambda}_x\mathbf{A}'^{-1}\mathbf{A}'\mathbf{B}^{\mathrm{H}}\bar{\mathbf{A}}^{-1} + \bar{\mathbf{A}}\boldsymbol{\Lambda}_n\bar{\mathbf{A}}^{-1}\right)^{-1}\bar{\mathbf{A}}\mathbf{y} \\ &= \mathbf{S}_x d_x d_y \mathbf{H}_{\mathrm{D}}^{*}\mathbf{S}_{\mathrm{D}}^{\mathrm{H}}(\boldsymbol{\delta})\mathbf{F}^{\mathrm{T}}\left(\mathbf{F}\mathbf{S}_{\mathrm{D}}(\boldsymbol{\delta})\mathbf{H}_{\mathrm{D}}\mathbf{S}_x d_x d_y \mathbf{H}_{\mathrm{D}}^{*}\mathbf{S}_{\mathrm{D}}^{\mathrm{H}}(\boldsymbol{\delta})\mathbf{F}^{\mathrm{T}} + \mathbf{S}_n\right)^{-1}\mathbf{Y} \end{aligned} \tag{67}$$

From inspection of (67), it appears as though transforming the problem to the frequency domain has not made the inversion of $\mathbf{B}\boldsymbol{\Lambda}_x\mathbf{B}^H + \boldsymbol{\Lambda}_n$ any easier. However, by decomposing the block diagonal matrices into their sub-matrices, one obtains

$$\mathbf{M}_{x|y} = d_x d_y \left[\mathbf{S}_x \mathbf{H}_D^* \mathbf{S}_D^*(\boldsymbol{\delta}_0)\mathbf{F}_0^T \quad \mathbf{S}_x \mathbf{H}_D^* \mathbf{S}_D^*(\boldsymbol{\delta}_1)\mathbf{F}_1^T \quad \cdots \quad \mathbf{S}_x \mathbf{H}_D^* \mathbf{S}_D^*(\boldsymbol{\delta}_{P-1})\mathbf{F}_{P-1}^T \right] \cdot$$

$$\begin{bmatrix} \mathbf{F}_0 \mathbf{S}_x |\mathbf{H}_D|^2 \mathbf{F}_0^T + \sigma^2 \mathbf{I} & \mathbf{F}_0 \mathbf{S}_x |\mathbf{H}_D|^2 \mathbf{S}_D^*(\boldsymbol{\delta}_1)\mathbf{F}_1^T & \cdots & \mathbf{F}_0 \mathbf{S}_x |\mathbf{H}_D|^2 \mathbf{S}_D^*(\boldsymbol{\delta}_{P-1})\mathbf{F}_{P-1}^T \\ \mathbf{F}_1 \mathbf{S}_x |\mathbf{H}_D|^2 \mathbf{S}_D(\boldsymbol{\delta}_1)\mathbf{F}^T & \mathbf{F}_1 \mathbf{S}_x |\mathbf{H}_D|^2 |\mathbf{S}_D(\boldsymbol{\delta}_1)| \mathbf{F}_1^T + \sigma^2 \mathbf{I} & \cdots & \mathbf{F}_1 \mathbf{S}_x |\mathbf{H}_D|^2 \mathbf{S}_D(\boldsymbol{\delta}_1)\mathbf{S}_D^*(\boldsymbol{\delta}_{P-1})\mathbf{F}^T \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{F}_{P-1} \mathbf{S}_x |\mathbf{H}_D|^2 \mathbf{S}_D(\boldsymbol{\delta}_{P-1})\mathbf{F}_0^T & \mathbf{F}_{P-1} \mathbf{S}_x |\mathbf{H}_D|^2 \mathbf{S}_D(\boldsymbol{\delta}_{P-1})\mathbf{S}_D^*(\boldsymbol{\delta}_1)\mathbf{F}_1^T & \cdots & \mathbf{F}_{P-1} \mathbf{S}_x |\mathbf{H}_D|^2 |\mathbf{S}_D(\boldsymbol{\delta}_{P-1})| \mathbf{F}_{P-1}^T + \sigma^2 \mathbf{I} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{Y}_0 \\ \mathbf{Y}_1 \\ \vdots \\ \mathbf{Y}_{P-1} \end{bmatrix}$$

(68)

where each matrix of the form $\mathbf{F}_i \mathbf{S}_x |\mathbf{H}_D|^2 \mathbf{S}_D(\boldsymbol{\delta}_i)\mathbf{S}_D^*(\boldsymbol{\delta}_j)\mathbf{F}_j^T$ is a *diagonal matrix* of size $M_L N_L$ x $M_L N_L$. By performing row and column operations, we can convert the P x P collection of diagonal $M_L N_L$ x $M_L N_L$ submatrices into a block diagonal matrix consisting of $M_L N_L$ P x P submatrices.

Let us define a set k_m consisting of the integers generated by

$$\left\{ k_m : k_m = jN_L + iM_L N_H + \left\lfloor \frac{m}{N_L} \right\rfloor N_H + m \bmod N_L \right\} \tag{69}$$

where

$$\begin{aligned} i &= 0,1,\ldots,d_y - 1 \\ j &= 0,1,\ldots,d_x - 1 \end{aligned} \tag{70}$$

Now, we define the P x P matrix $\mathbf{V}_m$

$$\mathbf{V}_m = \begin{bmatrix} \sum_{i \in k_m} \left[\mathbf{S}_x |\mathbf{H}_D|^2 + d_x d_y \sigma^2 \mathbf{I} \right]_i & \sum_{i \in k_m} \left[\mathbf{S}_x |\mathbf{H}_D|^2 \mathbf{S}_D^*(\boldsymbol{\delta}_1) \right]_i & \cdots & \sum_{i \in k_m} \left[\mathbf{S}_x |\mathbf{H}_D|^2 \mathbf{S}_D^*(\boldsymbol{\delta}_{P-1}) \right]_i \\ \sum_{i \in k_m} \left[\mathbf{S}_x |\mathbf{H}_D|^2 \mathbf{S}_D(\boldsymbol{\delta}_1) \right]_i & \sum_{i \in k_m} \left[\mathbf{S}_x |\mathbf{H}_D|^2 |\mathbf{S}_D(\boldsymbol{\delta}_1)|^2 + d_x d_y \sigma^2 \mathbf{I} \right]_i & \cdots & \sum_{i \in k_m} \left[\mathbf{S}_x |\mathbf{H}_D|^2 \mathbf{S}_D(\boldsymbol{\delta}_1)\mathbf{S}_D^*(\boldsymbol{\delta}_{P-1}) \right]_i \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i \in k_m} \left[\mathbf{S}_x |\mathbf{H}_D|^2 \mathbf{S}_D(\boldsymbol{\delta}_{P-1}) \right]_i & \sum_{i \in k_m} \left[\mathbf{S}_x |\mathbf{H}_D|^2 \mathbf{S}_D(\boldsymbol{\delta}_{P-1}) S_D^*(\boldsymbol{\delta}_1) \right]_i & \cdots & \sum_{i \in k_m} \left[\mathbf{S}_x |\mathbf{H}_D|^2 |\mathbf{S}_D(\boldsymbol{\delta}_{P-1})|^2 + d_x d_y \sigma^2 \mathbf{I} \right]_i \end{bmatrix}$$

(71)

In (71), the notation $[\mathbf{A}]_i$ selects the i^{th} diagonal of matrix $\mathbf{A}$. Let $\mathbf{V}$ be a block diagonal matrix consisting of the P x P submatrices defined as,

$$\mathbf{V} = \text{blockdiag}(\mathbf{V}_i) \quad i = 0,1,\ldots,M_L N_L - 1 \tag{72}$$

Now, let us define a matrix $\mathbf{U}$ recursively as follows:

$$\mathbf{U} = \begin{bmatrix} \dot{\mathbf{U}}_0 \\ \dot{\mathbf{U}}_1 \\ \vdots \\ \dot{\mathbf{U}}_{d_y-1} \end{bmatrix} \tag{73}$$

where

$$\dot{\mathbf{U}}_i = \text{blockdiag}\left(\ddot{\mathbf{U}}_{i,j}\right) \quad j = 0,1,\ldots M_L - 1 \tag{74}$$

$$\ddot{\mathbf{U}}_{i,j} = \begin{bmatrix} \dddot{\mathbf{U}}_{i,j,0} \\ \dddot{\mathbf{U}}_{i,j,1} \\ \vdots \\ \dddot{\mathbf{U}}_{i,j,d_z-1} \end{bmatrix} \tag{75}$$

$$\dddot{\mathbf{U}}_{i,j,k} = \text{diag}\left(\ddddot{\mathbf{U}}_{i,j,k,m}\right) \quad m = 0,1,\ldots N_L - 1 \tag{76}$$

$$\ddddot{\mathbf{U}}_{i,j,k,m} = \left[\mathbf{S}_x \mathbf{H}_D^*\right]_n \left[\left[\mathbf{S}_D^*(\boldsymbol{\delta}_0)\right]_n \quad \left[\mathbf{S}_D^*(\boldsymbol{\delta}_1)\right]_n \quad \cdots \quad \left[\mathbf{S}_D^*(\boldsymbol{\delta}_{P-1})\right]_n\right] \tag{77}$$

and diag( ) forms a diagonal matrix and

$$n = \left(iM_L + j\right)N_H + kN_L + m \tag{78}$$

With these definitions, the conditional mean in the frequency domain is given by

$$\mathbf{M}_{x|y} = d_x d_y \mathbf{U}\mathbf{V}^{-1}\mathbf{Y}' \tag{79}$$

and the conditional covariance in the frequency domain is given by

$$\mathbf{S}_{x|y} = \mathbf{S}_x - \mathbf{U}\mathbf{V}^{-1}\mathbf{U}^{\mathrm{H}} \tag{80}$$

This completes the derivation of the E-step in the frequency domain.

References

[1] T.S. Huang and R. Y. Tsai, "Multiple frame image reconstruction and registration," in *Advances in Computer Vision and Image Processing*, T.S. Huang, Ed. Greenwich, CT: JAI, 1984, ch. 7.

[2] S. P. Kim, N. K. Bose and H. M. Valenzuela, "Recursive Reconstruction of High Resolution Image From Noisy Undersampled Multiframes," *IEEE Trans. Acoust., Speech, Signal Proc.*, vol. 38, no.6, pp. 1013-1027, June 1990.

[3] S. Kim and W.Y. Su, "Recursive high-resolution reconstruction of blurred multiframe images," *IEEE Trans. Image Processing*, vol. 2, pp. 534-539, Oct. 1993.

[4] M. Irani and S. Peleg, "Improving resolution by image registration," *CVGIP: Graph., Models, Image Process.*, vol. 53, pp. 231-239, May 1991.
[5] H. Stark and P. Oskoui, "High-resolution image recovery from image-plane arrays, using convex projections," *J. Opt. Soc. Amer. A,* vol. 6, pp. 1715-1726, 1989.
[6] A. M. Tekalp, M. K. Ozkan, and M. I. Sezan, "High-resolution image reconstruction from lower-resolution image sequences and space-varying image restoration," in *Proc. IEEE Int. Conf. Accoust., Speech, Signal Processing,* San Francisco, CA, 1992, pp. III-169 to III-172.
[7] A. J. Patti, M. I. Sezan, and A. M. Tekalp, "Superresolution video reconstruction with arbitrary sampling lattices and non-zero aperture time," *IEEE Trans. Image Processing*, vol. 6, pp. 1064-1076, Aug. 1997.
[8] M. Elad and A. Feuer, "Restoration of a single superresolution image from several blurred, noisy, and undersampled measured images," *IEEE Trans. Image Processing*, vol. 6, pp. 1646-1658, Dec. 1997.
[9] R. R. Schultz and R. L. Stevenson, "Extraction of high-resolution frames from video sequences," *IEEE Trans. Image Processing,* vol 5, no. 6, pp. 996-1011, June 1996.
[10] R. R. Schultz and R. L. Stevenson, "A Bayesian approach to image expansion for improved definition," *IEEE Trans. Image Processing*, vol. 3, no. 3, pp. 233-242, 1994.
[11] S. Borman and R. L. Stevenson, "Simultaneous Multi-frame MAP Super-Resolution Video Enhancement using Spatio-temporal Priors," *1999 IEEE International Conf. on Image Processing*, vol. 3, pp. 469-473.
[12] R. C. Hardie, K. J. Barnard, and E.E. Armstrong, "Joint MAP Registration and High-Resolution Image Estimation Using a Sequence of Undersampled Images," *IEEE Transactions on Image Processing*, vol. 6, no.12, pp. 1621-1633, Dec. 1997.
[13] B. C. S. Tom and A. K. Katsaggelos, "Reconstruction of a High Resolution Image from Multiple-Degraded and Misregistered Low-Resolution Images", *Proc. 1994 SPIE Conf. on Visual Communications and Image Processing*, SPIE Vol. 2308, pp. 971-981, Chicago, IL, Sept. 1994.
[14] B. C. S. Tom, A. K. Katsaggelos, and N. P. Galatsanos, "Reconstruction of a High Resolution Image from Registration and Restoration of Low Resolution Images", *Proc. 1994 IEEE International Conf. on Image Processing*, pp. III-553-557, Austin, TX, Nov. 1994.
[15] B. C. S. Tom and A. K. Katsaggelos, "Reconstruction of a High-Resolution Image by Simultaneous Registration, Restoration, and Interpolation of Low-Resolution Images," *Proc. 1995 IEEE International Conf. on Image Processing*, pp. II-539-542, Washington, DC, Oct. 1995.
[16] N. Nguyen, P. Milanfar, and G. Golub, "A computationally efficient super-resolution image reconstruction algorithm," *IEEE Transactions on Image Processing*, vol. 10, no. 4, pp 573-583, April 2001.
[17] M. Elad and Y. Hel-Or, "A Fast Super-Resolution Reconstruction Algorithm for Pure Translational Motion and Common Space Invariant Blur", *the IEEE Trans. on Image Processing,* Vol. 10, Number 8, pp. 1187-1193, August 2001
[18] M. Kang and E. Lee, "Regularized adaptive high-resolution image reconstruction considering inaccurate subpixel registration," *IEEE Transactions on Image Processing*, vol. 12, no. 7, pp 826-837, July 2003.
[19] S. Park, M. Park, and M. Kang, "Super-resolution image reconstruction: a technical overview," *IEEE Signal Processing Magazine*, vol. 20, no. 3, pp 21-36, May 2003.
[20] C. A. Segall, R. Molina, and A. K. Katsaggelos, "High-resolution images from low-resolution compressed video," *IEEE Signal Processing Magazine*, vol. 20, no. 3, pp 37-48, May 2003.
[21] M. K. Ng and N. K. Bose, "Mathematical Analysis of Super-Resolution Methodology," *IEEE Signal Processing Magazine*, vol. 20, no. 3, pp 62-74, May 2003.

[22] A. K. Katsaggelos and K. T. Lay, "Identification and Restoration of Images Using the Expectation Maximization Algorithm," in *Digital Image Restoration*, A. K. Katsaggelos, editor, Springer-Verlag, ch. 6, pp. 143-176, 1991.

[23] A. D. Dempster, N. M. Laird, and D. B. Rubin, "Maximum Likelihood from Incomplete Data via the EM algorithm," *J. Roy. Stat. Soc.*, vol B39, pp 1-37, 1977.

[24] 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, no. 3, July 1993.

[25] D. Dungeon and R. M. Mersereau, *Multidimensional Digital Signal Processing,* Prentice Hall, 1984.

[26] B. D. Ripley, *Spatial Statistics*, Wiley, New York, 1981.

[27] R. Molina, B. D. Ripley, "Using spatial models as priors in astronomical images analysis", *J. Appl. Stat.*, vol.16, pp.193-206, 1989.

[28] 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.

[29] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C*, Cambridge University Press, New York, 1992.

[30] W. S. Hoge, "A Subspace Identification Extension to the Phase Correlation Method," *IEEE Transactions of Medical Imaging*, Vol. 2, No. 2, pp. 277-280, February 2003.

[31] M. J. Beal, *Variational Algorithms for Approximate Bayesian Inference*, Ph.D. dissertation, University of Cambridge, U.K., 1998.

[32] R. M. Neal and G. E. Hinton, "A view of the E-M algorithm that justifies incremental, sparse and other variants," in *Learning in Graphical Models*, M. I. Jordan, Ed. Cambridge, MA: MIT Press, 1998, pp. 355–368.

[33] A. Likas and N. Galatsanos, "A Variational Approach for Bayesian Blind Deconvolution", *IEEE Trans. on Signal Processing*, Vol. 52, No. 8, pp. 2222-2233, August 2004.

[34] K. T. Lay and A. K. Katsaggelos, "Image Identification and Restoration Based on The Expectation-Maximization Algorithm," *Optical Engineering*, vol. 29, pp. 436-445, May 1990.

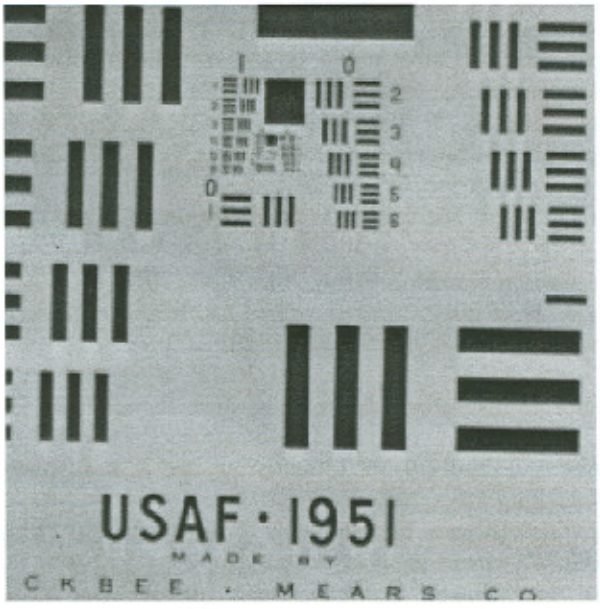

Fig. 1 Original resolution chart image, 256 x 256 pixels

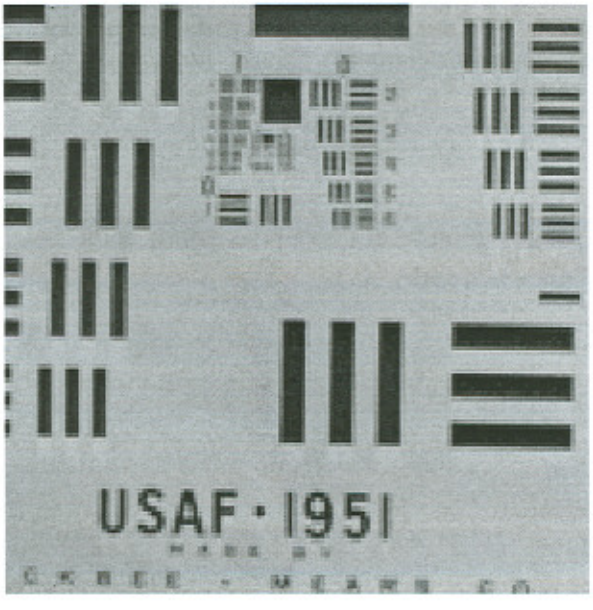

Fig. 2a Experiment 1, Zero-order hold of one of the 128 x 128 observations

Fig. 2b Experiment 1, Worst restoration produced by Algorithm 1

Fig. 2c Experiment 1, Restoration produced by Algorithm 2 (MAP)

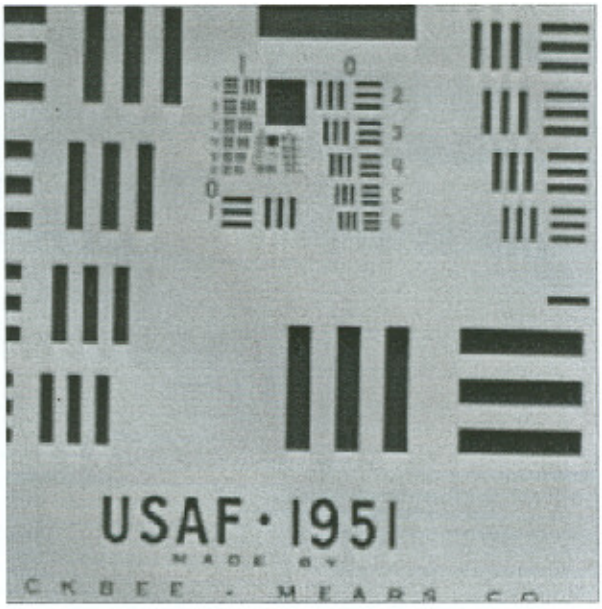

Fig. 2d Experiment 1, Restoration produced by Algorithm 3 (Bayesian)

Fig. 3a Experiment 2, Zero-order hold of one of the 128x128 observations

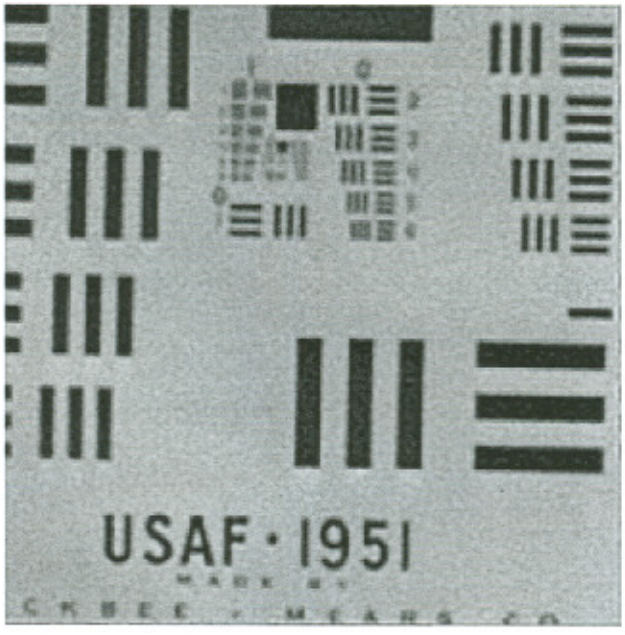

Fig. 3b Experiment 2, Best restoration produced by Algorithm 1

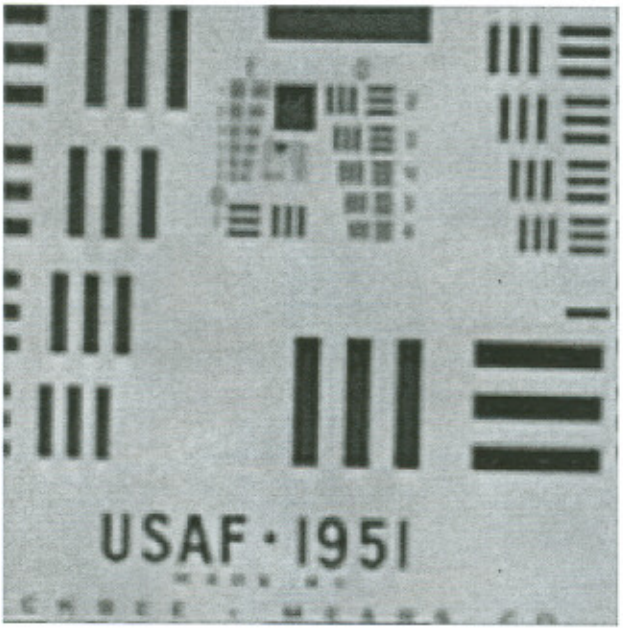

Fig. 3c Experiment 2, Restoration Produced by Algorithm 2 (MAP)

Fig. 3d Experiment 2, Restoration Produced by Algorithm 3 (Bayesian)

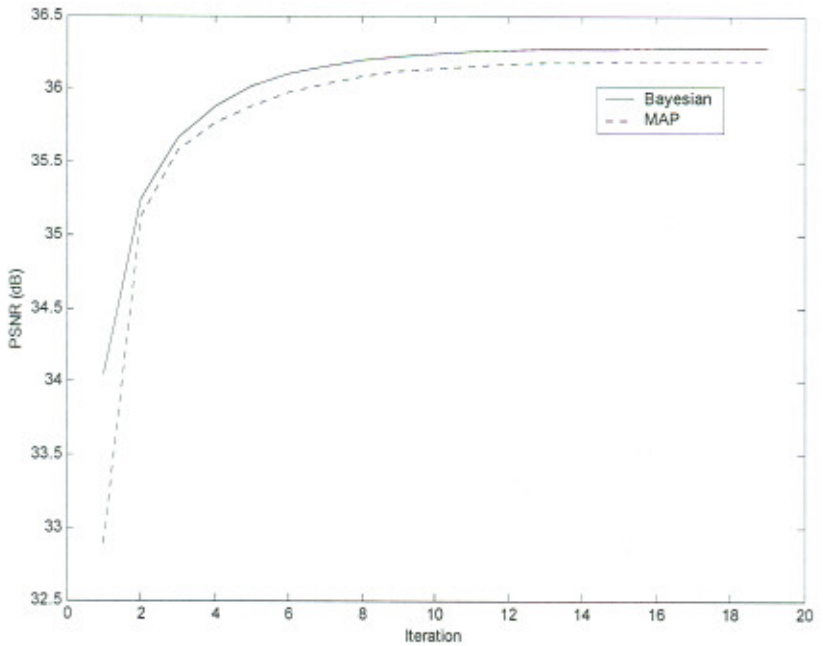

Fig. 4 PSNR evolution versus the number of iterations, Experiment 1, SNR = 40 dB

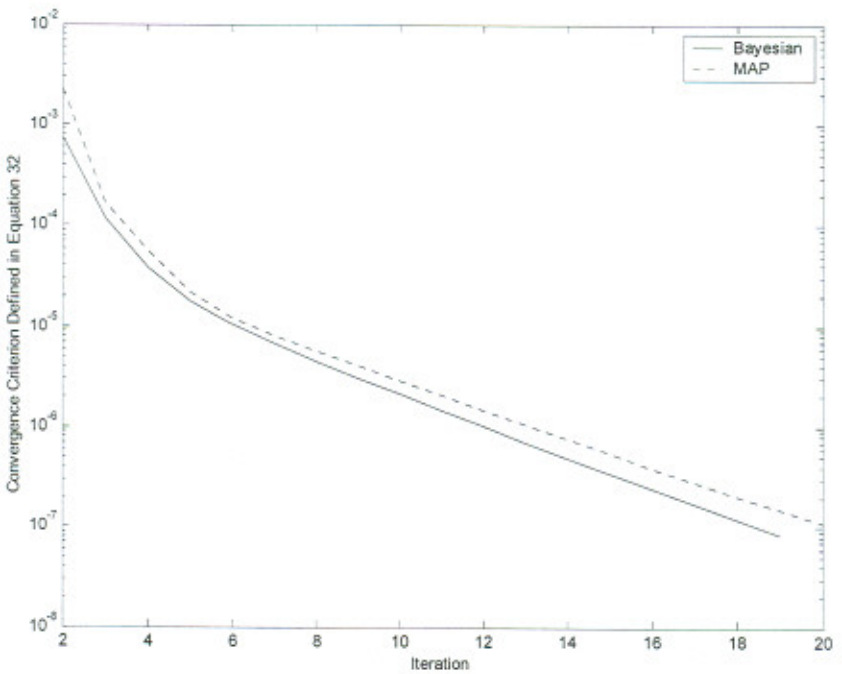

Fig. 5 Convergence Criterion versus the number of iterations, Experiment 1, SNR = 40 dB

TABLE I
EXPERIMENT 1: UNIFORM 2X2 BLUR
PARAMETER ESTIMATION RESULTS

	SNR: 10 dB	SNR: 20 dB	SNR: 30 dB	SNR:40 dB
Algorithm 1 Mean magnitude of registration error, $\mu_{\lvert\delta-\hat{\delta}\rvert}$	9.6665	1.3590	1.9465	0.0304
Algorithm 1 Standard dev. of magnitude of registration error, $\sigma_{\lvert\delta-\hat{\delta}\rvert}$	8.6717	3.3636	5.2946	0.0307
Algorithm 1 Mean of estimated noise variance	2135.2	410.08	515.14	3.061
Algorithm 1 Standard dev. of estimated noise variance	597.32	690.64	645.07	0.7557
Algorithm 1 Mean of estimated covariance parameter	0.1062	0.0415	0.0384	0.0307
Algorithm 1 Standard dev. of estimated covariance parameter	$359.3 \cdot 10^{-4}$	$429.8 \cdot 10^{-4}$	$131.9 \cdot 10^{-4}$	$9.29 \cdot 10^{-4}$
Algorithm 2 Mean magnitude of registration error, $\mu_{\lvert\delta-\hat{\delta}\rvert}$	--	0.0108	0.0141	0.0173
Algorithm 2 Standard dev. of magnitude of registration error, $\sigma_{\lvert\delta-\hat{\delta}\rvert}$	--	0.0076	0.0123	0.0101
Algorithm 2 Mean of estimated noise variance	--	25.80	1.885	0.9079
Algorithm 2 Standard dev. of estimated noise variance	--	1.176	0.5777	0.3539
Algorithm 2 Mean of estimated covariance parameter	--	0.0890	0.0476	0.0495
Algorithm 2 Standard dev. of estimated covariance parameter	--	$50.3 \cdot 10^{-4}$	$59.7 \cdot 10^{-4}$	$42.3 \cdot 10^{-4}$
Algorithm 3 Mean magnitude of registration error, $\mu_{\lvert\delta-\hat{\delta}\rvert}$	0.0144	0.0098	0.0137	0.0144
Algorithm 3 Standard dev. of magnitude of registration error, $\sigma_{\lvert\delta-\hat{\delta}\rvert}$	0.0228	0.0194	0.0184	0.0200
Algorithm 3 Mean of estimated noise variance	304.41	30.48	4.602	2.413
Algorithm 3 Standard dev. of estimated noise variance	2.4964	0.8406	0.8066	0.7407
Algorithm 3 Mean of estimated covariance parameter	0.0183	0.0230	0.0283	0.0310
Algorithm 3 Standard dev. of estimated covariance parameter	$1.555 \cdot 10^{-4}$	$4.828 \cdot 10^{-4}$	$9.415 \cdot 10^{-4}$	$10.68 \cdot 10^{-4}$

TABLE II
EXPERIMENT 2: CASCADE OF UNIFORM 2X2 BLUR AND GAUSSIAN BLUR WITH β^2=0.5
PARAMETER ESTIMATION RESULTS

	SNR: 10 dB	SNR: 20 dB	SNR: 30 dB	SNR:40 dB
Algorithm 1, Mean magnitude of registration error, $\mu_{\lvert\delta-\hat{\delta}\rvert}$	6.6715	2.0865	0.6716	0.0285
Algorithm 1, Standard dev. of magnitude of registration error, $\sigma_{\lvert\delta-\hat{\delta}\rvert}$	7.3326	5.7795	2.2253	0.0208
Algorithm 1, Mean of estimated noise variance	1787.5	546.84	234.72	2.299
Algorithm 1, Standard dev. of estimated noise variance	337.18	581.34	493.68	0.7210
Algorithm 1, Mean of estimated covariance parameter	0.0774	0.0358	0.0294	0.0230
Algorithm 1, Standard dev. of estimated covariance parameter	0.0211	0.0228	0.0169	0.0011
Algorithm 2 Mean magnitude of registration error, $\mu_{\lvert\delta-\hat{\delta}\rvert}$	--	0.0194	0.017	0.0197
Algorithm 2 Standard dev. of magnitude of registration error, $\sigma_{\lvert\delta-\hat{\delta}\rvert}$	--	0.0086	0.0109	0.0101
Algorithm 2 Mean of estimated noise variance	--	43.16	3.5438	1.2310
Algorithm 2 Standard dev. of estimated noise variance	--	1.1114	0.5285	0.4267
Algorithm 2 Mean of estimated covariance parameter	--	0.1748	0.0750	0.0607
Algorithm 2 Standard dev. of estimated covariance parameter	--	$32.71 \cdot 10^{-4}$	$39.77 \cdot 10^{-4}$	$59.47 \cdot 10^{-4}$
Algorithm 3, Mean magnitude of registration error, $\mu_{\lvert\delta-\hat{\delta}\rvert}$	0.0228	0.0194	0.0184	0.0199
Algorithm 3, Standard dev. of magnitude of registration error, $\sigma_{\lvert\delta-\hat{\delta}\rvert}$	0.0160	0.0092	0.0105	0.0097
Algorithm 3, Mean of estimated noise variance	314.62	32.33	4.638	1.951
Algorithm 3, Standard dev. of estimated noise variance	2.0908	0.4147	0.6738	0.6007
Algorithm 3, Mean of estimated covariance parameter	0.0168	0.0186	0.0211	0.0228
Algorithm 3, Standard dev. of estimated covariance parameter	$2.403 \cdot 10^{-4}$	$1.509 \cdot 10^{-4}$	$7.040 \cdot 10^{-4}$	$10.81 \cdot 10^{-4}$

TABLE III
EXPERIMENT 1, PSNR RESTORATION RESULTS AVERAGED OVER ALL TRIALS

	SNR = 10 dB	SNR = 20 dB	SNR = 30 dB	SNR = 40 dB
Algorithm 1	16.56 dB	27.72 dB	28.35 dB	35.35 dB
Algorithm 2	--	31.54 dB	33.91 dB	35.27 dB
Algorithm 3	27.56 dB	31.74 dB	34.57 dB	35.68 dB
Gain Provided by Algorithm 3 over Algorithm 1, Experiment 1	11.00 dB	4.02 dB	6.22 dB	0.33 dB
Gain Provided by Algorithm 3 over Algorithm 2, Experiment 1	--	0.20 dB	0.66 dB	0.41 dB

TABLE IV
EXPERIMENT 2, PSNR RESTORATION RESULTS AVERAGED OVER ALL TRIALS

	SNR = 10 dB	SNR = 20 dB	SNR = 30 dB	SNR = 40 dB
Algorithm 1	16.20 dB	25.54 dB	27.72 dB	33.25 dB
Algorithm 2	--	28.40 dB	32.49 dB	33.37 dB
Algorithm 3	26.54 dB	30.30 dB	32.46 dB	33.36 dB
Gain Provided by Algorithm 3 over Algorithm 1	10.34 dB	4.76 dB	4.74 dB	0.11 dB
Gain Provided by Algorithm 3 over Algorithm 2	--	1.90 dB	~ 0 dB	~ 0 dB