Super-Resolution Based on Fast Registration and Maximum *a Posteriori* Reconstruction

Giannis K. Chantas, Nikolaos P. Galatsanos, and Nathan A. Woods

Abstract—In this paper, we propose a maximum *a posteriori* framework for the super-resolution problem, i.e., reconstructing high-resolution images from shifted, rotated, low-resolution degraded observations. The main contributions of this work are two; first, the use of a new locally adaptive edge preserving prior for the super-resolution problem. Second an efficient two-step reconstruction methodology that includes first an initial registration using only the low-resolution degraded observations. This is followed by a fast iterative algorithm implemented in the discrete Fourier transform domain in which the restoration, interpolation and the registration subtasks of this problem are preformed simultaneously. We present examples with both synthetic and real data that demonstrate the advantages of the proposed framework.

Index Terms—Maximum *a posteriori* (MAP), registration, spatially varying regularization, super-resolution.

I. INTRODUCTION

THE problem of super-resolution is defined as obtaining an image with enhanced resolution from a set of lower resolution degraded images. The super-resolution problem has a long history. In this paper, we will not attempt to fully overview it; for this purpose, the interested reader is referred to the recent surveys articles [1] and [2] and the edited books [3] and [23]. Many methodologies have been applied to the super-resolution problem. An important category of them formulates this problem as an ill-posed image reconstruction problem [6] and introduces prior information (regularization) to find the super-resolved image [1]. However, super-resolution viewed as learning problem has also been recently proposed [4] and [5].

Recent efforts after the surveys in [1] and [2] based on the regularized reconstruction methodology for the super-resolution problem are the works in [7]–[13]. The work in [9] uses a methodology based on the theory of projections onto convex sets [19]. In the rest of this work, we will concentrate on the regularized reconstruction point of view. Regularized reconstruction can be also viewed as a maximum *a posteriori* (MAP) approach by assuming an appropriate probability density for the error in the assumed imaging model and an appropriate prior

for the image [6]. Thus, in what follows we will not distinguish between these two approaches.

In [7], the problem of reconstructing high-resolution frames from compressed video is examined using a Bayesian formulation based on a Gaussian simultaneously autoregressive (SAR) stationary image prior. In [10], color images and demosaicing are considered, and regularization (image priors) based on the L_1-norm which is proposed in order to avoid the shortcomings of L_2-norm based regularization. Furthermore, non-Gaussian measurement errors are considered. More specifically, it was also shown that L_1-norm minimization yields better results in the case of inaccuracies in the imaging model. In [8], a computationally fast method is proposed based on the L_1-norm assuming known integer pixel displacements between frames. However, in [8] and [10] the parameters that define the regularization term are chosen empirically. In [11], an expectation-maximization (E-M) algorithm and a MAP algorithm are presented for simultaneous registration, restoration and interpolation for super-resolution. Nevertheless, a stationary SAR prior is used in both formulations in [11]. In [12], different degradations are assumed in each low-resolution observation. However, L_2-norm-based stationary regularization is used. In [13], an interesting statistical performance analysis is presented that offers insight into the fundamental bottlenecks limiting the performance of super-resolution algorithms.

The first contribution of this work is that we utilize for the first time for the super-resolution problem a new image prior, which is based on a hierarchical two-level model. The first level of this model captures the correlations while the second level provides a description of the local image edge structure in different directions. Thus, it is possible using this prior model to reconstruct the images without smoothing edges or ringing artifacts in the vicinity of edges. Furthermore, this prior has been applied successfully to the image restoration problem in [14] and [15].

It is interesting also to note that an algorithm similar to the herein proposed MAP algorithm can be obtained using a completely different deterministic principle; half-quadratic regularization [18], with the appropriate potential function and parameters selected [15]. In [18], a very elegant theory and a convergence proof for this class of algorithms are given. However, it is not specified how to select the appropriate potential function and parameters for a given data set.

The second contribution of this work is a new two-step reconstruction algorithm. In the work in [11], the imaging model assumed only shifts and did not incorporate rotations. In spite of this, the registration task was extremely slow because registration was performed using the high-resolution image as it was iteratively reconstructed. Furthermore, it was based on a method that used only 1st derivatives. The first stage of

the herein proposed methodology is a preprocessing step that approximately registers the degraded low-resolution observations. These "almost-registered" low-resolution observations are used subsequently by an iterative algorithm which simultaneously reconstructs the high-resolution images and finds their registration parameters. We propose this suboptimal two-stage approach in order to speed up the super-resolution algorithm. Thus, the MAP functional is maximized based on coarse estimation of rotation and translation between image pairs. We have found that such coarse estimation provides enough accuracy to effectively remove the rotational and coarse (super-pixel) translational motion between image pairs. This algorithm is implemented entirely in the discrete Fourier transform (DFT) domain. Furthermore, the registration subtask is based on the Newton–Raphson (NR) algorithm that utilizes analytically calculated first and second derivatives and converges rapidly since NR algorithms display quadratic convergence [16]. The purpose of the preprocessing step is to also ameliorate one of the main difficulties of NR methods which are known to diverge unless initialized close to the solution.

The rest of this paper is organized as follows. In Sections II and III, we present the imaging model and the proposed image prior models, respectively. In Section IV, we describe the preregistration step. In Section V, we present the MAP based restoration algorithm. In Section VI, we present experiments with synthetic and real data that demonstrate the properties of our algorithm. Finally, in Section VII, we provide conclusions and thoughts for future research.

II. IMAGING MODEL

A linear imaging model is assumed. We denote as d the integer decimation factor. In other words, the imaging model assumes a high-resolution image of size $N_H \times 1$, where $N_H = dN$. This model also assumes as observations P low-resolution images of size $N \times 1$ by applying the $PN \times N_H$ degradation operator $\mathbf{B}$ to the high-resolution image. Then, white noise is added at each observation. Let $\mathbf{y}$ be a $PN \times 1$ vector, containing the P low-resolution observed images $\mathbf{y}_i$

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}_1^T & \mathbf{y}_2^T & \cdots & \mathbf{y}_P^T \end{bmatrix}^T$$

where $\mathbf{y}_i$ is a $N \times 1$ vector, representing a low-resolution image. Using this notation, the observations are given by

$$\mathbf{y} = \mathbf{B}\mathbf{x} + \mathbf{n} \qquad (1)$$

where $\mathbf{x}$ the (unknown) original $N_H \times 1$ high-resolution image to be estimated, $\mathbf{B}$ is a $PN \times N_H$ degradation matrix and $\mathbf{n} = [\mathbf{n}_1^T \ \mathbf{n}_2^T \ \cdots \ \mathbf{n}_P^T]^T$ a $PN \times 1$ vector consisting of $PN \times 1$ additive white noise vectors. We assume Gaussian statistics for the noise given by $\mathbf{n}_i \sim N(\mathbf{0}, \beta_i^{-1}\mathbf{I}), i = 1, \ldots, P$, where $\mathbf{0}$ is a $N \times 1$ vector with zeros, $\mathbf{I}$ the $N \times N$ identity matrix, respectively, and $\beta_i^{-1}, i = 1, \ldots P$, are the noise variances of the observations that are assumed unknown and statistically independent with each other. The degradation operator $\mathbf{B}$ is given by

$$\mathbf{B} = [\mathbf{B}_1^T \ \cdots \ \mathbf{B}_P^T]^T$$

where $\mathbf{B}_i = \mathbf{D}\mathbf{H}_i\mathbf{S}(\delta_i)\mathbf{R}(\theta_i)$ for $i = 1, \ldots P$. The matrix $\mathbf{D}$ is the known $N \times N_H$ decimation matrix. $\mathbf{H}_i, i = 1 \ldots P$ are the shift-invariant $N_H \times N_H$ blurring convolutional operators, and $\mathbf{S}(\delta_i)$, for $i = 1, \ldots, P$, are the $N_H \times N_H$ shift-invariant shifting operators. Each δ_i is a scalar which represents translation (with respect to the first image) and is assumed unknown. The shift operator, $\mathbf{S}(\delta_i)$, is the Shannon interpolation operator which is shift invariant [3]. The impulse response of the shift operator is given by

$$S_{\text{shift}}(m; \delta_i) = \frac{\sin(\pi(m - \delta_i))}{\pi(m - \delta_i)}, \quad m = 1, \ldots, N.$$

The shift invariant operators are assumed circulant. This is very useful for computational purposes because such matrices can be easily diagonalized in the DFT domain. One difficulty that arises in the super-resolution problem is the decimation operator which is not square and, thus, not circulant. In this work, we take advantage of the simple form of this matrix, and, despite its noncirculant nature, we obtain tractable calculations in the DFT domain.

Last, the $N_H \times N_H$ matrix $\mathbf{R}(\theta_i)$ represents the rotation of each observation relative to the unknown ideal image $\mathbf{x}$. The imaging model assumes that image i is a rotated (as well as shifted) version of the first image, with angle θ_i. Using all the above definitions, (1) can be rewritten as such

$$\mathbf{y}_i = \mathbf{B}_i\mathbf{x} + \mathbf{n}_i = \mathbf{D}\mathbf{H}_i\mathbf{S}(\delta_i)\mathbf{R}(\theta_i)\mathbf{x} + \mathbf{n}_i$$
$$\text{for } i = 1, \ldots, P. \quad (2)$$

III. IMAGE PRIOR MODEL

Since we utilize a MAP algorithm, a prior for the image is necessary. The prior used here is nonstationary and has been used with success in other image processing problems [14] and [15]. This image prior model assumes that the first-order differences of the image $\mathbf{x}$ in four directions, 0^0, and 90^0 respectively, are given by

$$\varepsilon^1(i, j) = \mathbf{x}(i, j) - \mathbf{x}(i, j + 1)$$
$$\varepsilon^2(i, j) = \mathbf{x}(i, j) - \mathbf{x}(i + 1, j) \qquad (3)$$

with $\varepsilon^k(i, j)k = 1, 2$, the difference residuals for the image location (i, j). The above equations can be also written in matrix vector form for the entire image as $\mathbf{Q}^k\mathbf{x} = \varepsilon^k, k = 1, 2$ where $\mathbf{Q}^k$ are the $N_H \times N_H$ directional difference operators for $N_H \times 1$ images. Without loss of generality, in what follows, for convenience, we will use 1-D notation; in other words, we assume $\varepsilon^k = [\varepsilon_1^k \ \varepsilon_2^k \ \ldots \ \varepsilon_{N_H}^k]^T$. We also assume that the differences have Gaussian statistics according to $\varepsilon_i^k \sim N(0, (a_i^k)^{-1})$, for $i = 1, \ldots, N_H$ and $k = 1, 2$ where a_i^k is the inverse variance of ε_i^k.

For the inverse variances (i.e., the a_i^ks), we introduce the notation $\mathbf{A}^k = \text{diag}\{a_1^k, a_2^k, \ldots a_N^k\}$ a $N_H \times N_H$ diagonal matrix, $\tilde{\mathbf{A}} = \text{diag}\{\mathbf{A}^1, \mathbf{A}^2\}$ a $2N_H \times 2N_H$ diagonal matrix and $\tilde{\mathbf{a}} = [(\mathbf{a}^1)^T, (\mathbf{a}^2)^T]^T$ a $2N_H \times 1$ vector, consisting of two vectors $\mathbf{a}^k = [a_1^k, a_2^k, \ldots a_{N_H}^k]^T$. Also, for the differences we use the notation $\tilde{\varepsilon} = [(\varepsilon^1)^T(\varepsilon^2)^T]^T$. We assume that the differences in each direction and at each pixel location are independent. This

assumption makes subsequent calculations tractable. Thus, the joint density for the errors is Gaussian and is given as

$$p(\tilde{\boldsymbol{\varepsilon}}; \tilde{\mathbf{a}}) \propto \left(\prod_{k=1}^{2} \prod_{i=1}^{N_H} (a_i^k)^{1/2} \right) \exp \left(-0.5((\varepsilon^k)^{\mathrm{T}} \mathbf{A}^k \varepsilon^k) \right) =$$
$$\left(\prod_{k=1}^{2} \prod_{i=1}^{N_H} (a_i^k)^{1/2} \right) \exp \left(-0.5(\tilde{\varepsilon}^{\mathrm{T}} \tilde{\mathbf{A}} \tilde{\varepsilon}) \right).$$

To relate $\tilde{\boldsymbol{\varepsilon}}$ with the image $\mathbf{x}$ we define the $2N_H \times N$ operator $\tilde{\mathbf{Q}} = [(\mathbf{Q}^1)^{\mathrm{T}}, (\mathbf{Q}^2)^{\mathrm{T}}]^{\mathrm{T}}$. Then, the relation between the image and the differences is $\tilde{\varepsilon} = \tilde{\mathbf{Q}} \mathbf{x}$. Based on this relation and $p(\tilde{\varepsilon}; \tilde{\mathbf{a}})$ we can define an *improper prior* (one that does not integrate to 1) for the image $\mathbf{x}$ [15]. This prior is given by

$$p(\mathbf{x}; \tilde{\mathbf{a}}) \propto \prod_{k=1}^{2} \prod_{i=1}^{N_H} (a_i^k)^{1/4} \exp(-0.5((\tilde{\mathbf{Q}} \mathbf{x})^{\mathrm{T}} \tilde{\mathbf{A}} \tilde{Q} x))$$
$$= \prod_{k=1}^{2} \prod_{i=1}^{N_H} (a_i^k)^{1/4} \exp(-0.5((\mathbf{Q}^k \mathbf{x})^{\mathrm{T}} \mathbf{A}^k \mathbf{Q}^k \mathbf{x})). \quad (4)$$

The role of the parameters a_i^k is to capture the directional variation structure of the image. More specifically, a large variance (small a_i^k) indicates the presence of a large variation along the direction of the difference, in other words an edge perpendicular to this direction. The introduction of the spatially varying a_i^k scales down the differences of adjacent pixels in regions of image discontinuities. As a result this prior maintains edges and suppresses noise in smooth areas of the image.

The drawback of this prior, as described thus, far is that it introduces $2N_H$ parameters a_i^k that have to be estimated from PN observations. This is clearly not a desirable situation from an estimation point of view. To address this, we employ the Bayesian paradigm and consider a_i^k as random variables (instead of parameters) and introduce Gamma hyper-priors for them. In the case of a stationary model where all a_i^k are equal, the over-parameterization problem does not exist, and it is rather straightforward to obtain good estimates for the unknown parameters using maximum likelihood (ML).

We consider the following parameterization for the Gamma hyper-prior:

$$p\left(a_i^k; m_k, l_k\right) \propto \left(a_i^k\right)^{\frac{l_k-2}{2}} \exp \left\{ -m_k(l_k - 2) a_i^k \right\}. \quad (5)$$

For such a representation, the mean and variance of the Gamma pdf are given by $E[a_i^k] = l_k(2m_k(l_k - 2))^{-1}$, and $\mathrm{Var}[a_i^k] = l_k(2m_k^2(l_k - 2)^2)^{-1}$ respectively. This representation is used because the value of the parameter l_k can be also interpreted as the level of confidence to the prior knowledge provided by the Gamma hyper prior. More specifically, as $l_k \to \infty, E[a_i^k] \to (2m_k)^{-1}$ and $\mathrm{Var}[a_i^k] \to 0$. In other words, the prior becomes very informative and restrictive, resulting in $a_i^k = (2m_k)^{-1} \forall i$. In contrast, when $l_k \to 2$ then both $E[a_i^k] \to \infty$ and $\mathrm{Var}[a_i^k] \to \infty$, thus, in this case, the prior becomes uninformative and does not influence the values of the a_i^ks.

IV. PREPROCESSING STEP OF THE SUPER-RESOLUTION ALGORITHM

For this imaging model, the noncircular nature of the rotation matrix $\mathbf{R}$ renders computationally impractical simultaneous registration and restoration for large images. In contrast, all other matrices used in both the imaging and image prior model have characteristics that can be exploited in the DFT domain to render both tasks computationally very efficient. Particularly, the blurring $\mathbf{H}$ and shift matrices $\mathbf{S}$ are circulant, hence, diagonal in the DFT domain. As mentioned before, the decimation matrix $\mathbf{D}$, which is not circulant, has a convenient structure in the DFT domain that helps bypass computational difficulties. Finally, matrices $\mathbf{Q}^k$ and $\mathbf{A}^k$ of the image prior are circulant and diagonal, respectively. For this combination, one can exploit the diagonal structure by *alternating* calculations in the DFT and spatial domain.

To bypass the problems with the rotation, a preprocessing step is performed before the super-resolution algorithm. In this step, we estimate the registration parameters between the low-resolution observations. At this point it is important to notice that the rotations between the degraded low-resolution and the high-resolution images of the imaging model in (2) are the same. However, the shifts of the low-resolution images must be also multiplied by the decimation factor. Thus, these parameters in the preprocessing step will be called δ_i' and $\theta_i, i = 2, \ldots, P$, for translation and rotation, respectively. Using this notation, we assume that image $\mathbf{y}_i$ resulted by applying both translation and rotation with respect to the first image $\mathbf{y}_1$ (or the reverse). In other words, we have

$$\mathbf{y}_i = \mathbf{S}'(\delta_i') \mathbf{R}'(\theta_i) \mathbf{y}_1 \quad \text{or} \quad \mathbf{y}_1 = \mathbf{S}'(-\delta_i') \mathbf{R}'(-\theta_i) \mathbf{y}_i$$

where $\mathbf{R}'$ and $\mathbf{S}'$ are the $N \times N$ rotation and shift matrices respectively, smaller than their respective $N_H \times N_H$ matrices $\mathbf{R}$ and $\mathbf{S}$. Thus, image $\mathbf{y}_1$ is considered as the reference image.

We define the vector that represents the difference between the registered image i and the reference image to be

$$\mathbf{L}_i = \mathbf{S}'(-\delta_i') \mathbf{R}'(-\theta_i) \mathbf{y}_i - \mathbf{y}_1, \quad \text{for } i = 2, \ldots, P.$$

In this registration preprocessing step, we aim to estimate the registration parameters by minimizing the quantity in the following equation:

$$[\hat{\delta}_i', \hat{\theta}_i] = \underset{[\delta_i', \theta_i]}{\arg \min} \|\mathbf{L}_i\|_2^2, \quad \text{for } i = 2, \ldots, P.$$

The minimization is achieved using the simplex search method [17]. Having computed the registration parameters δ_i' and θ_i, at the end of the preprocessing step the low-resolution observations $\mathbf{y}_i$ are replaced by the "almost-registered" low-resolution images given by

$$\mathbf{z}_i = \mathbf{S}'(\mathrm{int}[\hat{\delta}_i' + 0.5]) \mathbf{R}'(\hat{\theta}_i) \mathbf{y}_i, \quad i = 2, \ldots, P$$

where $\mathrm{int}[\cdot]$ denotes the integer part of the real number. This is intentional because low-resolution images that are shifted by a

fraction of a pixel are required in order to achieve super-resolution reconstruction [3]. For the rest of the paper, we assume as observations the registered versions $\mathbf{z}_i$ of the initially observed images $\mathbf{y}_i$. We also define as $\mathbf{z}$ the vector that contains all the $\mathbf{z}_i$ as

$$\mathbf{z} = \begin{bmatrix} \mathbf{z}_1^{\mathrm{T}} & \mathbf{z}_2^{\mathrm{T}} & \cdots & \mathbf{z}_P^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}}.$$

In this way, the rotation is removed from the observations. Thus, the rotation matrices can be omitted from the imaging model used for super-resolution reconstruction, described in the next section.

V. MAXIMUM A POSTERIORI (MAP) RECONSTRUCTION

The super-resolution image $\mathbf{x}$ is estimated from the observations $\mathbf{z}$, (after the preprocessing step), utilizing a MAP approach in which we estimate simultaneously $\tilde{\mathbf{a}}, \mathbf{x}$, and $\zeta = [\zeta_2 \cdots \zeta_P]^{\mathrm{T}}$ where the registration parameters have changed to according to $\delta_i = \delta_i' + \zeta_i$. At this point, we must note that even in the absence of noise

$$\mathbf{D}\mathbf{H}_i\mathbf{S}(\zeta_i)\mathbf{x} \neq \mathbf{z}_i = \mathbf{S}'(\mathrm{int}[\hat{\delta}_i' + 0.5])\mathbf{R}'(\hat{\theta}_i)\mathbf{y}_i, \text{ for } i = 1, \ldots, P.$$

To correct this, we make the assumption that the coarsely registered $\mathbf{z}_i$, using rotation and translation, satisfies the equation $\mathbf{z}_i = \mathbf{D}\mathbf{H}_i\mathbf{S}(\delta_i)\mathbf{x} + \mathbf{n}_i'$ where $\mathbf{n}_i'$ is an error term. Thus, the imaging model that is finally solved by the MAP algorithm is

$$\mathbf{z}_i = \mathbf{D}\mathbf{H}_i\mathbf{S}(\zeta_i)\mathbf{x} + \mathbf{w}_i, \quad \text{for } i = 1, \ldots, P$$

where $\mathbf{w}_i = \mathbf{n}_i' + \mathbf{n}_i$ the new error term which is assumed WGN with precision b_i.

MAP estimation is based on maximization of the posterior probability. Thus, based on Bayes' theorem, we have

$$p(\mathbf{x}, \tilde{\mathbf{a}}|\mathbf{z}; \mathbf{b}, \mathbf{m}, \mathbf{l}, \zeta) \propto p(\mathbf{z}, \mathbf{x}, \tilde{\mathbf{a}}; \mathbf{b}, \mathbf{m}, \mathbf{l}, \zeta)$$
$$= p(\mathbf{z}|\mathbf{x}, \tilde{\mathbf{a}}; \mathbf{b}, \zeta)p(\mathbf{x}|\tilde{\mathbf{a}})p(\tilde{\mathbf{a}}; \mathbf{m}, \mathbf{l})$$

where

$$\mathbf{m} = [m_1, m_2, m_3, m_4]^{\mathrm{T}}, \mathbf{l} = [l_1, l_2, l_3, l_4]^{\mathrm{T}}, \mathbf{b} = [b_1 \cdots b_P].$$

Maximizing the quantity $p(\mathbf{x}, \tilde{\mathbf{a}}|\mathbf{z}; \mathbf{b}, \mathbf{m}, \mathbf{l}, \zeta)$ with respect to $\mathbf{x}, \tilde{\mathbf{a}}$, and ζ is equivalent to minimizing the negative logarithm

$$\begin{aligned}
J_{\mathrm{MAP}}(\mathbf{x}, \tilde{\mathbf{a}}, \zeta) \\
&= -\log p(\mathbf{z}, \mathbf{x}, \tilde{\mathbf{a}}; \mathbf{b}, \mathbf{m}, \mathbf{l}, \zeta) = -\log p(\mathbf{z}|\mathbf{x}, \tilde{\mathbf{a}}; \mathbf{b}, \zeta) \\
&\quad -\log p(\mathbf{x}|\tilde{\mathbf{a}}) - \log p(\tilde{\mathbf{a}}; \mathbf{m}, \mathbf{l}) \\
&= -\frac{N}{2}\sum_{i=1}^{P}\log b_i + \frac{1}{2}\sum_{i=1}^{P}b_i\|\mathbf{B}_i(\zeta_i)\mathbf{x} - \mathbf{z}_i\|_2^2 \\
&\quad -\frac{1}{4}\sum_{k=1}^{2}\sum_{i=1}^{N}\log a_i^k + \frac{1}{2}\sum_{k=1}^{2}\sum_{i=1}^{N}(\mathbf{Q}^k\mathbf{x})^{\mathrm{T}}\mathbf{A}^k\mathbf{Q}^k\mathbf{x} \\
&\quad -\sum_{k=1}^{2}\left(\frac{l_k - 2}{2}\sum_{i=1}^{N}\log a_i^k\right) \\
&\quad + \sum_{k=1}^{2}\left(m_k(l_k - 2)\sum_{i=1}^{N}a_i^k\right).
\end{aligned} \quad (6)$$

To minimize the above function with respect to $\mathbf{x}, \tilde{\mathbf{a}}$, and ζ, we adopt an iterative scheme that sets alternatively the gradient with respect to $\mathbf{x}, \tilde{\mathbf{a}}$, and ζ equal to zero.

Setting $\nabla_{\tilde{\mathbf{a}}}J_{\mathrm{MAP}}(\mathbf{x}, \tilde{\mathbf{a}}, \zeta) = 0$ yields

$$\left(a_i^k\right)^* = \frac{\left(\frac{1}{4} + \frac{1}{2}(l_k - 2)\right)}{\left(\frac{1}{2}\left(\varepsilon_i^k\right)^2 + m_k(l_k - 2)\right)}. \quad (7)$$

The observation of the previous section that the parameters l_k express the degree of confidence to the prior can be viewed from another point when looking (7), the MAP estimates of the (a_i^k). More specifically, when $l_k \to \infty, (a_i^k)^* = (2m_k)^{-1}\forall i$; thus, the $(a_i^k)^*\forall i$ are equal, and the image model becomes *stationary*. In contrast, when $l_k \to 2, (a_i^k)^* = ((\varepsilon_i^k)^2)^{-1}\forall i$; thus, the $(a_i^k)^*$s are completely unaffected from the moderating effect of the Gamma hyper-prior and only follow the data. In this case, the image model can be viewed as the "most nonstationary."

Setting $\nabla_{\mathbf{x}}J_{\mathrm{MAP}}(\mathbf{x}, \tilde{\mathbf{a}}, \zeta) = 0$ yields

$$\begin{aligned}
\mathbf{x}^* = \left(\sum_{i=1}^{P} b_i\mathbf{B}_i^{\mathrm{T}}(\zeta_i)\mathbf{B}_i(\zeta_i) + \sum_{k=1}^{2}(\mathbf{Q}^k)^{\mathrm{T}}\mathbf{A}^k\mathbf{Q}^k\right)^{-1} \\
\times \sum_{i=1}^{P} b_i\mathbf{B}_i^{\mathrm{T}}(\zeta_i)\mathbf{z}_i. \quad (8)
\end{aligned}$$

Equation (8) cannot be solved in closed form since analytical inversion of $\sum_{i=1}^{P} b_i\mathbf{B}_i^{\mathrm{T}}(\zeta_i)\mathbf{B}_i(\zeta_i) + \sum_{k=1}^{2}(\mathbf{Q}^k)^{\mathrm{T}}\mathbf{A}^k\mathbf{Q}^k$ is not possible due to the noncircular nature of the matrices $\mathbf{B}_i$ and $\mathbf{A}^k$. Thus, we resort to a numerical solution using a conjugate gradient algorithm [16]. In this algorithm, the space and DFT domains are alternated when expressions with circulant and diagonal matrices are computed. More specifically, multiplications with circulant matrices (convolutions) are performed in the DFT domain while multiplications with diagonal matrices are performed in the space domain.

In the case of the registration parameters, it is not possible to find in closed form the ζ that make the gradient $\nabla_{\zeta}J_{\mathrm{MAP}}$ equal to zero, or equivalently to minimize the quantity J_{MAP} with respect to ζ

$$\zeta^* = \arg\min_{\zeta} J_{\mathrm{MAP}}(\mathbf{x}, \tilde{\mathbf{a}}, \zeta)$$

which can also be written as

$$\zeta_i^* = \arg\min_{\zeta_i} J_{\mathrm{MAP}}(\zeta_i) = \arg\min_{\zeta_i}\|\mathbf{B}_i(\zeta_i)\mathbf{x} - \mathbf{z}_i\|_2^2 \quad (9)$$

where $J_{\mathrm{MAP}}(\zeta_i)$ denotes the part of $J_{\mathrm{MAP}}(\mathbf{x}, \tilde{\mathbf{a}}, \zeta)$ that depends on ζ_i. Since ζ_i^* cannot be found in closed form we resort to the NP algorithm. This method is chosen due to its convergence speed [16]. Registration is equivalent to the minimization task in (9). By the definition of the matrix $\mathbf{B}_i$ with $\mathbf{R}(\theta_i) = \mathbf{I}$, in (9) is

$$\begin{aligned}
J_{\mathrm{MAP}}(\zeta_i) = 2\mathbf{z}_i^{\mathrm{T}}\mathbf{D}\mathbf{S}(\zeta_i)\mathbf{H}_i\mathbf{x} \\
+ \mathbf{x}^{\mathrm{T}}\mathbf{H}_i^{\mathrm{T}}\mathbf{S}^{\mathrm{T}}(\zeta_i)\mathbf{D}^{\mathrm{T}}\mathbf{D}\mathbf{S}(\zeta_i)\mathbf{H}_i\mathbf{x} + \mathbf{z}_i^{\mathrm{T}}\mathbf{z}_i. \quad (10)
\end{aligned}$$

The DFT domain is used to evaluate (10), since it allows easy analytic calculations of the first and second derivatives of the objective function. Since the shift parameters are independent

with each other, it is sufficient to demonstrate the derivatives for one ζ_i. The details of the derivative calculation of $J_{\mathrm{MAP}}(\zeta_i)$ are given in the Appendix. With the derivatives calculated, the update scheme of the NR algorithm is

$$\zeta_i^{n+1} = \zeta_i^n - \frac{\partial J_{\mathrm{MAP}}\left(\zeta_i^n\right)}{\partial \zeta_i} \left(\frac{\partial^2 J_{\mathrm{MAP}}\left(\zeta_i^n\right)}{\partial \zeta_i^2}\right)^{-1}. \quad (11)$$

The shift parameters are initialized as $\zeta_{\mathrm{init}} = \hat{\boldsymbol{\delta}}' \cdot d$, where $\hat{\boldsymbol{\delta}}' = [\hat{\delta}_2', \ldots, \hat{\delta}_P']^{\mathrm{T}}$ are the shift parameters estimated in the preprocessing step (Section IV) and d is the decimation factor. This initialization provides starting values close to the solution, which is essential for the convergence of the NP algorithm to the correct solution [16].

The shift parameters are initialized as $\boldsymbol{\delta}_{\mathrm{init}} = \hat{\boldsymbol{\delta}}' \cdot d$, where $\hat{\boldsymbol{\delta}}' = [\hat{\delta}_2', \ldots, \hat{\delta}_P']^{\mathrm{T}}$ are the shift parameters estimated in the preprocessing step (Section IV) and d is the decimation factor. This initialization provides starting values close to the solution, which is essential for the convergence of the NR algorithm to the correct solution [16].

VI. Experiments

In order to test the proposed methodology, we used both artificially generated and real data. We compared the new MAP super-resolution algorithm with the nonstationary prior with the E-M super-resolution algorithm in [11] that uses a stationary prior. We also compared our super-resolution algorithm with one that uses total variation (TV) regularization [8]. For this comparison, a gradient descent algorithm was used given by

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - \beta \left\{ \mathbf{B}^T \left(\mathbf{B}\mathbf{x}^{(k)} - \mathbf{z}\right) \right. \\ \left. + \lambda \left((\mathbf{Q}^1)^T \mathbf{v}_1^{(k)} + (\mathbf{Q}^2)^T \mathbf{v}_2^{(k)}\right) \right\} \quad (12)$$

where the superscript (k) denotes the iteration number, $\mathbf{v}_m(i) = \mathrm{sign}(\varepsilon^m(i))$, with $\boldsymbol{\varepsilon}^1 = \mathbf{Q}^1\mathbf{x}$ and $\boldsymbol{\varepsilon}^2 = \mathbf{Q}^2\mathbf{x}$, the first-order horizontal and vertical differences of the image, λ the regularization parameter and β the step of the algorithm. In the following experiments, the λ and β parameters were selected by trial and error to provide the best possible results. This is a difficult task. However, in general, as λ increases the image becomes blurrier and the algorithm converges for smaller step β. For all methods, we used the same registration algorithm. The results generated by (12) are not a comparison with the methodology presented in [8] since although similar priors are used the other aspects of the super-resolution algorithm (registration, chosen PSF) are different. Nevertheless, the authors of [8] have published results with the herein used data sets in [21] and [22] where the interested reader can resort. In the preprocessing step of the herein proposed algorithm, the interpolation algorithm in [20] was used for rotation in order to handle boundary artifacts.

In the first experiment, eight 128×128 low-resolution images were generated by performing translation and rotation to the well-known "Cameraman" image of size 256×256, before blurring and then down-sampling by a factor of 2. The PSF of the blur was uniform 5×5. Last, noise was added, corresponding to $\mathrm{SNR} = 20$ (the same for all images). This metric is

Fig. 1. Low-resolution degraded observation.

defined as $\mathrm{SNR} = 10\log\left(\|\mathbf{x}\|_2^2\right)/(N_H\sigma^2)$ where σ^2 is the variance of the additive noise and N_H is the size of the image $\mathbf{x}$.

The mean square error (MSE) metric between the restored image and the original was used to evaluate the performance of the algorithm. The MSE is defined as $\mathrm{MSE} = (\|\mathbf{x} - \hat{\mathbf{x}}\|_2^2)/(N_H)$, where $\mathbf{x}$ and $\hat{\mathbf{x}}$ are the original and estimated images, respectively.

Fig. 1 shows one of the observed low-resolution degraded image.

In Fig. 2(a)–(c), we show the super-resolved images and the corresponding MSEs, using the stationary prior in [11], TV regularization as implemented in (12), and the new algorithm based on the nonstationary prior, respectively. Also, to demonstrate the robustness of the proposed registration methodology, we show the true and the estimated registration parameters in Table I. We observed in all the experiments we performed with simulated data that the proposed preprocessing step estimated the rotation parameters with an accuracy of almost four decimal digits in degrees. The reconstructed super-resolved images assuming knowledge of the registration parameters are almost identical to their reconstructed counterparts using the estimated parameters. From these experiments, we can draw two conclusions. First, the proposed nonstationary prior improves the reconstruction of the high-resolution images. Indeed, the MSE using the nonstationary model is significantly lower apart from the difference in the visual quality of the images. Second, the proposed two step registration methodology seems very accurate (when the image formation model is correct).

We also used the proposed super-resolution algorithm on two real data sets. The first contains 20 low-resolution degraded images. In Fig. 3, one of these images is shown. Their original size was slightly smaller than 64×64, so they were padded with zeros, extending their size exactly to 64×64 pixels. In this data set the low-resolution images were only translated and did not contain any rotations. Super-resolved images of double size $(2x)$ are shown in Fig. 4(a)–(c) using the stationary, TV regularization and the nonstationary priors, respectively.

The second set includes four low-resolution degraded images that contain both translations and rotations and one of them is shown in Fig. 5(a). Each low-resolution image is of size 128×128. In order to test the ability of the proposed priors to reconstruct beyond the resolution of the available data, we quadrupled $(4x)$ the size of the reconstructed super-resolved

(a)

(b)   (c)
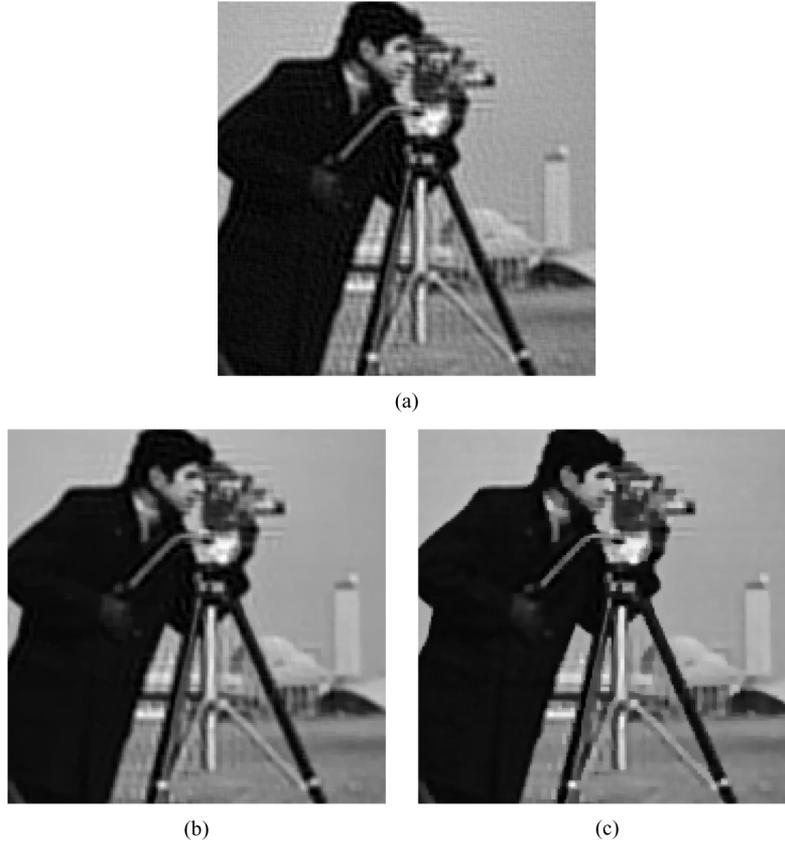
Fig. 2.   (a) Stationary 2x [11], $(\mathrm{MSE} = 195.2)$; (b) total variation, (12) $(\mathrm{MSE} = 182.1)$; (c) nonstationary MAP 2x super-resolved image $(\mathrm{MSE} = 162.4)$.

TABLE I
ORIGINAL AND ESTIMATED PARAMETERS FOR
THE ARTIFICIALLY GENERATED IMAGES

	θ	$\hat{\theta}$	(δ^x, δ^y)	$(\hat{\delta}^x, \hat{\delta}^y)$
Im 2	1.00	0.99	(0.30, -0.20)	(0.31, -0.21)
Im 3	2.00	2.00	(0.10, -0.30)	(0.14, -0.32)
Im 4	3.00	3.00	(-0.20, 0.10)	(-0.15, 0.06)
Im 5	1.50	1.50	(-0.15, 0.25)	(-0.13, 0.23)
Im 6	0.50	0.48	(0.00, 0.10)	(0.00, 0.09)
Im 7	-1.00	-1.00	(0.05, 0.12)	(0.04, 0.14)
Im 8	-2.00	-2.02	(0.14, 0.32)	(0.13, 0.35)
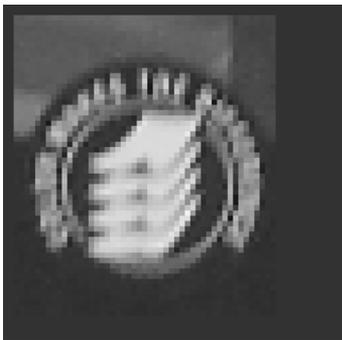
Fig. 3.   Sample of low-resolution observation.

images. The $4x$, super-resolved images with the stationary, TV regularization, and nonstationary prior are shown in Fig. 6(a)–(c), respectively.

In estimating the shape of the blur for the real data sets, a Gaussian-shaped blur was assumed. This choice was motivated by the observation that Gaussian shaped functions are smooth and have good approximation properties. The width of each blur was experimentally estimated using trial and error experiments. The width is captured by the variance of Gaussian PSF. For the first set, the values of the variances of the Gausian shaped PSFs were found in the range [2.5–4] pixels and for the second the variance was set equal to 4.

To facilitate learning the proposed image model, we used equal b_i^{-1} for all i (additive noise variances) and equal m_k for all k obtained by learning a stationary SAR model [11]. The parameters m_k were obtained as $m_k = 1/(2a_{\mathrm{STAT}})$ where a_{STAT} the image model parameter of the stationary SAR model. The parameters l_k were selected to be equal to $l = 2.1$ for the reconstruction of both the real data and synthetic data. This value was found by trial and error experiments. We observed that as $l \to 2$ the reconstructed images assume a "cartoon" like appearance where large edges are preserved and areas with small variations are flattened out. When $l \to \infty$, as also explained previously, the reconstructed images assume the appearance of images that were reconstructed by a stationary prior model. In other words, at the expense of ringing in edges and noise amplification in smooth areas, textured areas can be better reconstructed. The selection of $l = 2.1$ reflects our subjective choice between these two opposing trends. For the case of the TV regularization, the algorithm's parameters were also found by trial and error. We set for the first experiment $\beta = 0.05, \lambda = 1$, for the second $\beta = 0.01, \lambda = 1$, and for the last one $\beta = 0.1, \lambda = 2$.

(a)

(b)

(c)
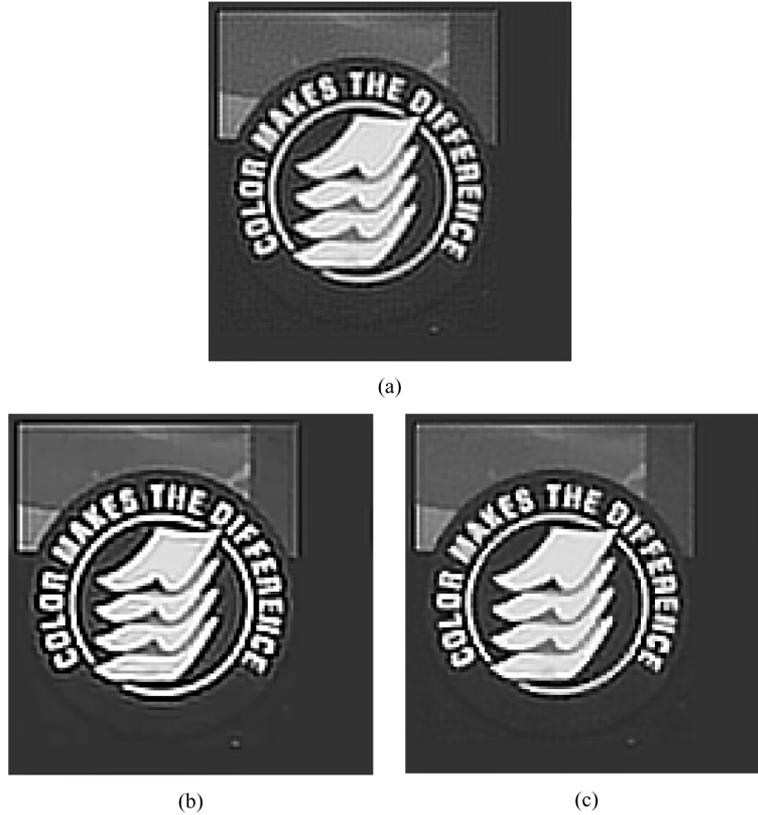
Fig. 4   Super-resolved images: (a) $2x$ stationary [11]; (b) $2x$ total variation, (12); (c) $2x$ MAP nonstationary.

Fig. 5.   Sample of low-resolution observations.

The super-resolution estimates of $\mathbf{x}, a_i^k$, and ζ_i were found by iterating between (7), (8), and (11) till convergence. In the presented experiments, the convergence criterion was $(\|\mathbf{x}^t - \mathbf{x}^{t+1}\|_2)/(\|\mathbf{x}^t\|_2) < ((10^{-3})/(\overline{b}))$ where t denotes the iteration number and $\overline{b}$ is the average of the inverse noise variances $\overline{b} = (1/P)\sum_{i=1}^{P} b_i$.

Finally, we would like to note that the MAP function in (6), although derived using a completely different principle, can be viewed as a half-quadratic function that is generated using a φ_{HL}-like potential function [18, Table II, pp. 302] with appropriate choice of parameters, for details see [15]. The convergence of alternating direction minimization of half quadratic

functions has been rigorously shown in [18]. It has been shown that if the generating potential function is strictly convex, and the null spaces of matrices $\mathbf{B}$ and $\mathbf{Q}^k$ do not intersect, the MAP function is convex. However, the φ_{HL}-like potential function used herein is not convex; thus, the proposed alternating direction minimization converges to a local minimum. For this reason, good initialization of the algorithm is important.

VII. CONCLUSIONS AND FUTURE WORK

Inspection of the super-resolved images in Figs. 2(a)–(c), 4(a)–(c), and 6(a)–(c) reveal that the resolution in every case has significantly been improved. The letters in the super resolved images [Figs. 4(a)–(c) and 6(a)–(c)] are now easily legible.

Furthermore, the images reconstructed using the proposed nonstationary prior, Figs. 2(c), 4(c), and 6(c), are visually more pleasant and display less ringing at the edges as compared to both stationary and TV based super-resolution reconstruction. The MSE for the reconstructed images using nonstationary prior is also smaller than both the stationary and the TV based models. It is worth noticing that for the first experiment, the MSE results when using the real registration parameters are almost identical to that when the registration parameters are estimated. This demonstrates the robustness of the proposed algorithm regarding the registration parameters.

In what follows, we report implementation times for the "Cameraman" experiment. Registration in the preprocessing step requires 26 min. One iteration of the stationary model based algorithm requires 4–5 s with almost 4 s the time for fast subpixel registration. One iteration of the nonstationary MAP

(a)

(b)

(c)
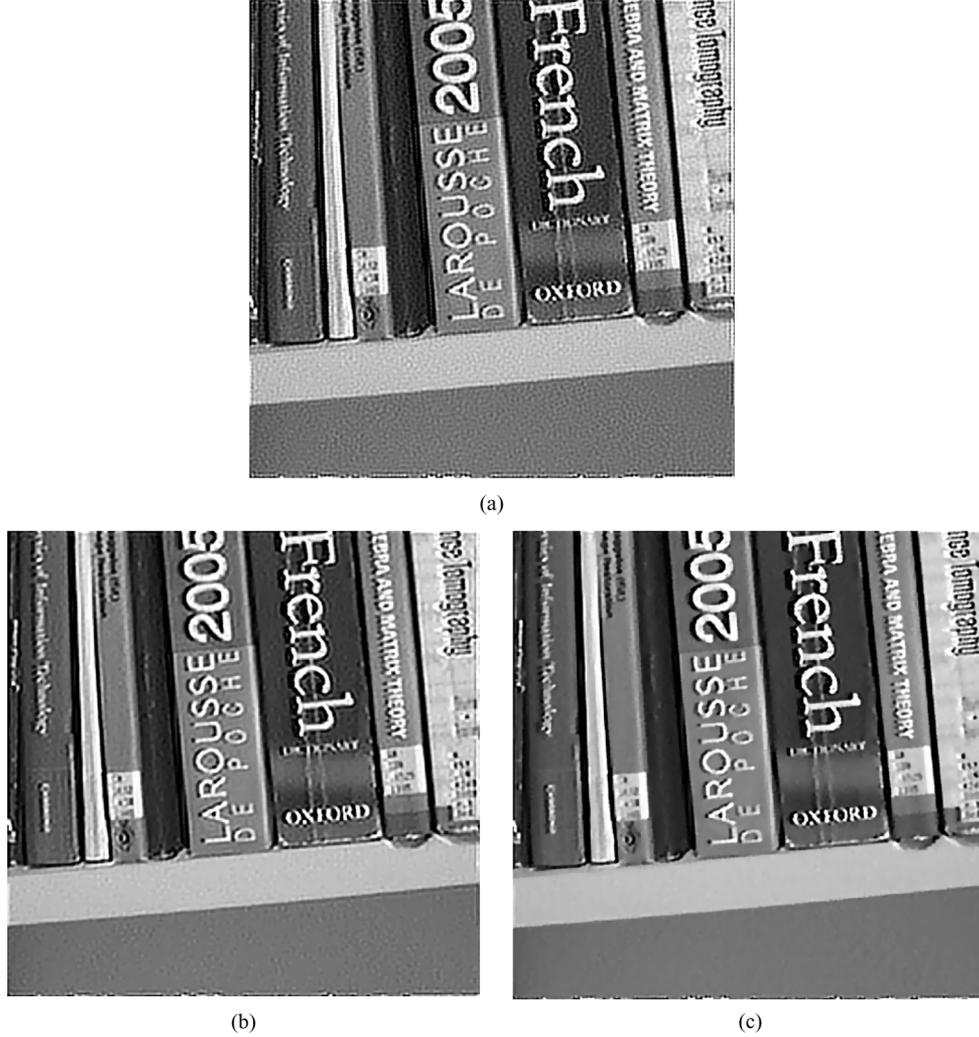
Fig. 6   Super-resolved images: (a) $4x$ stationary [11]; (b) $4x$ total variation, (12); (c) $4x$ MAP nonstationary.

algorithm requires about 38 s, out of which 4 s are required for fast subpixel registration, and the rest for 40 iterations of the conjugate gradient algorithm in (8). The TV algorithm requires about 1 s per iteration of the gradient algorithm in (12). These times we obtained using a Pentium 4 3.4-GHz PC and a Matlab implementation.

In the future, we plan to include a PSF estimation step in the formulation of this problem, as well as faster rotation estimation in the preprocessing step. Furthermore, methodologies to better model the statistical nature of the errors if the imaging model is not accurate will be considered.

APPENDIX
DERIVATIVES OF $J_{\mathrm{MAP}}(\zeta_i)$ IN THE DFT DOMAIN

Assume the $N_H \times N_H$ DFT matrix $\mathbf{W}_1$ and the $N \times N$ DFT matrix $\mathbf{W}_2$. Then $\mathbf{X} = \mathbf{W}_1\mathbf{x}$, and $\mathbf{Z}_i = \mathbf{W}_2\mathbf{z}_i$ are the DFTs of the vectors $\mathbf{x}$ and $\mathbf{z}_i$, respectively. The matrices $\Lambda_{\mathbf{S}} = \mathbf{W}_1\mathbf{S}(\delta_i)\mathbf{W}_1^{-1}$ and $\Lambda_{\mathbf{H}} = \mathbf{W}_1\mathbf{H}_i\mathbf{W}_1^{-1}$ are diagonal due to the

circulant nature of the matrices $\mathbf{S}(\delta_i)$ and $\mathbf{H}_i$. It can also be shown that

$$\Lambda_{\mathbf{D}} = \mathbf{W}_2\mathbf{D}\mathbf{W}_1^{-1} = [\mathbf{I}_1, \quad \mathbf{I}_2, \quad \dots, \mathbf{I}_d]/d \qquad (A1)$$

is a $N \times N_H$ block matrix that contains d identity matrices of size $N \times N$. Then, we can write

$$J_{\mathrm{MAP}}(\zeta_i) = 2\,\mathrm{real}\left\{\mathbf{Z}_i^{\mathrm{H}}\Lambda_{\mathbf{D}}\Lambda_{\mathbf{S}}\Lambda_{\mathbf{H}}\mathbf{X}\right\}$$
$$+\mathbf{X}^{\mathrm{H}}\Lambda_{\mathbf{H}}^*\Lambda_{\mathbf{S}}^*\Lambda_{\mathbf{D}}^{\mathrm{T}}\Lambda_{\mathbf{D}}\Lambda_{\mathbf{S}}\Lambda_{\mathbf{H}}\mathbf{X} + \mathbf{Z}_i^{\mathrm{H}}\mathbf{Z}_i \quad (A2)$$

where the symbol H denotes the Hermitian and * denotes the conjugate. For simplicity, the diagonal element of a matrix is denoted as $[m]$. Then, we can write

$$J_{\mathrm{MAP}}(\delta_i) \propto 2\,\mathrm{real}\left\{\sum_{m=1}^{N}\mathbf{Z}_i^*[m]\mathbf{T}_i[m]\right\} + \sum_{m=1}^{N}\mathbf{T}_i^*[m]\mathbf{T}_i[m]$$
$$(A3)$$

where $\mathbf{T}_i[m]$ are the elements of the $N \times 1$ vector $\mathbf{T}_i$ and they are

$$
\mathbf{T}_i[m]
$$
$$
\equiv \frac{\sum_{n=0}^{d-1} \left(\Lambda_{\mathbf{S}} \left[m + \frac{nN}{d} \right] \Lambda_{\mathbf{H}} \left[m + \frac{nN}{d} \right] \mathbf{X} \left[m + \frac{nN}{d} \right] \right)}{d}. \tag{A4}
$$

The evaluation of the first and second derivatives of (A2) is very convenient in the DFT domain since the parameter ζ_i is only in the diagonal elements of the matrix $\Lambda_{\mathbf{S}}$. These elements, see for example [3], are equal to

$$
\Lambda_{\mathbf{S}}[m] = \exp\{-2j\pi\zeta_i(m-1)/N\}
$$
$$
\text{for } m = 1, \ldots, N/2
$$

where $j^2 = -1$. The remaining elements are a "mirrored" version of the previous ones; in other words

$$
\Lambda_{\mathbf{S}}[m] = \exp\{-2j\pi\zeta_i(N-m+1)/N\}
$$
$$
\text{for } m = \frac{N}{2} + 1, \ldots, N.
$$

The first and second derivatives for the first half are, respectively

$$
\frac{\partial \Lambda_{\mathbf{S}}[m]}{\partial \zeta_i} = \frac{-2j\pi(m-1)}{N} \exp\{-2j\pi\zeta_i(m-1)/N\}
$$
$$
\text{for } m = 1, \ldots, N/2 \tag{A5}
$$
$$
\frac{\partial^2 \Lambda_{\mathbf{S}}[m]}{(\partial \zeta_i)^2} = \frac{-4\pi^2}{N^2}(m-1)^2 \exp\{-2j\pi\zeta_i(m-1)/N\}
$$
$$
\text{for } m = 1, \ldots, N/2 \tag{A6}
$$

and for the second half

$$
\frac{\partial \Lambda_{\mathbf{S}}[m]}{\partial \zeta_i}
$$
$$
= \frac{-2j\pi(N-m+1)}{N} \exp\{-2j\pi\zeta_i(N-m+1)/N\}\text{for}
$$
$$
\text{for } m = \frac{N}{2} + 1, \ldots, N \tag{A7}
$$
$$
\frac{\partial^2 \Lambda_{\mathbf{S}}[m]}{(\partial \zeta_i)^2}
$$
$$
= \frac{-4\pi^2}{N^2}(N-m+1)^2 \exp\{-2j\pi\zeta_i(N-m+1)/N\}
$$
$$
\text{for } m = \frac{N}{2} + 1, \ldots, N. \tag{A8}
$$

The derivative of the terms in (A2) are given by applying (A5)–(A8)

$$
\frac{\partial J_{\mathrm{MAP}}(\zeta_i)}{\partial \zeta_i} = \sum_{m=1}^{N} \mathbf{Z}_i^*[m] \frac{\partial \mathbf{T}_i[m]}{\partial \zeta_i} + \sum_{m=1}^{N} \mathbf{Z}_i[m] \frac{\partial \mathbf{T}_i^*[m]}{\partial \zeta_i}
$$
$$
+ \sum_{m=1}^{N} \mathbf{T}_i^*[m] \frac{\partial \mathbf{T}_i[m]}{\partial \zeta_i} + \sum_{m=1}^{N} \frac{\partial \mathbf{T}_i^*[m]}{\partial \zeta_i} \mathbf{T}_i[m].
$$

From the definition of $\mathbf{T}_i$, it is

$$
\frac{\partial \mathbf{T}_i[m]}{\partial \zeta_i}
$$
$$
= \sum_{n=0}^{d-1} \frac{\partial \Lambda_{\mathbf{S}} \left[m + \frac{nN}{d} \right]}{\partial \zeta_i} \Lambda_{\mathbf{H}} \left[m + \frac{nN}{d} \right] \mathbf{X} \left[m + \frac{nN}{d} \right]
$$

and

$$
\frac{\partial \mathbf{T}_i^*[m]}{\partial \zeta_i}
$$
$$
= \sum_{n=0}^{d-1} \frac{\partial \Lambda_{\mathbf{S}}^* \left[m + \frac{n}{d}N \right]}{\partial \zeta_i} \Lambda_{\mathbf{H}}^* \left[m + \frac{n}{d}N \right] \mathbf{X}^* \left[m + \frac{n}{d}N \right].
$$

Similarly, the second derivative is

$$
\frac{\partial^2 J_{\mathrm{MAP}}(\zeta_i)}{\partial \zeta_i^2}
$$
$$
= \sum_{m=1}^{N} \mathbf{Z}_i^*[m] \frac{\partial^2 \mathbf{T}_i[m]}{\partial \zeta_i^2} + \sum_{m=1}^{N} \mathbf{Z}_i[m] \frac{\partial^2 \mathbf{T}_i^*[m]}{\partial \zeta_i^2}
$$
$$
+ \sum_{m=1}^{N} \mathbf{R}_i^*[m] \frac{\partial^2 \mathbf{T}_i[m]}{\partial \zeta_i^2}
$$
$$
+ 2\sum_{m=1}^{N} \left| \frac{\partial \mathbf{T}_i[m]}{\partial \zeta_i} \right|^2 + \sum_{m=1}^{N} \frac{\partial^2 \mathbf{T}_i^*[m]}{\partial \zeta_i^2} \mathbf{T}_i[m].
$$

To be precise, in our application, we deal with 2-D signals where here are two translations parameters per image $\vec{\zeta}_i = \left(\zeta_i^x, \zeta_i^y \right)$. Thus, in the NR update equation, (10) $(\partial J_{\mathrm{MAP}}(\vec{\zeta}_i^n))/(\partial \vec{\zeta}_i)$ is a 2×1 gradient vector and $((\partial^2 J_{\mathrm{MAP}}(\vec{\zeta}_i^n))/(\partial \vec{\zeta}_i^2))$ a 2×2 Hessian matrix involved. However, the inversion of a 2×2 matrix is easily found in closed form; hence, the 2-D version of the registration algorithm is also very fast.

ACKNOWLEDGMENT

REFERENCES

[1] S. Park, M. Park, and M. Kang, "Super-resolution image reconstruction: A technical overview," *IEEE Signal Process. Mag.*, vol. 20, no. 3, pp. 21–36, May 2003.
[2] S. Farsiu, D. Robinson, M. Elad, and P. Milanfar, "Advances and challenges in super-resolution," *Int. J. Imag. Syst. Technol.*, vol. 14, no. 2, pp. 47–57, Aug. 2004.
[3] S. Chaudhuri, Ed., "Super-resolution imaging," in *Reconstruction of a High Resolution Image from Multiple Low Resolution Images*. B. Tom, N. Galatsanos, and A. Katsaggelos, Eds. Norwell, MA: Kluwer, 2001, ch. 4, pp. 71–105.

[4] W. T. Freeman, T. R. Jones, and E. C. Pasztor, "Example-based super-resolution," *IEEE Comput. Graph. Appl.*, vol. 22, no. 2, pp. 56–65, Mar.–Apr. 2002.

[5] M. V. Joshi, S. Chaudhuri, and R. Panuganti, "A learning-based method for image super-resolution from zoomed observations," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 35, no. 3, pp. 527–537, Jun. 2005.

[6] G. Demoment, "Image restoration and reconstruction: Overview of common estimation structures and problems," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 12, pp. 2024–2036, Dec. 1989.

[7] C. A. Segall, A. K. Katsaggelos, R. Molina, and J. Mateos, "Bayesian resolution enhancement of compressed video," *IEEE Trans. Image Process.*, vol. 13, no. 7, pp. 898–911, Jul. 2004.

[8] S. Farsiu, M. D. Robinson, M. Elad, and P. Milanfar, "Fast and robust multi-frame super resolution," *IEEE Trans. Image Process.*, vol. 13, no. 10, pp. 1327–1344, Oct. 2004.

[9] T. Akgun, Y. Altunbasak, and R. M. Mersereau, "Super-resolution reconstruction of hyperspectral images," *IEEE Trans. Image Process.*, vol. 14, no. 11, pp. 1860–1875, Nov. 2005.

[10] S. Farsiu, M. Elad, and P. Milanfar, "Multiframe demosaicing and super-resolution of color images," *IEEE Trans. Image Process.*, vol. 15, no. 1, pp. 141–159, Jan. 2006.

[11] N. A. Woods, N. P. Galatsanos, and A. K. Katsaggeloss, "Stochastic methods for joint restoration, interpolation and registration of multiple under sampled images," *IEEE Trans. Image Process.*, vol. 15, no. 1, pp. 201–213, Jan. 2006.

[12] H. He and L. P. Kondi, "An image super-resolution algorithm for different error levels per frame," *IEEE Trans. Image Process.*, vol. 15, no. 3, pp. 592–603, Mar. 2006.

[13] D. Robinson and P. Milanfar, "Statistical performance analysis of super-resolution," *IEEE Trans. Image Process.*, vol. 15, no. 6, pp. 1413–1428, Jun. 2006.

[14] G. Chantas, N. Galatsanos, and A. Likas, "Maximum *a posteriori* image restoration based on a new directional continuous edge image Prior," in *Proc. IEEE Int. Conf. Image Processing*, Genoa, Italy, Sep. 11–14, 2006, vol. 1, pp. 941–944.

[15] G. Chantas, N. Galatsanos, and A. Likas, "Bayesian restoration using a new non-stationary edge preserving image prior," *IEEE Trans. Image Process.*, vol. 15, no. 10, pp. 2987–2997, Oct. 2006.

[16] S. Nash and A. Sofer, *Linear and Nonlinear Programming*.   New York: Mc Graw-Hill, 1996.

[17] J. C. Lagarias, J. A. Reeds, M. H. Wright, and P. E. Wright, "Convergence properties of the Nelder-Mead simplex method in low dimensions," *SIAM J. Optim.*, vol. 9, no. 1, pp. 112–147, 1998.

[18] P. Charbonnier, L. Blanc-Feraud, G. Aubert, and M. Barlaud, "Deterministic edge preserving regularization in computed imaging," *IEEE Trans. Image Process.*, vol. 6, no. 2, pp. 298–311, Feb. 1997.

[19] H. Stark and Y. Yang, *Vector Space Projections Methods a Numerical Approach to Signal and Image Processing Neural Networks and Optics*.   New York: Wiley, 1998.

[20] L. Yaroslavsky, "Boundary effect free and adaptive discrete signal sinc-interpolation algorithms for signal and image resampling," *Appl. Opt.*, vol. 42, no. 20, pp. 4166–4175, July 2003.

[21] S. Farsiu, D. Robinson, M. Elad, and P. Milanfar, "Robust shift and add approach to super-resolution," in *Proc. SPIE Int. Symp. Optical Science and Technology*, San Diego, CA, Aug. 2003, vol. 5203, pp. 121–130.

[22] S. Farsiu, M. Elad, and P. Milanfar, "Video-to-video dynamic superresolution for grayscale and color sequences," *EURASIP J. Appl. Signal Process.*, vol. 2006, Jan. 2006, Article ID 61859.

[23] A. K. Katsagellos, R. Mollina, and J. Mateos, *Super-Resolution of Images and Video*.   San Rafael, CA: Morgan & Claypool, 2007.

Giannis K. Chantas received the Diploma degree in computer science and the M.S. degree from the University of Ioannina, Ioannina, Greece, in 2002 and 2004, respectively. He is currently pursuing the Ph.D. degree at the Department of Computer Science, University of Ioannina.

His research interests are in the area of statistical image processing and include Bayesian image restoration, blind deconvolution, and super-resolution.

Nikolaos P. Galatsanos received the Diploma of Electrical Engineering from the National Technical University of Athens, Athens, Greece, in 1982, and the MSEE and Ph.D. degrees from the Electrical and Computer Engineering Department, University of Wisconsin, Madison, in 1984 and 1989, respectively.

He was on the faculty of the Electrical and Computer Engineering Department, Illinois Institute of Technology, Chicago, from 1989 to 2002. Presently, he is on the faculty of the Department of Computer Science, University of Ioannina, Ioannina, Greece.
He co-edited a book titled *Image Recovery Techniques for Image and Video Compression and Transmission* (Kluwer, 1998). His research interests center on Bayesian methods for image processing and statistical learning problems with applications for medical imaging and visual communications.

Dr. Galatsanos has served as an Associate Editor for the IEEE TRANSACTIONS ON IMAGE PROCESSING, the IEEE SIGNAL PROCESSING MAGAZINE, the *SPIE Journal of Electronic Imaging*, and currently serves as Associate Editor for the IEEE SIGNAL PROCESSING LETTERS.

Nathan A. Woods received the B.S.E.E. and M.S.E.E. degrees from Case Western Reserve University, Cleveland, OH, in 1994, and the Ph.D. degree in electrical engineering from the Illinois Institute of Technology, Chicago, in 2004.

From 1994 to 1996, he was a Senior Electrical Engineer with Motorola Cellular Infrastructure, Arlington Heights, IL. From 1996 to 2000, he was a Senior Staff Engineer with Advanced Digital Designs, an engineering consultancy. He is currently a Principal at Binary Machines, Inc., Schaumburg, IL, an engineering consultancy specializing in the design of custom DSP hardware and software. His present research interests include image restoration, blind image deconvolution, super-resolution, and image compression.