MAXIMUM A POSTERIORI SUPER RESOLUTION BASED ON SIMULTANEOUS NON-STATIONARY RESTORATION, INTERPOLATION AND FAST REGISTRATION

Giannis Chantas[1], Nikolaos Galatsanos[2], and Nathan Woods[3]

[1,2]Department of Computer Science, University of Ioannina, 45110, Ioannina, Greece
phone[1]: +30 2651098873, email[1]: chanjohn@cs.uoi.gr, phone[2]: +30 2651098861, email[2]: galatsanos@cs.uoi.gr
[3]Binary Machines, Inc, 1320 Tower Rd., 60173, Schaumburg, IL, e-mail: nathan@BinaryMachinesInc.com

ABSTRACT

In this paper we propose a maximum a posteriori (MAP) framework for the super resolution problem, i.e. reconstructing high-resolution images from shifted, low-resolution degraded observations. In this framework the restoration, interpolation and registration subtasks of this problem are preformed simultaneously. The main novelties of this work are the use of a new hierarchical non-stationary edge adaptive prior for the super resolution problem, and an efficient implementation of this methodology in the discrete Fourier transform (DFT) domain. We present examples with real data that demonstrate the advantages of this methodology.

1. 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 overview it, for this purpose the interested reader is referred to the recent surveys articles [1] and [2] and the edited book [7]. In this work, we introduce a maximum a posteriori framework that bypasses certain shortcomings of our previous efforts in [3].

More specifically, in [3] a maximum a posteriori (MAP) and a Bayesian methodology, based on the expectation-maximization (EM) algorithm, was presented which solved simultaneously the restoration, registration and interpolation subtasks of the super resolution problem. The main shortcomings of the work in [3] that we address herein are: the stationary simultaneously autoregressive (SAR) prior image model used in [3] and the computationally demanding approach used to perform the registration subtask. Stationary models are convenient; however, they cannot capture effectively the local image structure especially in the vicinity of image edges.

In this work we utilize for the first time in the super resolution problem a new non-stationary, directional prior, that uses a continuously valued model for the image edge structure. The Bayesian model used in this prior enforces sparseness in the directional differences among neighboring pixels and has been applied successfully to the image restoration problem in [4] and [5]. We must note that a non-stationary prior has already been used in [8]. However, the major nov-

elty in this work is the use of a directional prior, which is an extension of the prior used in [8].

The second novelty of this work is a fast implementation of the registration subtask. In [3], the registration subtask was computationally very demanding. Herein we propose a new method to estimate the registration parameters. This algorithm is implemented in the discrete Fourier transform (DFT) domain and is based on the Newton-Raphson algorithm. Furthermore, it utilizes analytically calculated 1st and 2nd derivatives. Thus, it is fast to implement and converges rapidly since Newton-Raphson algorithms display quadratic convergence [6].

The rest of this paper is organized as follows; in section 2 and 3 we present the imaging and the proposed image prior models, respectively. In section 4 we present the MAP based restoration algorithm. In section 5 we present the method for estimating the registration parameters. In section 6 we present experiments with real data that demonstrate the properties of our algorithm. Finally in section 7 we provide conclusions and thoughts for future research.

2. 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 images $\mathbf{y}_i$ such that:

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

where $\mathbf{y}_i$ is a $N \times 1$. Then the observations are given by:

$$\mathbf{y} = \mathbf{Bx} + \mathbf{n}, \qquad (1)$$

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 $\mathbf{n} = \begin{bmatrix} \mathbf{n}_1^T & \mathbf{n}_2^T & \cdots & \mathbf{n}_P^T \end{bmatrix}^T$ a $PN_H \times 1$ vector consisting of P $N_H \times 1$, additive white noise vectors. We assume Gaussian statistics for the noise given by $\mathbf{n}_i \sim N(\mathbf{0}, \beta_i^{-1}\mathbf{I})$, $i = 1,...P$, where $\mathbf{0}$ is a $N_H \times 1$ vector with zeros and $\mathbf{I}$ the

$N_H \times N_H$ identity matrix respectively, and β_i^{-1}, $i=1,...P$, are the noise variance at each image that are assumed unknown and statistically independent with each other. The degradation operator $\mathbf{B}$ is given by:

$$\mathbf{B} = \left[\left(\mathbf{DS}(\delta_1) \mathbf{H}_1 \right)^{\mathrm{T}} \quad \cdots \quad \left(\mathbf{DS}(\delta_P) \mathbf{H}_P \right)^{\mathrm{T}} \right]^{\mathrm{T}},$$

where matrix $\mathbf{D}$ is the known $N \times N_H$ decimation matrix. $\mathbf{H}_i$, $i=1...P$, are the shift-invariant $N_H \times N_H$ blurring convolutional operators, and $\mathbf{S}(\delta_i)$, $i=1,...P$, are the $N_H \times N_H$ shift-invariant shifting operators. Each δ_i is a scalar which represents translation 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:

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

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 non-circulant nature, we obtain tractable calculations in the DFT domain.

3. IMAGE PRIOR MODEL

Since we utilize a MAP algorithm, a prior for the image is necessary. The prior used here is non stationary and has been used with success in other image processing problems [4] and [5]. This image prior model assumes that the first order differences of the image $\mathbf{x}$ in four directions, 0^0, 90^0, 45^0 and 135^0 respectively, are given by:

$$\varepsilon^1(i,j) = \mathbf{x}(i,j) - \mathbf{x}(i,j+1),$$
$$\varepsilon^2(i,j) = \mathbf{x}(i,j) - \mathbf{x}(i+1,j),$$
$$\varepsilon^3(i,j) = \mathbf{x}(i,j) - \mathbf{x}(i+1,j+1),$$
$$\varepsilon^4(i,j) = \mathbf{x}(i,j) - \mathbf{x}(i-1,j+1), \tag{2}$$

with $\varepsilon^k(i,j)$ $k=1,2,3,4$, 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 $\mathbf{Q}^k \mathbf{x} = \varepsilon^k$, $k=1,2,3,4$, where $\mathbf{Q}^k$ the $N_H \times N_H$ directional difference operators for $N_H \times 1$ images. Without loss of generality, in what follows, for convenience, in what follows we will use one dimensional notation, in other words, we assume $\varepsilon^k = \left[\varepsilon_1^k, \varepsilon_2^k, \ldots \varepsilon_{N_H}^k \right]^T$. We assume that the residuals have Gaussian statistics according to $\varepsilon_i^k \sim N\left(0, \left(a_i^k\right)^{-1}\right)$, for $i=1,\ldots,N_H$ and $k=1,2,3,4$, where a_i^k the inverse variance of is ε_i^k and N_H the size of the image.

For the inverse variances a_i^k's we introduce the notation $\mathbf{A}^k = diag\left\{ a_1^k, a_2^k, \ldots a_N^k \right\}$ a $N_H \times N_H$ diagonal matrix and $\tilde{\mathbf{A}} = diag\left\{ \mathbf{A}^1, \mathbf{A}^2, \mathbf{A}^3, \mathbf{A}^4 \right\}$ a $4N_H \times 4N_H$ diagonal matrix and $\tilde{\mathbf{a}} = \left[\mathbf{a}^1, \mathbf{a}^2, \mathbf{a}^3, \mathbf{a}^4 \right]^{\mathrm{T}}$ a $4N_H \times 1$ vector. Also for the errors we use the notation $\tilde{\boldsymbol{\varepsilon}} = \left[\boldsymbol{\varepsilon}^1, \boldsymbol{\varepsilon}^2, \boldsymbol{\varepsilon}^3, \boldsymbol{\varepsilon}^4 \right]^{\mathrm{T}}$. We assume that the errors in each direction and at each pixel location are independent. This is based on the assumption that at each pixel location an edge can occur at any direction independently of what happens in adjacent pixels. This assumption makes subsequent calculations tractable. Thus, the joint density for the errors is Gaussian and is given as

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

To relate $\tilde{\boldsymbol{\varepsilon}}$ with the image $\mathbf{x}$ we define the $4N_H \times N_H$ operator $\tilde{\mathbf{Q}} = \left[\left(\mathbf{Q}^1\right)^{\mathrm{T}}, \left(\mathbf{Q}^2\right)^{\mathrm{T}}, \left(\mathbf{Q}^3\right)^{\mathrm{T}}, \left(\mathbf{Q}^4\right)^{\mathrm{T}} \right]^{\mathrm{T}}$. Then, the relation between the image and the residuals is $\tilde{\boldsymbol{\varepsilon}} = \tilde{\mathbf{Q}} \mathbf{x}$. Based on this relation and $p(\tilde{\boldsymbol{\varepsilon}}; \tilde{\mathbf{a}})$ we can define an *improper prior* for the image $\mathbf{x}$. This prior is given by:

$$p(\mathbf{x}; \tilde{\mathbf{a}}) \propto \prod_{k=1}^{4} \prod_{i=1}^{N_H} \left(a_i^k \right)^{1/8} \exp\left(-0.5\left(\left(\tilde{\mathbf{Q}}\mathbf{x}\right)^{\mathrm{T}} \tilde{\mathbf{A}} \tilde{\mathbf{Q}}\mathbf{x} \right)\right)$$
$$= \prod_{k=1}^{4} \prod_{i=1}^{N_H} \left(a_i^k \right)^{1/8} \exp\left(-0.5\left(\left(\mathbf{Q}^k \mathbf{x}\right)^{\mathrm{T}} \mathbf{A}^k \mathbf{Q}^k \mathbf{x} \right)\right). \tag{3}$$

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 the proposed prior is that it introduces $4N_H$ 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. For this purpose 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 even maximum likelihood (ML).

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

$$p\left(a_i^k; m_k, l_k \right) \propto \left(a_i^k \right)^{\frac{l_k - 2}{2}} \exp\left\{ -m_k (l_k - 2) a_i^k \right\}. \tag{4}$$

For such a representation the mean and variance of Gamma are given by $E\left[a_i^k\right] = l_k\left(2m_k\left(l_k - 2\right)\right)^{-1}$, and $Var\left[a_i^k\right] = l_k\left(2m_k^2\left(l_k - 2\right)^2\right)^{-1}$ 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 $l_k \to \infty$, $E\left[a_i^k\right] \to \left(2m_k\right)^{-1}$ and $Var\left[a_i^k\right] \to 0$. In other words, the prior becomes very informative and restrictive resulting in $a_i^k = \left(2m_k\right)^{-1} \forall i$. This also implies that the image model becomes stationary. In contrast, when $l_k \to 2$ then both $E\left[a_i^k\right] \to \infty$ and $Var\left[a_i^k\right] \to \infty$, thus the prior becomes uninformative and does not influence at all the values of the a_i^k's. In other words, the a_i^k's are free from the moderating influence of the prior and are allowed to "vary wildly" following the data. In such case the image model becomes "highly non stationary". As a result, the value of the parameter l_k can be also viewed as a way to adjust for the *degree of non stationarity* of the image model.

4. MAXIMUM A POSTERIORI (MAP) ESTIMATION

The super-resolution image $\mathbf{x}$ is estimated from the observations $\mathbf{y}$, utilizing a MAP approach in which we infer simultaneously $\tilde{\mathbf{a}}$, $\mathbf{x}$ and $\boldsymbol{\delta} = \left[\delta_2 \quad \cdots \quad \delta_P\right]^{\mathrm{T}}$. This is based on maximization of the posterior probability. Thus we have:

$$p\left(\mathbf{x}, \tilde{\mathbf{a}} \mid \mathbf{g}; \beta, \mathbf{m}, \mathbf{l}, \boldsymbol{\delta}\right) \propto p\left(\mathbf{y}, \mathbf{x}, \tilde{\mathbf{a}}; \beta, \mathbf{m}, \mathbf{l}, \boldsymbol{\delta}\right) =$$
$$= p\left(\mathbf{y} \mid \mathbf{x}, \tilde{\mathbf{a}}; \beta, \boldsymbol{\delta}\right) p\left(\mathbf{x} \mid \tilde{\mathbf{a}};\right) p\left(\tilde{\mathbf{a}}; \mathbf{m}, \mathbf{l}\right),$$

where:

$$\mathbf{m} = \left[m_1, m_2, m_3, m_4\right]^{\mathrm{T}}, \mathbf{l} = \left[l_1, l_2, l_3, l_4\right]^{\mathrm{T}}.$$

Maximizing the quantity $p\left(\mathbf{x}, \tilde{\mathbf{a}} \mid \mathbf{g}; \beta, \mathbf{m}, \mathbf{l}, \boldsymbol{\delta}\right)$ with respect to $\mathbf{x}$ and $\tilde{\mathbf{a}}$ is equivalent to minimizing the negative logarithm:

$$J_{MAP}\left(\mathbf{x}, \tilde{\mathbf{a}} \mid \mathbf{y}; \beta, \mathbf{m}, \mathbf{l}, \boldsymbol{\delta}\right) \propto -\log p\left(\mathbf{y}, \mathbf{x}, \tilde{\mathbf{a}}; \beta, \mathbf{m}, \mathbf{l}, \boldsymbol{\delta}\right) =$$
$$-\log p\left(\mathbf{y} \mid \mathbf{x}, \tilde{\mathbf{a}}; \beta, \boldsymbol{\delta}\right) + \log p\left(\mathbf{x} \mid \tilde{\mathbf{a}}\right) + \log p\left(\tilde{\mathbf{a}}; \mathbf{m}, \mathbf{l}\right) =$$
$$= -\frac{N}{2}\log\beta + \frac{1}{2}\beta\left\|\mathbf{B}(\boldsymbol{\delta})\mathbf{x} - \mathbf{y}\right\|^2 - \frac{1}{8}\sum_{k=1}^{4}\sum_{i=1}^{N}\log a_i^k +$$
$$+\frac{1}{2}\sum_{k=1}^{4}\sum_{i=1}^{N}\left(\mathbf{Q}^k\mathbf{x}\right)^{\mathrm{T}}\mathbf{A}^k\mathbf{Q}^k\mathbf{x} - \sum_{k=1}^{4}\left(\frac{l_k - 2}{2}\sum_{i=1}^{N}\log a_i^k\right) +$$
$$+\sum_{k=1}^{4}\left(m_k\left(l_k - 2\right)\sum_{i=1}^{N}a_i^k\right).$$
$$(5)$$

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

Setting $\nabla_{\tilde{\mathbf{a}}} J_{MAP}\left(\mathbf{x}, \tilde{\mathbf{a}} \mid \mathbf{y}; \beta, \mathbf{m}, \mathbf{l}, \boldsymbol{\delta}\right) = 0$ yields:

$$\left(a_i^k\right)^* = \frac{\left(\frac{1}{8} + \frac{1}{2}\left(l_k - 2\right)\right)}{\left(\frac{1}{2}\left(\varepsilon_i^k\right)^2 + m_k\left(l_k - 2\right)\right)}. \qquad (6)$$

Setting $\nabla_{\mathbf{x}} J_{MAP}\left(\mathbf{x}, \tilde{\mathbf{a}} \mid \mathbf{y}; \beta, \mathbf{m}, \mathbf{l}, \boldsymbol{\delta}\right) = 0$ yields:

$$\mathbf{x}^* = \left(\mathbf{B}^{\mathrm{T}}(\boldsymbol{\delta})\mathbf{B}(\boldsymbol{\delta}) + \beta^{-1}\sum_{k=1}^{4}\left(\mathbf{Q}^k\right)^{\mathrm{T}}\mathbf{A}^k\mathbf{Q}^k\right)^{-1}\mathbf{B}^{\mathrm{T}}\mathbf{y}. \qquad (7)$$

Equation (7) cannot be solved in closed form since analytical inversion of $\left[\mathbf{B}^{\mathrm{T}}(\boldsymbol{\delta})\mathbf{B}(\boldsymbol{\delta}) + \beta^{-1}\sum_{k=1}^{4}\left(\mathbf{Q}^k\right)^{\mathrm{T}}\mathbf{A}^k\mathbf{Q}^k\right]$ is not possible due to the non-circulant nature of matrices $\mathbf{B}$ and $\mathbf{A}^k$. Thus, we resort to a numerical solution using a conjugate gradient algorithm [6].

Estimation of the $\boldsymbol{\delta} = \left[\delta_2 \quad \cdots \quad \delta_P\right]^{\mathrm{T}}$ is equivalent to registration and is performed in a similar manner. The registration parameters $\boldsymbol{\delta}$ are estimated by minimizing the quantity J_{MAP} with respect to $\boldsymbol{\delta}$:

$$\boldsymbol{\delta}^* = \arg\min_{\boldsymbol{\delta}} J_{MAP}\left(\boldsymbol{\delta}\right) = \arg\min_{\boldsymbol{\delta}}\left\|\mathbf{B}(\boldsymbol{\delta})\mathbf{x}^* - \mathbf{y}\right\|^2. \qquad (8)$$

The implementation details of this minimization task are described in detail in section 5.

The observation of the previous section that the parameters l_k control the degree of non-stationarity of the model can be verified from Eq. (6), the MAP estimates of the $\left(a_i^k\right)$. More specifically, when $l_k \to \infty$, $\left(a_i^k\right)^* = \left(2m_k\right)^{-1} \forall i$, and the image model becomes stationary. In contrast, when $l_k \to 2$, $\left(a_i^k\right)^* = \left(\left(\varepsilon_i^k\right)^2\right)^{-1} \forall i$, thus the $\left(a_i^k\right)^*$'s are completely unaffected from the moderating effect of the Gamma hyper-prior and only follow the data. For example, in smooth areas of the image where the local residual in the denominator of Eq. (6) tend to zero, it holds that $\left(a_i^k\right)^* \to \infty$.

5. FAST REGISTRATION IN THE DFT DOMAIN

In this section a fast implementation of the registration task based on the Newton-Raphson algorithm is described. This method is chosen as the method of preference due its convergence speed [6]. Registration requires the minimization in Eq. (8). The DFT domain is used since it allows easy analytic calculations of the first and second derivatives of the objective function. By the definition of the matrix $\mathbf{B}$, the norm in Eq. (8) is written as such:

$$\left\|\mathbf{B}(\boldsymbol{\delta})\mathbf{x} - \mathbf{y}\right\|^2 = 2\sum_{i=1}^{P}\mathbf{y}_i^{\mathrm{T}}\mathbf{DS}(\delta_i)\mathbf{H}_i\mathbf{x} + \qquad (9)$$
$$+\sum_{i=1}^{P}\mathbf{x}^{\mathrm{T}}\mathbf{H}_i^{\mathrm{T}}\mathbf{S}^{\mathrm{T}}(\delta_i)\mathbf{D}^{\mathrm{T}}\mathbf{DS}(\delta_i)\mathbf{H}_i\mathbf{x} + C = \sum_{i=1}^{P}J_{MAP}^i(\delta_i) + C,$$

where:

$$J_{MAP}^i(\delta_i) = 2\mathbf{y}_i^{\mathrm{T}}\mathbf{DS}(\delta_i)\mathbf{H}_i\mathbf{x} + \mathbf{x}^{\mathrm{T}}\mathbf{H}_i^{\mathrm{T}}\mathbf{S}^{\mathrm{T}}(\delta_i)\mathbf{D}^{\mathrm{T}}\mathbf{DS}(\delta_i)\mathbf{H}_i\mathbf{x}$$

and $C = \mathbf{y}^{\mathrm{T}}\mathbf{y}$ is constant.

It is sufficient to demonstrate the derivatives for one δ_i. 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 $\mathbf{Y}_i = \mathbf{W}_2\mathbf{y}_i$ for $i = 1, ..., P$,

the DFTs of the desired high resolution and the low resolution observations, respectively. Then, we can write:

$$J^i_{MAP}(\delta_i) \propto 2\mathbf{y}^T_i \mathbf{DS}(\delta_i)\mathbf{H}_i\mathbf{x} + \mathbf{x}^T\mathbf{H}^T_i\mathbf{S}^T(\delta_i)\mathbf{D}^T\mathbf{DS}(\delta_i)\mathbf{H}_i\mathbf{x} =$$

$$= 2real\left\{\mathbf{Y}^H_i\mathbf{\Lambda_D}\mathbf{\Lambda_{S(\delta_i)}}\mathbf{\Lambda_{H_i}}\mathbf{X}\right\} + \mathbf{X}^H\mathbf{\Lambda}^*_{\mathbf{H}_i}\mathbf{\Lambda}^*_{\mathbf{S}(\delta_i)}\mathbf{\Lambda}^T_{\mathbf{D}}\mathbf{\Lambda_D}\mathbf{\Lambda_{S(\delta_i)}}\mathbf{\Lambda_{H_i}}\mathbf{X}$$

$$= 2real\left\{\sum^N_{m=1}\mathbf{Y}^*_i[m]\mathbf{R}_i[m]\right\} + \sum^N_{m=1}\mathbf{R}^*_i[m]\mathbf{R}_i[m], \quad (10)$$

where:

$$\mathbf{R}_i[m] = \frac{\sum^{d-1}_{n=0}\left(\mathbf{\Lambda_{S(\delta_i)}}\left[m+\frac{nN}{d}\right]\mathbf{\Lambda_{H_i}}\left[m+\frac{nN}{d}\right]\mathbf{X}\left[m+\frac{nN}{d}\right]\right)}{d}.$$

The symbol 'H' denotes the Hermitian and '*' the conjugate. The matrices $\mathbf{\Lambda_{S(\delta_i)}} = \mathbf{W}_1\mathbf{S}(\delta_i)\mathbf{W}^{-1}_1$ and $\mathbf{\Lambda_{H_i}} = \mathbf{W}_1\mathbf{H}_i\mathbf{W}^{-1}_1$ for $i = 1,...,P$, are diagonal due to the circulant nature of the matrices $\mathbf{S}(\delta_i)$ and $\mathbf{H}_i$. For simplicity, the notation $[m]$ for the matrices denotes their diagonal element. It can be shown that

$$\mathbf{\Lambda_D} = \mathbf{W}_2\mathbf{DW}^{-1}_1 = [\mathbf{I}_1 \quad \mathbf{I}_2 \quad ... \quad \mathbf{I}_d]/d$$

is a $N \times N_H$ block matrix that contains d identity matrices of size $N \times N$.

The evaluation of the first and second derivatives of Eq. (10) is very convenient in the DFT domain since the parameter δ_i is only in the diagonal elements of the matrix $\mathbf{\Lambda_{S(\delta_i)}}$. The diagonal elements of matrix $\mathbf{\Lambda_{S(\delta_i)}}$, see for example [7], are equal to:

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

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

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

For convenience, the derivative calculations are done only for the first half. The first and second derivatives are:

$$\frac{\partial\mathbf{\Lambda_{S(\delta_i)}}[m]}{\partial\delta_i} = \frac{-2j\pi(m-1)}{N}\exp\{-2j\pi\delta_i(m-1)/N\} \quad (11)$$

$$\frac{\partial^2\mathbf{\Lambda_{S(\delta_i)}}[m]}{(\partial\delta_i)^2} = \frac{-4\pi^2}{N^2}(m-1)^2\exp\{-2j\pi\delta_i(m-1)/N\}. \quad (12)$$

The derivative of the term in Eq. (10) is given by applying Eq. (11) as:

$$\frac{\partial J^i_{MAP}(\delta_i)}{\partial\delta_i} = \sum^N_{m=1}\mathbf{Y}^*_i[m]\frac{\partial\mathbf{R}_i[m]}{\partial\delta_i} + \sum^N_{m=1}\mathbf{Y}_i[m]\frac{\partial\mathbf{R}^*_i[m]}{\partial\delta_i} +$$

$$+\sum^N_{m=1}\mathbf{R}^*_i[m]\frac{\partial\mathbf{R}_i[m]}{\partial\delta_i} + \sum^N_{m=1}\frac{\partial\mathbf{R}^*_i[m]}{\partial\delta_i}\mathbf{R}_i[m],$$

where:

$$\frac{\partial\mathbf{R}_i[m]}{\partial\delta_i} = \sum^{d-1}_{n=0}\frac{\partial\mathbf{\Lambda_{S(\delta_i)}}\left[m+\frac{nN}{d}\right]}{\partial\delta_i}\mathbf{\Lambda_{H_i}}\left[m+\frac{nN}{d}\right]\mathbf{X}\left[m+\frac{nN}{d}\right],$$

and:

$$\frac{\partial\mathbf{R}^*_i[m]}{\partial\delta_i} = \sum^{d-1}_{n=0}\left(\frac{\partial\mathbf{\Lambda}^*_{\mathbf{S}(\delta_i)}\left[m+\frac{n}{d}N\right]}{\partial\delta_i}\times \\ \times\mathbf{\Lambda}^*_{\mathbf{H}_i}\left[m+\frac{n}{d}N\right]\mathbf{X}^*\left[m+\frac{n}{d}N\right]\right).$$

Similarly, the second derivative is:

$$\frac{\partial^2 J_{MAP}(\delta_i)}{\partial\delta^2_i} = \sum^N_{m=1}\mathbf{Y}^*_i[m]\frac{\partial^2\mathbf{R}_i[m]}{\partial\delta^2_i} + \sum^N_{m=1}\mathbf{Y}_i[m]\frac{\partial^2\mathbf{R}^*_i[m]}{\partial\delta^2_i} +$$

$$+\sum^N_{m=1}\mathbf{R}^*_i[m]\frac{\partial^2\mathbf{R}_i[m]}{\partial\delta^2_i} + 2\sum^N_{m=1}\left|\frac{\partial\mathbf{R}_i[m]}{\partial\delta_i}\right|^2 + \sum^N_{m=1}\frac{\partial^2\mathbf{R}^*_i[m]}{\partial\delta^2_i}\mathbf{R}_i[m].$$

In both cases the evaluation of the first and second derivatives of $\mathbf{R}$ is straightforward. Using these results, the application of the Newton-Raphson iterations, is straightforward. To be precise, we note that for 2-d signals, like images, there are two translations parameters per image. Thus, in the above update equation there is a 2×1 gradient vector and a 2×2 Hessian matrix involved. However, the inversion of a 2×2 matrix is easily found in closed form; hence the 2-d registration algorithm is very fast.

6. EXPERIMENTS

In order to test the proposed MAP methodology, we used it on a real set of 20 grayscale images of size 57×49, which were extended by padding with zeros to 64×64. A subset of these images is shown in *Fig. 1a-d*. Using this data we produced high resolution images of size 128×128, shown in *Fig. 2* and *3*.

To facilitate learning the proposed image model we used the β^{-1}_i for all i (additive noise variances) and equal m_k for all k obtained by learning a stationary SAR model [3]. The parameters m_k were obtained as $m_k = 1/(2a_{STAT})$ where a_{STAT} the image model parameter of the stationary SAR model. The parameters l_k were selected to be equal to $l = 2.1$. This value was found by trial and error experiments. The stationary restored image, implementing the method in [3], is shown in *Fig. 2*.

In estimating the shape of the blur for each low resolution image a Gaussian-shaped blur was assumed. The width of each blur was experimentally estimated in trial and error restoration experiments with each low resolution image. The values of the widths were found in the range [0.5, 4] pixels.

The restoration algorithm estimates δ_i, $\mathbf{x}$ and a^k_i iterating between Eq. (6), (7) and (8) till convergence. The shifts found between the low resolution image in 1(a) and the rest are also given in *Fig. 1a-d*. In the presented experiments, the convergence criterion was the likelihood function.

7. CONCLUSIONS

Inspection of the super-resolved images in *Fig. 2* and *3* reveals that the resolution of the reconstructed image has significantly been improved. The letters in the super resolved images are now easily legible. Furthermore, the image re-

constructed using the proposed non-stationary prior is visually more pleasant and displays less ringing in the edges.

Each iteration of the proposed algorithm required 15-20 seconds per iteration in spite of the iterative conjugate gradient used to compute $\mathbf{x}$ in Eq. (7). This is a significant improvement in speed as compared to the algorithm in [3] that requires ~60 seconds per iteration, even though an estimate in closed form for $\mathbf{x}$ is available. As noted in [3] most of that delay was due to the numerical search for the registration parameters in the M step of the algorithm. In the future we plan to include a PSF estimation step in the formulation of this problem. A Bayesian algorithm for the non-stationary model is also being developed.

ACKNOWLEDGMENTS

REFERENCES

[1] 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.

[2] S. Farsui, D. Robinson, M. Elad, and P. Milanfar, "Advances and Challenges in Super-Resolution," *International Journal of Imaging Systems and Technology*, vol. 14, no. 2, pp. 47-57, August 2004.

[3] N.A. Woods, N.P. Galatsanos and A.K. Katsaggeloss, "Stochastic Methods for Joint Restoration, Interpolation and Registration of Multiple Undersampled Images," *IEEE Transactions on Image Processing*, vol. 15, no. 1, pp. 201-213, January 2006.

[4] G. Chantas, N. Galatsanos and A. Likas, "Maximum a Posteriori Image Restoration Based on a New Directional Continuous Edge Image Prior," *in Proc. IEEE ICIP 2005,* Genova, Italy, vol. 1, September 11-14, 2006, pp. 941-944.

[5] G. Chantas, N. Galatsanos, and A. Likas, "Bayesian Restoration Using a new Hierarchical Directional Continuous Edge Image Prior", *IEEE Transactions on Image Processing,* to appear.

[6] S. Nash and A. Sofer, *Linear and Nonlinear Programming*, Mc Graw Hill, 1996.

[7] S. Chaudhuri, Editor, *Super-Resolution Imaging*, Chapter 4, "Reconstruction of a High Resolution Image from Multiple Low Resolution Images", by B. Tom, N. Galatsanos and A. Katsaggelos, pp. 71-105, Kluwer, 2001.

[8] N.A. Woods and N.P. Galatsanos, "Non-stationary approximate Bayesian super-resolution using a hierarchical prior model", *in Proc. IEEE ICIP 2005,* Genova, Italy, vol. 1, pp. 37-40, September 11-14, 2006.
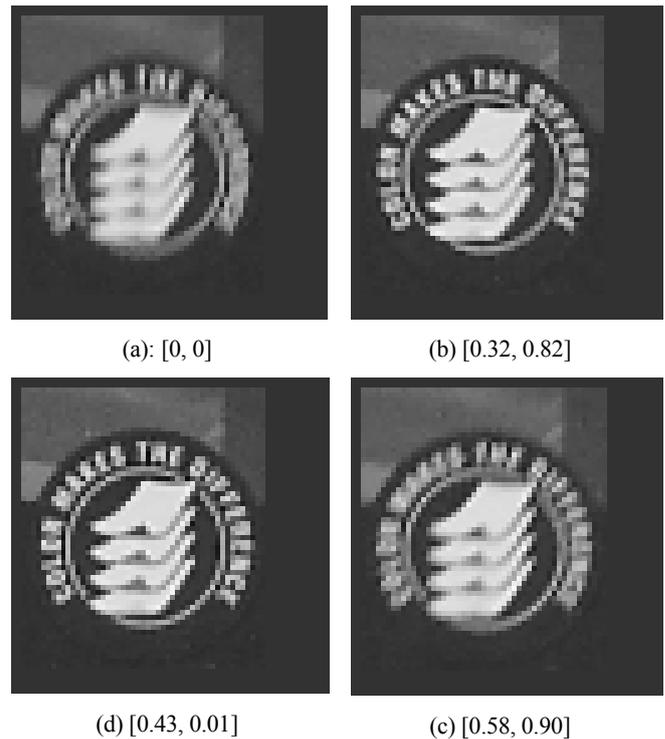

(a): [0, 0] (b) [0.32, 0.82] (d) [0.43, 0.01] (c) [0.58, 0.90]

Figure 1: Degraded low-resolution images, shifts between the image in (a) and the rest are given.

Figure 2: Stationary [3] super-resolution image.

Figure 3: Non-stationary MAP super-resolution image.