A Computationally Efficient Superresolution Image Reconstruction Algorithm

Nhat Nguyen, Peyman Milanfar, *Senior Member, IEEE*, and Gene Golub

***Abstract*—Superresolution reconstruction produces a high-resolution image from a set of low-resolution images. Previous iterative methods for superresolution [9], [11], [18], [27], [30] had not adequately addressed the computational and numerical issues for this ill-conditioned and typically underdetermined large scale problem. We propose efficient block circulant preconditioners for solving the Tikhonov-regularized superresolution problem by the conjugate gradient method. We also extend to underdetermined systems the derivation of the generalized cross-validation method for automatic calculation of regularization parameters. Effectiveness of our preconditioners and regularization techniques is demonstrated with superresolution results for a simulated sequence and a forward looking infrared (FLIR) camera image sequence.**

***Index Terms*—Circulant preconditioners, generalized cross-validation, superresolution, underdetermined systems.**

I. Introduction

PHYSICAL constraints limit image resolution quality in many imaging applications. These imaging systems yield aliased and undersampled images if their detector array is not sufficiently dense. This is particularly true for infrared imagers and some charge-coupled device (CCD) cameras. Several papers have developed direct and iterative techniques for superresolution: the reconstruction of a high-resolution (HR) unaliased image from several low-resolution (LR) aliased images. Proposed direct methods include Fourier domain approaches by Tsai and Huang [34], Tekalp *et al.* [33], [27], [26], and Kim *et al.* [22], [21] where high-frequency information is extracted from low-frequency data in the given LR frames. Several methods [29], [1] have used an interpolation-restoration combination approach in the image domain. Ur and Gross [35] and Shekarforoush and Chellappa [31] considered superresolution as generalized sampling problem with periodic samples. In this paper, we are mainly interested in the computational issue for iterative methods. Projection type methods have been used by several researchers. Patti *et al.* [27] and Stark and Oskoui [32] proposed projection onto convex sets (POCS) algorithms, which defined sets of closed convex constraints whose intersection contains the HR estimate and successively projected an arbitrary initial estimate onto these constraint sets. Others [4], [20], [23] adopted a related method, the iterative back-projection method, frequently used in computer aided tomography. Cheeseman *et al.* [9] used the standard Jacobi's method, and Hardie *et al.* [18] proposed a steepest descent based algorithm in combination with block matching to compute simultaneously the HR image and the registration parameters. Although they are usually robust to noise and allow some modeling flexibility, projection-based algorithms are also known for their low rate of convergence. More recently, Hardie *et al.* [18], Connolly and Lane [10] and Chan *et al.* [6] have considered conjugate gradient (CG) methods for Tikhonov regularized superresolution. However, to a large degree, computational and numerical difficulties of superresolution have not been addressed adequately. For instance, Hardie *et al.* [18] and Connolly and Lane [10] did not consider preconditioning for their CG algorithm, and Chan *et al.* [6] preconditioner is only applicable for multisensor arrays. Superresolution is a computationally intensive problem typically involving tens of thousands unknowns. For example, superresolving a sequence of 50×50 pixel LR frames by a factor of 4 in each spatial dimension involves $200 \times 200 = 40\,000$ unknown pixel values in the HR image. Furthermore, the matrix system is typically underdetermined and ill-conditioned, which can exacerbate system noise and blurring effects. In this paper, we present efficient circulant block preconditioners that take advantage of the inherent structures in the superresolution system matrix to accelerate CG. We adopt the generalized cross-validation (GCV) method, which is often used to calculate regularization parameters for Tikhonov-regularized *overdetermined* least squares problems without accurate knowledge of the variance of noise, to our *underdetermined* problem. Golub *et al.* [14] suggested that GCV can be used for underdetermined problems, and McIntosh and Veronis [24] successfully implemented GCV for their underdetermined tracer inverse problems. However, since the derivation for the GCV formulation by Golub, Heath, and Wahba [14] applies only for overdetermined least squares problems, we extend their derivation in this paper to the underdetermined case. We also propose an efficient technique to approximate the GCV expression. Finally, we show reconstruction results from two test image sequences to demonstrate the effectiveness of our preconditioners and regularization techniques.

The rest of the paper is organized as follows. In Section II, we describe our model for the relationship between high- and low-resolution images in superresolution. Section III outlines our approach to regularization and formulation for GCV. The

Manuscript received June 8, 1999; revised November 16, 2000. This work was supported in part by the National Science Foundation under Grant CCR-9984246. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Michael R. Frater.

N. Nguyen is with KLA-Tencor Corporation, Milpitas, CA 95035 USA (e-mail: Nhat.Nguyen@kla-tencor.com).

P. Milanfar is with the Department of Electrical Engineering, University of California, Santa Cruz, CA 95064 USA (e-mail: milanfar@cse.ucsc.edu).

G. Golub is with Scientific Computing and Computational Mathematics Program, Stanford University, Stanford CA 94305-9025 USA (e-mail: golub@sccm.stanford.edu).

Publisher Item Identifier S 1057-7149(01)01664-5.

preconditioners and approximate convergence bounds are detailed in Section IV. We present the experimental results in Section V and draw some conclusions in Section VI.

II. The Model

Conceptually, superresolution, multichannel, and multisensor data fusion are very similar problems. The goal is to combine information about the same scene from different sources. In superresolution, in particular, the LR frames typically represent different "looks" at the same scene from slightly different directions. Each frame contributes new information used to interpolate subpixel values. To get different looks at the same scene, some relative scene motions must be recorded from frame to frame. These scene motions can be due to controlled motions in the imaging system, e.g., images acquired from orbiting satellites, or uncontrolled motions within the scene itself, e.g., objects moving within view of a surveillance camera. If these scene motions are known or can be estimated within subpixel accuracy, superresolution is possible.

We model each LR frame as a noisy, uniformly down-sampled version of the HR image which has been shifted and blurred [11]. More formally

$$\mathbf{b}_k = D_k C_k F_k \mathbf{x} + \mathbf{n}_k, \qquad 1 \le k \le p \tag{1}$$

where p is the number of available frames, $\mathbf{b}_k$ is an $N \times 1$ vector representing the kth $m \times n$ ($N = mn$ pixels) LR frame in lexicographic order. If l is the resolution enhancement factor in each direction, $\mathbf{x}$ is an $l^2N \times 1$ vector representing the $lm \times ln$ HR image in lexicographic order, F_k is an $l^2N \times l^2N$ shift matrix that represents the relative motions between frames k and a reference frame, C_k is a blur matrix of size $l^2N \times l^2N$, D_k is the $N \times l^2N$ uniform down-sampling matrix, and $\mathbf{n}_k$ is the $N \times 1$ vector representing additive noise.

Fig. 1 illustrates our model conceptually. A pixel value in an LR frame is a weighted "average" value over a box of pixels in the HR image. In the figure, the (1,1) pixel of the LR frame to the right is a weighted "average" over the dashed box, while the (1,1) pixel in another frame is a weighted "average" over the solid box. The relative motion from the dashed box to the solid box is 1 HR pixel down and to the right. Each LR frame contributes new and different information about the HR image. Combining the equations in (1), we have

$$\begin{bmatrix} \mathbf{b}_1 \\ \vdots \\ \mathbf{b}_p \end{bmatrix} = \begin{bmatrix} D_1 C_1 F_1 \\ \vdots \\ D_p C_p F_p \end{bmatrix} \mathbf{x} + \begin{bmatrix} \mathbf{n}_1 \\ \vdots \\ \mathbf{n}_p \end{bmatrix}$$
$$\mathbf{b} = H\mathbf{x} + \mathbf{n}. \tag{2}$$

In this paper, we consider only shifts of integral multiples of one HR pixel. A nonintegral shift is replaced with the nearest integral shift. In this case, at most r^2 nonredundant LR frames are possible for superresolution with enhancement factor r in each dimension. If all possible combinations of subpixel horizontal and vertical shifts are available, the above linear system is square and reduces to essentially a deblurring problem. In general, this is not the case, and the system of equations above is underdetermined. The techniques presented here can be extended to the more general framework to allow for any shifts. This extension is addressed in the first author's thesis [25].

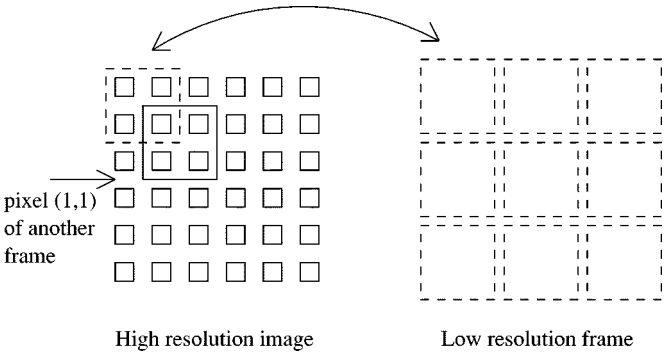

Fig. 1. Superresolution model.

For each frame k, we approximate the relative motions between that frame and a reference frame by a single motion vector. In the case where the scene motions are controlled, the motion vectors are known. Otherwise, they may be estimated by some image registration algorithm; see survey by Brown [5]. For completeness, we include in Appendix A a simple algorithm we used to compute scene motion vectors. We will assume that the point spread function (PSF) which generates the blurring operator is known and spatially invariant. Blind superresolution or superresolution without knowledge of the PSF is a challenging topic under development which will be reported in a subsequent submission.

III. Regularization

The PSF is derived from the discretization of a compact operator (i.e., the image of every L^2-bounded sequence of functions has at least one converging subsequence [16]), so H is ill-conditioned [2]. Thus, even small changes in $\mathbf{b}$ can result in wild oscillations in approximations to $\mathbf{x}$ when (2) is solved directly. To obtain a reasonable estimate for $\mathbf{x}$ we reformulate the problem as a regularized minimization problem

$$\min_{\mathbf{x}} \ \alpha \|\mathbf{b} - H\mathbf{x}\|_2^2 + \mathbf{x}^T Q \mathbf{x} \tag{3}$$

where Q is some symmetric, positive definite matrix, and α is related to the Lagrange multiplier. In this formulation, Q serves as a stabilization matrix, and the new system is better conditioned. While a simple and effective regularization matrix can be the identity I, Q can also incorporate some prior knowledge of the problem, e.g., degree of smoothness [16]. Since Q is symmetric, positive definite, we have the Cholesky decomposition $Q = L^T L$, where L is an upper triangular matrix. Letting $\mathbf{y} = L\mathbf{x}$, $A = HL^{-1}$, then reduces (3) to the standard form

$$\min_{\mathbf{y}} \ \alpha \|\mathbf{b} - A\mathbf{y}\|_2^2 + \|\mathbf{y}\|_2^2. \tag{4}$$

The solution of the underdetermined least squares problem (4) above is

$$\mathbf{y} = A^T (AA^T + \lambda I)^{-1} \mathbf{b}, \quad \lambda = \frac{1}{\alpha}$$

which gives us

$$\mathbf{x} = Q^{-1}H^T(HQ^{-1}H^T + \lambda I)^{-1}\mathbf{b} \tag{5}$$

and when $Q = I$

$$\mathbf{x} = H^T(HH^T + \lambda I)^{-1}\mathbf{b}. \tag{6}$$

In the above formulation, λ is the regularization parameter. A larger λ corresponds to a better conditioned system, but the new system is also farther away from the original system we wish to solve. We will adopt GCV (cf. [14]), a technique popular in overdetermined least squares, for calculating the regularization parameter.

A. Cross-Validation

The idea of cross-validation is simple. Namely, we divide the data set into two parts; one part is used to construct an approximate solution, and the other is used to validate that approximation. For example, the validation error by using the jth pixel as the validation set is

$$\mathrm{CV}_j(\lambda) = \|b_j - \mathbf{a}_j\mathbf{y}^{\not j}(\lambda)\|_2^2$$

where the notation $\not j$ is used to indicate variables corresponding to the system without the jth pixel/row and

$$\mathbf{y}^{\not j}(\lambda) = A_{\not j}^T(A_{\not j}A_{\not j}^T + \lambda I)^{-1}\mathbf{b}_{\not j} \tag{7}$$

is the regularized underdetermined least squares solution of the original system without the jth pixel, $A_{\not j}\mathbf{y}^{\not j} = \mathbf{b}_{\not j}$, with

$$A_{\not j} = \begin{bmatrix} \mathbf{a}_1 \\ \vdots \\ \mathbf{a}_{j-1} \\ \mathbf{a}_{j+1} \\ \vdots \\ \mathbf{a}_K \end{bmatrix}, \quad \mathbf{b}_{\not j} = \begin{bmatrix} b_1 \\ \vdots \\ b_{j-1} \\ b_{j+1} \\ \vdots \\ b_K \end{bmatrix}.$$

The cross-validated regularization parameter is the solution to

$$\lambda_{\mathrm{CV}} = \arg\min_{\lambda} \sum_{j=1}^{K} \mathrm{CV}_j(\lambda). \tag{8}$$

GCV is simply cross-validation applied to the original system after it has undergone a unitary transformation. GCV is also known to be less sensitive to large individual equation errors than cross-validation [24]. For overdetermined systems, it has been shown that the asymptotically optimum regularization parameter according to GCV is given by [14]

$$\lambda_{\mathrm{GCV}} = \arg\min_{\lambda} \frac{\|(AA^T + \lambda I)^{-1}\mathbf{b}\|_2}{\mathrm{tr}((AA^T + \lambda I)^{-1})}. \tag{9}$$

In the next subsection, we derive this same expression for our underdetermined system and in Appendix B, we describe an efficient way to obtain a suboptimal solution for (9).

B. Generalized Cross-Validation for Underdetermined Systems

Here we derive the closed-form for regularization parameter by GCV for *underdetermined* systems $A\mathbf{y} = \mathbf{b}$. We start by examining the following expression for the regularized underdetermined least squares solution to the equation above

$$\begin{aligned}\mathbf{y} &= A^T(AA^T + \lambda I)^{-1}\mathbf{b} \\ &= \begin{bmatrix} A_{\not j}^T & \mathbf{a}_j^T \end{bmatrix}\left(\begin{bmatrix} A_{\not j} \\ \mathbf{a}_j \end{bmatrix}\begin{bmatrix} A_{\not j}^T & \mathbf{a}_j^T \end{bmatrix} + \lambda I\right)^{-1}\begin{bmatrix} \mathbf{b}_{\not j} \\ b_j \end{bmatrix} \\ &= \begin{bmatrix} A_{\not j}^T & \mathbf{a}_j^T \end{bmatrix}\begin{bmatrix} A_{\not j}A_{\not j}^T + \lambda I & A_{\not j}\mathbf{a}_j^T \\ \mathbf{a}_j A_{\not j}^T & \mathbf{a}_j\mathbf{a}_j^T + \lambda I \end{bmatrix}^{-1}\begin{bmatrix} \mathbf{b}_{\not j} \\ b_j \end{bmatrix}.\end{aligned}$$

Let

$$\begin{bmatrix} \mathbf{z}_{\not j} \\ z_j \end{bmatrix} = \begin{bmatrix} A_{\not j}A_{\not j}^T + \lambda I & A_{\not j}\mathbf{a}_j^T \\ \mathbf{a}_j A_{\not j}^T & \mathbf{a}_j\mathbf{a}_j^T + \lambda I \end{bmatrix}^{-1}\begin{bmatrix} \mathbf{b}_{\not j} \\ b_j \end{bmatrix}$$

then

$$\mathbf{y} = A_{\not j}^T\mathbf{z}_{\not j} + \mathbf{a}_j^T z_j \tag{10}$$

and

$$(A_{\not j}A_{\not j}^T + \lambda I)\mathbf{z}_{\not j} + A_{\not j}\mathbf{a}_j^T z_j = \mathbf{b}_{\not j} \tag{11}$$

$$\mathbf{a}_j A_{\not j}^T\mathbf{z}_{\not j} + (\mathbf{a}_j\mathbf{a}_j^T + \lambda I)z_j = b_j. \tag{12}$$

Multiplying both sides of (11) by $(A_{\not j}A_{\not j}^T + \lambda I)^{-1}$

$$(A_{\not j}A_{\not j}^T + \lambda I)^{-1}\mathbf{b}_{\not j} = \mathbf{z}_{\not j} + (A_{\not j}A_{\not j}^T + \lambda I)^{-1}A_{\not j}\mathbf{a}_j^T z_j.$$

Therefore, from (7)

$$\mathbf{y}^{\not j} = A_{\not j}^T\mathbf{z}_{\not j} + A_{\not j}^T(A_{\not j}A_{\not j}^T + \lambda I)^{-1}A_{\not j}\mathbf{a}_j^T z_j$$

and by (10)

$$\mathbf{y}^{\not j} = \mathbf{y} - (I - A_{\not j}^T(A_{\not j}A_{\not j}^T + \lambda I)^{-1}A_{\not j})\mathbf{a}_j^T z_j.$$

Note that for $\lambda > 0$, $I - A_{\not j}^T(A_{\not j}A_{\not j}^T + \lambda I)^{-1}A_{\not j} = \lambda(A_{\not j}A_{\not j}^T + \lambda I)^{-1}$. So

$$\mathbf{y}^{\not j} = \mathbf{y} - \lambda(A_{\not j}^T A_{\not j} + \lambda I)^{-1}\mathbf{a}_j^T z_j. \tag{13}$$

Next, note that

$$A_{\not j}^T A_{\not j} = A^T A - \mathbf{a}_j^T\mathbf{a}_j.$$

Let $d_j = \mathbf{a}_j(A^T A + \lambda I)^{-1}\mathbf{a}_j^T$. By Sherman–Morrison–Woodbury formula [15]

$$\begin{aligned}&(A_{\not j}^T A_{\not j} + \lambda I)^{-1} \\ &\quad = (A^T A + \lambda I)^{-1} \\ &\qquad + (A^T A + \lambda I)^{-1}\mathbf{a}_j^T(1 - d_j)^{-1}\mathbf{a}_j(A^T A + \lambda I)^{-1}\end{aligned}$$

which leads to

$$\begin{aligned}\mathbf{a}_j(A_{\not j}^T A_{\not j} + \lambda I)^{-1}\mathbf{a}_j^T &= d_j + d_j(1 - d_j)^{-1}d_j \\ &= d_j(1 - d_j)^{-1}.\end{aligned} \tag{14}$$

From (10) and (12) we get

$$\mathbf{a}_j\mathbf{y} + \lambda z_j = b_j. \tag{15}$$

By (13)

$$b_j - \mathbf{a}_j\mathbf{y}^{\dot{-}} = b_j - \mathbf{a}_j\mathbf{y} + \lambda\mathbf{a}_j(A_{\dot{-}}^T A_{\dot{-}} + \lambda I)^{-1}\mathbf{a}_j^T z_j$$

which, using (14) and (15), leads to

$$\begin{aligned} b_j - \mathbf{a}_j\mathbf{y}^{\dot{-}} &= \lambda z_j + \lambda\mathbf{a}_j(A_{\dot{-}}^T A_{\dot{-}} + \lambda I)^{-1}\mathbf{a}_j^T z_j \\ &= \lambda\left(1 + d_j(1-d_j)^{-1}\right) z_j \\ &= \lambda(1-d_j)^{-1} z_j \\ &= (1-d_j)^{-1}(b_j - \mathbf{a}_j\mathbf{y}). \end{aligned} \tag{16}$$

Define $D(\lambda) = \text{diag}(I - A(A^TA + \lambda I)^{-1}A^T)$ to be the diagonal matrix with the same diagonal entries as $I - A(A^TA + \lambda I)^{-1}A^T$. Again, using the fact that $I - A(A^TA + \lambda I)^{-1}A^T = \lambda(AA^T + \lambda I)^{-1}$ for $\lambda > 0$, we can rewrite $D = \lambda\text{diag}(AA^T + \lambda I)^{-1}$. From (8), the optimal cross-validation regularization parameter is

$$\begin{aligned} &\arg\min_\lambda \sum_{j=1}^{K} \text{CV}_j(\lambda) \\ &= \arg\min_\lambda \sum_{j=1}^{K} \|b_j - \mathbf{a}_j\mathbf{y}^{\dot{-}}\|_2^2 \\ &= \arg\min_\lambda \sum_{j=1}^{K} \|(1-d_j)^{-1}(b_j - \mathbf{a}_j\mathbf{y})\|_2^2 \\ &= \arg\min_\lambda \|D^{-1}(I - AA^T(AA^T + \lambda I)^{-1})\mathbf{b}\|_2^2 \\ &= \arg\min_\lambda \lambda^2\|D^{-1}(AA^T + \lambda I)^{-1}\mathbf{b}\|_2^2. \end{aligned} \tag{17}$$

We have just derived the matrix formulation of cross-validation for underdetermined systems. Generalized cross-validation is simply a rotation-invariant form of cross-validation. Following [14], consider the singular value decomposition of A

$$A = U\Sigma V^T.$$

Now let W be the matrix representing the Fourier transform, that is,

$$(W)_{jk} = \frac{1}{\sqrt{n}} e^{2\pi ijk/n}, \qquad j, k = 1, 2, \cdots, n.$$

The GCV estimation for regularization parameter λ can be thought of as cross-validation on the transformed system:

$$\begin{aligned} \tilde{\mathbf{b}} &= WU^T\mathbf{b} \\ &= W\Sigma V^T\mathbf{y} + WU^T\epsilon \\ &= \tilde{A}\mathbf{y} + WU^T\epsilon. \\ (\tilde{A}\tilde{A}^T + \lambda I)^{-1}\tilde{\mathbf{b}} &= WU^T(AA^T + \lambda I)^{-1}\mathbf{b}. \end{aligned}$$

Note that $\tilde{A}\tilde{A}^T + \lambda I = W(\Sigma\Sigma^T + \lambda I)W^T$ is a circulant matrix. Hence, $\tilde{D} = \lambda\text{ diag}((\tilde{A}\tilde{A}^T + \lambda I)^{-1})$ is a multiple of the identity. Thus,

$$\begin{aligned} \tilde{D} &= \frac{\lambda}{m}\text{tr}(\tilde{A}\tilde{A}^T + \lambda I)^{-1})I \\ &= \frac{\lambda}{m}\text{tr}((AA^T + \lambda I)^{-1})I. \end{aligned}$$

Hence, from (17) we can formulate GCV as follows:

$$\begin{aligned} &\arg\min_\lambda \lambda^2\|\tilde{D}^{-1}(\tilde{A}\tilde{A}^T + \lambda I)^{-1}\tilde{\mathbf{b}}\|_2 \\ &= \arg\min_\lambda \frac{\|(\tilde{A}\tilde{A}^T + \lambda I)^{-1}\tilde{\mathbf{b}}\|_2}{\text{tr}((AA^T + \lambda I)^{-1})} \\ &= \arg\min_\lambda \frac{\|WU^T(AA^T + \lambda I)^{-1}\mathbf{b}\|_2}{\text{tr}((AA^T + \lambda I)^{-1})} \\ &= \arg\min_\lambda \frac{\|(AA^T + \lambda I)^{-1}\mathbf{b}\|_2}{\text{tr}((AA^T + \lambda I)^{-1})}. \end{aligned} \tag{18}$$

Not surprisingly, this formulation has the same form as that of the overdetermined case.

IV. Preconditioning for Conjugate Gradient

As we described earlier, superresolution is computationally intensive. The number of unknowns, the same as the number of pixels in the HR image, is typically in the tens or hundreds of thousands. The convergence rate for CG [36] is dependent on the distribution of eigenvalues of the system matrix. The method works well on matrices that are either well-conditioned or have just a few distinct eigenvalues; see [15, p. 525]. Preconditioning is a technique used to transform the original system into one with the same solution, but which can be solved by the iterative solver more quickly [28, p. 245]. For CG, we seek preconditioners with preconditioned system having eigenvalues clustering around 1. In these situations, CG converges very rapidly. Our preconditioners approximate H by exploiting its structure. To see this, we reorder the columns of H and the elements of $\mathbf{x}$, correspondingly, as follows. We partition the HR image $\mathbf{x}$ into N regions each of size $l \times l$, and we enumerate the pixels in each region in lexicographic order 1 to l^2. The desired ordering is then $q_1^{(1)}, \cdots, q_N^{(1)}, q_1^{(2)}, \cdots, q_N^{(2)}, \cdots, q_1^{(l^2)}, \cdots, q_N^{(l^2)}$, where $q_i^{(j)}$ is the jth pixel in the ith region. From our spatial invariance assumption of the PSF, the reordered matrix $\mathcal{T}$ has the following form:

$$\mathcal{T} = \begin{bmatrix} \mathcal{T}_{11} & \mathcal{T}_{12} & \cdots & \mathcal{T}_{1l^2} \\ \mathcal{T}_{21} & \mathcal{T}_{22} & \cdots & \mathcal{T}_{2l^2} \\ \vdots & \vdots & \ddots & \vdots \\ \mathcal{T}_{p1} & \mathcal{T}_{p2} & \cdots & \mathcal{T}_{pl^2} \end{bmatrix} \tag{19}$$

where each block $\mathcal{T}_{ij}$ is an $N \times N$ "nearly"[1] Toeplitz upper band matrix, that is, $\mathcal{T}_{ij}$ only has nonzero entries on a single superdiagonal. For example, in the simple case of superresolving

[1]Almost all entries along the diagonals of $\mathcal{T}_{ij}$ are constant.

a sequence of four, 2×4 pixel, LR frames by a factor of two in each dimension, $\mathcal{T}_{23}$ has the following structure:

$$\mathcal{T}_{23} = \begin{bmatrix} 0 & t & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & t & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & t & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & t & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & t & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & t \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

We first approximate $\mathcal{T}$ by a block matrix $\tilde{\mathcal{T}} = (\tilde{\mathcal{T}}_{ij})$, whose blocks $\tilde{\mathcal{T}}_{ij}$ are Toeplitz. We construct $\tilde{\mathcal{T}}_{ij}$ from $\mathcal{T}_{ij}$ by filling in the zero entries along nonzero diagonals, so that $\tilde{\mathcal{T}}_{ij}$ is just a low rank change from $\mathcal{T}_{ij}$. For example, the approximation to $\mathcal{T}_{23}$ would be

$$\tilde{\mathcal{T}}_{23} = \begin{bmatrix} 0 & t & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & t & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & t & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & t & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & t & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & t & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & t \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

To describe the preconditioners for the block Toeplitz matrix $\tilde{\mathcal{T}}$, we first describe the corresponding preconditioners for standard banded Toeplitz matrices in the following subsection. Extensions of these preconditioners and their convergence properties to the block case are then straightforward [7].

A. Circulant Preconditioners

The first preconditioner, originally developed by Strang [8], completes a Toeplitz matrix T by copying the central diagonals. For an upper triangular banded Toeplitz matrix T

$$T = \begin{bmatrix} t_0 & \cdots & t_b & & \\ & \ddots & & \ddots & \\ & & t_0 & \cdots & t_b \\ & & & \ddots & \vdots \\ & & & & t_0 \end{bmatrix}$$

the preconditioner C_S is simply

$$C_S = \begin{bmatrix} t_0 & \cdots & \cdots & t_b & & \\ & \ddots & & & \ddots & \\ & & t_0 & \cdots & \cdots & t_b \\ t_b & & & t_0 & \cdots & t_{b-1} \\ \vdots & \ddots & & & \ddots & \vdots \\ t_1 & \cdots & t_b & & & t_0 \end{bmatrix}.$$

For a block matrix $T = (T_{ij})$ with Toeplitz blocks T_{ij}, the block version of the preconditioner is $C = (C_{ij})$, where each block C_{ij} is Strang's circulant approximation to T_{ij}. For systems arising from the regularized underdetermined least squares problem, $TT^T + \lambda I$, with block Toeplitz matrix T, we precondition with $CC^T + \lambda I$, where C is the block preconditioner as described above.

Our second preconditioner, developed by Hanke and Nagy [17], is an approximate inverse preconditioner for an upper banded Toeplitz matrix $T_{N\times N}$ with bandwidth less than or equal to b. It is constructed as follows. First we embed T into an $(N+b)\times(N+b)$ circulant matrix C_{HN} according to the form

$$C_{HN} = \begin{bmatrix} T & T_{12} \\ T_{21} & T_{22} \end{bmatrix}$$

where

$$T_{22} = \begin{bmatrix} t_0 & t_1 & \cdots & t_{b-1} \\ & t_0 & \cdots & t_{b-2} \\ & & \ddots & \vdots \\ & & & t_0 \end{bmatrix},$$

$$T_{21} = [L \quad 0 \quad 0], \quad T_{12} = \begin{bmatrix} 0 \\ 0 \\ L \end{bmatrix}, \quad L = \begin{bmatrix} t_b & & \\ \vdots & \ddots & \\ t_1 & \cdots & t_b \end{bmatrix}.$$

Next, we partition C_{HN}^{-1} as

$$C_{HN}^{-1} = \begin{bmatrix} M & M_{12} \\ M_{21} & M_{22} \end{bmatrix}$$

where M is the $N \times N$ leading principal submatrix. The matrix M is the approximate inverse preconditioner for T.

For the regularized underdetermined least squares matrices of the form $TT^T + \lambda I$, we embed T into a circulant block matrix C, with each block C_{ij} being a circulant extension of T_{ij} as described above. For $\lambda > 0$, $CC^T + \lambda I$ is nonsingular. We use M, the submatrix with the same set of rows and columns in $(CC^T + \lambda I)^{-1}$ as the rows and columns of entries of T in C, as the approximate inverse preconditioner to $TT^T + \lambda I$.

In practice, aside from the identity, $Q = \Delta$, the Poisson operator, or some operators derived from discrete approximation to the kth derivative are also popular choices. With a proper arrangement of the rows, these operators also have banded block Toeplitz–Toeplitz block structure, and we can extend the proposed techniques to precondition (5).

B. Complexity, Convergence, and Implementation

A preconditioner C for a matrix T should satisfy the following criteria [3, p. 253]:

- cost of computing C should be low;
- computational cost of solving a linear system with coefficient matrix C should be low;
- iterative solver should converge much faster with $C^{-1}T$ than with T.

We will demonstrate that our preconditioners satisfy the three criteria described above. In particular, circulant matrices have the useful property that they can be diagonalized by discrete Fourier transforms (cf. [8]). The eigendecomposition of a circulant matrix C can be written as follows:

$$C = F^* \Lambda F$$

where F is the unitary discrete Fourier transform matrix and Λ is a diagonal matrix containing the eigenvalues of C. We can compute the eigenvalues of C by taking the FFT of its first column.

Using this special property, we do not need to construct our preconditioners explicitly. We only need to store the entries of their first columns. Therefore, the cost of constructing the preconditioners is negligible. Additionally, operations involving circulant matrices can be done efficiently by FFTs. The cost of solving a linear system with a circulant coefficient matrix is two FFTs. For a block matrix with circulant blocks such as our preconditioners, we need to solve a linear system with a block coefficient matrix with diagonal blocks in addition to the two FFTs. The asymptotic computational complexity for the FFT is $\mathcal{O}(N \log N)$, where N is the dimension of the matrix, and correspondingly, for the linear solver with a block diagonal coefficient matrix, the complexity is $\mathcal{O}(BN)$, where B is the number of blocks. Thus, the cost of solving a linear system with our preconditioners as the coefficient matrix is inexpensive. To study the convergence behavior of preconditioned CG described here, we have the following results, the proofs of which are omitted, but may be found in [25].

Theorem 1: Let T be an upper banded Toeplitz matrix with bandwidth less than or equal to b, C_{HN} be the nonsingular extension of T, and C_S be the circulant approximation to T. If M is either C_S^{-1} or the $N \times N$ leading principal submatrix of C_{HN}^{-1} then

$$MT = I + K$$

where $rank(K) \leq b$.

The theorem above means that at most b eigenvalues of the preconditioned system are not equal to 1. So for any circulant preconditioned banded Toeplitz matrix with bandwidth b, at most $b+1$ preconditioned CG iterations are needed for convergence. This result is one of the reasons we chose our preconditioners over other circulant preconditioners, which can only claim eigenvalues of the preconditioned system "clustering" around one [7]. We can also bound the amount of work to solve a banded Toeplitz system by CG to $\mathcal{O}(bN \log(N))$ with Strang's circulant preconditioner and $\mathcal{O}(b(N+b) \log(N+b))$ with the approximate inverse preconditioner. The next theorem bounds the bandwidth of each block in the superresolution system matrix. Finally, Theorem 3 bounds the number of preconditioned CG iterations needed to solve a block matrix system with banded Toeplitz blocks using the proposed circulant preconditioners.

Theorem 2: The matrix $\mathcal{T}$ in (19) and its block Toeplitz approximate $\tilde{\mathcal{T}}$ have blocks with bandwidths bounded by $n+1$, where n is the width of an LR frame.

Theorem 3: Let T be a block matrix with upper banded Toeplitz blocks

$$T = \begin{bmatrix} T_{11} & T_{12} & \cdots & T_{1l} \\ T_{21} & T_{22} & \cdots & T_{2l} \\ \vdots & \vdots & \ddots & \vdots \\ T_{k1} & T_{k2} & \cdots & T_{kl} \end{bmatrix}$$

with the bandwidths of the blocks bounded by some constant b. If M is either the circulant or approximate inverse preconditioner to $TT^T + \lambda I$ as described in Section IV-A, then

$$M(TT^T + \lambda I) = I + K$$

where $rank(K) \leq \min(k, l)b$.

Combining Theorems 2 and 3, we get $p(n+1)$ as an approximate upper bound on the number of preconditioned CG iterations for superresolution, with p being the number of frames and n the width of an LR frame. In our experience, as we will demonstrate in Section V, this bound is quite loose, and within at most ten iterations or so, we have effective convergence. Furthermore, in practice, the two preconditioners achieve comparable results.

V. Experiments

The first test sequence consists of artificially generated LR frames. In this experiment, we compare the quality of superresolution against the original image. We blur a single 172×172 pixels image with a 4×4 Gaussian PSF with standard deviation of one and down-sample to produce 16 43×43 LR frames. Using nine (randomly chosen) out of the complete set of 16 frames, we reconstruct an estimate for the original HR image. Fig. 2 presents the results from our superresolution algorithm. The top left portion displays a sample LR frame, the top right the result of bilinearly interpolating one LR frame by a factor of four in each dimension, the bottom left the result from superresolution after four iterations, and the bottom right the original image. We stop the algorithm when the relative residual[2] tolerance of 10^{-2} is reached. We use regularization parameter $\lambda = 0.001$ calculated with our approximate GCV criterion as described in Section III-B. In Fig. 3, we compare convergence rates for preconditioned CG versus unpreconditioned CG and steepest descent. To reach tolerance threshold, four iterations of preconditioned CG are required for either preconditioner while 31 iterations are required for unpreconditioned CG and 101 iterations for steepest descent. The runtime for preconditioned CG for this simulated sequence on a Sun Sparc-20 is 17.7 s versus 42.1 s for unpreconditioned CG. For a qualitative comparison, we show in Fig. 4 reconstruction results from steepest descent, unpreconditioned, and preconditioned CG after exactly four iterations. These experiments demonstrate the advantage of using preconditioned CG over unpreconditioned CG and steepest descent. Our experiments show that in the first few iterations, steepest descent and unpreconditioned CG have similar convergence rate. However, steepest descent exhibits oscillatory convergence behavior as the number of iterations increases.

The low-resolution FLIR images in our second test sequence are provided courtesy of B. Yasuda and the FLIR research group in the Sensors Technology Branch, Wright Laboratory, WPAFB, OH. Results using this data set are also shown in [18]. Each image is 64×64 pixels, and a resolution enhancement factor of five is sought. The objects in the scene are stationary, and 16 frames are acquired by controlled movements of a FLIR imager described in [18]. Fig. 5 has similar subplot arrangements as in Fig. 2 except now the bottom right shows the relative residual graphs for steepest descent, unpreconditioned and preconditioned runs. For this sequence, we again set the relative residual tolerance to 10^{-2} and use regularization parameter $\lambda = 0.0076$ as calculated with our approximate GCV criterion. Six

[2]Relative residual is defined as the ratio $(\|\mathbf{r}_k\|/\|\mathbf{r}_0\|)$, where $\mathbf{r}_0$ is the initial residual and $\mathbf{r}_k$ is the current residual after k iterations. The residual is defined as $\mathbf{r}_k = \mathbf{b} - A\mathbf{x}_k$, and the vector $\mathbf{x}_k$ is the current estimate of the solution.

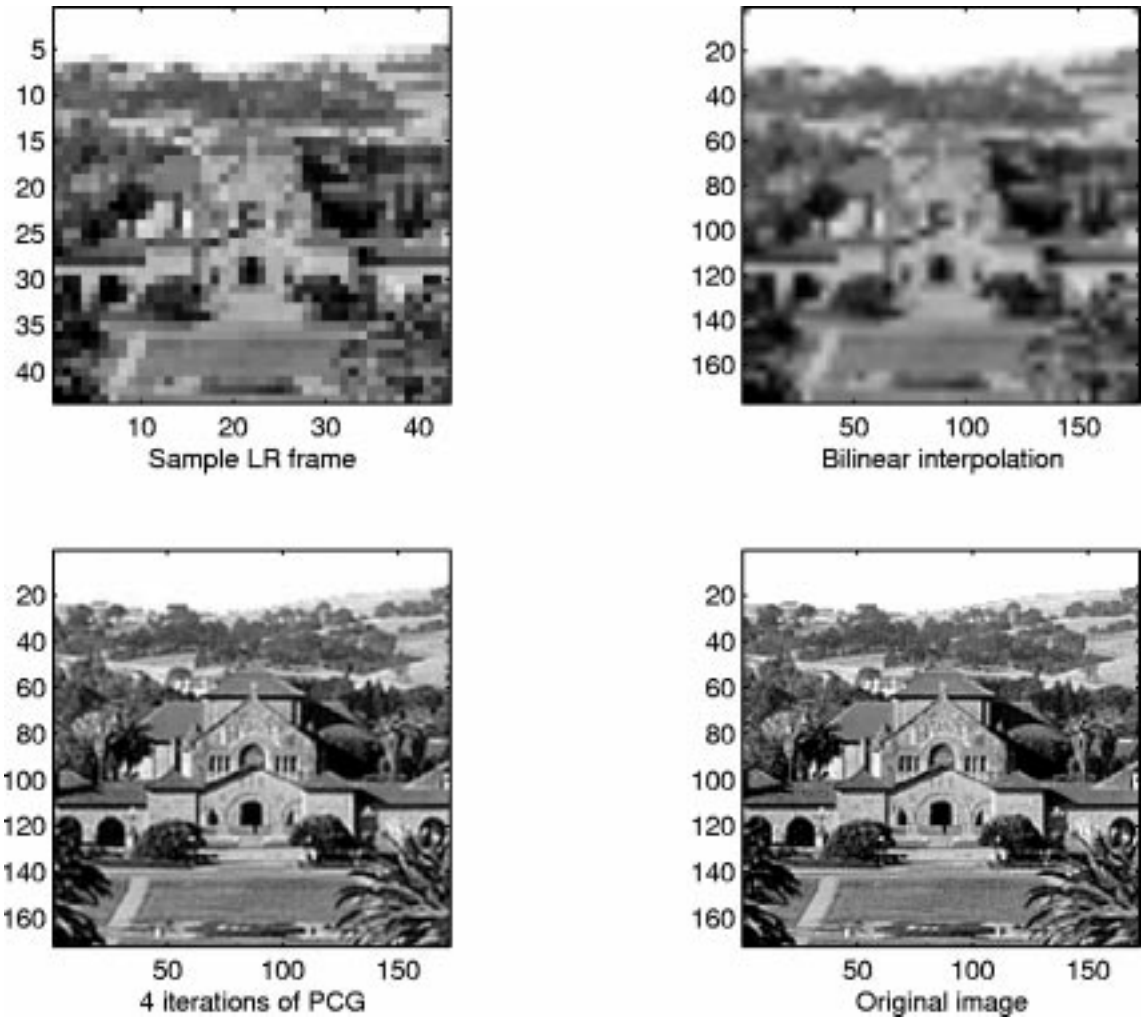

Fig. 2. Superresolution on simulated sequence.

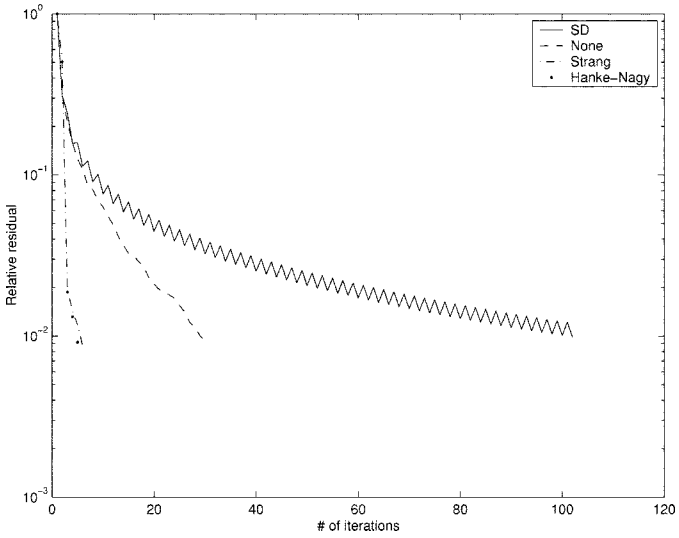

Fig. 3. Convergence plot for Stanford sequence.

iterations are required for preconditioned CG with Strang's preconditioner and eight iterations for Hanke and Nagy's approximate inverse preconditioner versus 20 for unpreconditioned CG, to reach the residual threshold. Again, we see the oscillatory behavior of steepest descent for superresolution. Preconditioned CG runtime for this FLIR sequence on our Sparc 20 is 93.6 s versus 111.3 s for unpreconditioned conjugate gradient.

VI. Conclusion and Future Work

In this paper, we presented an efficient and robust algorithm for image superresolution. The contributions in this work are twofold. First, our robust approach for superresolution reconstruction employs Tikhonov regularization. To automatically calculate the regularization parameter, we adopt the generalized cross-validation criterion to our underdetermined systems. Although generalized cross-validation is a well-known technique for parameter estimation for overdetermined least squares problems, to our knowledge, the derivation for *underdetermined* problems is new.

Secondly, to accelerate CG convergence, we proposed circulant-type preconditioners based on previous work by Strang, Hanke and Nagy. These preconditioners can be easily constructed, operations involving these preconditioners can be done efficiently by FFTs, and most importantly, the number of CG iterations is dramatically reduced. In practice, we observed that preconditioned CG takes at most $1/3$ the number of iterations of unpreconditioned CG, leading to significant

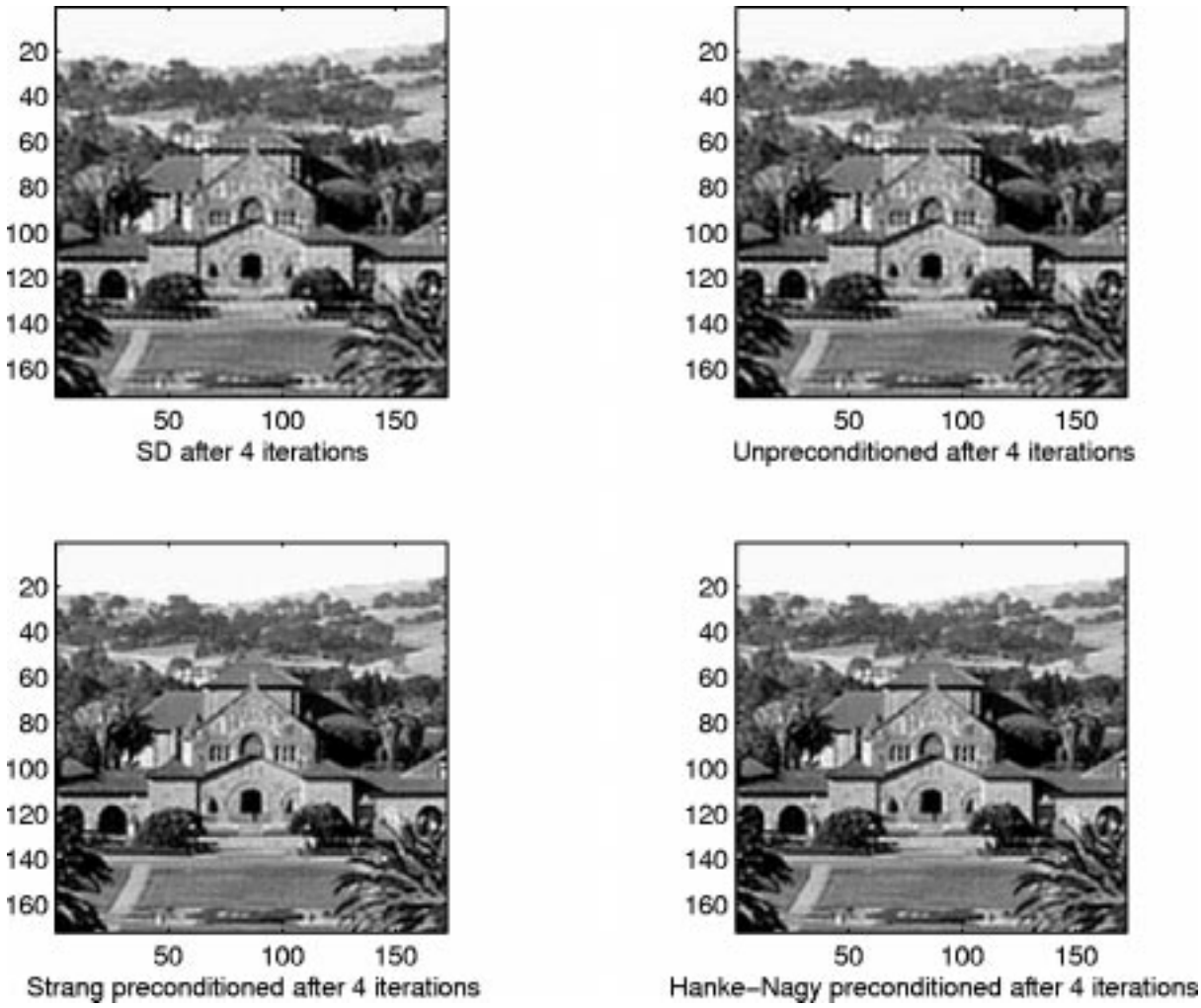

Fig. 4. Comparison of reconstruction quality.

improvement in runtime. Typically, we stop after five preconditioned CG iterations because results obtained thereafter are not significantly different visually. By these experiments, we have demonstrated that with the use of appropriate preconditioners, image superresolution is computationally much more tractable.

Image superresolution can be generalized to video superresolution [12], [13] where a sequence of superresolved images is obtained from a sequence of LR video frames. Under such conditions, the computational advantage of our preconditioners is compounded.

There is a strong relationship between the size of the regularization parameter, the condition number of the regularized system, the number of iterations of CG (unpreconditioned and preconditioned) required to solve the system and time savings with preconditioning. As we increase the regularization parameter λ, the condition number of the system decreases, leading to a faster convergence rate for both preconditioned and unpreconditioned CG. However, as mentioned before, a larger regularization parameter also moves the regularized system farther away from the original system we wish to solve. The result is a more blurry HR estimate. As we decrease λ, the system becomes more ill-conditioned, and the condition number increases. Preconditioned CG is less affected by ill-conditioned systems than the unpreconditioned version. The ratio of the number of unpreconditioned iterations over the number of preconditioned iterations increases for smaller regularization parameters. Thus, time savings with preconditioning increase with under-regularization.

An important and practical extension of the algorithm is the implementation of the positivity constraint within preconditioned CG. We note that each CG iteration computes an estimate for the term $(HH^T + \lambda I)^{-1}\mathbf{b}$ in (6). However, the constraint should be applied to the HR estimate $\mathbf{x}$. An interesting topic of future research would be an efficient algorithm to incorporate positivity constraint into our framework. We have assumed in this work that the parameters for the camera's PSF are known. In many applications, this is not necessarily the case. Blind superresolution or superresolution without accurate knowledge of the camera parameters is a challenging topic. A blind superresolution algorithm must reconstruct estimates for both the PSF and the HR image. In order to make this problem feasible, some constraints can be placed on the PSF, e.g., finite support, symmetry. Another important issue in image superresolution is the accuracy of the motion estimation process. Although a simple algorithm such as the one described in Appendix A would be adequate in most cases, more accurate algorithms are needed for higher resolution enhancement. We

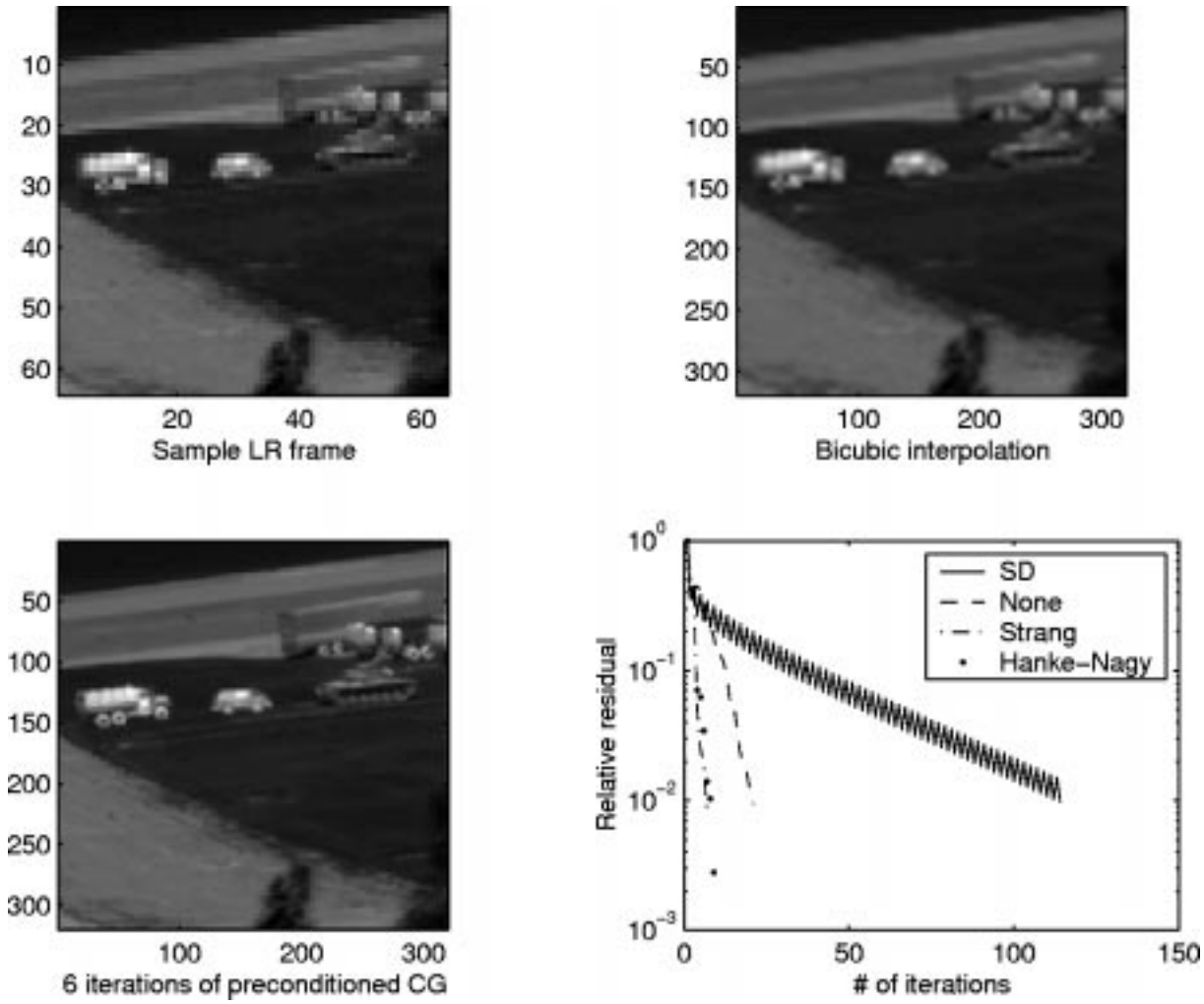

Fig. 5. Superresolution on FLIR sequence.

are currently working on these problems and will report results in subsequent submissions.

Appendix A
Motion Estimation

In the case of uncontrolled frame to frame motions, we need to estimate these motions as a precursor to superresolution. We assume that the motion is smooth, and apply Taylor's series to compute its approximation. Let $f(x, y, t)$ be the continuous frame sequence. By Taylor's expansion, we get

$$f(x+\Delta x, y+\Delta y, \Delta t) \approx f(x, y, 0) + \nabla f(x, y, 0)^T \Delta$$

where

$$\Delta = \begin{pmatrix} \Delta x \\ \Delta y \\ \Delta t \end{pmatrix}.$$

Following Irani and Peleg [20], for consecutive frames, $f(x, y)$ and $g(x, y)$, we write

$$g(x, y) \approx f(x, y) + \frac{\partial f(x, y)}{\partial x}\Delta x + \frac{\partial f(x, y)}{\partial y}\Delta y$$

and we solve the following least squares problem for Δx and Δy

$$\min_{\Delta x, \Delta y} \sum_{x, y} \left(g(x, y) - f(x, y) - \frac{\partial f(x, y)}{\partial x}\Delta x - \frac{\partial f(x, y)}{\partial y}\Delta y \right)^2$$

which leads to a 2×2 system

$$\begin{bmatrix} \sum_{x,y} \left(\frac{\partial f(x,y)}{\partial x}\right)^2 & \sum_{x,y} \frac{\partial f(x,y)}{\partial x}\frac{\partial f(x,y)}{\partial y} \\ \sum_{x,y} \frac{\partial f(x,y)}{\partial x}\frac{\partial f(x,y)}{\partial y} & \sum_{x,y} \left(\frac{\partial f(x,y)}{\partial y}\right)^2 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} = \begin{bmatrix} \sum_{x,y} (g(x,y) - f(x,y)) \frac{\partial f(x,y)}{\partial x} \\ \sum_{x,y} (g(x,y) - f(x,y)) \frac{\partial f(x,y)}{\partial y} \end{bmatrix}.$$

TABLE I
VALUES CALCULATED USING SUBOPTIMAL GCV AND GCV: WITH MISSING FRAMES

Noise Std	0	5	10
Suboptimal λ	0.001	0.0085	0.0250
Exact λ	0.001	0.0083	0.0247

TABLE II
VALUES CALCULATED USING SUBOPTIMAL GCV AND GVC: WITH ALL FRAMES

Noise Std	0	5	10
Suboptimal λ	0.001	0.0109	0.0305
Exact λ	0.001	0.0110	0.0304

Note that this idea is a simplification of the gradient constraint equation often used in optical flow calculations

$$\frac{\partial f(x,y,t)}{\partial x}\Delta x+\frac{\partial f(x,y,t)}{\partial y}\Delta y+\frac{\partial f(x,y,t)}{\partial t}=0.$$

In Irani and Peleg's formulation used above, the partial derivative with respect to time is crudely approximated by the difference between the given frames.

APPENDIX B
COMPUTING THE REGULARIZATION PARAMETER

Evaluating (18) as it stands requires intensive computation. We instead approximate λ_{GCV} by replacing A by its preconditioner C_S in (18). The alternate formulation is

$$\tilde{\lambda}_{\mathrm{GCV}}=\min_{\lambda}\frac{\|(C_SC_S^T+\lambda I)^{-1}\mathbf{b}\|_2}{\mathrm{tr}((C_SC_S^T+\lambda I)^{-1})}.$$

The motivation here is that since C_S approximates A well, $\tilde{\lambda}_{\mathrm{GCV}}$ should be close to λ_{GCV}. Even so, calculating the term $\mathrm{tr}((C_SC_S^T+\lambda I)^{-1})$ exactly is still infeasible. Therefore, we use the unbiased trace estimator proposed by Hutchinson [19]. Let U be a discrete random variable which takes the values -1 and $+1$ each with probability $1/2$, and let $\mathbf{u}$ be a vector whose entries are independent samples from U. Then the term $\mathbf{u}^T(C_SC_S^T+\lambda I)^{-1}\mathbf{u}$ is an unbiased estimator of $\mathrm{tr}((C_SC_S^T+\lambda I)^{-1})$.

Since C_S is a block matrix with circulant blocks, we can decompose $C_S=F^*\Lambda F$, where Λ is a block matrix with diagonal blocks, and F is the block discrete Fourier transform matrix which diagonalizes the blocks of C_S. The minimization problem above becomes

$$\min_{\lambda}\frac{\|F^*(\Lambda\Lambda^*+\lambda I)^{-1}F\mathbf{b}\|_2}{\mathbf{u}^TF^*(\Lambda\Lambda^*+\lambda I)^{-1}F\mathbf{u}}.$$

Let $\hat{\mathbf{b}}=F\mathbf{b}$ and $\hat{\mathbf{u}}=F\mathbf{u}$, since $\|\cdot\|_2$ is invariant under unitary transformation, we have

$$\min_{\lambda}\frac{\|(\Lambda\Lambda^*+\lambda I)^{-1}\hat{\mathbf{b}}\|_2}{\hat{\mathbf{u}}^*(\Lambda\Lambda^*+\lambda I)^{-1}\hat{\mathbf{u}}}.$$

We solve this minimization problem with Matlab's CONSTR subroutine with 0.001 being the lower bound. In our experiments, we found that this formulation produces reasonable regularization parameters for both the simulated and FLIR image sequence. Tables I and II list the values for our suboptimal regularization parameter $\tilde{\lambda}_{\mathrm{GCV}}$ along with the optimal GCV value λ_{GCV} for the Stanford sequence example. Results under various noise conditions for ten frames are compiled in Table I and with all 16 frames in Table II.

REFERENCES

[1] K. Aizawa, T. Komatsu, and T. Saito, "Acquisition of very high resolution images using stereo cameras," in *Proc. SPIE Visual Communications Image Processing '91*, vol. 1605, Boston, MA, Nov. 1991, pp. 318–328.

[2] H. Andrews and B. Hunt, *Digital Image Restoration*. Englewood Cliffs, NJ: Prentice-Hall, 1977.

[3] O. Axelsson, *Iterative Solution Methods*. New York: Cambridge Univ. Press, 1994.

[4] B. Bascle, A. Blake, and A. Zisserman, *Motion Deblurring and Super-Resolution from an Image Sequence*: EECV, Apr. 1996, pp. 573–581.

[5] L. Brown, "A survey of image registration techniques," *ACM Comput. Surv.*, vol. 24, pp. 325–376, Dec. 1992.

[6] R. Chan, T. Chan, M. Ng, W. Tang, and C. Wong, "Preconditioned iterative methods for high-resolution image reconstruction from multisensors," *Adv. Signal Process. Algorithms, Architectures, Implementations*, vol. 3461, pp. 348–357, 1998.

[7] R. Chan and M. Ng, "Conjugate gradient methods for Toeplitz systems," *SIAM Rev.*, vol. 38, pp. 427–482, Sept. 1996.

[8] R. Chan and G. Strang, "Toeplitz equations by conjugate gradient with circulant preconditioners," *SIAM J. Sci. Statist. Comput.*, vol. 10, pp. 104–119, Jan. 1989.

[9] P. Cheeseman, B. Kanefsky, R. Kraft, J. Stutz, and R. Hanson, "Super-resolved surface reconstruction from multiple images," NASA Ames Research Center, Moffet Field, CA, Tech. Rep. FIA-94-12 NASA, Dec. 1994.

[10] T. Connolly and R. Lane, "Gradient methods for superresolution," in *Proc. Int. Conf. Image Processing*, vol. 1, 1997, pp. 917–920.

[11] M. Elad and A. Feuer, "Restoration of a single superresolution image from several blurred, noisy, and undersampled measured images," in *IEEE Trans. Image Processing*, vol. 6, Dec. 1997, pp. 1646–1658.

[12] ——, "Super-resolution reconstruction of continuous image sequence," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 21, pp. 817–834, Sept. 1999.

[13] ——, "Super-resolution restoration of an image sequence: Adaptive filtering approach.," *IEEE Trans. Image Processing*, vol. 8, pp. 387–395, Mar. 1999.

[14] G. Golub, M. Heath, and G. Wahba, "Generalized cross-validation as a method for choosing a good ridge parameter," *Technometrics*, vol. 21, pp. 215–223, 1979.

[15] G. Golub and C. van Loan, *Matrix Computations*, 2nd ed. Baltimore, MD: Johns Hopkins Univ. Press, 1989.

[16] M. Hanke and P. Hansen, "Regularization methods for large-scale problems," *Surv. Math. Ind.*, vol. 3, pp. 253–315, 1993.

[17] M. Hanke and J. Nagy, "Toeplitz approximate inverse preconditioner for banded Toeplitz matrices," *Numer. Alg.*, vol. 7, pp. 183–199, 1994.

[18] R. Hardie, K. Barnard, and E. Armstrong, "Joint MAP registration and high-resolution image estimation using a sequence of undersampled images," *IEEE Trans. Image Processing*, vol. 6, pp. 1621–1633, Dec. 1997.

[19] M. Hutchinson, "A stochastic estimator of the trace of the influence matrix for Laplacian smoothing splines," *Commun. Statist., Simul., and Comput.*, vol. 19, pp. 433–450, 1990.

[20] M. Irani and S. Peleg, "Improving resolution by image registration," *CVGIP: Graph. Models Image Process.*, vol. 53, pp. 324–335, Dec. 1993.

[21] S. Kim, N. Bose, and H. Valenzuela, "Recursive reconstruction of high resolution image from noisy undersampled multiframes," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, no. 6, June 1990.

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

[23] S. Mann and R. Picard, "Virtual bellows: Construction high quality stills from video," in *Proc. IEEE Int. Conf. Image Processing*, Austin, TX, Dec. 1994.

[24] P. Mcintosh and G. Veronis, "Solving underdetermined tracer inverse problems by spatial smoothing and cross validation," *J. Phys. Oceanogr.*, vol. 23, pp. 716–730, Apr. 1993.

[25] N. Nguyen, "Numerical Techniques for Image Superresolution," Ph.D. dissertation, Stanford Univ., Stanford, CA, Apr. 2000.
[26] A. Patti, M. Sezan, and A. Tekalp, "Image sequence restoration and de-interlacing by motioncompensated Kalman filtering," *Proc. SPIE*, vol. 1903, 1993.
[27] A. Patti, M. Sezan, and A. Tekalp, "Superresolution video reconstruction with arbitrary sampling lattices and nonzero aperture time," *IEEE Trans. Image Processing*, vol. 6, pp. 1064–1076, Aug. 1997.
[28] Y. Saad, *Iterative Methods for Sparse Linear Systems*. Boston, MA: PWS, 1996.
[29] K. Sauer and J. Allebach, "Iterative reconstruction of band-limited images from nonuniformly spaced samples," *IEEE Trans. Circuits Syst.*, vol. CAS-34, pp. 1497–1505, 1987.
[30] R. Schultz and R. Stevenson, "Extraction of high-resolution frames from video sequences," *IEEE Trans. Image Processing*, vol. 5, pp. 996–1011, June 1996.
[31] H. Shekarforoush and R. Chellappa, "Data-driven multi-channel superresolution with application to video sequences," *J. Opt. Soc. Amer. A*, vol. 16, pp. 481–492, Mar. 1999.
[32] 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.
[33] A. Tekalp, M. Ozkan, and M. Sezan, "High-resolution image reconstruction from lower resolution image sequences and space-varying image restoration," in *Proc. Int. Conf. Acoustics, Speech, Signal Processing '92*, vol. 3, San Francisco, CA, Mar. 1992, pp. 169–172.
[34] R. Tsai and T. Huang, "Multiframe image restoration and registration," in *Advances in Computer Vision and Image Processing*. Greenwich, CT: JAI, 1984, vol. 1.
[35] H. Ur and D. Gross, "Improved resolution from subpixel shifted pictures," *CVGIP: Graph. Models Image Process.*, vol. 54, pp. 181–186, Mar. 1992.
[36] A. van der Sluis and H. van der Vorst, "The rate of convergence of conjugate gradients," *Numer. Math.*, vol. 48, pp. 543–560, 1986.

Nhat Nguyen received B.S. degrees in mathematics and computer science from the California Institute of Technology, Pasadena, in 1994. He recently completed his Ph.D. dissertion with the Scientific Computing and Computational Mathematics Program, Stanford University, Palo Alto, CA.

He is currently a Software Engineer with the Surfscan Division, KLA-Tencor Corporation, Milpitas, CA. His technical interests are in signal and image processing algorithms and numerical linear algebra.

Peyman Milanfar (S'90–M'93–SM'98) received the B.S. degree in engineering mathematics from the University of California, Berkeley, in 1988, and the S.M., E.E., and Ph.D. degrees in electrical engineering from the Massachusetts Institute of Technology, Cambridge, in 1990, 1992, and 1993, respectively.

From 1993 to 1994, he was a Member of Technical Staff at Alphatech, Inc., Burlington, MA. From 1994 to 1999, he was a Senior Research Engineer at SRI International, Menlo Park, CA. He is currently an Assistant Professor of electrical engineering at the University of California, Santa Cruz. He was a Consulting Assistant Professor of computer science at Stanford University from 1998 to 2000. His technical interests are in statistical and numerical methods for signal and image processing, and more generally inverse problems, particularly those involving moments.

Dr. Milanfar was awarded a National Science Foundation CAREER grant in January 2000. He is an Associate Editor for the IEEE SIGNAL PROCESSING LETTERS.

Gene Golub was born in Chicago, IL, in 1932. He received the Ph.D. degree from the University of Illinois, Urbana, in 1959.

Subsequently, he held an NSF Fellowship at the University of Cambridge, Cambridge, U.K. He joined the faculty of Stanford University, Stanford, CA, in 1962, where he is currently the Fletcher Jones Professor of computer science and the Director of the Program in Scientific Computing and Computational Mathematics. He served as Chairman of the Computer Science Department from 1981 until 1984. He is noted for his work in the use of numerical methods in linear algebra for solving scientific and engineering problems. This work resulted in a book, *Matrix Computations*, co-authored with Charles Van Loan.

He is the Founding Editor of the *SIAM Journal of Scientific and Statistical Computing* and the *SIAM Journal of Matrix Analysis and Applications*. He is the originator of na-net.

Dr. Golub holds several honorary degrees. He is a member of The National Academy of Engineering, the National Academy of Sciences and the American Academy of Sciences. He has served as President of the Society of Industrial and Applied Mathematics (1985–1987) and