Advance Access publication on April 12, 2007 doi:10.1093/comjnl/bxm007

Optimal Registration Of Aliased Images Using Variable Projection With Applications To Super-Resolution

DIRK ROBINSON[1,*], SINA FARSIU[2] AND PEYMAN MILANFAR[2]
[1]*Ricoh Innovations Inc., 2882 Sand Hill Rd, Suite 115, Menlo Park, CA 94025, USA*
[2]*Electrical Engineering Department, University of California Santa Cruz, Santa Cruz, CA 95064, USA*
**Corresponding author: dirkr@rii.ricoh.com*

Accurate registration of images is the most important and challenging aspect of multiframe image restoration problems such as super-resolution. The accuracy of super-resolution algorithms is quite often limited by the ability to register a set of low-resolution images. The main challenge in registering such images is the presence of aliasing. In this paper, we analyse the problem of jointly registering a set of aliased images and its relationship to super-resolution. We describe a statistically optimal approach to multiframe registration which exploits the concept of variable projections to achieve very efficient algorithms. Finally, we demonstrate how the proposed algorithm offers accurate estimation under various conditions when standard approaches fail to provide sufficient accuracy for super-resolution.

Keywords: Multi-frame motion estimation, variable projection, super-resolution, aliasing

Received 14 May 2006; revised 25 October 2006

1. INTRODUCTION

The problem known as super-resolution, namely combining a set of aliased, noisy, low-resolution and blurry images to produce a higher resolution image or image sequence, has received much attention in recent years. We refer the interested readers to [1–3] for a broad review of recent algorithmic development in this area. Perhaps, the most fundamental component to multi-frame super-resolution is the accurate registration of aliased images.

The difficulty of registration for super-resolution is the presence of aliasing inherent to most modern digital imaging systems. In traditional sampling of electronic waveforms, analog engineers have a fair amount of control over the bandwidth properties of electric signals. Because of this, one rarely encounters the problem of aliasing in traditional one-dimensional sampling. When sampling optical fields, however, control over the spatial frequency bandwidth is limited by the nature of optics.

Designing optical low-pass filters using lenses and filters require finding a delicate balance between excessively blurring an optical image and introducing spatial aliasing artifacts into the lens system. The typical approach to designing optical systems involves first designing high-quality optics with large enough apertures to increase the amount of light, and hence signal, sampled by a focal plane array (FPA). Suppose for the time being that the optical lens is a diffraction-limited lens with a modest aperture size corresponding to an F# of 4.0. In the visible spectrum ($\sim$500 nm wavelength), such optical system has diffraction limits of $\sim$500 line pairs per millimeter (lp/mm) [4]. The dashed curve of Fig. 1 shows the modulation transfer function (MTF) of such an optical system. In current technologies, FPAs have pixel dimensions ranging from 3 to 15 μm, which correspond to Nyquist sampling rates ranging from $\sim$160 to $\sim$35 lp/mm.

Thus, we observe that the sampling rate of modern FPAs arrays is significantly lower than the spatial frequencies passed by optical systems. That is, the apertures associated with these pixels are often insufficient to eliminate higher spatial frequencies. To make matters worse, the effective aperture or fill-factor of modern pixels is rarely the same as the pixel dimension. For example, low-cost CMOS detectors, without the benefit of expensive microlens technology, may have fill-factors as low as 40–50%. Furthermore, by adding a color filter array (CFA), the effective fill-factor is reduced by a factor of two. For example, the dash-dot line in Fig. 1 shows the effective MTF of a 4 μm square pixel having a fill-factor of 50%. The solid line shows the effective imaging MTF for the system. We observe that the imaging system passes

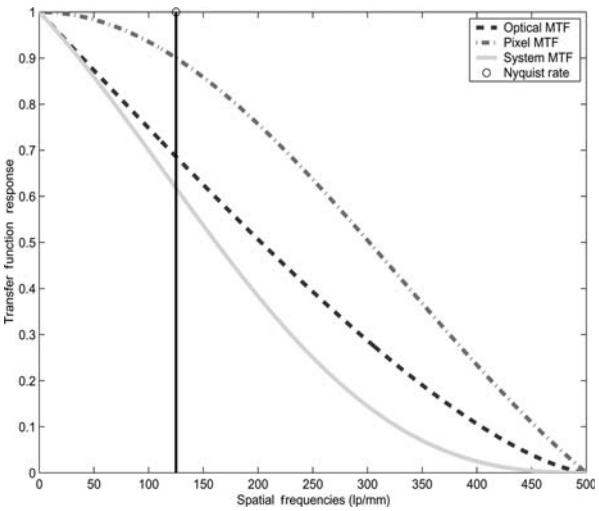

FIGURE 1. Graph showing the diffraction-limited lens system. The dashed curve shows the optical MTF of an F# 4 diffraction-limited lens system at 500 nm wavelength light. The dash-dot curve shows the pixel transfer function for a 4 μm square detector with 50% fill-factor. The solid line represents the total imaging system transfer function. The black line indicates the Nyquist rate associated with the detector. In this case, we observe that the imaging system is undersampled by a factor of $M = 4$ without any zero-crossings up to the diffraction limit.

many spatial frequencies above the Nyquist rate (indicated by the black line) associated with the detector. Even in expensive optical systems when optical low-pass filters (often based on birefringent optical materials [5]) are added to the system, a significant amount of spatial frequencies above the Nyquist rate is passed by the optical system.

Because the imaging system almost invariably passes spatial frequencies above the Nyquist, almost all digital images contain varying amounts of aliasing artifacts. In fact, it is these aliasing artifacts which enable multiframe super-resolution enhancement. Multiple captured images which contain phase variations allow super-resolution algorithms to *unwrap* these aliasing artifacts, reconstructing the high-resolution image. While there exist motion free super-resolution algorithms, which utilize variations in the amplitude of the optical transfer function [6, 7], the more common form of super-resolution relies on variations in the phase of the optical transfer function in the form of inter-frame subpixel shifts contained in a set of low-resolution images. It is phase information, which allows super-resolution algorithms to reconstruct the high-resolution image.

The standard mathematical model describing a set of captured images is given by

$$\boldsymbol{y}_k = \boldsymbol{DHF}(\boldsymbol{v}_k)\boldsymbol{x} + \boldsymbol{e}_k, \tag{1}$$

where $\boldsymbol{x}$ represents the unknown $N_{\mathrm{H}} \times N_{\mathrm{H}}$, bandlimited high-resolution image lexicographically ordered as a $N_{\mathrm{H}}^2 \times 1$ vector. The bandlimited nature of the high-resolution image is assured due to the diffraction limit of the optical system. The warping operator $\boldsymbol{F}(\boldsymbol{v}_k)$ of size $N_{\mathrm{H}}^2 \times N_{\mathrm{H}}^2$ is comprised of shifted periodic *sinc* functions. The shifts for each pixel in the image are defined by the vector of motion parameters $\boldsymbol{v}_k$ for the kth frame. Matrix $\boldsymbol{H}$ of size $N_{\mathrm{H}}^2 \times N_{\mathrm{H}}^2$ represents the blurring operator due to the combined effect of the imaging system optical and pixel point spread functions (PSF). The downsampling operator $\boldsymbol{D}$ of size $N_{\mathrm{L}}^2 \times N_{\mathrm{H}}^2$ captures the undersampling of the detector by an integer factor M. The $N_{\mathrm{L}}^2 \times 1$ vector $\mathbf{y}_k$ is created by lexicographically ordering the kth $N_{\mathrm{L}} \times N_{\mathrm{L}}$ low-resolution image. We assume that super-resolution is applied given $K + 1$ such low-resolution images. Finally, $\mathbf{e}_k$ of size $N_{\mathrm{L}}^2 \times 1$ represents the noise inherent in the analog-to-digital conversion. For our purposes, we assume this noise to be uncorrelated zero-mean Gaussian noise with standard deviation σ. Without loss of generality, we assume that the initial image $\boldsymbol{y}_0$ defines the coordinate system of the high-resolution image so that $\boldsymbol{y}_0 = \boldsymbol{DHx} + \boldsymbol{e}_0$ and hence, we only have to estimate K unknown motion vector parameters $\boldsymbol{v}_k$ during the super-resolution process for a given set of $K + 1$ low resolution frames. The set of all such unknown motion vectors is represented as $\underline{\mathrm{v}} = [\boldsymbol{v}_1, \ldots \boldsymbol{v}_K]^{\mathrm{T}}$.

Super-resolution algorithms for estimating the unknown high-resolution image can generally be divided into two classes of approaches. In the first, the problem is broken up into separate tasks of registering the low-resolution images with respect to the coordinate system of the desired high-resolution image (estimating $\underline{\mathrm{v}}$), followed by a reconstruction or fusion of the low-resolution data combined with deblurring and interpolation (estimating $\boldsymbol{x}$). The majority of the effort in this class has tended to focus on the reconstruction and restoration aspects of the super-resolution problem. In much of these works, it is assumed that relatively high quality estimations of the inter-frame sampling offsets can be estimated to an arbitrary degree of accuracy. In the second class of algorithms, researchers have attempted to solve both estimation problems of image registration and super-resolution in a joint fashion [8–11].

The first class makes the implicit assumption that registration of aliased images is possible without concern for the original high resolution $\boldsymbol{x}$ from which the captured low-resolution images are derived. After estimating the motion parameters for a set of images, a second step uses these estimates to reconstruct an estimate of the high-resolution image $\hat{\boldsymbol{x}}$ [12–14]. The simplest, and most common, approach for estimating the set of motion parameters is to apply a pairwise registration technique between a reference frame ($\boldsymbol{y}_0$) and the other low-resolution frames. The pairwise subpixel registration is often based on either standard registration techniques such as gradient-based approaches [15] or algorithms especially designed for aliased images [16–18].

In [19], it was shown that from a maximum likelihood (ML) perspective (i.e. no prior information), pairwise image

registration of aliased images is ill-posed. Consequently, all pairwise registration algorithms will exhibit certain amounts of estimator bias. This estimator bias relates to either the implicit or explicit assumptions made when constructing the pairwise registration algorithm. For instance, applying standard registration techniques designed for non-aliased images (e.g. gradient-based methods [15]) assumes intensity is conserved over the set of images. This assumption obviously breaks down in the presence of aliasing. Again, the robust phase estimation algorithms of [16, 17] make assumptions about the spectral decay of a typical image. The effects of such bias in pairwise algorithms have been noticed motivating robust approaches to minimize the effects of poor registration estimates on the final image reconstruction [13]. While the pairwise estimation approach has proven to be a computationally tractable and reasonably effective approach, it is important to understand its inherent limitations.

Another approach to registering low-resolution images attempts to estimate the set of motion parameters using the entire collection of observed images [20–23]. These papers operate under the assumption that the combinatorial set of pairwise estimated motion vector fields are constrained to lie in a space whose geometry and structure is conveniently described via an algebraic group structure (a Lie algebra). Such constraints dictate that the operator describing the motion between any pair of frames must be the composition of the operators between another pair of frames. These models, however, do not consider the presence of aliasing in the captured images, limiting their applicability to image registration for super-resolution [23].

The second class of algorithms combine the registration problem into the super-resolution estimation problem [8, 11, 24]. In [19], such a joint method was shown to be statistically optimal. To date, all of these algorithms use some form of cyclic coordinate descent or a form of the Expectation–Maximization (EM) algorithm [11, 24]. Such algorithms iterate between estimating the high-resolution image $\boldsymbol{x}$ (expectation step) and the set of unknown motion parameters (maximization step). Often the methods rely on either an initial guess of the high-resolution image $\boldsymbol{x}$ using interpolation of a single low-resolution frame, or by using a standard pairwise registration algorithm to find an initial guess for the set of motion parameters. While this second approach is more principled than the two-stage approach, we show that the use of cyclic coordinate descent is not the most efficient search strategy. It has been shown [25], that optimization using the coordinate descent approach is numerically less well conditioned and likely to become stuck in local minima.

In this paper, we propose an alternate framework for solving the joint registration/reconstruction aspect of super-resolution. Our approach is a multiframe registration solution based on the principle of variable projections [25]. In Section 2, we describe how to avoid the shortcomings of the cyclic coordinate descent using this variable projection formulation and present an efficient implementation for the case of translational motion. We use this estimator as a prototypical registration algorithm to analyse the performance of general aliased image registration. In Section 3, we conduct a series of experiments aimed at describing the applicability of this approach to various registration scenarios where the standard approaches to registration perform poorly. Finally, in Section 4, we conclude and show some future directions of this work.

2. REGISTERING ALIASED IMAGES

In this section, we describe a method for efficiently registering a set of aliased images using the concept of variable projections. We first introduce the multiframe estimation problem in the ML framework where no prior information about the high-resolution image is known. Later, we describe how the ML solution can be extended to incorporate prior information [Maximum *a posteriori* (MAP) estimation] in an efficient manner. We conclude this section by describing a fast implementation for the case of translational motion.

2.1. Multi-frame ML registration

Given the forward imaging model of Equation (1), the natural ML estimate of the high-resolution image $\boldsymbol{x}$ and the set of motion parameters $\underline{\mathrm{v}} = [\boldsymbol{v}_1, \ldots, \boldsymbol{v}_K]^{\mathrm{T}}$ is a function of the entire set of captured low-resolution images. Since the noise terms are uncorrelated Gaussian random variables, the ML estimate of the unknown parameters is defined as the minimum of a non-linear least squares (LS) cost function

$$J_{\mathrm{ML}}(\boldsymbol{x}, \underline{\mathrm{v}}) = \frac{1}{\sigma^2} \sum_k \|\boldsymbol{y}_k - \boldsymbol{DHF}(\boldsymbol{v}_k)\boldsymbol{x}\|^2. \tag{2}$$

To simplify the notation for the moment, we can think of building one single measurement vector $\boldsymbol{y} = [\boldsymbol{y}_0^{\mathrm{T}}, \ldots, \boldsymbol{y}_K^{\mathrm{T}}]^{\mathrm{T}}$ and one large system matrix $\boldsymbol{B}(\underline{\mathrm{v}}) = [(\boldsymbol{DH})^{\mathrm{T}}, \ldots, (\boldsymbol{DHF}(\boldsymbol{v}_K))^{\mathrm{T}}]^{\mathrm{T}}$. Using this formulation we can rewrite Equation (2) as

$$J_{\mathrm{ML}}(\boldsymbol{x}, \underline{\mathrm{v}}) = \frac{1}{\sigma^2} \|\boldsymbol{y} - \boldsymbol{B}(\underline{\mathrm{v}})\boldsymbol{x}\|^2. \tag{3}$$

The cost function of Equation (3) falls into the class of what are known as separable non-linear LS problems [26]. Ignoring the possible ill-posedness of the system for the time being, the solution to the ML reconstruction problem (were the motion parameters $\underline{\mathrm{v}}$ known) would be given by the LS

$$\hat{\boldsymbol{x}} = (\boldsymbol{B}(v)^{\mathrm{T}}\boldsymbol{B}(\underline{\mathrm{v}})^{-1}\boldsymbol{B}(\underline{\mathrm{v}})^{\mathrm{T}}\boldsymbol{y}. \tag{4}$$

Plugging the estimate of Equation (4) back into the original ML cost function of Equation (3) we obtain

$$J_{\mathrm{ML}}(v) = \|\boldsymbol{y} - \boldsymbol{B}(\underline{\mathrm{v}})(\boldsymbol{B}(\underline{\mathrm{v}})^{\mathrm{T}}\boldsymbol{B}(\underline{\mathrm{v}}))^{-1}\boldsymbol{B}(\underline{\mathrm{v}})^{\mathrm{T}}\boldsymbol{y}\|^2 = \|(\boldsymbol{I} - \boldsymbol{B}(\underline{\mathrm{v}})(\boldsymbol{B}(\underline{\mathrm{v}})^{\mathrm{T}}\boldsymbol{B}(\underline{\mathrm{v}}))^{-1}\boldsymbol{B}(\underline{\mathrm{v}}))^{\mathrm{T}}\boldsymbol{y}\|^2. \quad (5)$$

After some additional algebraic manipulation, we see that minimizing Equation (5) is equivalent to maximizing the functional

$$J_{\mathrm{ML}}(\underline{\mathrm{v}}) = \boldsymbol{y}^{\mathrm{T}}\boldsymbol{B}(\underline{\mathrm{v}})(\boldsymbol{B}(\underline{\mathrm{v}})^{\mathrm{T}}\boldsymbol{B}(\underline{\mathrm{v}}))^{-1}\boldsymbol{B}(\underline{\mathrm{v}})^{\mathrm{T}}\boldsymbol{y}. \quad (6)$$

After minimizing Equation (5) or maximizing Equation (6), the estimates for the motion parameters v can be used to reconstruct the estimate of the high-resolution image $\boldsymbol{x}$ using Equation (4).

The form of Equation (5) is known as the variable projection functional [25] as the operator $\boldsymbol{B}(\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B})^{-1}\boldsymbol{B}^{\mathrm{T}}$ is a projection operator. This form of the cost function has several advantages over the original cost function of Equation (3). Namely, the search space is a function of far fewer unknowns. In fact, in the case of any parametric motion model, often the dimension of the unknown motion parameters $\underline{\mathrm{v}}$ is much smaller than the dimension of the unknown image $\boldsymbol{x}$. More importantly, it can be shown that minimizing the variable projections cost function of Equation (5) converges in far fewer iterations than trying to minimize Equation (3) directly or via cyclic coordinate descent. Furthermore, the minima of Equation (5) and hence Equation (6) are better defined. This model was further supported by work in [27] published while our paper was under review. We refer the interested reader to [25] for more information on variable projections.

While efficient minimization of Equation (5) for large images is difficult due to the shear size of the matrices involved, the problem is tractable for small images. In fact, it is precisely such scenarios (where the observed images are very small or the undersampling factors M are very large) that the traditional approaches to aliased image registration break down.

2.2. Multi-frame MAP registration

If we have some prior information about the unknown high-resolution image $\boldsymbol{x}$, we may incorporate this into the cost function to regularize the estimation problem. For instance, one common assumption about the unknown signal is that it is a Gaussian random variable with a covariance matrix $\boldsymbol{C}_{\boldsymbol{x}}$. Assuming for the time being that the signal is zero-mean, the MAP estimate of the high-resolution image and motion parameters is the minimum of the cost function[1]

$$J_{\mathrm{MAP}}(\boldsymbol{x}, \underline{\mathrm{v}}) = \frac{1}{\sigma^2}\|\boldsymbol{y} - \boldsymbol{B}(\underline{\mathrm{v}})\boldsymbol{x}\|^2 + \lambda\boldsymbol{x}^{\mathrm{T}}\boldsymbol{C}_{\boldsymbol{x}}^{-1}\boldsymbol{x}. \quad (7)$$

[1]Note that it is also possible to consider priors on the motion vectors [28].

It is this equation that the authors of [8, 11, 24] attempt to minimize using iterative cyclic coordinate descent type algorithms.

Using the variable projection approach described for the ML solution, we can find the analogous non-linear cost function for the MAP estimator. For a given estimate of the motion parameters $\underline{\mathrm{v}}$, the MAP estimate of the high-resolution image is given by

$$\hat{\boldsymbol{x}} = (\boldsymbol{B}(\underline{\mathrm{v}})^{\mathrm{T}}\boldsymbol{B}(\underline{\mathrm{v}}) + \lambda\boldsymbol{C}_{\boldsymbol{x}}^{-1})^{-1}\boldsymbol{B}(\underline{\mathrm{v}})^{\mathrm{T}}\boldsymbol{y}. \quad (8)$$

Plugging this into the MAP function of Equation (7), we obtain a variable projection maximization functional of the form

$$J_{\mathrm{MAP}}(v) = \boldsymbol{y}^{\mathrm{T}}\boldsymbol{B}(v)(\boldsymbol{B}(\underline{\mathrm{v}})^{\mathrm{T}}\boldsymbol{B}(\underline{\mathrm{v}}) + \lambda\boldsymbol{C}_{\boldsymbol{x}}^{-1})^{-1}\boldsymbol{B}(\underline{\mathrm{v}})^{\mathrm{T}}\boldsymbol{y}. \quad (9)$$

The form of this functional is very similar to the ML function of Equation (6).

Addition of the term $\lambda\boldsymbol{C}_{\boldsymbol{x}}^{-1}$ not only act as a form of regularization, but also stabilizes the search for a global minimum. In fact, an unexplored area of research is in adapting the tuning parameter λ during optimization to define a trust region in a type of Levenberg–Marquardt optimization [29] to more efficiently and accurately find the global optimum of Equation (9).

2.3. The case of translational motion and space invariant blur

We examine the case of simple translational motion or shift where the unknown motion parameter has shifts in the two directions, $\boldsymbol{v} = [v_1, v_2]^{\mathrm{T}}$. This is perhaps the most commonly utilized motion model for aliased image registration algorithms [16]. Using this assumption, we can simplify the problem and produce a very fast algorithm. The translational model is often fairly well justified for small image patches where even complicated global motion fields will appear as approximately translational. Also, the blurring model for such local patches is very likely to be spatially invariant.

Given that both the motion and the blurring operators are spatially invariant, $\boldsymbol{H}$ and $\boldsymbol{F}_k$ are block circulant matrices and therefore commute. A natural transformation of the unknown image $\boldsymbol{x}$ emerges as $\boldsymbol{z} = \boldsymbol{H}\boldsymbol{x}$ [12, 13], which maps the unknown high-resolution image into a blurry version of the high-resolution image $\boldsymbol{z}$. The modified signal model becomes

$$\boldsymbol{y}_k = \boldsymbol{D}\boldsymbol{F}(\boldsymbol{v}_k)\boldsymbol{z} + \boldsymbol{e}'_k. \quad (10)$$

Such an approach is justified by the invariance property of the ML estimator [26]. For the remainder of this Section, we will ignore the subsequent step of restoring the high-resolution image $\boldsymbol{x}$ from an estimate of the reconstructed blurry image

z and instead focus on the problem of registering the observed aliased images with the intent to reconstruct the image z.

Thinking about the problem in this way offers a statistically justifiable approach to separating the problem of estimating $\boldsymbol{x}$ from that of jointly estimating the blurred high-resolution image z and the unknown motion parameters $\boldsymbol{v}$. Under certain scenarios, this presents a natural alternative to the more complex coordinate descent approach of [11], where the coordinate descent optimization is employed to estimate the motion parameters, high-resolution image, blur operator and noise power. As shown in [19], when a sufficient number of frames are available, the estimation of $\{z, v\}$ is often well conditioned. In other words, adding prior information offers very little benefit when estimating $\{z, v\}$. This suggests that the more complicated step of estimating the high-resolution image, blur and noise power can justifiably be considered after estimating $\{z, v\}$, which simplifies the search space by reducing the number of unknowns. Furthermore, such a two-stage approach would benefit from allowing much more powerful (yet highly non-linear) priors such as [30] to be incorporated into the second step, resulting in accurate reconstruction of the high-resolution image $\boldsymbol{x}$.

2.4. Efficient implementation

In this section, we describe an efficient implementation of the variable projection functionals of Equations (5) and (9) for the case of translational motion with spatially invariant blur. The algorithmic efficiency is achieved due to fast and memory efficient inversion of the matrices in the Fourier domain. We note that the efficiency of treating the problem in the Fourier domain has been noted previously [11, 14, 31].

To distinguish the Fourier representation from the spatial domain representation, we utilize the tilde ($\sim$) to indicate matrices and vectors in the Fourier domain. Thus, the simplified model of (Equation 10) is represented in the frequency domain as:

$$\tilde{\boldsymbol{y}}_k = \tilde{\boldsymbol{D}}\tilde{\boldsymbol{F}}(\boldsymbol{v}_k)\tilde{\boldsymbol{z}} + \tilde{\boldsymbol{e}}_k. \tag{11}$$

The matrices and vectors in Equation (11) have the same dimensions as those in the spatial domain. By virtue of their spatial-invariance, these operators are block-diagonalized in the Fourier domain. The downsampling and shifting operators become

$$\tilde{\mathbf{D}} = (\boldsymbol{I}_M^{\mathrm{T}} \otimes \boldsymbol{I}_{N_{\mathrm{L}}}) \otimes (\boldsymbol{I}_M^{\mathrm{T}} \otimes \boldsymbol{I}_{N_{\mathrm{L}}}), \tag{12}$$

$$\tilde{\mathbf{F}} = \mathrm{diag}\{e^{j(v_1\underline{\theta}_1 + v_2\underline{\theta}_2)}\}, \tag{13}$$

where $\otimes$ represents the Kronecker tensor product (`kron` in Matlab), $\boldsymbol{I}_{N_{\mathrm{L}}}$ represents the identity matrix of dimension N_{L} and $\boldsymbol{I}_M$ represents an M-dimensional vector of all ones. The θ vectors represent the spatial frequency indices.

Computing the ML variable projections functionals of Equation (6), involves two steps. First, we must compute the vector

$$\tilde{\boldsymbol{g}}(\underline{\mathrm{v}}) \equiv \tilde{\boldsymbol{B}}^{\mathrm{H}}(\underline{\mathrm{v}})\tilde{\boldsymbol{y}} = \sum_k \tilde{\boldsymbol{F}}^{\mathrm{H}}(\boldsymbol{v}_k)\tilde{\boldsymbol{D}}^{\mathrm{H}}\tilde{\boldsymbol{y}}_k, \tag{14}$$

where $\tilde{\boldsymbol{B}}^{\mathrm{H}}$ denotes the Hermitian transpose of the matrix $\tilde{\boldsymbol{B}}$. Computing $\tilde{\boldsymbol{g}}(\underline{\mathrm{v}})$ is very simple given that $\tilde{\boldsymbol{F}}$ is a diagonal matrix and $\tilde{\boldsymbol{D}}^{\mathrm{H}}$ is simply an M^2 replication of the image $\tilde{\boldsymbol{y}}_k$. Second, we must invert the matrix $\tilde{\boldsymbol{B}}^{\mathrm{H}}(\underline{\mathrm{v}})\tilde{\boldsymbol{B}}(\underline{\mathrm{v}})$. As noted in [14, 31], the spatial frequency components are decoupled in the Fourier domain. This can be understood by looking at the Kronecker tensor product used to build the matrices in Equations (12) and (13). Because of these Kronecker products, inverting the $N_{\mathrm{H}}^2 \times N_{\mathrm{H}}^2$ matrix $\tilde{\boldsymbol{B}}^{\mathrm{H}}(\underline{\mathrm{v}})\tilde{\boldsymbol{B}}(\underline{\mathrm{v}})$ is equivalent to inverting an $M^2 \times M^2$ matrix of the form

$$\tilde{\boldsymbol{Q}}(v) = \sum_{k=0}^{K} e^{(jv_2)/M\boldsymbol{A}_M} \otimes e^{(jv_1)/M\boldsymbol{A}_M} = \sum_{k=0}^{k} \tilde{\boldsymbol{S}}(\boldsymbol{v}_k). \tag{15}$$

The matrix $\boldsymbol{A}_M$ is an $M \times M$ matrix capturing the spectral folding of the downsampling operation. For example, Equations (16)–(18) show the $\boldsymbol{A}_M$ matrix for $M = 2, 3, 4$.

$$\boldsymbol{A}_2 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \tag{16}$$

$$\boldsymbol{A}_3 = \begin{pmatrix} 0 & -1 & 1 \\ 1 & 0 & 2 \\ -1 & -2 & 0 \end{pmatrix}, \tag{17}$$

$$\boldsymbol{A}_4 = \begin{pmatrix} 0 & -1 & 2 & 1 \\ 1 & 0 & 3 & 2 \\ -2 & -3 & 0 & -1 \\ -1 & -2 & 1 & 0 \end{pmatrix}. \tag{18}$$

We compute the LS estimate of the high-resolution image $(\tilde{\boldsymbol{B}}^{\mathrm{H}}(\underline{\mathrm{v}})\tilde{\boldsymbol{B}}(\underline{\mathrm{v}}))^{-1}\tilde{\boldsymbol{g}}(\underline{\mathrm{v}})$ by building an $M^2 \times (N/M)^2$ matrix $\tilde{\boldsymbol{G}}(\underline{\mathrm{v}})$ composed of the set of $\tilde{\boldsymbol{g}}(\underline{\mathrm{v}})$ vector elements combined during the spectral folding process. We represent the operations involved in constructing such a matrix by the operator $\mathcal{P}$. For example, if $M = 2$ the matrix $\tilde{\boldsymbol{G}} = \mathcal{P}\{\tilde{\boldsymbol{g}}\}$ is constructed as

$$\tilde{\boldsymbol{G}} = \begin{bmatrix} \{\tilde{\boldsymbol{g}}\}_1 & \{\tilde{\boldsymbol{g}}\}_2 & \cdots & \{\tilde{\boldsymbol{g}}\}_{N_{\mathrm{H}}(N_{\mathrm{L}}-1)+N_{\mathrm{L}}} \\ \{\tilde{\boldsymbol{g}}\}_{N_{\mathrm{L}}+1} & \{\tilde{\boldsymbol{g}}\}_{N_{\mathrm{L}}+2} & \cdots & \{\tilde{\boldsymbol{g}}\}_{N_{\mathrm{H}}N_{\mathrm{L}}} \\ \{\tilde{\boldsymbol{g}}\}_{N_{\mathrm{H}}N_{\mathrm{L}}+1} & \{\tilde{\boldsymbol{g}}\}_{N_{\mathrm{H}}N_{\mathrm{L}}+2} & \cdots & \{\tilde{\boldsymbol{g}}\}_{N_{\mathrm{H}}^2-N_{\mathrm{L}}} \\ \{\tilde{\boldsymbol{g}}\}_{(N_{\mathrm{H}}+1)N_{\mathrm{L}}+1} & \{\tilde{\boldsymbol{g}}\}_{(N_{\mathrm{H}}+1)N_{\mathrm{L}}+2} & \cdots & \{\tilde{\boldsymbol{g}}\}_{N_{\mathrm{H}}^2} \end{bmatrix} \tag{19}$$

Such permutation can be very efficiently implemented using the Matlab command `im2col`. For more information on this spatial frequency decoupling, we refer the reader to [14].

By defining $\tilde{\boldsymbol{W}}(\underline{v}) \equiv \tilde{\boldsymbol{Q}}^{-1}(\underline{v})\ \tilde{\boldsymbol{G}}(v)$, we compute the cost function of Equation (6) as

$$J_{\mathrm{ML}}(\underline{v}) = \mathcal{R}e\left\{\sum_{a,b} \tilde{\boldsymbol{G}}(\underline{v}) \odot \tilde{W}(\underline{v})\right\}, \tag{20}$$

where $\odot$ represents the elementwise multiplication of two complex matrices with indices a, b and $\mathcal{R}e$ represents the real part of the complex number. The gradient of this cost function is given by

$$\frac{\partial J_{\mathrm{ML}}(\underline{v})}{\partial \boldsymbol{v}_k} = \mathcal{R}e\begin{bmatrix} 2j\sum_{a,b} \Theta_1 \odot (\tilde{\boldsymbol{p}}(\boldsymbol{v}_k) \odot \tilde{\boldsymbol{W}}(\underline{v}) \\ -\tilde{\boldsymbol{W}}(\underline{v}) \odot \tilde{\boldsymbol{S}}(\boldsymbol{v}_k)\tilde{\boldsymbol{W}}(\underline{v})) \\ 2j\sum_{a,b} \Theta_2 \odot (\tilde{\boldsymbol{p}}(\boldsymbol{v}_k) \odot \tilde{\boldsymbol{W}}(\underline{v}) \\ -\tilde{\boldsymbol{W}}(\underline{v}) \odot \tilde{\boldsymbol{S}}(\boldsymbol{v}_k)\tilde{\boldsymbol{W}}(\underline{v})) \end{bmatrix}, \tag{21}$$

where

$$\Theta_i = \mathcal{P}\{\mathrm{diag}\{\theta_i\}\} \tag{22}$$

$$\tilde{\boldsymbol{p}}(\boldsymbol{v}_k) = \mathcal{P}\{\tilde{\boldsymbol{F}}^{\mathrm{H}}(\boldsymbol{v}_k)\tilde{\boldsymbol{D}}^{\mathrm{H}}\tilde{\boldsymbol{y}}_k\} \tag{23}$$

In this way we can very efficiently and quickly compute the ML cost function and its derivative.

We use Matlab's generic non-linear optimization routine `fminunc` to maximize the variable projection functional of Equation (6). We find that the rate of convergence depends on the downsampling factor M. We find that the number of required iterations scales approximately linearly with the downsampling factor M. For example, we also find that only 20–30 iterations are needed to converge for the $M = 2$ case, whereas 80–100 iterations are needed when $M = 4$. As predicted in [19], as the number of frames K increases, the problem typically becomes better conditioned leading to faster convergence albeit with increased complexity per iteration.

Up to this point, we have described an efficient method for computing and optimizing the ML cost function. The same computational tricks described above do not readily transfer when we consider the MAP cost function of Equation (9). Applying the same ideas to computation of the MAP cost function is difficult because of the typical form of the matrix $\tilde{\boldsymbol{C}}_{\boldsymbol{x}}$. Typically, the signal $\boldsymbol{x}$ (and hence $\boldsymbol{z}$) is assumed to be a wide sense stationary random process. In this case, the Fourier domain terms $\tilde{\boldsymbol{C}}_{\tilde{\boldsymbol{x}}}$ and $\tilde{\boldsymbol{C}}_{\tilde{\boldsymbol{z}}}$ are diagonal matrices

$$\tilde{\boldsymbol{C}}_{\tilde{\boldsymbol{x}}} = \mathrm{diag}\{\mathcal{X}(\theta_1, \theta_2)\}, \tag{24}$$

$$\tilde{\boldsymbol{C}}_{\tilde{\boldsymbol{z}}} = \tilde{\boldsymbol{H}}\tilde{\boldsymbol{C}}_{\tilde{\boldsymbol{x}}}\tilde{\boldsymbol{H}}^{\mathrm{H}}, \tag{25}$$

where $\mathcal{X}(\theta_1, \theta_2)$ represents the power spectral density (PSD) of the random signal $\boldsymbol{x}$. For example, in [3], the signal $\boldsymbol{x}$ is assumed to come from a distribution with a PSD implicitly defined by the Laplacian image operator. The implicit PSD is given by

$$\mathcal{X}(\theta_1, \theta_2) = \frac{\lambda}{|10 - 4\cos(\theta_1) - 4\cos(\theta_2) - \cos(\theta_1 + \theta_2) - \cos(\theta_1 - \theta_2)|^2} \tag{26}$$

In the ML case, we capitalized on the fact that the matrix $\tilde{\boldsymbol{B}}^{\mathrm{H}}(\underline{v})\tilde{\boldsymbol{B}}(\underline{v})$ could efficiently be represented by a smaller matrix $\tilde{\boldsymbol{Q}}(\underline{v})$. The diagonal elements of $\tilde{\boldsymbol{C}}_{\tilde{\boldsymbol{z}}}$ vary smoothly as a function of spatial frequency, eliminating the possibility for compact representation of $\tilde{\boldsymbol{B}}^{\mathrm{H}}(\underline{v})\tilde{\boldsymbol{B}}(\underline{v}) + \tilde{\boldsymbol{C}}_{\tilde{\boldsymbol{z}}}^{-1}$.

We propose approximating the full PSD function with a piecewise constant function. For example, the solid line of Fig. 2 shows an example of a one-dimensional version of the Laplacian-based diagonal prior matrix $\tilde{\boldsymbol{C}}_{\tilde{\boldsymbol{x}}}^{-1}$. The other curves show piece-wise constant approximations to the signal PSD with decreasing resolution. We denote the accuracy of the piecewise constant approximation by the number of tiles used to approximate the PSD. Assuming that the PSD is approximated using R^2 such tiles, computing the MAP cost function of Equation (9) requires R^2 inversions of a $M^2 \times M^2$ compressed matrix of the form

$$\tilde{\boldsymbol{Q}}(\underline{v}) + \lambda\tilde{\boldsymbol{L}}_r^{-1}, \tag{27}$$

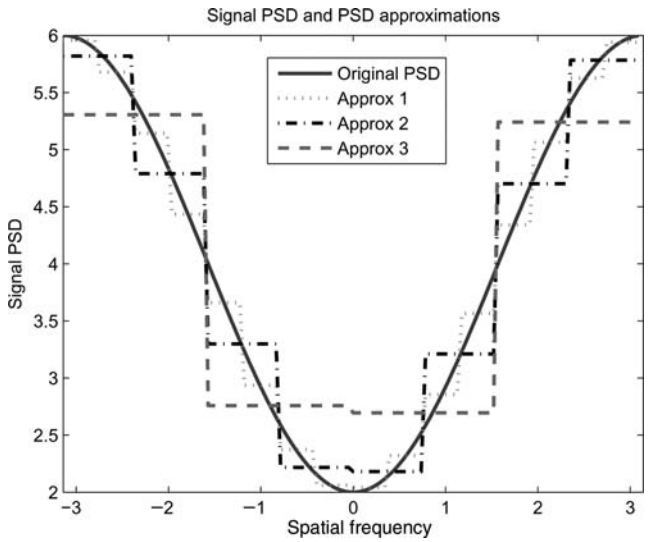

FIGURE 2. An example of the one-dimensional Laplacian-based signal PSD. The solid line shows the prior information term $\tilde{\boldsymbol{C}}_{\tilde{\boldsymbol{x}}}^{-1}$ provided for a one-dimensional version of the Laplacian-based prior. The additional dotted and dash-dotted lines show the successive piecewise constant approximations have each time reducing the number of piecewise constant approximation tiles by a factor of two. We verify that such PSD approximations have very little effect on the registration accuracy of the multi-frame algorithm.

where $\tilde{\boldsymbol{L}}_r^{-1}$ represents the elements of the matrix $\tilde{\boldsymbol{C}}^{-1}$ corresponding to the rth tile in the approximated PSD. To compute the MAP functional of Equation (9), we construct R^2 matrices $\tilde{\boldsymbol{G}}_r$ associated with each of the approximation tiles. In this way, we can achieve similar computational efficiencies in the MAP case as in the ML case with only a linear increase in complexity.

As we shall show momentarily, this piecewise constant approximation of the prior matrix does not significantly affect the image registration accuracy. The accuracy of the prior information plays a much more important role in reconstructing the high-resolution image. We find that we can use these simple PSD approximations during multi-frame registration followed by reconstruction using the full PSD matrix form of Equation (8).

3. EXPERIMENTS

In this section, we describe different conditions where the traditional approaches to registering aliased images become inadequate. We demonstrate that by using the multi-frame registration algorithm described above, we can effectively handle these difficult conditions. The matlab software used in the experiments may be downloaded from www.soe.ucsc.edu/∼dirkr/Publications/PublicSRCode.zip.

In all of our experiments, we use root mean-square error (RMSE) as a measure of performance, for both image registration parameters $\underline{\mathrm{v}}$ and the unknown image $\boldsymbol{z}$. The RMSE when estimating the translation parameters v is defined as

$$\mathrm{RMSE}(\hat{v}) = \sqrt{\frac{1}{2TK}\|\hat{v} - \underline{\mathrm{v}}\|^2}, \tag{28}$$

where T is the number of Monte Carlo simulations used during the experiment. Similarly, the RMSE when estimating the high-resolution image is defined as

$$\mathrm{RMSE}(\hat{z}) = \sqrt{\frac{1}{TN_{\mathrm{H}}^2}\|\hat{\boldsymbol{z}} - \boldsymbol{z}\|^2} \tag{29}$$

This type of measurement gives a measure of the image quality over the entire image.

3.1. Performance *vs* image size

For large images with reasonable Signal-to-noise ratio (SNR), most pairwise image registration algorithms will provide an acceptable level of registration accuracy. As the captured image size decreases, however, the ability of traditional registration algorithms to accurately estimate motion is severely compromised. The reduced number of data points found in a pair of images becomes insufficient for estimation.

To demonstrate this tendency, we compare the registration performance of several pairwise estimation algorithms as the size of the captured images shrinks. The left images of Fig. 3 show the original high-resolution images from which sets of 10 low-resolution images were generated. The images are constructed by low-pass filtering with a raised cosine filter [32] followed by downsampling, forming a dyadic pyramid. The size of the square images are $N_{\mathrm{H}} \times N_{\mathrm{H}}$ where $N_{\mathrm{H}} = 480$, 240, 120, 60. In this way, we ensure that the image PSD is the same across image scales. At each scale, a set of 10 low-resolution images were generated by first shifting the source images by a randomly selected set of shift vectors v drawn from a uniform distribution over $[0, M]^K$. The set of translations is shown in the right graph of Fig. 3. The images

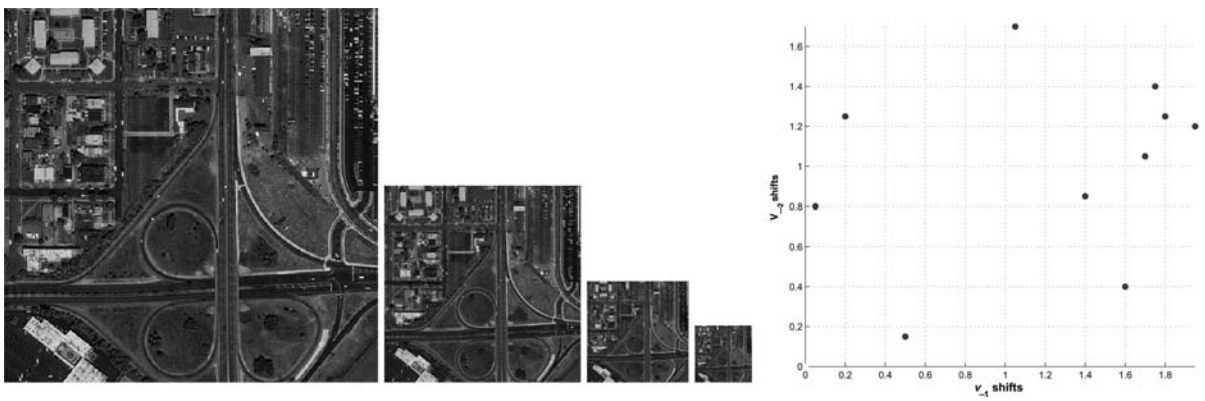

FIGURE 3. An example of high- and low-resolution test images.
The images on the left show the set of high-resolution test images used to evaluate the registration performance of different estimators on images of different sizes.
The set of images forms a dyadic pyramid where the image content at each scale is approximately equivalent. The graphic on the right shows the set of $K = 10$ image shifts v used to construct the test set of low-resolution images $\{\boldsymbol{y}_k\}$.

were then downsampled by a factor of $M = 2$. With so many frames, the registration problem is well defined and prior information is unnecessary for either registration or reconstruction. In the process of creating the low-resolution images from the high resolution ones, we did not incorporate any blurring into this experiment. At each scale, a different amount of uncorrelated Gaussian random noise was added to the images so that the SNR was 20 dB ($\text{SNR} = 20\log_{10}^{(\sigma_x)/\sigma}$ where σ_x is the standard deviation of the high-resolution image signal $\boldsymbol{x}$). This corresponds to standard deviations of roughly $\sigma = 4, 4, 3, 2$, respectively. In this case, we performed only a single experiment.

We applied the proposed multiframe registration algorithm using Equation (6) as the cost function. We compared this algorithm with the relative phase approach [16], a standard iterative gradient-based approach for estimating global translation [15] and a cyclic coordinate descent type algorithm as described earlier similar in spirit to [11, 24]. The tuning parameters were modified to provide the best MSE possible at each image scale. Both the cyclic-coordinate descent and the variable projection algorithms used the gradient-based estimate of the motion parameters as an initial guess.

The cyclic coordinate descent used about five optimization iterations over the motion parameters for each estimate of the high resolution image. For all image sizes, the algorithm converged after 20 total coordinate descent cycles for a total of $\sim$100–150 optimization iterations. The multi-frame algorithm also used the gradient-based estimate as the initial guess of the motion vectors. The proposed multi-frame algorithm based on the variable projections functional required less than 40 total iterations.

Table 1 shows the registration RMSE of the three algorithms. We observe that all of the algorithms provide a high degree of accuracy when applied to large images. The pairwise registration exhibits accuracy down to about 1/20 of a pixel. As the size of the captured images decreases, however, the pairwise registration algorithms begin to fail. At the smallest image size, both of the pairwise registration techniques provide unacceptable estimates of the shift parameters. In each case, the registration performance worsens as the size of the image decreases. However, both multi-frame registration algorithms not only continue to offer better estimates, but they also degrade more slowly than the pairwise algorithms. The variable projection-based algorithm offers improved performance possibly due to better numerical conditioning.

TABLE 1. RMSE registration accuracy in units of pixels (averaged over the number of frames) as a function of low-resolution image size N_L.

Algorithm	$N_L = 240$, $\sigma = 4$	$N_L = 120$, $\sigma = 4$	$N_L = 60$, $\sigma = 3$	$N_L = 30$, $\sigma = 2$
Relative phase [16]	0.0442	0.0681	0.215	0.618
Gradient-based [15]	0.054	0.0685	0.203	0.455
Coordinate descent	0.011	0.0051	0.021	0.034
Variable projections	0.0018	0.0031	0.0090	0.0090

TABLE 2. Image reconstruction RMSE performance in terms of gray levels as a function of the low-resolution image size N_L.

Algorithm	$N_L = 240$, $\sigma = 4$	$N_L = 120$, $\sigma = 4$	$N_L = 60$, $\sigma = 3$	$N_L = 30$, $\sigma = 2$
Relative phase [16]	1.763	1.879	2.161	3.569
Gradient-based [15]	1.838	1.989	2.859	3.027
Coordinate descent	1.687	1.687	1.348	1.217
Variable projections	1.684	1.673	1.310	1.077

Next, we took the set of motion estimates and reconstructed a high-resolution image via Equation (4) using the estimates $\hat{v}$ produced by the different algorithms. Table 2 compares the image reconstruction performance of the three algorithms as a function of image size. Again, we observe that for large images, the pairwise registration performance is accurate enough to provide approximately equivalent reconstruction performance regardless of the registration algorithm. As the size of the low-resolution images decreases, however, only the multi-frame registration techniques continue to offer good performance[2].

Finally, Table 3 shows the length of time required for each of the different algorithms as a function of the image size N_L. The experiments were performed on a Pentium 4, 2.0 GHz machine. We observe from Table 3 that the iterative algorithms require more cycles to achieve the greater level of accuracy. The faster algorithm based on the variable projections is about 30% faster than a coordinated cyclic descent type algorithm.

From this experiment, we observe that while the traditional approaches to aliased image registration, and hence super-resolution, work well for large images, as the image size is decreased, only the joint methods provide adequate performance.

3.2. Effect of prior information on registration performance

In this section, we explore the performance of image registration as a function of prior information or regularization. First, by way of experimentation, we verify that the piecewise constant prior information approximation strategy described in the previous section affects the performance of the multi-frame MAP estimation very little. As a test image, we cropped the smallest image shown in the image pyramid of Fig. 3 to be 48×48 pixels. We estimated the prior image

[2]Lower reconstruction RMSE values for the smaller images are due to incorporating less additive noise than that of the larger images to ensure constant SNR over image size.

TABLE 3. Timing requirenments (seconds) for the different algorithms as a function of low-resolution image size N_L.

Algorithm	$N_L = 240$	$N_L = 120$	$N_L = 60$	$N_L = 30$
Relative phase [16]	0.431	0.110	0.041	0.021
Gradient-based [15]	0.791	0.180	0.060	0.030
Coordinate descent	254.6	54.19	18.41	8.45
Variable projections	103.2	37.78	11.07	3.70

PSD using a periodogram method [32] applied to the high resolution large image shown in Fig. 3. Using this PSD estimate, we constructed the diagonal prior information matrix $\tilde{\boldsymbol{C}}_{\tilde{z}}$. We then computed a set of piecewise constant approximations to this prior information matrix for approximation factors of $R = 2, 4, 6, 8, 12, 24$. Then, we drew 100 random sets of $K = 3$ sampling shifts from a uniform distribution over $[0,2]^K$. For each set of motion parameters, we simulated the set of low-resolution images for a downsampling factor of $M = 2$ and noise standard deviation of $\sigma = 3$ (SNR of 18 db). We then computed the RMSE performance (RMSE($\hat{v}$)) when using the complete prior matrix $\tilde{\boldsymbol{C}}_{\tilde{z}}$ as well as when approximating the prior matrices with quantization factors. Figure 4 describes the average difference between the registration RMSE performance when using the MAP estimator with the complete prior information matrix and the RMSE performance when using the approximate prior information matrix. The difference is negligible when the quantization factor is greater than $R = 2$. Using the approximate prior information, matrices speeds-up registration performance significantly.

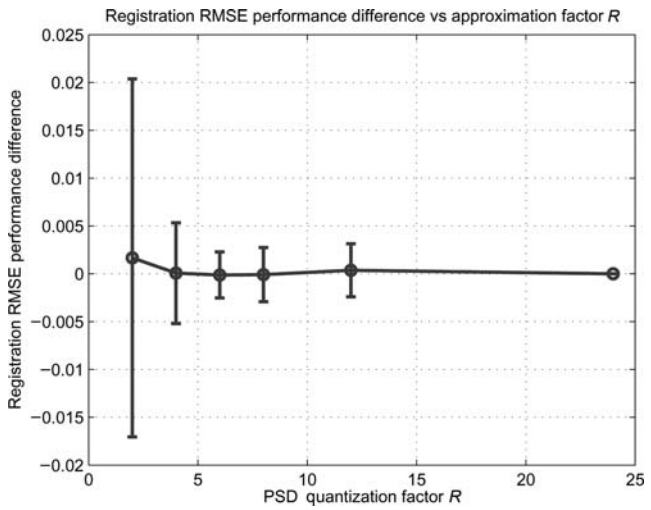

FIGURE 4. Graph showing the average difference between the registration performance RMSE($\hat{v}$) when using the MAP estimator with the complete prior information matrix $\tilde{\boldsymbol{C}}_{\tilde{x}}$ and the MAP estimator when using the approximate prior information matrix with different quantization defined by R. The errorbars show the standard deviation between the RMSE performance using the quantized prior matrices and the full prior matrix. The performance loss when using the quantization factor of $R = 4$ or greater is small enough to justify using the much more computationally efficient method.
The performance results were averaged for each value of R over 100 random motion sets with $K = 3$, $M = 2$, $\sigma = 3$, for the smallest image shown in Fig. 3.

We also conducted an extensive Monte Carlo experiment measuring both the reconstruction and registration performance of the MAP estimator as a function of the number of captured frames. We used the same 48×48 pixel image as in the previous experiments. For each value of K we randomly selected 100 motion vector sets v from a uniform distribution over [0, 2] and simulated the captured images by adding noise with $\sigma = 1$ (SNR = 28.2 db) for a downsampling factor of $M = 2$. The fast MAP algorithm uses the approximate prior information matrix quantized to $R^2 = 16$. We also computed estimates varying the prior information tuning parameter to be $\lambda = 0.1$, 1.0 and 10. This entire process was repeated for an increased noise power of $\sigma = 5$ (SNR = 14.5 db). The graph on the left side of Fig. 5 shows the registration RMSE averaged over the set of random motions for each value of $K + 1$. Also shown is the RMSE performance of the gradient-based pairwise registration algorithm used to initialize the MAP estimation. After obtaining estimates of the registration parameters, the image $\boldsymbol{z}$ was reconstructed using either the ML estimate of Equation (4) or the MAP estimate of Equation (8). The complete prior information matrix $\tilde{\boldsymbol{C}}_{\tilde{z}}$ was used when reconstructing the image using the MAP form. The graph on the right side of Fig. 5 similarly displays the reconstruction performance of the MAP estimator as a function of the number of frames $K + 1$. In this case, only the reconstruction performance for $\lambda = 1$ is shown. Both graphs in Fig. 5 demonstrate the superiority of the multiframe registration technique to the pairwise gradient-based approach. Both graphs also show the prior information to be unnecessary as the number of frames increases. This is explained by well posedness of the registration/reconstruction when sufficiently diverse sampling offsets are present. As the number of frames increases, the probability that the set of sampling offsets is degenerate decreases [19]. This suggests a two-stage approach that first estimates the motion parameters using the ML formulation for its superior computational efficiency followed by a more computationally demanding MAP formulation. Overall, these performance curves demonstrate the effectiveness of the multi-frame approach for both registration and reconstruction.

3.3. Uncontrolled experiment

In this section, we apply the proposed multi-frame algorithm on a very small portion of the well-known infrared tank sequence generously provided by the US Air Force research laboratory. We cropped the image to an $N_L = 20$ square image patch around the personnel carrier. The image sequence has 15 frames with unknown, yet apparently translational motion. We applied the variable projections cost function

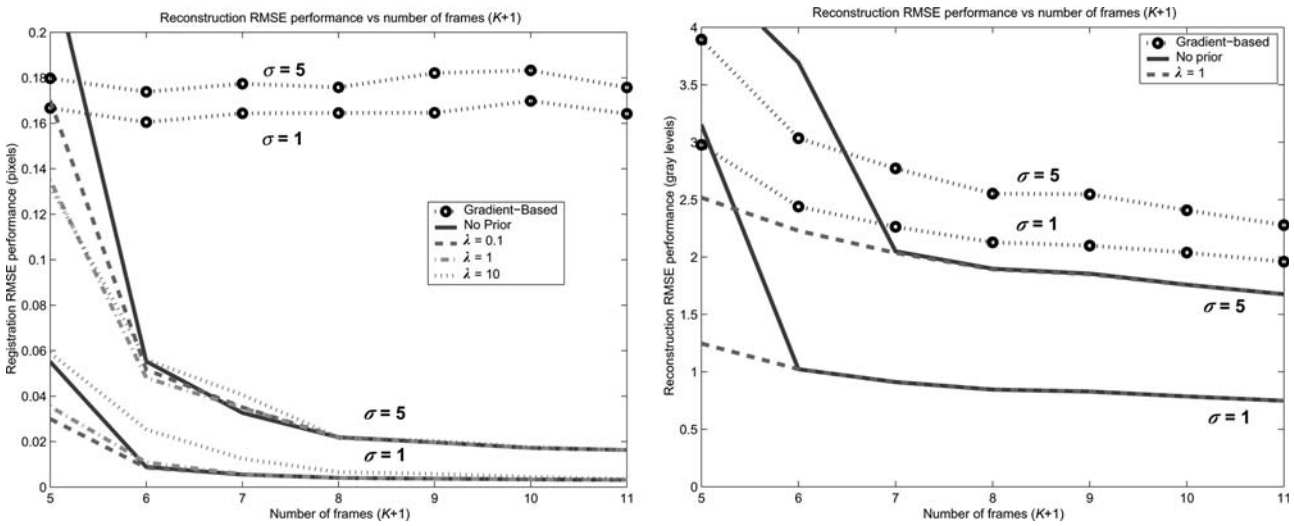

FIGURE 5. Comparison of registration and reconstruction RMSE performances. The graph on the left shows the registration RMSE as a function of the number of frames $K + 1$ while adding noise with standard deviation $\sigma = 1$ and 5. We observe that the prior information is very useful when the number of frames are low. When sufficient number of frames become available, the prior information is unnecessary. The graph on the right compares the corresponding reconstruction RMSE performance showing a similar phenomenon.

using 30 iterations. We initialized the translation estimates using gradient-based estimation. To build an image prior PSD, we randomly selected 50 $N_L \times N_L$ patches from the fine-scale satellite image shown in Fig. 3 and estimated the PSD using periodogram estimation. This PSD estimate was used to construct the prior applied during our MAP registration and reconstruction.

As a first step, we assumed a downsampling factor of $M = 2$. Given 15 frames, we could apply the very efficient ML version of the estimator without any prior information. While 15 frames should theoretically be sufficient for $M = 3$, we found in practice that the estimator converged to more reasonable estimates when incorporating the prior information in a MAP formulation. This is explained by the fact that the motions vectors do not span the entire range [0, M] as M increases, leading to more degenerate matrices.

The graph on the left side of Fig. 6 shows a scatter plot of the motion vectors in units of the captured low-resolution images. The similarity of the motion estimates for each downsampling factor M suggests an efficient strategy for registering a set of images. Since the optimization requires far fewer iterations for smaller downsampling factors, one could start with small downsampling factors and gradually increase M as needed. For example, one could apply the ML version of the

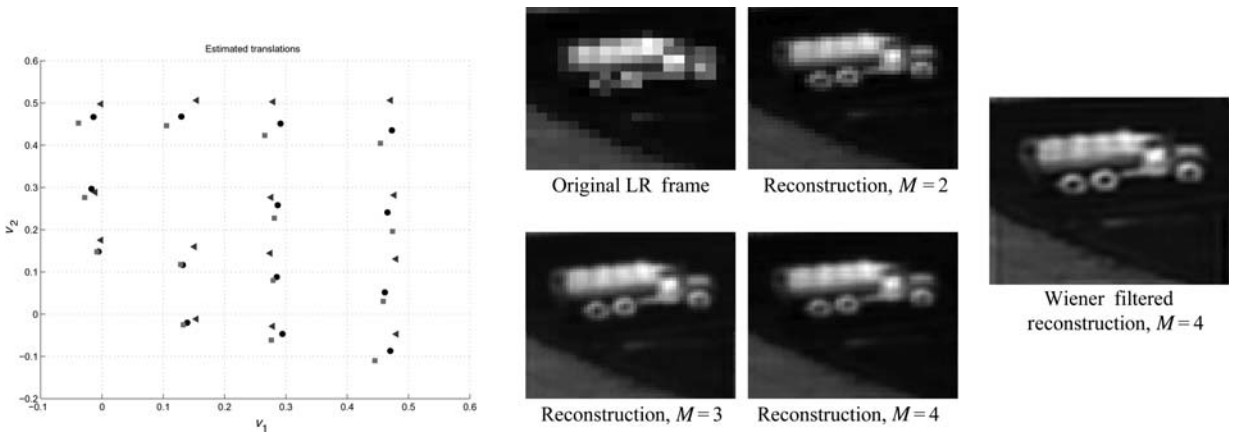

FIGURE 6. Scatter plotter and low-resolution images of the motion vectors. The graph on the left shows a scatterplot of the estimated sampling offsets for $M = 2$ (filled square box), $M = 3$ (filled circle), $M = 4$ (filled triangle) in units of the low-resolution captured image. We observe that the estimates at each downsampling factor are similar. The images on the right show the reconstructed images at each downsampling factor. Also shown is a sharpened version of the reconstructed image for $M = 4$. The effectiveness of this simple restoration method demonstrates the ability to jointly register and reconstruct the captured images independently of the restoration.

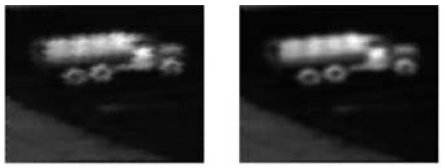

FIGURE 7. The images depict the reconstruction estimates for a downsampling factor $M = 4$ using the pairwise registration estimates provided by the relative phase 16 and gradient-based 15 approaches. While the gradient-based approach is clearly superior, both reconstructions exhibit slightly jagged edges due to inconsistent motion estimates.

estimator for $M = 2$ (assuming non-degenerate motion) and use the scaled version of the estimates as a reasonable initial guess for $M = 3$.

The images on the right side graph of Fig. 6 show the resulting reconstructed images for each downsampling factor. For each enhancement factor, the image was reconstructed using the translation estimates for that particular level. The prior information PSD described above was applied at each scale to regularize the solution. We observe that at each enhancement factor, the resolution of the reconstructed image is improved. The gain in resolution between $M = 3$ and 4 is, however, slight.

As a point of reference, Fig. 7 shows the reconstructed images using both the relative phase [16] and gradient-based [15] pairwise registration algorithms. While the gradient-based approach is clearly superior to the relative phase approach, both reconstructions exhibit slightly jagged edges due to inconsistent motion estimates. Such artifacts would be much more pronounced for images having more high frequency image content.

Applying our reconstruction algorithm provides an estimate of the blurry high resolution $\hat{\boldsymbol{z}}$. To obtain an estimate of the original high resolution $\boldsymbol{x}$ one must apply a deblurring algorithm. The image in the far right of Fig. 6 shows a very simple deconvolution filter applied to the image based on Matlab's `deconvwnr` function assuming a Gaussian PSF with standard deviation of 1.5 pixels. While one might argue that the resulting restored image is not optimal, the efficacy of this simple method 21 suggests that much more powerful, possibly blind, deconvolution strategies such as [11] might be applied with improved results.

4. CONCLUSION

In this paper, we described an efficient approach to solving both ML and MAP registration of a set of aliased images. By exploiting the idea of variable projection functionals, we decoupled the registration and resolution enhancement problems, increasing the computational efficiency of the iterative approach. We demonstrated the further efficiency gains for the case of translational motion wherein the more difficult problem of image restoration may be decoupled from that of registration and reconstruction. In this way, fast registration based on simple quadratic penalty functions may be utilized without need to incorporate the non-linear signal priors shown to be successful for image restoration. We analysed the conditions under which the performance of such accurate motion estimation technique is considerably better than the competing techniques. Our experiments showed the proposed method to be effective even for traditionally challenging conditions such as small image size, low SNR and significant aliasing.

REFERENCES

[1] Farsiu, S., Robinson, D., Elad, M. and Milanfar, P. (2004) Advances and challenges in super-resolution. *Int. J. Imag. Syst. Technol.*, **14**, 47–57.

[2] Borman, S. and Stevenson, R.L. (1998) Super-resolution from image sequences—a review. *Proc. 1998 Midwest Symposium on Circuits and Systems*, Notre Dame, IN, April, pp. 374–378.

[3] Park, S., Park, M. and Kang, M. G. (2003) Super-resolution image reconstruction, a technical overview. *IEEE Signal Process. Mag.*, **20**, 21–36.

[4] Goodman, J. (2005) *Introduction to Fourier Optics* (3rd edn). Roberts and Company Publishers, Greenwood Village, CO.

[5] Grievenkamp, J. (1990) Color dependent optical prefilter for suppression of aliasing artifacts. *Appl. Opt.*, **29**, 676–684.

[6] Chaudhuri, S. and Manjunath, J. (2005) *Motion-Free Superresolution*. Springer, New York, NY.

[7] Rajagopalan, A. and Kiran, P. (2003) Motion-free superresolution and the role of relative blur. *J. Opt. Soc. Am.*, **20**, 2022–2032.

[8] Hardie, R., Barnard, K. and Armstrong, E. (1997) Joint MAP registration and high-resolution image estimation using a sequence of undersampled images. *IEEE Trans. Image Process.*, **6**, 1621–1633.

[9] Vandewalle, P., Sbaiz, L., Susstrunk, S. and Vetterli, M. (2004) How to take advantage of aliasing in bandlimited signals. *Proc. IEEE Conf. Acoustics, Speech and Signal Processing*, Montreal, Canada, May, pp. 948–951.

[10] Tom, B. and Katsaggelos, A. (2001) Resolution enhancement of monochrome and color video using motion compensation. *IEEE Trans. Image Process.*, **10**, 278–287.

[11] Woods, N., Galatsanos, N. and Katsaggelos, A. (2006) Stochastic methods for joint registration, restoration, and interpolation of multiple undersampled images. *IEEE Trans. Image Process.*, **15**, 201–213.

[12] Elad, M. and Hel-Or, Y. (2001) A fast super-resolution reconstruction algorithm for pure translational motion and common space invariant blur. *IEEE Trans. Image Process.*, **10**, 1186–1193.

[13] Farsiu, S., Robinson, D., Elad, M. and Milanfar, P. (2004) Fast and robust multi-frame super-resolution. *IEEE Trans. Image Process.*, **13**, 1327–1344.

[14] Kim, S., Bose, N. and Valenzuela, H. (1990) Recursive reconstruction of high resolution image from noisy undersampled multiframes. *IEEE Trans. Accoustics Speech Signal Process.*, **38**, 1013–1027.

[15] Lucas, B.D. and Kanade, T. (1981) An iterative image registration technique with an application to stereo vision. *DARPA81*, pp. 121–130.

[16] Stone, H.S., Orchard, M., Chang, E.-C. and Martucci, S. (2001) A fast direct Fourier-based algorithm for subpixel registration of images. *IEEE Trans. Geosci. Remote Sens.*, **39**, 2235–2243.

[17] Vandewalle, P., Susstrunk, S. and Vetterli, M. (2006) A frequency domain approach to registeration of aliased images with application to super-resolution. *EURASIP J. Appl. Signal Process.*, p. 14.

[18] Foroosh, H., Zerubia, J. and Berthod, M. (2002) Extension of phase correlation to subpixel registration. *IEEE Trans. Image Process.*, **11**, 188–200.

[19] Robinson, D. and Milanfar, P. (2006) Statistical performance analysis of super-resolution. *IEEE Trans. Image Process.*, **15**, 1413–1428.

[20] Sawhney, H., Hsu, S. and Kumar, R. (1998) Robust video mosaicing through topology inference and local to global alignment. *ECCV*, Freiburg, Germany, pp. 103–119.

[21] Govindu, V. (2004) Lie-algebraic averaging for globally consistent motion estimation. *Proc. Int. Conf. Computer Vision and Pattern Recognition (CVPR)*, Washington, DC, July, pp. 684–691.

[22] Sheikh, Y., Zhai, Y. and Shah, M. (2004) An accumulative framework for the alignment of an image sequence. *Proc. Asian Conference on Computer Vision (ACCV)*, Jeju, Korea, January.

[23] Farsiu, S., Elad, M. and Milanfar, P. (2005) Constrained, globally optimal, multi-frame motion estimation. *Proc. 2005 IEEE Workshop on Statistical Signal Processing*, Bordeaux, France, July, pp. 1396–1401.

[24] Tom, B.C., Katsaggelos, A.K. and Galatsanos, N.P. (1994) Reconstruction of a high resolution image from registration and restoration of low resolution images. *Proc. Int. Conf. Image Processing (ICIP)*, Austin, TX, **vol. 3**, pp. 553–557.

[25] Golub, G. and Pereyra, V. (2003) Separable nonlinear least squares: the variable projection method and its applications. *Inst. Phys. Inverse Probl.*, **19**, R1–R26.

[26] Kay, S.M. (1993) *Fundamentals of Statistical Signal Processing: Estimation Theory*. Prentice Hall Inc., Upper Saddle River, NJ.

[27] Chung, J., Haber, E. and Nagy, J. (2006) Numerical methods for coupled super-resolution. *Inverse Probl.*, **22**, 1261–1272.

[28] Patras, I., Worring, M. and van den Boomgaard, R. (2004) Dense motion estimation using regularization constraints on local parametric models. *IEEE Trans. Image Process.*, **13**, 1432–1443.

[29] Marquardt, D. (1963) An algorithm for least-squares estimation of nonlinear parameters. *SIAM J. Appl. Math.*, **11**, 431–441.

[30] Takeda, H., Farsiu, S. and Milanfar, P. (2006) Robust kernel regression for restoration and reconstruction of images from sparse noisy data. *Proc. Int. Conf. Image Processing (ICIP)*, Atlanta, GA, October, pp. 1257–1260.

[31] Tsai, R. and Huang, T. (1984) Multiframe image restoration and registration. *Adv. Comput. Vis. Image Process.*, **1**, 317–339.

[32] Porat, B. (1997) *A Course in Digital Signal Processing*. John Wiley and Sons, Inc., New York.