Fundamental Performance Limits in Image Registration

Dirk Robinson, *Student Member, IEEE,* and Peyman Milanfar, *Senior Member, IEEE*

***Abstract*—The task of image registration is fundamental in image processing. It often is a critical preprocessing step to many modern image processing and computer vision tasks, and many algorithms and techniques have been proposed to address the registration problem. Often, the performances of these techniques have been presented using a variety of relative measures comparing different estimators, leaving open the critical question of overall optimality. In this paper, we present the fundamental performance limits for the problem of image registration as derived from the Cramer–Rao inequality. We compare the experimental performance of several popular methods with respect to this performance bound, and explain the fundamental tradeoff between variance and bias inherent to the problem of image registration. In particular, we derive and explore the bias of the popular gradient-based estimator showing how widely used multiscale methods for improving performance can be explained with this bias expression. Finally, we present experimental simulations showing the general rule-of-thumb performance limits for gradient-based image registration techniques.**

***Index Terms*—Bias, Cramer–Rao bound, error analysis, Fisher information, gradient methods, image registration, motion estimation, optical flow, performance limits.**

I. Introduction

IMAGE registration is a fundamental inverse problem in imaging. It represents a critical preprocessing step to many modern image processing tasks such as motion compensated video compression, multiframe image enhancement, remote sensing, and many computer vision tasks, such as three-dimensional (3-D) shape estimation and object identification. The problem of image registration is a specific case of the more general problem of estimating motion in an image sequence wherein the observed data follows the form

$$z(x,y,t) = f(x - v_1(x,y,t), y - v_2(x,y,t)) + \epsilon(x,y,t) \quad (1)$$

where $f(x,y,t)$ is the image function, $\epsilon(x,y,t)$ is additive white Gaussian noise with variance σ^2, and $\vec{v}(x,y,t) = [v_1(x,y,t), v_2(x,y,t)]^T$ is an unknown vector field characterizing the evolution of the image sequence in time. In this paper, the problem is constrained to that of estimating the relative shift contained in a pair of frames. This generally nonlinear estimation problem is often referred to as image registration. This model ignores other factors influencing the dynamics of the images, such as variation in the illumination or specular reflections.

In practice, we are given only sampled versions of the image wherein the spatial sample spacing is T and the temporal sampling is determined by the frame rate of the imaging system. For the remainder of the paper, we will use the indices m, n to refer to the sampled functions $f(mT, nT)$, formulating the data producing model as

$$z_1(m,n) = f(m,n) + \epsilon_1(m,n) \quad (2)$$

$$z_2(m,n) = f(m - v_1(m,n), n - v_2(m,n)) + \epsilon_2(m,n). \quad (3)$$

Here, the z_1 and z_2 refer to a pair of images of the sequence $z(x,y,t)$ observed at times t_1 and t_2. In this paper, we focus on image motion which is translational in nature, where the unknown vector field is of the form $\mathbf{v} = [v_1, v_2]^T$ where $v_1(m,n) = v_1$ and $v_2(m,n) = v_2$ are constants. While this model for image motion is very simple, we will suggest how our analysis can be extended to more complex models of image dynamics.

The overall goal of this paper is to quantify bounds on performance in estimating image translation between a pair of images. Because this problem is of such fundamental importance, many registration algorithms have been developed over the years. In fact, there have been fairly comprehensive survey papers describing and comparing the performance of such algorithms [1]–[3]. Unfortunately, the benchmarks comparing the performance of such algorithms tend to vary as widely as the techniques themselves, and the typical performance measures fail to address the problem in a proper statistical framework. These performance measures have ranged from geometric error criteria such as the mean angular error [1], to that of visual inspection of the vector field for situations where ground truth is not available. While these measures have been very useful in advancing the methodologies of motion estimation, they fail to evaluate estimator performance from a statistically meaningful perspective. Furthermore, the performance evaluation has relied on comparison between different algorithms leaving open the important question of how close the algorithms come to achievable limits.

The problem of image registration for translational motion estimation is analogous to the classical problem of time delay estimation (TDE), as found in the signal processing literature [4]. For the TDE problem, performance is measured based on the mean square error (MSE) of a given estimator. We propose that performance of image registration should be evaluated using the

Manuscript received May 5, 2003; revised December 22, 2003. This work was supported in part by the National Science Foundation under Grant CCR-9984246 and in part by AFOSR under Grant F49620-03-1-0387. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Christoph Stiller.

The authors are with the Department of Electrical Engineering, University of California at Santa Cruz, Santa Cruz, CA 95064 USA (e-mail: dirkr@ee.ucsc.edu; milanfar@ee.ucsc.edu).

Digital Object Identifier 10.1109/TIP.2004.832923

same measure. By using MSE, we can explore the fundamental performance bounds using the Cramer–Rao inequality. Surprisingly, while the Cramer–Rao inequality has been used widely in the field of time delay estimation in communication, radar, and Ssonar, except for a few isolated attempts [5], [6], it has not been utilized to understand the problem of image registration in general. In this paper, we analyze the form of the Cramer–Rao inequality as it relates to the specific problem of registering translated images.

Developing such performance bounds provides a mechanism for critically comparing the performance of algorithms. We will show how a great deal of the heuristic knowledge used in motion estimation can be explained by examining this performance bound. Furthermore, understanding these fundamental limitations provides better understanding of the limitations inherent to the class of image processing problems that require image registration as a preprocessing step. In addition, analyzing the details of the bound offers insight into the very nature of the problem itself, thereby suggesting methods for improvement. Particularly, we will present the inherent performance tradeoff between bias and variance for several popular motion estimators. While estimator bias is often difficult to express, we will derive such bias expressions for the popular gradient-based estimator. While the bias for gradient-based estimators has been addressed in previous works [7]–[11], these works make overly simplified generalizations about the bias. In this paper, we present and analyze more precise expressions for the estimator bias. We will show that this bias limits accurate registration for typical imaging systems. Finally, we will use this bias function to propose a rule-of-thumb limit (based on our analytical results) for image registration accuracy using gradient-based estimators.

The organization of the paper is as follows. In Section II, we derive the performance bounds in registering translated images, based on the Cramer–Rao inequality. We show how these bounds depend on image content by analyzing the FIM. We show the inherent problem of bias for the problem of image registration. We present experimental evidence of such bias for several popular registration algorithms. In Section III, we derive and experimentally verify a specific bias expression for the general class of gradient-based estimators. We present extensive analysis of this bias function and show how it tends to dominate the MSE performance limit for common imaging systems. In Section IV, we present experimental results suggesting typical performance limits for image registration. We conclude by suggesting possible future extensions to the work derived in this paper.

II. MSE Bounds for Image Registration

In this section, we quantify the fundamental MSE performance bounds for registering images utilizing the Cramer–Rao lower bound (CRLB) [12]. Essentially, the CRLB characterizes, from an information theoretic standpoint, the "difficulty" with which a set of parameters can be estimated by examining the given data model. In general, the CRLB provides the lower bound on the mean square error of *any* method used to estimate a deterministic parameter vector $\mathbf{\Phi}$ from a given set of data. Specifically, the Cramer–Rao bound on the error correlation matrix $E[(\hat{\mathbf{\Phi}}-\mathbf{\Phi})(\hat{\mathbf{\Phi}}-\mathbf{\Phi})^T]$ for any estimator is given by

$$\mathrm{MSE}(\mathbf{\Phi}) \geq \frac{\partial E[\hat{\mathbf{\Phi}}]}{\partial \mathbf{\Phi}} J^{-1}(\mathbf{\Phi}) \frac{\partial E[\hat{\mathbf{\Phi}}]}{\partial \mathbf{\Phi}}^T + (E[\hat{\mathbf{\Phi}}]-\mathbf{\Phi})(E[\hat{\mathbf{\Phi}}]-\mathbf{\Phi})^T \quad (4)$$

where the matrix $J(\mathbf{\Phi})$ is referred to as the FIM, and $E[\hat{\mathbf{\Phi}}]-\mathbf{\Phi}$ represents the bias of the estimator [13]. We refer to the error correlation matrix as $\mathrm{MSE}(\mathbf{\Phi})$ since the diagonal terms of $E[(\hat{\mathbf{\Phi}}-\mathbf{\Phi})(\hat{\mathbf{\Phi}}-\mathbf{\Phi})^T]$ represent the MSE. The inequality indicates that the difference between the MSE (left side) and the CRLB (right side) will be a positive semidefinite matrix. From this formulation, we see that the mean square error bound is comprised of two terms corresponding to a variance term and a term which is the square or outer product of the of the bias associated with the estimator.

Ideally, we wish to have unbiased estimates. Assuming such an estimator exists, the bound (4) simplifies to the more familiar

$$\mathrm{MSE}(\mathbf{\Phi}) \geq J^{-1}(\mathbf{\Phi}). \quad (5)$$

Thus, for any unbiased estimator, $J(\mathbf{\Phi})$ characterizes the minimum variance (and hence MSE) attainable. Because the FIM plays such a central role in bounding estimator variance for the classes of both biased and unbiased estimators, we now explore the details of the FIM for the problem of image registration.

A. Fisher Information for Image Registration

The FIM provides a measure of the influence an unknown parameter vector has in producing observable data. In our case, the unknown vector is the translation vector $\mathbf{v}=[v_1,\ v_2]^T$. The FIM is derived by looking at the expected concavity of the likelihood function. Intuitively, a likelihood maximizing estimator should have an easier time finding the maximum of a sharply peaked likelihood function than a rather flat one. The joint likelihood function for the data is represented by $\mathbf{P}(z;\mathbf{v})$ where the log of the likelihood function is given by

$$\log \mathbf{P}(z;\mathbf{v}) = \frac{-1}{2\sigma^2}\sum_{m,n}[z_1(m,n)-f(m,n)]^2 + [z_2(m,n)-f(m-v_1,n-v_2)]^2 + \mathrm{const.} \quad (6)$$

Specifically, the FIM measures the sharpness of likelihood peak where the matrix is defined as $[J(\mathbf{v})]_{ij} = -E[(\partial^2 \log \mathbf{P}(z;\mathbf{v}))/(\partial v_i \partial v_j)]$. In deriving the FIM, we first derive the form of the partial derivatives with respect to the log-likelihood function

$$\frac{\partial^2 \log \mathbf{P}(z;\mathbf{v})}{\partial v_i^2} = \frac{\partial}{\partial v_i}\left[\frac{1}{\sigma^2}\sum_{m,n}(z_2-\tilde{f})\frac{\partial \tilde{f}}{\partial v_i}\right] = \frac{1}{\sigma^2}\sum_{m,n}\left[(z_2-\tilde{f})\frac{\partial^2 \tilde{f}}{\partial v_i^2} - \left(\frac{\partial \tilde{f}}{\partial v_i}\right)^2\right]. \quad (7)$$

To simplify the notation, we refer to the transformed image $f(m-v_1, n-v_2)$ as $\tilde{f}$. Since only the term z_2 is random, the negative expectation of (7) for each term becomes

$$-E\left[\frac{\partial^2 \log \mathbf{P}(z;\mathbf{v})}{\partial v_1^2}\right] = \frac{1}{\sigma^2}\left(\frac{\partial \tilde{f}}{\partial v_1}\right)^2$$
$$-E\left[\frac{\partial^2 \log \mathbf{P}(z;\mathbf{v})}{\partial v_2^2}\right] = \frac{1}{\sigma^2}\left(\frac{\partial \tilde{f}}{\partial v_2}\right)^2$$
$$-E\left[\frac{\partial^2 \log \mathbf{P}(z;\mathbf{v})}{\partial v_2 \partial v_1}\right] = \frac{1}{\sigma^2}\left(\frac{\partial \tilde{f}}{\partial v_2}\right)\left(\frac{\partial \tilde{f}}{\partial v_1}\right).$$

Finally, we note that by way of the chain rule

$$\frac{\partial \tilde{f}}{\partial v_1} = \frac{\partial \tilde{f}}{\partial x} = f_x(m-v_1, n-v_2)$$
$$\frac{\partial \tilde{f}}{\partial v_2} = \frac{\partial \tilde{f}}{\partial y} = f_y(m-v_1, n-v_2).$$

Hence, we get the FIM $J(\mathbf{v}) = (1/\sigma^2)\left[\begin{smallmatrix} a_1 & a_2 \\ a_2 & a_3 \end{smallmatrix}\right]$ where

$$a_1 = \sum_{m,n} f_x^2(m-v_1, n-v_2)$$
$$a_2 = \sum_{m,n} f_x(m-v_1, n-v_2) f_y(m-v_1, n-v_2)$$
$$a_3 = \sum_{m,n} f_y^2(m-v_1, n-v_2).$$

The subscripts indicate the partial derivative in the x, y directions.

A comment is in order regarding these partial derivatives. The FIM and, hence, the performance bound, depend on the partial derivatives of the shifted version of the continuous image $f(x,y)$ evaluated at the sample locations. While this is simple to present theoretically, in practice, the partial derivatives of the image function are not available. In fact, only samples of the image function are available presenting a practical challenge when trying to compute the FIM. There are a few approximations that can be made in order to calculate the FIM depending on the information available prior to estimation. For instance, if a relatively noise-free image is available, preferably of higher resolution than the images being registered, then the partial derivatives might be approximated using derivative filters. For situations where the scene being observed is known prior to estimation, such as in industrial applications, a continuous image function could be constructed to represent the scene and differentiated analytically. Finally, if only the discrete images are available, then such an image function could be approximated directly from the samples. One such method assumes that the image is periodic and that

$$f(x,y) = \sum_{m}^{M}\sum_{n}^{N} F\left(\frac{2\pi m}{M}, \frac{2\pi n}{N}\right) e^{j2\pi\left(\frac{xm}{M}+\frac{yn}{N}\right)} \quad (8)$$

where $F((2\pi m)/(M), (2\pi n)/(N))$ are the coefficients of the discrete Fourier transform (DFT) of the image. It is this last assumption that we use throughout this paper for our experiments.

To gain further insight, we now consider the FIM in the Fourier domain. To do so, we first must make certain general assumptions about our underlying image function $f(x,y)$. In particular, we assume that the image function is bandlimited and is sampled at a rate greater than Nyquist. Then, the discrete time Fourier transform (DTFT) of the samples of the derivative function $f_x(m-v_1, n-v_2)$ can be written as $e^{j(v_1\theta_x + v_2\theta_y)} j\theta_x F(\theta_x, \theta_y)$ and similarly for the y partial derivative. With such an image model, we then can write the terms of the FIM using Parseval's relation

$$a_1 = \frac{1}{4\pi^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} |F(\theta_x,\theta_y)|^2 \theta_x^2 \, d\theta_x \, d\theta_y$$
$$a_2 = \frac{1}{4\pi^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} |F(\theta_x,\theta_y)|^2 \theta_x\theta_y \, d\theta_x \, d\theta_y$$
$$a_3 = \frac{1}{4\pi^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} |F(\theta_x,\theta_y)|^2 \theta_y^2 \, d\theta_x \, d\theta_y.$$

Examining the FIM using this formulation, we see that it does not depend on the unknown translation vector $\mathbf{v}$ and depends only on the image content. This observation depends on our assumption that the image is periodic outside the field of view.

It is interesting to note that one can explain the well-known *aperture problem* [1] by examining the FIM. This problem arises when the spectral content of the image is highly localized. An example of this occurs when all of the spectral energy is contained along a slice passing through the origin of the spectrum at an angle ψ_0. Equivalently, in the spatial domain, the texture of the image is one-dimensional (1-D) in nature. In polar coordinates, such a spectrum looks like

$$F(\psi,\rho) = \begin{cases} F(\psi_0,\rho), & \psi = \psi_0 \\ 0, & \text{else} \end{cases}. \quad (9)$$

The terms of the corresponding FIM in polar coordinates

$$a_1 = \frac{1}{4\pi^2}\int_{-\pi}^{\pi} |F(\psi_0,\rho)|^2 \rho^3 \cos^2\psi_0 \, d\rho$$
$$a_2 = \frac{1}{4\pi^2}\int_{-\pi}^{\pi} |F(\psi_0,\rho)|^2 \rho^3 \cos\psi_0 \sin\psi_0 \, d\rho$$
$$a_3 = \frac{1}{4\pi^2}\int_{-\pi}^{\pi} |F(\psi_0,\rho)|^2 \rho^3 \sin^2\psi_0 \, d\rho.$$

Since the determinant of the FIM is

$$\det J(v) = a_1 a_3 - a_2^2$$
$$= (\cos^2\psi_0 \sin^2\psi_0 - \cos^2\psi_0 \sin^2\psi_0)K = 0$$

(where K is a constant), $J(v)$ is, therefore, not invertible, and any unbiased estimator will have infinite variance. Essentially, there is not enough information with which to register the images.

Next, we further observe that the information contained in a pair of images depends only on the gradients or the texture of the image. The relationship between estimator performance and image content has been noted in previous works and used to select features to register [14]. This previous work, however, provided only the heuristic suggestion that features with high

Fig. 1. Experimental images (tree, face, office, and forest).

frequency content are better for tracking by looking at one specific estimator. Here, we suggest that the trace of the inverse FIM J^{-1} (which is the sum of the eigenvalues of J^{-1}) to be a scalar predictor of performance as it relates to image content. In general, as the trace of J^{-1} decreases, improved estimator performance is expected. Fig. 2 shows the square root (to maintain units of pixels) of the trace of J^{-1} versus image bandwidth for the images shown in Fig. 1. The image spectral bandwidth was controlled by filtering the images with a low-pass filter whose radial cutoff frequency θ_c was constructed to be a percentage of the full image bandwidth. All of the images were normalized in that they were cropped to the same size and scaled to have the same intensity range. As seen in Fig. 2, the trace of J^{-1} decreases as the image bandwidth increases. This corroborates the general intuition that highly textured images are easier to register.

Furthermore, we see from the left graph in Fig. 2 that, while the performance may continue to improve with greater frequency content, the improvement tapers off as the bandwidth increases beyond about a quarter of the full bandwidth. This observation can be explained by the $(1/\theta_c) = (1/\sqrt{\theta_x^2 + \theta_y^2})$ spectral amplitude decay commonly found in natural images [15]. This suggests that the trace of J^{-1} could be approximated by a term such as $(1/\theta_c)$ where θ_c is the radial cutoff frequency (or bandwidth of the image). The left graph in Fig. 2 exhibits a $\log(1/\theta_c)$ type behavior. These results also suggest that the inherent bandwidth limitations induced by the imaging system affect the fundamental performance limits for image registration. Since the spectral bandwidth of the image predicts the ability to register the image, the inherently bandlimited nature of imaging systems eventually dominates the achievable performance limits.

Another interesting way to explore the registration performance limits as a function of image content is by examining the bounds along particular directions. Instead of estimating both the v_1 and v_2 components of translation, we consider the linear combination $v_\psi = v_1 \cos\psi + v_2 \sin\psi = \mathbf{w}^T\mathbf{v}$ of the unknown parameters. The CRLB inequality (5) can be extended to bound the performance in estimating a linear combination of the unknown parameters. In particular, we have $\mathrm{Var}(\mathbf{w}^T\mathbf{v}) \geq \mathbf{w}^T J^{-1}(\mathbf{v})\mathbf{w}$. From this inequality, it becomes apparent that, for a particular image, certain angles have better inherent performance these optimal angles depending on the eigenvectors of the matrix J^{-1}. The right graph of Fig. 2 shows the variance bound on the estimation of the directional components of translation as a function of angular direction for the four example images in Fig. 1. The face image and, to a lesser extent, the office image, have specific directions in which estimates are most reliable. Specifically, the vertical bars in the face image provides large amounts of spectral energy in the x direction. This spectral signature correspondingly suggests small estimator variance in this angular direction. Similarly, the office image is rotated about 45°, so the dominant derivative energy is located around 45°.

B. Bias in Image Registration

In this section, we show that unbiased estimators generally do not exist for the inverse problem of image registration. This implies that the bound given by (5) is overly optimistic and the complete bound (4) must be used to accurately predict estimator performance.

To understand the inherent bias associated with any translational motion estimator, we look at the maximum likelihood (ML) estimators. Many image registration algorithms can be shown to produce approximate maximum likelihood solutions. To find the ML solution, we again look at the log likelihood function for the shift parameters

$$\log \mathbf{P}(z;\mathbf{v}) = \frac{-1}{2\sigma^2}\sum_{m,n}[z_1(m,n) - f(m,n)]^2 + [z_2(m,n) - f(m-v_1, n-v_2)]^2 + \text{const.}$$

Since only the second term depends on the unknown parameters, the maximization problem can be expressed as a minimization of the objective function

$$Q_{\mathrm{LS}}(\mathbf{v}) = \sum_{m,n}[z_2(m,n) - f(m-v_1, n-v_2)]^2. \quad (10)$$

This is the general nonlinear least squares objective function used in defining the ML solution. By expanding the quadratic in (10), we get

$$\sum_{m,n}[z_2^2(m,n) - 2z_2(m,n)f(m-v_1, n-v_2) + f^2(m-v_1, n-v_2)]. \quad (11)$$

Ignoring the first term since it does not depend on the parameter $\mathbf{v}$, and negating the entire function, we can rewrite the objective function as

$$\sum_{m,n} 2z_2(m,n)f(m-v_1, n-v_2) - \sum_{m,n} f^2(m-v_1, n-v_2). \quad (12)$$

By normalizing the entire cost function with respect to the energy in the image, the second term of (12), we obtain the direct correlator objective function

$$Q_{\mathrm{DC}}(\mathbf{v}) = \frac{\sum_{m,n} z_2(m,n)f(m-v_1, n-v_2)}{\sum_{m,n} f^2(m-v_1, n-v_2)}. \quad (13)$$

In general, minimizing/maximizing these two objective functions with respect to the unknown parameter $\mathbf{v}$ provides the ML

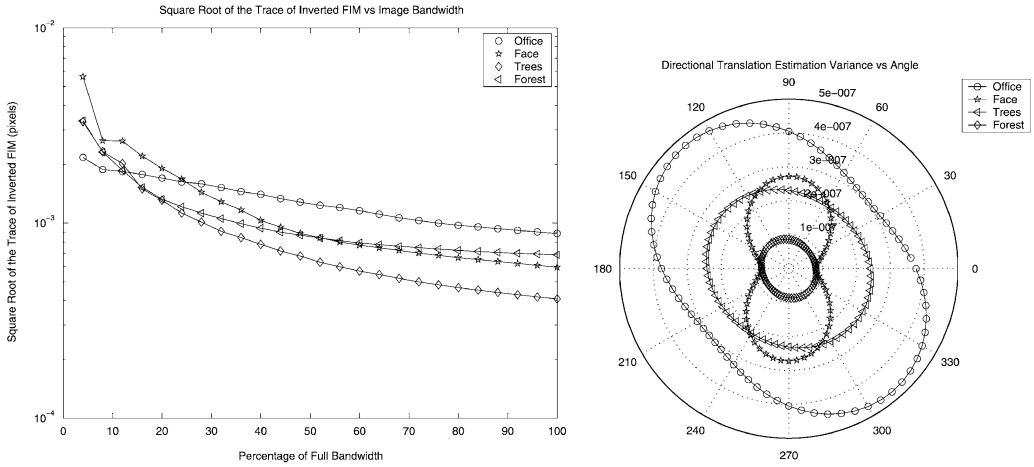

Fig. 2. Performance as it relates to image bandwidth and translational direction. (a) Trace of J^{-1} versus image bandwidth. (b) Angular performance as a function of image content.

solution. But, as previously noted, the function $f(m-v_1, n-v_2)$ is typically unknown. An approximate ML solution is found using an estimate of the unknown function, most commonly given by $z_1(m-v_1, n-v_2)$. It is easy to see that for very high SNR situations, this estimate should be very close to $f(m-v_1, n-v_2)$. Even in such high SNR (low noise) situations, however, the objective functions (10) and (13) can be evaluated for only integer values of v_1 and v_2, constraining the estimates to that of integer multiples of pixel motion. While some progress has been made to address this issue [4], [16], [3], the proposed algorithms often are based on overly simplified approximations that are known to produce biased estimates [17].

For many applications in image processing, accurate subpixel image registration is needed. To register images to subpixel accuracy, the image function $f(x, y)$ effectively must be reconstructed from the noisy samples of $z_1(m, n)$. In general, this reconstruction is an ill-posed problem. All estimators contain inherent prior assumptions about the space of continuous images under observation. These priors act to regularize the problem allowing solutions to be found. However, when the real underlying functions do not match the model assumptions, the estimators inevitably produce biased estimates. There is only a small class of images wherein the problem is not ill posed. The exception occurs when the underlying continuous image is *constructed* through the assumed process such as that of (8). Unfortunately, this requirement is almost never satisfied in practical image processing problems implying that all image registration algorithms are inherently biased.

To verify the presence of this bias in existing algorithms, we conduct a Monte–Carlo (MC) simulation computing actual estimator performance for a collection of image registration algorithms. The estimators used in the experiment are the following.

1) **Approximate Minimum Average Square Difference (ASD) (2-D version of** [4]). Samples of the average square difference function

$$\mathrm{ASD}(v_1, v_2) = \frac{1}{\mathrm{MN}} \sum_{m,n} (z_1(m-v_1, n-v_2) - z_2(m,n))^2 \tag{14}$$

[an approximation to (10)] are computed for pixel shift values of v_1 and v_2 in some range. Then, the subpixel shift is computed by finding the minimum of a quadratic fit about the minimum of the cost function given for integer pixel shifts.

2) **Approximate Maximum Direct Correlator (DC)** [3]. A sample correlation estimate is used to approximate (13). Essentially, the denominator of (13) is assumed to be approximately constant independent of the underlying image shift $\mathbf{v}$. Thus, the simplified sample correlation estimate

$$\mathrm{Cor}(v_1, v_2) = \frac{1}{\mathrm{MN}} \sum_{m,n} z_1(m-v_1, n-v_2) z_2(m,n) \tag{15}$$

is computed for integer pixel shifts. Then, the subpixel shift is estimated as the maximum of a quadratic fit about the maximum of the sample correlation function.

3) **Linear Gradient-Based Method (GB)** [18], [19]. Essentially, the differences between a pair of images is related to the spatial gradients of the image to produce the linear equation

$$z_2 - z_1 = f_x v_1 + f_y v_2 + \epsilon. \tag{16}$$

Then, a system of linear equations is constructed for a region within the image (or the entire image). These equations are often called the optical flow equations. The system is solved using least squares to produce an estimate of the translation $\mathbf{v}$. We will explore this model in more detail in the next section.

4) **Multiscale (Pyramid) Gradient-Based Method (Pyr)** [20]. The images are first decomposed into a multiscale (multiresolution) pyramid. The algorithm begins by estimating the translation in the coarsest images in the pyramid. Using this estimate, one of the images at the next coarsest level is warped according to the estimate. Essentially, this attempts to "undo" the motion. Then, the residual motion is estimated again using the GB method and combined with the previous motion estimate. This process continues down the pyramid in a multiscale iterative fashion. In our implementation, we use a cubic spline based resampling scheme at each level of the pyramid to

warp the images. In our experiments, we use a multiscale pyramid with three levels.

5) **Projection Gradient-Based Method (Proj-GB)** [21]. The images are integrated along the x and y axis to produce two sets of data $z^x_{1,2}(m) = \sum_n z_{1,2}(m,n)$ and $z^y_{1,2}(n) = \sum_m z_{1,2}(m,n)$. Then, a pair of equations similar to those in the 2-D GB method are generated

$$z^x_2 - z^x_1 = f^x_x v_1 + \epsilon^x$$
$$z^y_2 - z^y_1 = f^y_y v_2 + \epsilon^y.$$

Finally, two independent sets of linear equations are constructed and solved using least squares to obtain estimates of the components of $\mathbf{v}$.

6) **Projection Multiscale Gradient-Based Method (Pyr-Proj)** [22]. The image is first decomposed into a pyramid as in the multiscale gradient-based (Pyr) method. At each level of the pyramid, instead of using the 2-D GB method, the motion is estimated using the projection gradient-based (Proj-GB) method.

7) **Relative Phase (Phase)** [23]. Using the shift property of the Fourier transform, it is noted that $(F_1 F_2^*)/(|F_2|^2) = (F_1/F_2) = e^{j2\pi(\theta_x v_1 + \theta_y v_2)}$. The vector $\mathbf{v}$ is estimated by finding the solution to the set of linear equations of the phase function

$$\angle \frac{Z_1}{Z_2} = j2\pi(\theta_x v_1 + \theta_y v_2) \tag{17}$$

where $Z_{1,2}$ represents the DFT of the input images $z_{1,2}$ and $\angle$ indicates the measured phase angle. We used the implementation of [23] wherein the solution is found using weighted least squares.

To generate a pair of images for the experiment, we use the discrete Fourier transforms (DFT) approach following the method of [10]. This effectively generates an image pair assuming the continuous model is given by (8). The image used in this experiment was the trees image from [1]. White Gaussian noise was added to the image pair prior to estimation and the entire process was repeated 500 times at each SNR value. We explore SNR situations ranging from 0 dB (very noisy) to 70 dB (effectively noiseless). To capture a single representation of error, we computed the square root of the trace of the MSE matrix for each of the estimators and the bound of (5). The square root of the trace of the MSE matrix is a valid measure of the mean magnitude error and is useful for comparing with the performance bounds given by the CRLB [24]. Fig. 3 shows the actual estimator performance as a function of of SNR. The dashed line indicates the performance bound using (5) for the class of unbiased estimators. While this bound suggests continued improvement as the noise decreases, above certain SNR values, the performance of each estimator levels out. This flattening of the performance curves is indicative of the bias present in each of the estimators.

While we can see the effect of this bias experimentally, the actual bias function for a given estimator typically is very difficult to express. The bias is often a combination of both the deterministic modeling error and the statistical bias of the estimator. If the estimator is an ML estimator, the estimates should theoretically be asymptotically unbiased, leaving only the bias

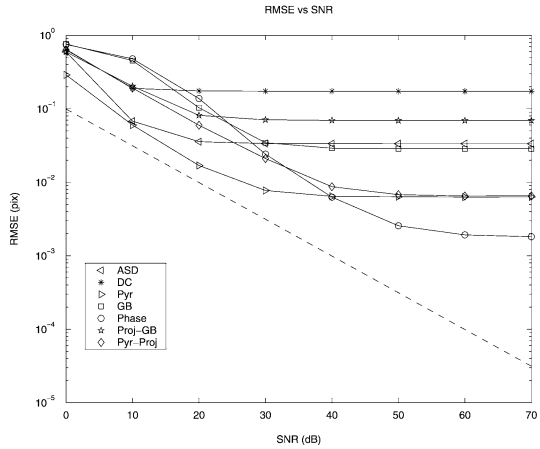

Fig. 3. Magnitude error performance versus SNR, $\mathbf{v} = [.5,\ .5]^T$.

stemming from modeling error. This appears to be the dominant bias for high SNR situations as seen in Fig. 3 where the bias is independent of the noise in the images. This modeling error has been only infrequently addressed in the image registration literature. In [25], the approximate direct correlator method (DC) produces biased estimates resulting from the quadratic approximation about the peak of the correlation function. Basically, the DC method using the quadratic approximation about the mean of the sample correlation function makes implicit assumptions about the underlying continuous function. In [25], and similarly in [17], the resulting bias is derived for situations where the likelihood function is not quadratic about its maximum as typically assumed. The gradient-based estimators have been studied in the context of bias as well [7]–[10]. Nevertheless, an accurate functional expression describing the estimator bias is not available. In the next section, we describe these attempts at understanding gradient-based estimator bias and derive and verify a new functional form of bias inherent to gradient-based estimators.

III. Bias in Gradient-Based Estimators

To understand more clearly the effect of estimator bias, we now derive the functional form of bias inherent to gradient-based estimators. We show how the approximations used to generate a simple linear estimator produce inherent estimator bias. We show that in almost all situations, the gradient-based estimator contains bias. To maintain focus and facilitate the exposition, we derive and later analyze the bias expression for the 1-D analog of the gradient-based image registration algorithm. The full derivations for the two-dimensional (2-D) case are included in the Appendix.

A. Gradient-Based Estimators

For the 1-D case, we suppose that the measured data is of the form

$$z_1(k) = f(k) + \epsilon_1(k) \tag{18}$$
$$z_2(k) = f(k - v) + \epsilon_2(k). \tag{19}$$

In the derivation of the gradient-based estimator, we must reformulate the data as $z(k) = z_1(k) - z_2(k) = f(k+v) - f(k) + \epsilon(k)$ where ϵ is a Gaussian white noise process with variance σ^2.

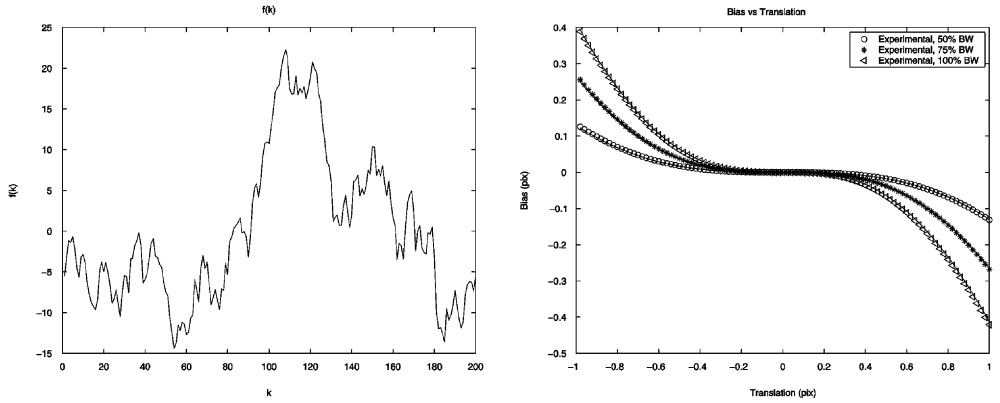

Fig. 4. (Left) Plot of $f(k)$ and (right) estimator bias; continuous is predicted.

Gradient-based methods solve this equation for v by linearizing the function $f(k+v)$ about a point $v=0$ in a Taylor series. This expansion looks like

$$f(k+v) - f(k) = vf'(k) + R(k,v) \tag{20}$$

where R is the remainder term in the Taylor expansion. This remainder has the form $R(k,v) = \sum_{r=2}^{\infty}(v^r/r!)f^{(r)}(k)$. Thus, the new data model becomes $z(k) = vf'(k) + R(k,v) + \epsilon(k)$. When the remainder term R is ignored, the linearized model of the data becomes $z(k) = f'(k)v + \epsilon(k)$. Using the derivative values, we obtain the linear estimator for the velocity v using least squares

$$\hat{v} = \frac{\sum f'(k)z(k)}{\sum (f'(k))^2}. \tag{21}$$

where the sum is taken to be over some region which may be the entire image. This type of estimator is commonly referred to as the gradient-based or differential estimation method [19], [18]. This estimator derivation assumes that in addition to the samples of $f(k)$, we also have samples of the derivative of the function $f'(k)$. Later, we show how this assumption is relaxed.

It is interesting to note that the variance of the gradient-based estimator is $\mathrm{var}(\hat{v}) = (\sigma^2)/(\sum (f'(k))^2)$ if, in fact, the remainder term is zero. The variance is almost exactly the same as the CRLB for unbiased estimators, which is $(\sigma^2)/(\sum (f'(k+v))^2)$. This relationship implies that the gradient-based estimator would be a maximum likelihood estimator for the case when the remainder term is, in fact, zero.

B. Bias From Series Truncation

One source of systematic error or bias in the gradient-based estimation method comes from the remainder term $R(k,v)$ in (20) originally ignored to construct a linear estimator.

When we include the remainder term in the estimator we obtain as the expected value of the estimator (21), $E[\hat{v}] = v + (\sum f'(k)R(k,v))/(\sum (f'(k))^2)$. So, unless the second term is zero, the higher order terms introduce a systematic bias into the estimator.

This is somewhat more informative in the frequency domain. First, we define the Fourier transform of the original function $f(x)$ as $F(\omega)$. Under the assumption that the function is sampled above the Nyquist rate, the DTFT of the derivative sequence $f'(k)$ can be represented as $j\theta F(\theta)$. By Parseval's relation, we can rewrite the estimator (21) as

$$\hat{v} = \frac{\int_{-\pi}^{\pi} j\theta F(\theta) Z^*(\theta) d\theta}{\int_{-\pi}^{\pi} |F|^2 \theta^2 \, d\theta}. \tag{22}$$

As a side note, we can also arrive at the same estimator form by modeling the data itself directly in the frequency domain, as follows. The shifted sequence $f(k+v)$ has a DTFT of $F(\theta)e^{jv\theta}$ and the DTFT of the data model becomes

$$Z(\theta) = F(\theta)[e^{jv\theta} - 1] + \xi(\theta). \tag{23}$$

If we again expand the exponential in a Taylor series $e^{jv\theta} = 1 + jv\theta - ((v\theta)^2)/(2) + \cdots$ and truncate after the linear term we obtain the linear relationship $Z(\theta) = [F(\theta)j\theta]v + \xi(\theta)$. From which we obtain the linear estimator as (22).

Returning to the case where the complete data model is used, we see that the expected value of the estimate is

$$\begin{aligned} E[\hat{v}] &= \frac{\int_{-\pi}^{\pi} |F(\theta)|^2 j\theta (e^{-jv\theta} - 1)\, d\theta}{\int_{-\pi}^{\pi} |F(\theta)|^2 \theta^2 \, d\theta} \\ &= \frac{\int_{-\pi}^{\pi} |F(\theta)|^2 \theta \sin(v\theta)\, d\theta}{\int_{-\pi}^{\pi} |F(\theta)|^2 \theta^2 \, d\theta} \\ &\quad + j\frac{\int_{-\pi}^{\pi} |F(\theta)|^2 \theta (\cos(v\theta) - 1)\, d\theta}{\int_{-\pi}^{\pi} |F(\theta)|^2 \theta^2 \, d\theta} \\ &= \frac{\int_{-\pi}^{\pi} |F(\theta)|^2 \theta \sin(v\theta)\, d\theta}{\int_{-\pi}^{\pi} |F(\theta)|^2 \theta^2 \, d\theta} \end{aligned} \tag{24}$$

where in the last equality we note that since $\mathrm{Im}[j\theta(e^{-jv\theta} - 1)] = \theta(\cos(v\theta) - 1)$ is an odd function, it integrates to zero. Using the expected value of the estimate, we obtain a bias function of the form

$$b(v) = E[\hat{v}] - v = \frac{\int_{-\pi}^{\pi} |F(\theta)|^2 (\theta \sin(v\theta) - v\theta^2)\, d\theta}{\int_{-\pi}^{\pi} |F(\theta)|^2 \theta^2 \, d\theta}. \tag{25}$$

To verify this bias function experimentally, we measure the bias in estimating translation for a randomly constructed function such that the actual derivative values were available to the

estimator. The actual function $f(k)$ used in the experiment is plotted in the left graph of Fig. 4. The magnitude spectrum for the function used was $F(\theta) \approx (1/\theta)$ modeled after the spectrum of natural images. The phase angle was drawn from a uniform distribution. To measure purely the deterministic bias, no noise was added to the data prior to estimation. Fig. 4 shows a plot of the experimental estimator bias as it depends on translation v. The plot shows three different curves which indicate the bias for the full bandwidth function $f(k)$ as well as two filtered versions of $f(k)$ wherein the functions were bandlimited to 50% and 75% of the full bandwidth. The continuous curves represent the predicted bias using (25).

The bias function appears to follow the bias expression almost exactly. Furthermore, Fig. 4 indicates that as the bandwidth of $f(k)$ increases, the bias becomes more severe. This conflicts with the unbiased CRLB which suggests that increased bandwidth will improve estimator variance. Here, we begin to see the tradeoff between bias and variance for the gradient-based estimators. We will examine this notion more closely later in Section IV.

C. Bias From Gradient Approximation

In the previous section, we assumed that the derivative values at the sample points were known prior to the estimation process. As mentioned previously, in most applications, the derivative information is not available. Another source of error in gradient-based estimation arises from the need to approximate the gradient or the derivatives of the signal $f(k)$. Instead of using the actual $f'(k)$ in (21), noisy approximations of the derivatives $f'(k) \approx z_2(k) * g(k) = [f_2(k) + \epsilon_2(k)] * g(k)$ are used instead (where $*$ represents a convolution operation). This suggests that the deterministic bias is a combination of the error in approximating $f'(k)$, as well as the error introduced by truncating the Taylor series and ignoring the remainder term $R(k, v)$.

The error resulting from such derivative approximation has been noted before in the literature. For instance, in [10], the bias function was derived only for the case when f is a single sinusoid function. In addition, the works of [7] and [8] explored the effect of approximation errors in estimating the gradient for local estimation. Much of the analysis in these works, however, start from the assumption that the optical flow model applies to the image sequence exactly, or that the remainder term is negligible. Specifically, in [8], the results qualitatively described estimator bias in terms of image spectral content and were based on overly simplified bias approximation by examining only the second-order approximation error specifically for the forward difference gradient approximation. The authors in [7] note that the gradient approximation error increases as the image function exhibits higher energy in the second derivatives $f''(k)$. Using this observation, they propose an estimator post-processing scheme which examines the second-order derivatives of the image and rejects specific estimates according to a thresholding scheme. Other works, such as [9], have noted that errors in the gradient approximation tend to produce biased estimates. In [9], however, it is assumed that these errors are completely random in nature and drawn from some simple distribution. They develop overly simplified statistical bias models based on these distributions for the gradient approximation errors. Recently, the work of [11] investigates a method for minimizing the bias associated with such random errors for an application in vehicle tracking. Instead of treating these errors as random, as we shall show, approximation errors resulting from deterministic systematic modeling error dominate the estimator bias for gradient-based estimators at typical imaging system SNRs.

When we use the gradient approximations, the estimator (22) becomes

$$\hat{v} = \frac{\int_{-\pi}^{\pi} jG(\theta)Z_2(\theta)Z^*(\theta)\,d\theta}{\int_{-\pi}^{\pi} |G(\theta)Z_2(\theta)|^2\,d\theta} = \frac{\int_{-\pi}^{\pi} jG(\theta)[F(\theta)+\xi_2(\theta)]Z^*(\theta)\,d\theta}{\int_{-\pi}^{\pi} |G(\theta)[F(\theta)+\xi_2(\theta)]|^2\,d\theta} \tag{26}$$

where $G(\theta)$ represents the DTFT of $g(k)$ and $\xi_2(\theta)$ represents the DTFT of the noise samples $\epsilon_2(k)$. In general, the derivative filter $g(k)$ is usually a symmetric, linear-phase, FIR filter and as such its transform $jG(\theta)$ can be written as a sum of sinusoids or $G(\theta) = \sum_l^L c_l \sin(l\theta)$. Unfortunately, taking the expectation of (26) is very difficult. To simplify the equation, we ignore the noise in the derivative approximation. This assumption is quite reasonable for high SNR situations where basically we are examining the deterministic bias from modeling error as opposed to statistical error. In Section V, we will show the SNR region where this model accurately describes estimator performance and that this SNR region is typical for imaging systems using commercial video cameras. Thus, we approximate the bias function as

$$b(v) \simeq \frac{\int_{-\pi}^{\pi} |F(\theta)|^2[G(\theta)\sin(v\theta) - vG^2(\theta)]\,d\theta}{\int_{-\pi}^{\pi} |G(\theta)F(\theta)|^2\,d\theta}. \tag{27}$$

We can see here that the this equation differs from the original (25) only in that the exact derivative operator $j\theta$ is replaced by the derivative filter with frequency response $jG(\theta)$.

To verify this approximation of the bias function, we measure the actual estimator bias using the gradient kernel $g(k) = [.1069 .2846\ 0\ -.2846\ -.1069]$ on the same function shown in Fig. 4. This derivative kernel comes from [26]. The left graph of Fig. 5 shows the results of the bias. The experimental bias again follows the bias predicted by (27) almost exactly. The measured bias functions shown in [10] also appear to follow this trend providing further validation of our bias expression. Again, we note that the increased signal bandwidth produces increased estimator bias.

IV. Analysis of Gradient-Based Estimator Bias

In this section, we further explore the deterministic bias approximation (27). We will show how the structure of the bias function explains much of the heuristic knowledge about gradient-based estimators and suggests methodologies for improving performance. In particular, we will explore how the image spectrum, translation, and gradient kernel affect the bias of the gradient-based estimator.

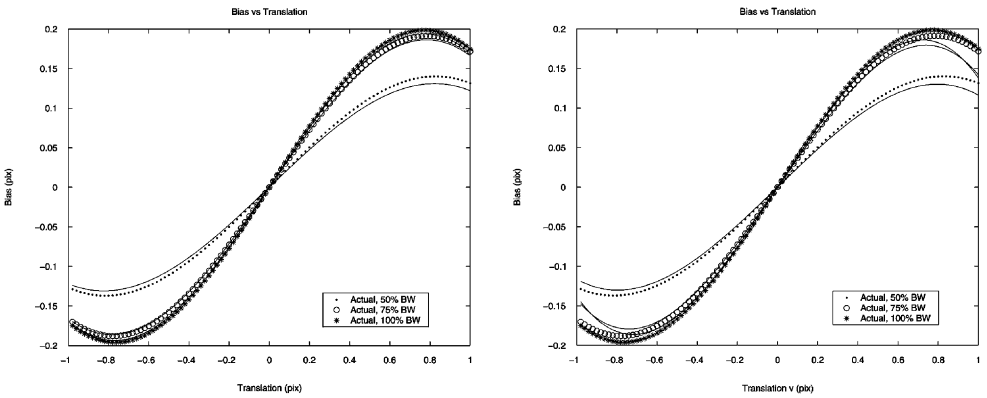

Fig. 5. Plot of actual estimator bias and predicted bias (solid lines) from (27) left graph and (32) right graph.

We begin by analyzing the bias function (25) wherein the exact derivatives are available to the estimator. To understand the bias, we expand the sin function in a Taylor series about $v = 0$ to get

$$b(v) = \frac{\int_{-\pi}^{\pi} |F(\theta)|^2 [v\Lambda_1(\theta) - v^3\Lambda_2(\theta) + v^5\Lambda_3(\theta)\cdots]\, d\theta}{\int_{-\pi}^{\pi} |F(\theta)|^2 \theta^2\, d\theta} \tag{28}$$

where the terms of the sequence are $\Lambda_1(\theta) = 0, \Lambda_2(\theta) = (\theta^4/3!), \Lambda_3(\theta) = (\theta^6/5!)$, and so on. Since the factorial in the denominator dominates these Λ functions, the coefficients of the Taylor approximation die off quickly. Only for very large translations, often larger than is found in typical registration problems, will these higher order terms affect the bias function. This suggests that for small v, the bias can be approximated as a cubic function of translation v according to

$$b(v) \approx -\frac{v^3}{3!} \frac{\int_{-\pi}^{\pi} |F(\theta)|^2 \theta^4\, d\theta}{\int_{-\pi}^{\pi} |F(\theta)|^2 \theta^2\, d\theta}. \tag{29}$$

This coefficient ratio can be interpreted as the energy in the second derivative over the energy in the first derivative of $f(x)$. In general, the Taylor series can be explained in the spatial domain as

$$b(v) = -\frac{v^3}{3!} \frac{\sum [f''(k)]^2}{\sum [f'(k)]^2} + \frac{v^5}{5!} \frac{\sum [f^{(4)}(k)]^2}{\sum [f'(k)]^2} - \cdots \tag{30}$$

Basically, these higher order terms depend on the smoothness of the function $f(x)$. For sufficiently smooth functions, the energy in these higher derivatives is negligible suggesting that the bias is well approximated by the cubic function given in (29). The accuracy of this bias approximation is evident in right graph of Fig. 4.

We repeat this analysis for the more complete bias function (27) expanding the function in a Taylor series about $v = 0$ to produce

$$b(v) = \frac{\int_{-\pi}^{\pi} |F(\theta)|^2 [v\Lambda_1(\theta) - v^3\Lambda_2(\theta) + v^5\Lambda_3(\theta)\cdots]\, d\theta}{\int_{-\pi}^{\pi} |G(\theta)F(\theta)|^2\, d\theta} \tag{31}$$

where the terms are of the sequence are $\Lambda_1(\theta) = \theta G(\theta) - G^2(\theta), \Lambda_2(\theta) = (\theta^3/3!)G(\theta), \Lambda_3(\theta) = (\theta^5/5!)G(\theta)$ and so on. From this approximation, we see that the polynomial coefficients depend on the relationship between the gradient kernel $G(\theta)$ and the image magnitude spectrum $|F(\theta)|$. Again, we simplify the bias expression by truncating the power series to that of a cubic function of v

$$b(v) \approx v \left(\frac{\int_{-\pi}^{\pi} |F(\theta)|^2 \Lambda_1(G,\theta)\, d\theta}{\int_{-\pi}^{\pi} |G(\theta)F(\theta)|^2\, d\theta} \right) - v^3 \left(\frac{\int_{-\pi}^{\pi} |F(\theta)|^2 \Lambda_2(G,\theta)\, d\theta}{\int_{-\pi}^{\pi} |G(\theta)F(\theta)|^2\, d\theta} \right). \tag{32}$$

In the right graph of Fig. 5, we show the same experimental bias curves as in the left graph of Fig. 5 this time using the cubic approximation of (32). We see that the approximation is quite close for the subpixel region of v.

A. Bias and Image Spectrum

The spectrum of the image/function plays an important role in the bias expression (27). One way to shape the image spectrum is through the use of image filters. For instance, it is well-known that presmoothing the images prior to estimation improves the performance of the gradient-based estimators [1], [26]. This presmoothing operation takes the form of a low-pass filter $H(\theta)$. To understand this, in left graph of Fig. 6 we plot the Λ functions found in (31), again using the gradient kernel $G(\theta)$ from [26].

Basically, the Λ functions and the $|G|^2$ (where $|G|^2 = |G(\theta)|^2$) term control numerator and denominator of the coefficients of the bias polynomial. Looking at the left graph of Fig. 6, we see that the $|G|^2$ term is larger than all of the Λ functions up to the frequency of about $(\pi/3)$ for $\Lambda_1, (\pi/2)$ for Λ_2 and about $(2\pi/3)$ for Λ_3. If the spectrum of the function were bandlimited such that the image contained no spectral energy outside these frequencies, we know that the bias coefficients of the bias function would be less than 1. Beyond these critical frequencies, the numerator Λ functions weight the spectrum more heavily than the denominator $|G|^2$ function, which has the effect of increasing the bias coefficients. As we will show, this explains the well-known assertion that presmoothing the images improves estimator performance. Intuitively, the image presmoothing has the effect of minimizing the high frequency

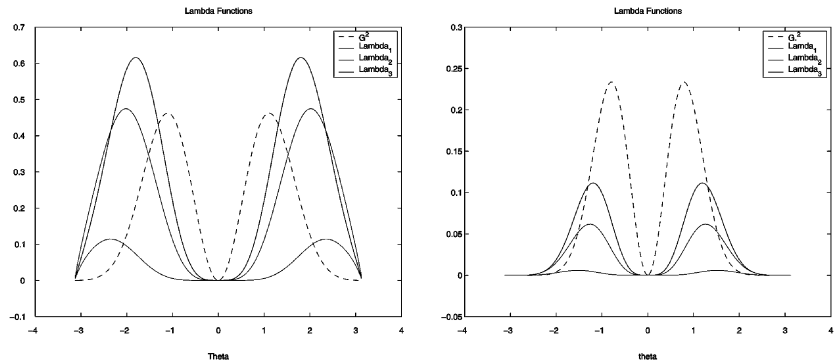

Fig. 6. (Left) Original and (right) filtered versions of Λ and $|G|^2$ functions. The filter function $h(k) = [0.035\ \ 0.248\ \ 0.432\ \ 0.248\ \ 0.035]$ is suggested in [26].

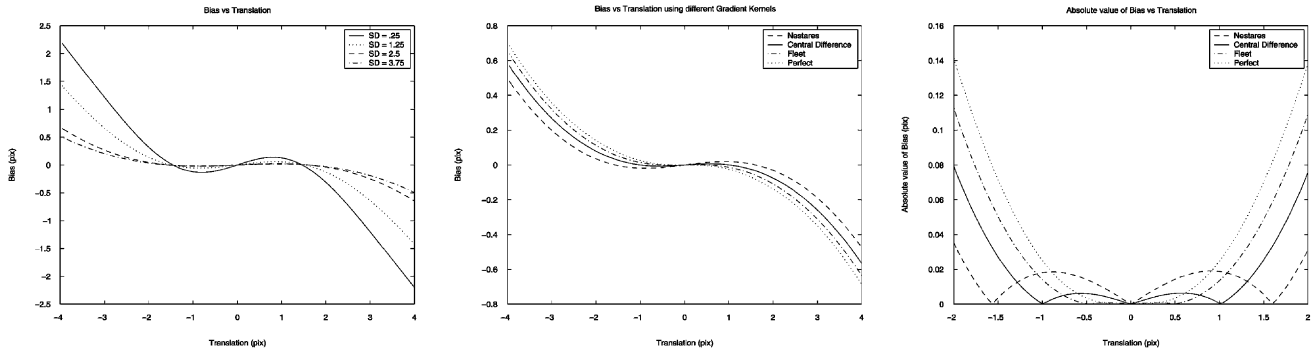

Fig. 7. Bias versus translation using different (left) prefilters and (middle and right) different gradient filters.

spectral components thereby reducing the numerator of the coefficients more than the denominator. Furthermore, since higher order terms place more emphasis on the high frequency information than the lower order terms, the presmoothing also has the affect of minimizing the higher order bias polynomial terms more than the lower order terms.

For instance, the authors in [26] suggest using a five-tap presmoothing low-pass filter $h(k)$. Effectively, this presmoothing changes the weighting functions into $|GH|^2, \Lambda_1|H|^2, \Lambda_2|H|^2$ and so on. In the right graph of Fig. 6, we show the filtered versions of the Λ functions. Unlike the original Λ functions, the smoothed versions have much smaller magnitude than the $|G|^2$ function and very small regions wherein the numerators would weight the spectrum F more than the denominator. This phenomenon tends to minimize the bias polynomial coefficients. For high SNR situations where the bias dominates MSE, presmoothing tends to minimize the bias in general. This is shown in Fig. 7, where the bias is plotted as a function of translation wherein the function in Fig. 4 is filtered by different presmoothing filters. Each of the filters was a Gaussian kernel with ten taps where the low-pass cutoff frequency was controlled by the standard deviation (SD) of the Gaussian. These low-pass filters were not designed in any optimal fashion, and yet we still see a significant reduction in bias. For this experiment, we extended the range of translation beyond subpixel translation to show the dramatic improvement for larger values of v.

Presmoothing an image has added the benefit of averaging, essentially decreasing the variance of the noise. Again, this presmoothing would, however, decrease the Fisher information by reducing the effective bandwidth of the signal. Interestingly, one could pose an optimization problem of finding the prefilter $H(\theta)$ that minimizes the bias in a sense similar to [27]. Of course, this optimization would only make sense for very high SNR situations as presmoothing the image would tend to minimize the FIM thereby making the estimator more sensitive to noise. We leave this interesting problem for future work.

B. Bias and Gradient Kernel

Another important ingredient in the bias function is the choice of gradient filters $G(\theta)$. The gradient kernel defines the shape of the Λ functions which in turn controls the bias coefficients. The middle graph of Fig. 7 exhibits the performance in estimating translation using three different filters from [1] and [26] and also the performance using the exact derivatives. The experimental setup was similar to previous experiments wherein the function used was shown in Fig. 4 and no noise was added to simulate infinite SNR.

Examining the bias curves, it might appear that the Nestares/Heeger filter minimizes the bias, even producing better estimates than when the exact derivatives were known prior to estimation. In the right graph of Fig. 7 we examine the curves more closely in the range $v \in [-2, 2]$, and display absolute value of the bias. In the subpixel range ($v \in [-1, 1]$), we see that the Nestares/Heeger filter, in fact, produces estimators with largest bias magnitude.

We see from these plots that there is a tradeoff in performance in estimating large and small translations. It appears that the tradeoff concerns the linear term in the bias polynomial approximation. The central difference and Fleet derivative filters of [1]

are the second- and the fourth-order optimal approximations to the infinite-ordered ideal derivative filter. Thus, these filters produce derivative estimates closer to the exact derivative than the filter of Nestares/Heeger. This more accurate derivative approximation tends to minimize the linear term of the bias polynomial leaving basically the cubic term as in the case of (29). The filter of Nestares/Heeger, however, is not an approximation to the ideal derivative filter and as such has a larger linear coefficient. This larger linear coefficient explains its poor performance around the subpixel range and yet produces a linear improvement for larger translations. Again, this phenomenon suggests a certain optimization framework similar to [27] where the gradient kernel may be optimized over some range of translations.

C. Bias and Translation

Finally, we examine how the bias varies with the unknown translation v. As expected, the first-order approximation used to generate the linear gradient-based estimator is accurate only for small translations. Thus, with perfect knowledge of the image derivatives, the magnitude of the bias tends to increase with the translation and the estimates are always biased toward zero, or underestimated. When the derivatives are only approximated using a gradient kernel, however, there are essentially two regions of operation wherein the estimates could be overestimated and underestimated. These regions are easy to identify when examining the cubic approximation of the bias (32). The roots of the cubic polynomial approximation are

$$\pm\tilde{v} = \left(\frac{\int_{-\pi}^{\pi} |F(\theta)|^2 \Lambda_1(G,\theta) d\theta}{\int_{-\pi}^{\pi} |F(\theta)|^2 \Lambda_2(G,\theta)\, d\theta} \right)^{\frac{1}{2}}. \tag{33}$$

Instead of biasing the estimates toward 0, as in the case where the derivatives were known exactly, the estimator produces estimates that are biased toward $\pm\tilde{v}$. Examination of the bias in the right graph of Fig. 7 shows that these values are around $\tilde{v} = 1.5$ for Nestares/Heeger, $\tilde{v} = 1$ for the central difference and $\tilde{v} = .5$ for the Fleet gradient filters. In fact, we found that these value of $\tilde{v}$ do not vary much across different images, for each derivative filter.

Whichever gradient kernel is used, if the kernel approximates the derivative, the magnitude of the bias will tend to worsen for values of $|v| > |\tilde{v}|$. In fact, the cubic approximation of bias suggests that even the relative bias $((b(v))/v)$ increases as a quadratic function of v. This partly explains the success of multiscale gradient-based methods in estimating large translations. The multiscale pyramids are constructed through a process of low-pass filtering and downsampling. We have already shown how the low-pass filtering improves estimator performance. The downsampling reduces the magnitude of the translation by the downsampling factor, the common factor being 2. Using this downsampling factor, the translation to be estimated at the qth level of the pyramid becomes $v_q = (v/2^q)$. This synthetic reduction in translation magnitude allows for estimation with smaller relative bias. The reduction in bias is most effective when the unknown translation is greater than a few pixels. In this case, the downsampling maps the translation into a range of reasonably small bias. In practice, the height of the pyramid Q is designed such that the expected downsampled velocity at the coarsest level is in $v_Q \in [-2, 2]$ pixels/frame where the magnitude of the relative bias is not very large.

The iterative nature of the multiscale pyramid raises an important question concerning the convergence in general of iterative gradient-based estimators. Iterative methods for gradient-based estimation have been used to improve performance [20], [26], [19]. These methods work by iteratively estimating motion, *undoing* this estimated motion, and estimating the residual motion not captured by the previous estimate. At very high SNR, the residual motion is dominated by the estimator bias. In practice, different methods are used to *undo* the previously estimated motion, often relying on some warping/resampling scheme. We would like to know if these iterative methods will converge and, if so, whether they will converge to an unbiased estimate of v.

To simplify the analysis, we assume that the warping methods work perfectly to synthesize a shifted version of the images (unlikely, however, given the ill-posed nature of image resampling). In fact, we see that the error in the gradient approximation could lead to oscillatory instability in the iterative gradient-based estimator. To see this, assume that an initial estimate of translation using the gradient based estimator was given by $\hat{v}_0 = v + b(v)$. After warping, the residual translation would simply be $r_0 = -b(v)$. The estimate of this residual motion will be $\hat{r}_0 = -b(v) + b(-b(v))$ such that the updated motion estimate becomes $\hat{v}_1 = \hat{v}_0 + \hat{r}_0 = v + b(-b(v))$. Thus, if $|b(v)| < |v|$ for all v, then $|b(-b(v))| < |b(v)|$ and so on suggesting convergence to an unbiased estimate. Practically speaking, we are only interested in this relationship for very small v since the residual motions are often within the range $[-\tilde{v}, \tilde{v}]$. In this region, we use the cubic approximation of (32) represented as

$$b(v) = \frac{\gamma_1}{\gamma_2} v - \frac{\gamma_3}{\gamma_2} v^3 \tag{34}$$

where the γ variables represent the numerator and denominators of the polynomial bias approximation. Because of the symmetry of the bias function, we must examine whether or not $|b(v)| < |v|$ for all $v \in [0, \tilde{v}]$. Since the function is a simple polynomial, we first examine the existence of a root to the equation $(b(v))/(v) - 1 = 0$, which after some algebraic manipulation gives the root as

$$v^* = \left(\frac{\gamma_1 - \gamma_2}{\gamma_3} \right)^{\frac{1}{2}}. \tag{35}$$

Thus, if $\gamma_2 \geq \gamma_1$, then we can safely assume that $|b(v)| < |v|$ for small translations assuring that the iterative method will converge to an unbiased estimate since the bias is reduced at every iteration. If, however, $\gamma_2 < \gamma_1$ then the estimator will oscillate between $\hat{v} = v \pm v^*$.

Since the condition of convergence depends on

$$\gamma_1 - \gamma_2 = \int_{-\pi}^{\pi} |F(\theta)|^2 [\theta G(\theta) - 2G^2(\theta)]\, d\theta \tag{36}$$

we plot $\theta G(\theta) - 2G^2(\theta)$ in left graph of Fig. 8. For the iterative estimator to converge, most of the spectral energy must be

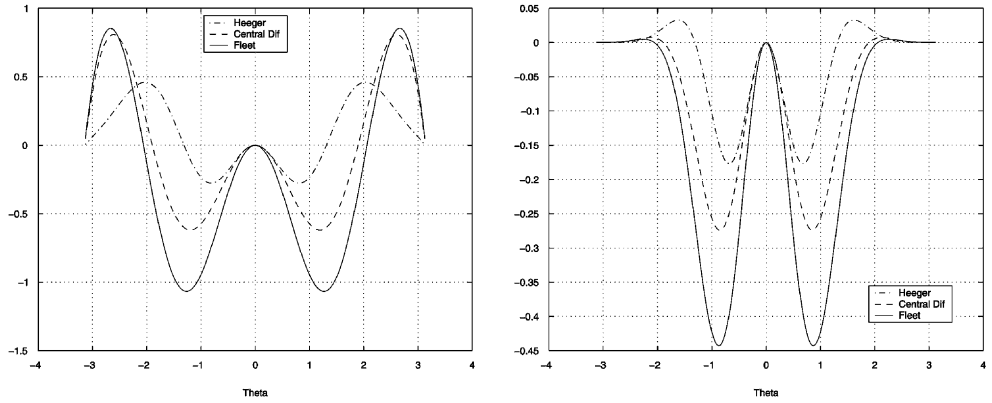

Fig. 8. (Left) Original and (right) filtered plot of $\theta G(\theta) - 2G^2(\theta)$.

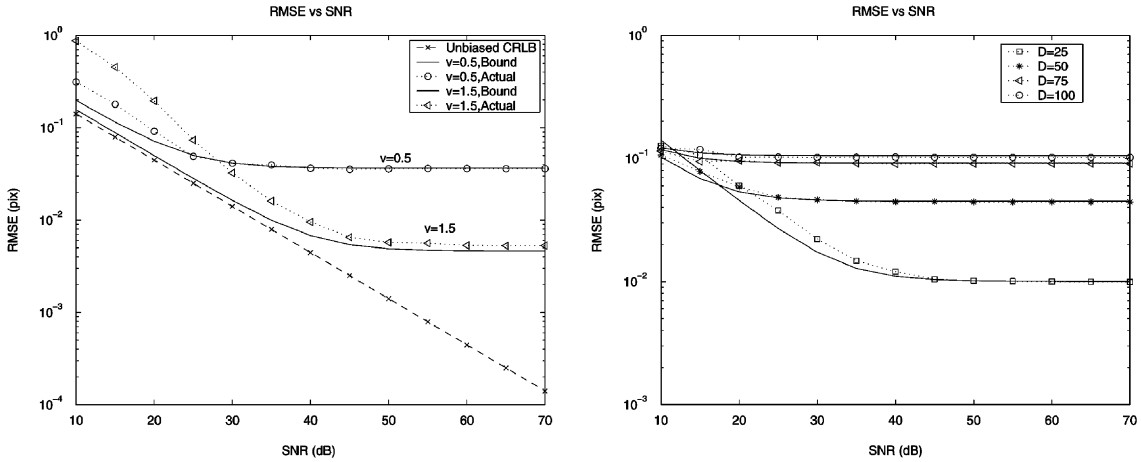

Fig. 9. Actual and predicted RMSE versus SNR as it relates to (left) translation and (right) bandwidth.

located in the low frequency range where the weighting function $\theta G(\theta) - 2G^2(\theta)$ applies negative weight. If too much high frequency content is present, the difference $\gamma_1 - \gamma_2$ will be positive and the algorithm will not converge to an unbiased estimate. Presmoothing the image minimizes the likelihood that $\gamma_1 - \gamma_2 > 0$, since most of the weighting function $\theta G(\theta) - 2G^2(\theta)$ is negative. In practice, multiscale iterative methods significantly decrease estimator bias as evidenced in Fig. 3, but may still contain estimator bias.

V. MSE Performance of the Gradient-Based Method

Armed with an approximate expression for the bias function, we can now examine the full performance bound given by (4) for the gradient-based estimators. In examining this bound, we find that the bias dominates the MSE performance for typical imaging systems with high SNR. Finally, we show experimental evidence justifying a general rule-of-thumb for performance of 2-D gradient-based image registration.

In order to use the performance bound given by (4), we must first examine the derivative of the bias function. Using the bias expression (27), we see that $b'(v) + 1 = (\int_{-\pi}^{\pi} |F(\theta)|^2 G(\theta)\theta \cos(v\theta) d\theta)/(\int_{-\pi}^{\pi} |G(\theta)F(\theta)|^2 \, d\theta)$. Using these expressions, we see that the complete MSE performance bound is given by

$$\mathrm{MSE}(v) \geq \frac{[b'(v)+1]^2}{J(v)} + b^2(v) \tag{37}$$

$$= \frac{J^{-1}\left(\int_{-\pi}^{\pi} |F(\theta)|^2 G(\theta)\theta\cos(v\theta)\, d\theta\right)^2}{\left(\int_{-\pi}^{\pi} |G(\theta)F(\theta)|^2\, d\theta\right)^2} + \frac{\left(\int_{-\pi}^{\pi} |F(\theta)|^2 [G(\theta)\sin(v\theta) - vG^2(\theta)]\, d\theta\right)^2}{\left(\int_{-\pi}^{\pi} |G(\theta)F(\theta)|^2\, d\theta\right)^2} \tag{38}$$

where the Fisher information is $J = (1/\sigma^2)\int_{-\pi}^{\pi} |F(\theta)|^2\theta^2\, d\theta$. In practice, we calculate the Fisher information using derivative approximations.

Here, we conduct a MC simulation to verify the accuracy of our complete MSE bound. Ideally, at high SNR, the complete bound given by (38) predicts actual estimator performance. We construct a bandlimited signal $f(k) = \sum_{d=1}^{D}(1/d)\sin((\pi k d)/(100) - \phi_d), k = 1 \ldots 100$ where ϕ_d is a fixed phase generated by drawing from a uniform distribution. We chose to use a closed-form expression for f so that that the exact values of the function derivative are available to calculating the FIM which we used to calculate the MSE bound (38). Actual estimator performance is measured by performing 500 MC runs at each value of SNR and averaging the error. The gradient kernel used by the estimator is the filter from [26]. The results of the simulation are shown in the left graph of Fig. 9, which compares the RMSE for the gradient based estimator with both the unbiased CRLB (5), which is just the inverse Fisher information,

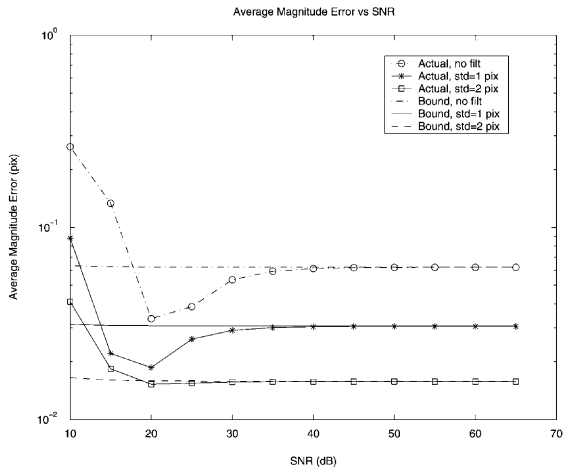

Fig. 10. Predicted and measured average performance measured by (39).

and the full bound (38). The actual estimator performance seems very close to the performance bound predicted by (38) at high SNR. This verifies that the bias function given by (27) is in fact accurate. For low SNR, however, both bounds are overly optimistic. This could be due in part to the approximation made in obtaining the simplified bias function. In general, nonlinear estimation problems suffer from what is known as the threshold effect [12]. This threshold effect is characterized by a significant departure from the CRLB as the SNR degrades. Furthermore, the apparent variation in performance between is predicted by the right graph of Fig. 7 where it suggests that the bias should be less for $v = 1.5 \approx \tilde{v}$ than for $v = 0.5$.

To understand the relationship between bandwidth and performance bound, we plot the expected performance bound for $v = 0.1$ for different values of D (which essentially encodes the bandwidth in the definition of f) in the right graph of Fig. 9. This figure shows the tradeoff between bias and variance as it relates to image bandwidth where D is the percentage of full bandwidth. As mentioned before, energy in higher frequencies tends to increase the Fisher information thereby improving estimator variance, but tends to worsen the affect of bias. Overall, it is apparent that bias dominates the MSE for images with much high frequency spectral energy.

Last, we extend this complete MSE performance bound for the case of 2-D image registration. The equations for bias derived in Appendix I were used to construct the lower bound MSE matrix. To provide a rule of thumb value for expected estimator performance, we use the following performance measure:

$$P = \frac{1}{4V_1V_2}\int_{-V_1}^{V_1}\int_{-V_2}^{V_2}\sqrt{\mathrm{Tr}(\mathrm{MSE}(\mathbf{v}))}\,dv_1\,dv_2. \qquad (39)$$

This provides a measure of the average performance limit for some range of unknown translations. We choose to examine estimator performance for subpixel translation where $\mathbf{v} \in [-1,1]\times[-1,1]$. Fig. 10 shows the performance predicted by (39) and actual performance using MC simulations for the tree image.

The tree image was again shifted synthetically as before using the method of [10]. For each value of SNR, 500 MC runs were performed and averaged to obtain the MSE matrices. To evaluate the improvement using image presmoothing, we apply a 9-tap Gaussian filter with standard deviation of 1 and 2 pixels. To compute the MSE bound, we estimate the spectrum F using the DFT coefficients. To take into account the noise reduction resulting from image presmoothing, we modified noise variance used to compute the FIM by $\tilde{\sigma}^2 = (\sigma^2)/(\sum h^2(k))$ where $h(k)$ are the coefficients of the Gaussian filter. Again, the gradient filter used was from [26]. The expected performance bound seems to approximate the estimator performance for high SNR situations. The estimator performance for SNRs at about 20–40 dB shows unexpected improvement over the high SNR situation. Most likely, this results from the statistical bias present in the estimator for low SNR situations. It was shown in [9] and [11] that the statistical bias for noisy images tends to produce underestimates of translation or negative bias. Since the deterministic bias using the [26] filter is positive for subpixel motion, we deduce that these two biases tend roughly to cancel one another out actually lessening estimator bias at low SNR. As significant low-pass filtering is applied to the image, estimator performance improves dramatically. Basically, the deterministic bias again dominates estimator bias and we have predictably improved estimator performance. This experiment presents the possibility of subpixel image registration accuracy down to almost one hundredth of a pixel for the gradient-based estimator under ideal situations. Again, this experiment correlates well with the results shown in Fig. 3. Thus, we can expect a rule of thumb performance bound limiting the performance of image registration under ideal situations accuracy above one hundredth of a pixel for non iterative gradient-based estimation.

VI. Conclusions and Future Work

This paper derives the fundamental performance limits for image registration using the Cramer–Rao bound to estimate MSE performance. We propose that MSE should be used as a standard performance measure to prevent unfair comparisons between algorithms and motivate statistically accurate analysis. We show that studying this performance bound as it relates to image registration provides much insight into the inherent tradeoffs between estimator variance and bias. We presented analysis as well as experimental evidence suggesting that in general, all estimators are biased. In particular, we derived the accurate expressions for the bias inherent to the very popular class of gradient-based image registration algorithms. Furthermore, in studying the form of this bias for gradient-based estimators, we explained much of the heuristic knowledge accumulated over the years.

This paper provides the foundation for much further work. For instance, we focused on the estimation of image translation. One could extend the analysis to more complex parametric motion models such as affine and bilinear motion. One could hope that this type of analysis would offer guidance to the practitioner choosing between complex motion models for large image regions or simple translational models for smaller or more local motion estimation. This type of performance analysis could be extended to many other situation such as imaging systems with sub-Nyquist sampling rates. Many applications such as image fusion and multiframe image resolution enhancement require accurate image registration as a critical preprocessing step. The performance bounds on image registration are necessary to explain performance limits for these higher level image processing tasks.

With respect to gradient-based motion estimation, we proposed several extensions using our bias expression. For instance, using the complete MSE bound as a cost function, one could optimize various operational parameters such as derivative filters and image presmoothers as in [27]. In general, understanding the bias function should suggest methods for eliminating such bias. While we examined the bias for the case of multiscale iterative gradient-based estimation, the derivation of the complete MSE bound for such iterative methods is yet to be done. The extension of such a derivation would provide insight into the performance of a variety of problems where simplified linearized estimators are improved using iterative methods. We hope that this type of analysis will establish a common framework for evaluating motion estimation and other inverse problems in imaging.

APPENDIX I
COMPLETE 2-D CRLB FOR GRADIENT-BASED ESTIMATION

In this section, we derive the bias equations for the 2-D case similar to Section III and incorporate this bias function into the complete CRLB bound in (4). Here, we use vector notation. Namely, $\mathbf{v} = [v_x, v_y]^T$ and $\underline{\theta} = [\theta_x, \theta_y]^T$ and $\mathbf{m} = [m, n]^T$. Thus, we write the data model as

$$z(\mathbf{m}) = f(\mathbf{m}+\mathbf{v}) - f(\mathbf{m}) + \epsilon(\mathbf{m}). \tag{40}$$

We proceed to derive the bias directly in the frequency domain. The shifted sequence $f(\mathbf{m}+\mathbf{v})$ has a DTFT of $F(\underline{\theta})e^{j(\underline{\theta}^T\mathbf{v})}$ and the DTFT of the data model becomes $Z(\underline{\theta}) = F(\underline{\theta})[e^{j(\underline{\theta}^T\mathbf{v})} - 1] + \xi(\underline{\theta})$. We expand the exponential in a Taylor series $e^{j(\underline{\theta}^T\mathbf{v})} = 1 + j(\underline{\theta}^T\mathbf{v}) - \cdots$ and truncate after the linear term to obtain the formula $Z(\underline{\theta}) = jF(\underline{\theta})\underline{\theta}^T\mathbf{v} + \xi(\underline{\theta})$ from which we obtain the linear estimator

$$\hat{\mathbf{v}} = \mathbf{Q}^{-1}\int |F(\underline{\theta})|^2 j\underline{\theta}Z^*(\underline{\theta})\,d\underline{\theta} \tag{41}$$

where $\mathbf{Q} = \int |F(\underline{\theta})|^2[\underline{\theta}\underline{\theta}^T]\,d\underline{\theta}$.

Similar to the 1-D case, the expected value of the estimate is $E[\hat{\mathbf{v}}] = \mathbf{Q}^{-1}\int |F(\underline{\theta})|^2\underline{\theta}\sin(\underline{\theta}^T\mathbf{v})\,d\underline{\theta}$. To obtain this form, we have made the same simplification as in Section III, wherein the imaginary portion of the integrand is removed as it is an odd function. Thus, we obtain the bias function

$$\mathbf{b}(\mathbf{v}) = \mathbf{Q}^{-1}\int |F(\underline{\theta})|^2\underline{\theta}\sin(\underline{\theta}^T\mathbf{v})\,d\underline{\theta} - \mathbf{v}. \tag{42}$$

To analyze this bias function, we approximate the sinusoid function within the integrand as a truncated Taylor series expansion about $v = 0$ as $\sin(\underline{\theta}^T\mathbf{v}) \approx \underline{\theta}^T\mathbf{v} - (1/6)(\underline{\theta}^T\mathbf{v})^3$. Noting that $\underline{\theta}^T\mathbf{v} = |\mathbf{v}|\underline{\theta}^T\mathbf{n}_\psi$ where $\mathbf{n}_\psi$ is the unit vector $[\cos(\psi)\sin(\psi)]$, we approximate the bias function as

$$\begin{aligned}\mathbf{b}(\mathbf{v}) &\approx \mathbf{Q}^{-1}\int |F(\underline{\theta})|^2\underline{\theta}[\underline{\theta}^T\mathbf{v} - (\underline{\theta}^T\mathbf{v})^3]\,d\underline{\theta} - \mathbf{v}\\ &= \mathbf{v} - \frac{1}{6}\mathbf{Q}^{-1}\int |F(\underline{\theta})|^2\underline{\theta}(\underline{\theta}^T\mathbf{v})^3\,d\underline{\theta} - \mathbf{v}\\ &= -\frac{|\mathbf{v}|^3}{6}\mathbf{Q}^{-1}\int |F(\underline{\theta})|^2\underline{\theta}(\underline{\theta}^T\mathbf{n})^3\,d\underline{\theta}\\ &= -\frac{|\mathbf{v}|^3}{6}\mathbf{Q}^{-1}\mathbf{s}\end{aligned} \tag{43}$$

where $\mathbf{s} = \int |F(\underline{\theta})|^2\underline{\theta}(\underline{\theta}^T\mathbf{n})^3\,d\underline{\theta}$. Thus, the bias behaves as a cubic function of the translation magnitude where the coefficient depends on the spectrum of the image.

As with the 1-D case, in practice we must approximate the gradients using gradient kernels $g_x(\mathbf{m})$ and $g_y(\mathbf{m})$ which have corresponding frequency representations $G_x(\underline{\theta})$ and $G_y(\underline{\theta})$ or in vector notation $\mathbf{G}(\underline{\theta})$. This produces the estimator

$$\hat{\mathbf{v}} = \mathbf{Q}^{-1}\int |F(\underline{\theta})|^2 j\mathbf{G}(\underline{\theta})Z^*(\underline{\theta})\,d\underline{\theta} \tag{44}$$

where now $\mathbf{Q} = \int |F(\underline{\theta})|^2[\mathbf{G}(\underline{\theta})\mathbf{G}(\underline{\theta})^T]\,d\underline{\theta}$. Using the same low-noise assumptions that we made in Section III, we examine only the deterministic bias which is

$$\mathbf{b}(\mathbf{v}) = \mathbf{Q}^{-1}\int |F(\underline{\theta})|^2\mathbf{G}(\underline{\theta})\sin(\underline{\theta}^T\mathbf{v})\,d\underline{\theta} - \mathbf{v}. \tag{45}$$

Using these equations for the bias, we can now derive the full CRLB for gradient-based estimation of 2-D translation. We first note that $(\partial E[\hat{\mathbf{v}}])/(\partial\mathbf{v}) = \mathbf{Q}^{-1}\int |F(\underline{\theta})|^2[\mathbf{G}(\underline{\theta})\underline{\theta}^T]\cos(\underline{\theta}^T\mathbf{v})\,d\underline{\theta} = \mathbf{A}$. Further, we use the following representation $\mathbf{B} = \mathbf{b}(\mathbf{v})\mathbf{b}(\mathbf{v})^T$. Using this equation, we obtain for the full CRLB bound

$$\mathrm{MSE}(\mathbf{v}) \geq \mathbf{A}J^{-1}\mathbf{A}^T + \mathbf{B}. \tag{46}$$

REFERENCES

[1] J. Barron, D. Fleet, S. Beauchemin, and T. Burkitt, "Performance of optical flow techniques," in *CVPR*, vol. 92, 1992, pp. 236–242.

[2] L. Brown, "A survey of image registration techniques," *ACM Comput. Surv.*, vol. 24, no. 4, pp. 325–376, Dec. 1992.

[3] Q. Tian and M. Huhns, "Algorithms for subpixel registration," *Comput. Vis., Graph., and Image Process.*, vol. 35, pp. 220–233, 1986.

[4] G. Jacovitti and G. Scarano, "Discrete time techniques for time delay estimation," *IEEE Trans. Signal Processing*, vol. 41, pp. 525–533, Feb. 1993.

[5] W. F. Walker and G. E. Trahey, "A fundamental limit on the performance of correlation based phase correction and flow estimation techniques," *IEEE Trans. Ultrason., Ferroelect., Freq. Contr.*, vol. 41, pp. 644–654, Sept. 1994.

[6] S. Auerbach and L. Hauser, "Cramer–Rao bound on the image registration accuracy," *Proc. SPIE*, vol. 3163, pp. 117–127, July 1997.

[7] J. Kearney, W. Thompson, and D. Boley, "Optical flow estimation: An error analysis of gradient-based methods with local optimization," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 9, pp. 229–244, Mar. 1987.

[8] J. W. Brandt, "Analysis of bias in gradient-based optical flow estimation," in *IEEE Asilomar Conf. Signals, Systems and Computers*, 1995, pp. 721–725.

[9] C. Fermuller, D. Shulman, and Y. Aloimonos, "The statistics of optical flow," *AP Comput. Vis. Image Understanding*, vol. 82, pp. 1–32, 2001.

[10] C. Q. Davis and D. M. Freeman, "Statistics of subpixel registration algorithms based on spatiotemporal gradients or block matching," *Opt. Eng.*, vol. 37, no. 4, pp. 1290–1298, 1998.

[11] H.-H. Nagel and M. Haag, "Bias-corrected optical flow estimation for road vehicle tracking," in *Proc. Int. Conf. Computer Vision*, 1998, pp. 1006–1011.

[12] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1993.

[13] H. L. V. Trees, *Detection, Estimation, and Modulation Theory, Part I*. New York: Wiley, 1968.

[14] J. Shi and C. Tomasi, "Good features to track," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, June 1994, pp. 593–600.

[15] D. J. Field, "Relations between the statistics of natural images and the response properties of cortical cells," *J. Opt. Soc. Amer. A*, vol. 4, no. 12, pp. 2379–2393, Dec. 1987.

[16] S. Dooley and A. Nandi, "Comparison of discrete subsample time delay estimation methods applied to narrowband signals," *IOP Meas. Sci. Technol.*, vol. 9, pp. 1400–1408, Sept. 1998.
[17] R. Moddemeijer, "On the determination of the position of extremum of sample correlators," *IEEE Trans. Signal Processing*, vol. 39, pp. 216–219, Jan. 1991.
[18] B. K. Horn, *Robot Vision*. Cambridge, MA: MIT, 1986.
[19] B. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *DARPA81*, 1981, pp. 121–130.
[20] J. R. Bergen, P. Anandan, K. J. Hanna, and R. Hingorani, "Hierachical model-based motion estimation," in *Proc. Eur. Conf. Computer Vision*, 1992, pp. 237–252.
[21] D. Robinson and P. Milanfar, "Accuracy and efficiency tradeoffs in using projections for motion estimation," in *Proc. 35th Asilomar Conf. Signals, Systems, and Computers*, Nov. 2001.
[22] ——, "Fast local and global projection-based methods for affine motion estimation," *J. Math. Imag. Vis.*, vol. 18, pp. 35–54, Jan. 2003.
[23] H. S. Stone, M. Orchard, and E.-C. Chang, "Subpixel registration of images," in *Proc. Asilomar Conf. Signals, Systems, and Computers*, Oct. 1999.
[24] A. Nehorai and M. Hawkes, "Performance bounds for estimating vector systems," *IEEE Trans. Signal Processing*, vol. 48, pp. 1737–1749, June 2000.
[25] V. Dvornychenko, "Bounds on (determinisitic) correlation functions with applications to registration," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 5, pp. 206–213, Mar. 1983.
[26] O. Nestares and D. Heeger, "Robust multiresolution alignment of MRI brain volumes," *Magn. Res. Med.*, vol. 43, pp. 705–715, 2000.
[27] M. Elad, P. Teo, and Y. Hel-Or, "On the design of optimal filters for gradient-based motion estimation," *Int. Int. J. Math. Imag. Vis.*, submitted for publication.

Dirk Robinson (S'01) received the B.S. degree in electrical engineering from Calvin College, Grand Rapids, MI, and the M.S. degree in computer engineering from the University of California, Santa Cruz (UCSC), in 1999 and 2001, respectively. He is currently pursuing the Ph.D. degree in electrical engineering at UCSC.

His technical interests include signal and image processing and machine learning.

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

Until 1999, he was a Senior Research Engineer at SRI International, Menlo Park, CA. He is currently Associate Professor of Electrical Engineering, University of California, Santa Cruz. He was a Consulting Assistant Professor of computer science at Stanford University, Stanford, CA, from 1998 to 2000, where he was also a Visiting Associate Professor from June to December 2002. His technical interests are in statistical signal and image processing and inverse problems.

Dr. Milanfar won a National Science Foundation CAREER award in 2000 and he was Associate Editor for the IEEE SIGNAL PROCESSING LETTERS from 1998 to 2001.