FUNDAMENTAL PERFORMANCE LIMITS IN IMAGE REGISTRATION

Dirk Robinson and Peyman Milanfar

Electrical Engineering Department, University of California Santa Cruz
Email: {dirkr, milanfar}@ee.ucsc.edu

ABSTRACT

While many algorithms have been developed to solve the problem of image registration, their performance has typically been evaluated only by comparing one method with another often in an ad-hoc manner. We propose a statistical performance measure based on the mean square error (MSE) and explore the performance bounds using the Cramer-Rao inequality. We show how these performance bounds depend on image content under observation. By analyzing these bounds we provide insight into the inherent tradeoff between bias and variance found in all image registration algorithms. Specifically, we derive a functional expression for the bias inherent in the popular class of gradient-based image registration algorithms.

1. INTRODUCTION

Image registration is a fundamental task in image processing. It is a critical preprocessing step to many modern image processing tasks such as, motion compensated video compression, multiframe image enhancement, as well as many computer vision tasks such as 3-D shape estimation and object identification. The problem of image registration is an example of the more general problem of estimating motion in an image sequence. In general, the problem of motion estimation is one wherein the observed data follows the form

$$z_1(m,n) = f(m,n) + \epsilon_1(m,n) \quad (1)$$

$$z_2(m,n) = f(m - v_1(m,n), n - v_2(m,n)) + \epsilon_2(m,n) \quad (2)$$

where $\epsilon_i(m,n)$ is typically modelled as white Gaussian noise with variance σ^2. We refer to the indices m, n as the sample indices for the sampled functions $f(mT, nT)$ where $f(x,y)$ represents the underlying continuous image. In this paper, we will restrict the class of unknown vector fields to those corresponding to translational motion where $\mathbf{v} = [v_1 \ v_2]^T$ is a constant.

Periodically, there have been fairly comprehensive survey papers describing and comparing the performance of many algorithms [1], [2],[3]. Unfortunately, the benchmarks comparing the performance of such algorithms tend to vary widely and fail to address the problem in a proper statistical framework. This type of performance characterization leaves open the important question of how close the algorithms in question are to being optimal.

We measure performance based on the mean square error (MSE) of a given estimator. Specifically, we will explore the MSE performance bounds using the Cramer-Rao inequality. Surprisingly, while the Cramer-Rao inequality has been widely used in the related field of time delay estimation in communication, Radar, and Sonar, it has not been utilized to understand the problem of motion estimation in general. Developing such performance bounds provides a mechanism for critically comparing the performance of algorithms. Finally, the details of such a performance bound can often provide insight into the problem itself and ideally suggest means for improving performance.

We will show how much of the common knowledge or heuristics used in motion estimation can be understood by studying the performance bounds. In particular, we will explain the inherent tradeoff between bias and variance for several popular estimators. In addition, we will derive expressions for the bias for the very popular gradient-based estimators and show how the bias function is incorporated into the fundamental performance limit. More information about this subject can be found in [4].

2. MSE BOUNDS FOR IMAGE REGISTRATION

In this section, we quantify the MSE performance bounds for registering images by utilizing the Cramer Rao lower bound (CRLB) [5]. The CRLB essentially characterizes the "difficulty" with which a set of parameters can be estimated using a given data model from an information theoretic standpoint. In general, the CRLB provides the lower bound on the mean square error matrix of *any* method used to estimate a deterministic parameter vector $\mathbf{\Phi}$ from a set of data. Specifically, the Cramer-Rao bound on mean square error $MSE(\mathbf{\Phi}) = E[(\hat{\mathbf{\Phi}} - \mathbf{\Phi})(\hat{\mathbf{\Phi}} - \mathbf{\Phi})^T]$ for any estimator

THIS WORK WAS SUPPORTED IN PART BY THE NSF GRANT CCR-9984246.

is given by

$$MSE(\mathbf{\Phi}) \geq \frac{\partial E[\hat{\mathbf{\Phi}}]}{\partial \mathbf{\Phi}} J^{-1}(\mathbf{\Phi}) \frac{\partial E[\hat{\mathbf{\Phi}}]}{\partial \mathbf{\Phi}}^T + (E[\hat{\mathbf{\Phi}}] - \mathbf{\Phi})(E[\hat{\mathbf{\Phi}}] - \mathbf{\Phi})^T \quad (3)$$

where the matrix $J(\mathbf{\Phi})$ is referred to as the Fisher Information matrix, and $E[\hat{\mathbf{\Phi}}] - \mathbf{\Phi}$ represents the bias of the estimator [5]. 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 Fisher Information matrix and the bias function are the two factors controlling the performance limits. If the class of estimators is assumed to be unbiased, the CRLB reduces to its more common form

$$MSE(\mathbf{\Phi}) \geq J^{-1}(\mathbf{\Phi}) \quad (4)$$

In this form, we see that the Fisher Information matrix plays a vital role in predicting estimator variance (MSE).

2.1. Fisher Information for Image Registration

The Fisher Information matrix provides a measure of influence an unknown set of parameters has in producing observable data. Understanding the Fisher Information matrix (FIM) for the problem of image registration provides direct insight into the estimation problem. The FIM for the problem of image registration is given by

$$J(\mathbf{v}) = \frac{1}{\sigma^2} \begin{bmatrix} a_1 & a_2 \\ a_2 & a_3 \end{bmatrix} \quad (5)$$

where

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

where the subscripts indicate the partial derivative in the x, y direction. From a theoretical standpoint, the FIM depends on the partial derivatives of the continuous image. Practically, these partial derivatives can only be approximated from a high resolution image using finite difference methods or by assuming some model for the continuous image based on the image samples.

By looking at the FIM in the Fourier domain, we can see that the FIM is independent of the unknown translation $\mathbf{v}$ and depends only on the image content [4]. The FIM depends on the spectral content as a two dimensional version of the mean angular bandwidth first introduced in [5]. To explore the FIM and hence the performance bound as a function of image content, we introduce the square root of the trace of the inverted FIM $\sqrt{tr(J^{-1})}$ as a concise measure. In the experimental section, we show that as the mean angular bandwidth of the image increases, the trace of J^{-1} decreases suggesting better estimator performance, as expected.

2.2. Bias in Image Registration

To understand the inherent bias associated with any estimator, we begin by looking at the class of maximum likelihood (ML) estimators. An ML estimator of the translational motion estimation problem should asymptotically achieve the CRLB. Most estimators attempt to find the ML estimate by minimizing an objective function of the form

$$Q(\mathbf{v}) = \sum_{m,n} [z_2(m, n) - f(m - v_1, n - v_2)]^2 \quad (6)$$

The objective function only includes z_2 since z_1 does not depend on the unknown shift vector $\mathbf{v}$. Since $f(x - v_1, y - v_2)$ is typically unknown, the function must be replaced with an estimate, most commonly $z_1(m - v_1, n - v_2)$. For very high SNR situations, this estimate will be accurate enough to find the ML estimate.

Even for high SNR situations, however, the ML estimator requires knowledge of the continuous function $f(x, y)$ to attain subpixel accuracy whereas only samples of the function $f(m, n)$ are known. Except for very limited situations, reconstructing a continuous function from a finite number of noisy samples is an ill-posed problem. Thus, all estimators contain implicit prior assumptions about the continuous image being observed, to "regularize" the problem. But, when the underlying functions do not match these assumptions, the estimators systematically produce biased estimates. In the experimental section, we compare the performance of several image registration algorithms and show that each method contains such bias. The bias of an estimator, however, is very difficult to express functionally. In the next section, we will describe the functional form of bias inherent to the class of gradient-based estimators.

3. BIAS IN GRADIENT-BASED ESTIMATORS

One common approach for registering images is the gradient-based method. While the bias inherent to such estimators has been described qualitatively in the past, here we provide a functional description of the bias. To simplify the presentation, we analyze the bias for the 1-D version of image registration. The 2-D analysis is available in [4]. In the 1-D case, we suppose that the measured data is of the form

$$z_1(k) = f(k) + \epsilon_1(k) \quad (7)$$

$$z_2(k) = f(k - v(k)) + \epsilon_2(k) \quad (8)$$

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

The iterative gradient-based method solves this equation by linearizing the function $f(k+v)$ about a point $v = 0$ in a Taylor series. The expansion looks like $f(k+v) - f(k) = vf'(k) + R(k, v)$ where the remainder term in the Taylor expansion R is ignored to produce a linear estimator for the velocity v,

$$\hat{v} = \frac{\sum f'(k)z(k)}{\sum (f'(k))^2} \quad (9)$$

This type of estimator is commonly referred to as the gradient-based, differential method, or the optical flow method.

One source of systematic bias in the gradient-based estimation comes from ignoring the remainder term $R(k, v)$ in the Taylor expansion. The expected value of the estimator (9) is $E[\hat{v}] = v + \frac{\sum f'(k)R(k,v)}{\sum (f'(k))^2}$. Thus, we see that the truncation of the higher order terms introduces a bias into the estimator. Using Parseval's relation and some algebraic manipulation [4] we can rewrite the bias of the gradient-based estimator as

$$b(v) = \frac{\int_{-\pi}^{\pi} |F(\theta)|^2 \left(\theta \sin(v\theta) - v\theta^2\right) d\theta}{\int_{-\pi}^{\pi} |F(\theta)|^2 \theta^2 d\theta} \quad (10)$$

where $|F(\theta)|$ is the amplitude spectrum of the function $f(x)$. This also explains why the bias is small only for small v since $\sin(v\theta) \approx v\theta$.

Another source of error arises in gradient-based estimation because of the need to approximate the gradient or the derivative of the signal function $f(x)$. Thus, instead of using the actual $f'(k)$ in (9), often noisy approximations of the derivatives $f_2'(k) = f(k) * g(k) + \epsilon_2(k) * g(k)$ are used instead where $*$ represents a convolution operation and $g(k)$ represents the derivative operator. We approximate the bias function for the linearized estimator as

$$b(v) \approx \frac{\int_{-\pi}^{\pi} |F(\theta)|^2 \left(G(\theta) \sin(v\theta) - vG^2(\theta)\right) d\theta}{\int_{-\pi}^{\pi} |G(\theta)|^2 |F(\theta)|^2 d\theta}. \quad (11)$$

Where $G(\theta)$ represents the DTFT of $g(k)$ and $\xi_2(\theta)$ represents the DTFT of the noise samples $\epsilon_2(k)$. This suggests that different gradient kernels should produce different estimator bias. Since we have a functional form for the bias of the estimator, we can use the bias to obtain a tighter MSE performance bound using (3). This will be experimentally verified in the next section.

4. RESULTS

In this section we present some experimental results verifying the previous analysis. For instance, Figure 1 shows $\sqrt{tr(J^{-1})}$ vs image bandwidth for a small set of images also shown in Figure (1). The trace of the inverted FIM (sum

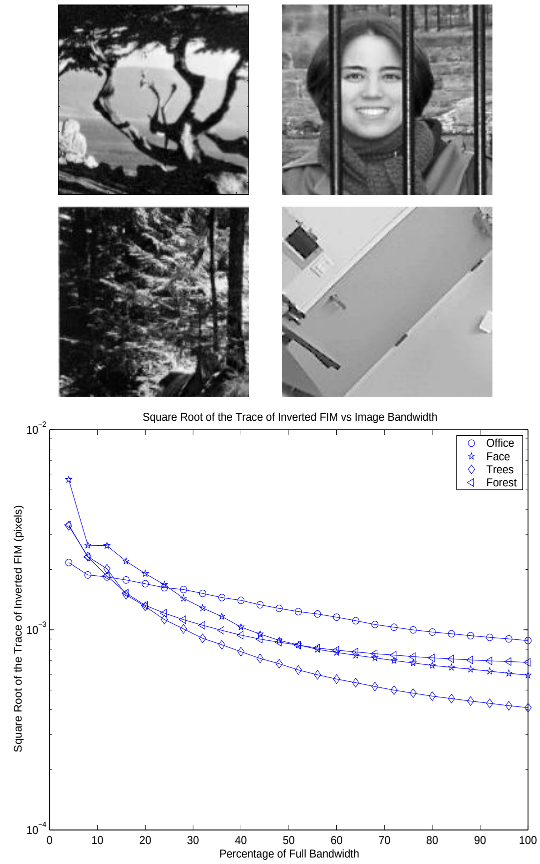

Fig. 1. Images (Trees, Face, Forest, Rotated Office) and $\sqrt{tr(J^{-1})}$ vs image bandwidth

of the eigenvalues) decreases as the image bandwidth and thus mean angular bandwidth increases. This corroborates the general intuition that highly textured images are easier to register. Furthermore, we see from Figure 1 that while the performance limit continues to improve with greater frequency content, the improvement tapers off as the bandwidth increases beyond about a quarter of the full bandwidth for all of the images examined.

Figure 2 shows the performance, in terms of the square root of the trace of the MSE matrix, ($\sqrt{tr(MSE)}$ is a type of RMSE measure), for a variety of approximate ML estimators [3][6][7][8][9]. The pair of images is generated by synthetically shifting the tree image of [1]. The results are obtained using Monte-Carlo simulations at various signal to noise ratios (SNR). As expected, above a certain SNR the performance of all the estimators flattens out while the unbiased CRLB of (4) suggests continued improvement. This flattening of the performance curves is due to the bias inherent to each of the estimators. Looking at the different estimators, it is apparent that often there is not one estima-

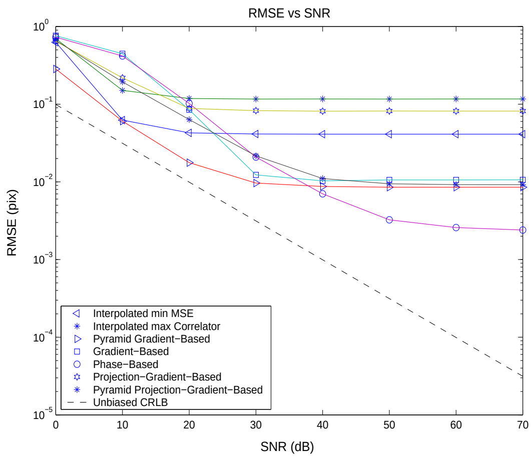

Fig. 2. RMSE Performance vs SNR, $v_1 = .8$, $v_2 = .2$

tor which is uniformly the best for all values of SNR. This type of performance analysis would be useful when selecting an algorithm for an application operating at a specific SNR.

To experimentally verify the bias function of (11) within the context of the complete CRLB of (3), we conducted a Monte-Carlo simulation for a signal with a bandlimited spectrum where $f(k) = \sum_{d=1}^{25} \frac{1}{d} \sin(\frac{\pi k d}{100} - \phi_d)$ and ϕ_d is a fixed phase initially drawn from a uniform distribution. Figure 3 compares the RMSE for the gradient based estimator with both the unbiased CRLB and the CRLB using (3) with the analytically derived bias function. The actual estimator performance seems very close to the performance bound predicted by (3). This verifies not only the bias function (11) but the complete CRLB as well.

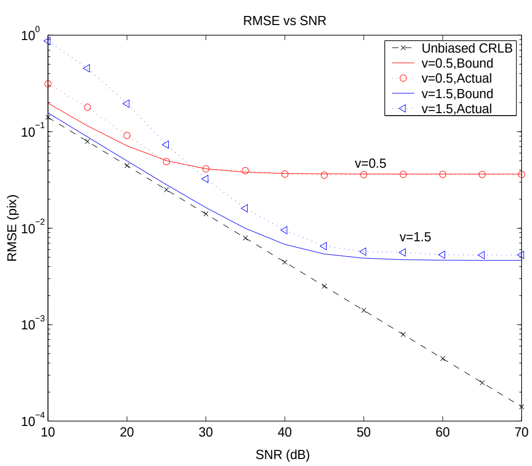

Fig. 3. RMSE vs SNR for Actual and Predicted estimator performance

5. CONCLUSION

In this summary, we have derived and verified the fundamental performance limits for the problem of image registration. In addition, we have derived and verified an expression for the bias for the class of gradient-based estimators. We have shown how these performance bounds can be useful in predicting estimator performance as well as comparing performance between algorithms. Armed with the characterization of the performance bounds, we might consider novel questions such as: What is the best image content for image registration? Or, how can we optimize estimator performance using the structure of the CRLB? These questions remain open for future investigation.

6. REFERENCES

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

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

[3] Q. Tian and M. Huhns, "Algorithms for subpixel registration," *Computer Vision, Graphics, and Image Processing*, vol. 35, pp. 220–233, 1986.

[4] D. Robinson and P. Milanfar, "Fundamental performance limits in image registration," *Submitted to IEEE Transactions on Image Processing*, 2003.

[5] H.L. Van Trees, *Detection, Estimation, and Modulation Theory, Part I*, Wiley, New York, 1968.

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

[7] J.R. Bergen, P. Anandan, K.J. Hanna, and R. Hingorani, "Hierachical model-based motion estimation," *Proceedings European Conference on Computer Vision*, pp. 237–252, 1992.

[8] D. Robinson and P. Milanfar, "Fast local and global projection- based methods for affine motion estimation," *Journal of Mathematical Imaging and Vision*, vol. 18, pp. 35–54, January 2003.

[9] H.S. Stone, M. Orchard, and E. Chang, "Subpixel registration of images," *Proceedings of the 1999 Asilomar Conference on Signals, Systems, and Computers*, October 1999.