MULTIFRAME SHIFT ESTIMATION

THESIS

Stephen A. Bruckart, Second Lieutenant, USAF

AFIT/GE/ENG/06-08

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

AFIT/GE/ENG/06-08

MULTIFRAME SHIFT ESTIMATION

THESIS

Presented to the Faculty

Department of Electrical and Computer Engineering

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

In Partial Fulfillment of the Requirements for the

Degree of Master of Science in Engineering Electrical Engineering

Stephen A. Bruckart, B.S.E.E.

Second Lieutenant, USAF

March 2006

AFIT/GE/ENG/06-08

MULTIFRAME SHIFT ESTIMATION

Stephen A. Bruckart, B.S.E.E.
Second Lieutenant, USAF

Approved:

/signed/

Stephen C. Cain (Chairman)	date

/signed/

Edward A. Watson (Member)	date

/signed/

Richard K. Martin (Member)	date

AFIT/GE/ENG/06-08

Abstract

The purpose of this research was to develop a fundamental framework for a new approach to multiframe translational shift estimation in image processing. This thesis sought to create a new multiframe shift estimator, to theoretically prove and experimentally test key properties of it, and to quantify its performance according to several metrics. The new estimator was modeled successfully and was proven to be an unbiased estimator under certain common image noise conditions. Furthermore its performance was shown to be superior to the cross correlation shift estimator, a robust estimator widely used in similar image processing cases, according to several criteria.

This research effort led to the derivation of a lower bound of estimation performance for the multiframe case. This valuable data analysis tool extends current boundary derivations to include prior information about the random shifting, thereby providing a more precise performance boundary.

To Rachel

Acknowledgments

I would like to express my gratitude to Dr. Steve Cain for his patient support and guidance through this thesis effort. I have learned many valuable professional skills as a direct result of his work. I would also like to thank Dr. Ed Watson and Dr. Rick Martin for their time and insights as committee members.

Stephen A. Bruckart

Table of Contents

List of Figures

MULTIFRAME SHIFT ESTIMATION

I. Introduction

All imaging systems mounted on moving platforms suffer from stabilization issues. Due to mounting imperfections a given imaging system is in a constant state of movement. Such movement is generally unpredictable so it can be modeled as a random vector. This means that the imaging system is pointing at a slightly different scene for each image acquisition period. These *global translational shift differences* among data frames can be represented by horizontal and vertical dimension value pairs.

The process of *image registration* involves estimating these random shifts and then removing them. Registration is an important step for post processing and analysis techniques such as sequential frame playback (video), frame averaging, and *microscanning*. Video playback can be a useful visual aid for identifying scene information in low light or high noise situations. Without registration the playback can appear *jittery* or *shaky*, depending on the degree of random shifts present. Registration can reduce this effect and *stabilize* the objects within the video. Frame averaging is a helpful tool for reducing noise corruption. Photon noise is typically modeled with a Poisson distribution with an average value of the scene without noise. Averaging a set of noisy frames together generally results in a scene estimate with less noise corruption than

any single frame. Without registration the misaligned frames average to a blurry result. Adding the registration step can properly line up the frames before averaging, yielding a result with more detail. Microscanning is a scene estimation technique for acquiring higher spatial frequency content than what the imaging system is able to produce in a single frame. It is particularly useful in situations of aliasing, where the scene contains frequency content that is too high for the imaging system to capture. Sub-pixel registration, combined with interpolation and appropriate frame combination, allows some of the higher frequency content to be extracted from the aliased data frames. Each of these post processing and analysis techniques work only as well as the registration algorithm allows. Accurate registration yields better playback, averaging, and microscanning results than inaccurate registration does, and registration performance is directly dependent on the shift estimation that drives it.

Existing shift estimation algorithms are based on a 2 frame approach. One frame is used as a reference and the other frame is registered to it. The result is a horizontal and vertical shift difference pair. This 2 value vector is generally assigned to the second frame and the reference frame is artificially forced to be the origin. Shift estimates for multiple frames are accomplished by repeating this process for different secondary frames, while always using the same reference frame (generally the first frame in the set). Since noise varies from frame to frame, this method yields estimates that are biased by the noise of the reference frame. Robinson and Milanfar discuss the *Approximate Minimum Average Square Difference*, *Approximate Maximum Direct Correlator*, *Linear Gradient-Based Method*, and *Multiscale (Pyramid) Gradient-Based Method* estimators,

all of which estimate translational shifts between two frames (8:1189-1190). Kuglin and Hines, as well as Gottesfeld Brown examine a *phase correlation* estimation technique which uses Fourier methods to determine the translational offset between two frames (4:163-165; 2:345-348). Frischholz and Spinnler discuss a *template matching* approach to estimating the shift difference between two data frames (1:50-59). Even a cursory survey of image registration literature strongly suggests that translational shift estimation theory is saturated by 2 frame based estimation approaches. While this is a reasonable starting point for multiframe estimation scenarios, it can lead to performance that is highly image dependent (i.e. if the reference frame is unusually corrupted then the estimates might be biased by the image). A more robust approach would seek to minimize such image dependence.

In this paper a new multiframe shift estimation method is presented. The algorithm is capable of sub-pixel estimation and is robust in the presence of noise. Mathematical modeling and justification for these claims are explained, and visual results are shown for further support. Furthermore, the total mean-square error lower performance bound for the multiframe case is derived and compared to the performance of this estimator. This tool offers justification for establishing points of diminishing returns for the number of frames used in this shift estimator.

II. Sensor Models

Imaging System Model

All imaging systems consist of several fundamental pieces. If they are properly mathematically modeled and assembled, the entire system can be simulated. The first piece is the true scene, or *object*, which represents what an ideal imaging system would see (no corruption due to noise, filtering, etc). Light travels from the object to the sensor through the atmosphere. Specific path conditions can vary, so an atmospheric transfer function is often computed. This transfer function acts as a filter on the light, attenuating content at different visual frequencies. The resulting light reaches the optical lens of the sensor. The geometric shape of the lens governs the *optical transfer function* (*OTF*), which low pass filters the visual frequency content. These two transfer functions are often combined (by the properties of convolution) to become a single transfer function (6:45).

After passing through the sensor optics the filtered light falls upon the charge-coupled device (CCD) array. It is comprised of many small detectors which each generate an electrical charge proportional to the amount of light on it. This charge is amplified and read out to a more permanent storage location where it can be accessed for post processing and analysis.

Since this is a digital signal processing problem, it is important to address the issue of proper sampling rates. The light that falls upon the CCD array has been filtered such that there is a cut off spatial frequency f_c above which no content exists. According to the Nyquist sampling criteria, for all existing frequencies to be kept, a sampling interval of at least $1/(2 f_c)$ must be used. Visual frequency content is measured in inverse distance terms, so the Nyquist criteria actually dictates the necessary physical dimensions of the CCD array for proper sampling. More specifically, the Nyquist criteria determines the necessary center to center spacings of the individual collection cells on the CCD array. For the remainder of this thesis it will be assumed that the Nyquist sampling criteria is met, and therefore no aliasing is present.

$$d(x,y) = \left[o(x,y) * h(x,y) \right] + n(x,y) \tag{1}$$

The data generated by an imaging system can be decomposed into several components. Equation 1 shows a very basic relationship between the true scene *o(x,y)* and the collected data *d(x,y)*. In this model *d(x,y)* is based on a photon counting process for a single frequency, and is therefore a matrix of positive integers. The total system noise is represented by *n(x,y)*, and the combined impulse responses of the path and optical system that the light traveled through is given by *h(x,y)*. The convolution of *o(x,y)* and *h(x,y)* represents the spatial frequency attenuation that occurs as a result of the atmospheric transfer function and the OTF, and is often denoted *i(x,y)* with the understanding that *o(x,y)* is the true scene without any loss of spatial frequency information, or *blurring*, and *i(x,y)* is the blurred version of *o(x,y)*. It is assumed that the

Nyquist sampling criteria is met with regards to the collector spacing on the CCD array and the maximum spatial frequency in *i(x,y)*. In other words, the CCD array is able to fully capture all spatial frequencies present in *i(x,y)*, thereby preventing any image aliasing. Equation 2 condenses equation 1 with this abbreviation and adds numerical subscript indices to indicate the data frame index and corresponding noise.

$$d_1(x,y) = i(x,y) + n_1(x,y) \tag{2}$$

A translational shift offset of (α,β) between two frames is accounted for by including shift values in the appropriate dimensions, as shown in equation 3. It is important to note that the noise is different for each frame, as indicated by the subscript notation.

$$d_2(x,y) = i(x-\alpha, y-\beta) + n_2(x,y) \tag{3}$$

Minimum Mean-Square Error Estimator

Having created a basic model for the data, it is necessary to estimate the shift differences between two data frames. The process for finding a minimum mean-square error (*MSE*) estimator involves creating the mean-square error cost function and then minimizing it with respect to the offset between the two frames. It is important to note that minimum MSE is referring to the minimization of error between the estimation and the truth, not to be confused with any other possible definitions of minimum MSE in

estimation theory. Equation 4 shows the mean-square error cost function associated with a CCD array of dimensions MxN, where (δ_x,δ_y) is the offset.

$$C\left(\delta_x,\delta_y\right)=\sum_{x=1}^{M}\sum_{y=1}^{N}\left(d_1\left(x,y\right)-d_2\left(\delta_x+x,\delta_y+y\right)\right)^2 \qquad (4)$$

Expanding this cost function permits further simplification, as shown in equation 5.

$$C\left(\delta_x,\delta_y\right)=\sum_{x=1}^{M}\sum_{y=1}^{N}\left[d_1^2\left(x,y\right)-2d_1\left(x,y\right)d_2\left(\delta_x+x,\delta_y+y\right)+d_2^2\left(\delta_x+x,\delta_y+y\right)\right] \qquad (5)$$

The first term can be eliminated from the minimization process since it is not dependent on (δ_x,δ_y), the parameters used to accomplish the maximization of *C*. The third term is technically dependent on (δ_x,δ_y), however it will not change significantly if we assume circular convolution, small shifts, or that new information entering the frame from the edges has similar intensity to existing information. The middle term remains as the only expression significantly dependent on (δ_x,δ_y). By removing the negative sign and the unnecessary scalar, it can be simplified such that the MMSE estimator is the expression that maximizes equation 6.

$$C_2\left(\delta_x,\delta_y\right)=\sum_{x=1}^{M}\sum_{y=1}^{N}d_1\left(x,y\right)d_2\left(\delta_x+x,\delta_y+y\right) \qquad (6)$$

By substituting equations 2 and 3, equation 6 can be rewritten as equation 7.

$$\begin{aligned} C_2\left(\delta_x,\delta_y\right) = \sum_{x=1}^{M}\sum_{y=1}^{N}[\ldots \\ & i\left(x,y\right)i\left(\delta_x+x-\alpha,\delta_y+y-\beta\right) \\ & +i\left(x,y\right)n_2\left(\delta_x+x,\delta_y+y\right) \\ & +i\left(\delta_x+x-\alpha,\delta_y+y-\beta\right)n_1\left(x,y\right) \\ & +n_1\left(x,y\right)n_2\left(\delta_x+x,\delta_y+y\right)] \end{aligned} \tag{7}$$

Assuming that the noise is spatially uncorrelated and has a zero mean, such as Gaussian noise, the expectation operator further simplifies the cost function. Equations 8, 9, and 10 are a result of this assumption.

$$E\left[i\left(x,y\right)n_2\left(\delta_x+x,\delta_y+y\right)\right] = i\left(x,y\right)E\left[n_2\left(\delta_x+x,\delta_y+y\right)\right] = 0 \tag{8}$$

$$E\left[i\left(\delta_x+x-\alpha,\delta_y+y-\beta\right)n_1\left(x,y\right)\right] = i\left(\delta_x+x-\alpha,\delta_y+y-\beta\right)E\left[n_1\left(x,y\right)\right] = 0 \tag{9}$$

$$E\left[n_1\left(x,y\right)n_2\left(\delta_x+x,\delta_y+y\right)\right] = E\left[n_1\left(x,y\right)\right]E\left[n_2\left(\delta_x+x,\delta_y+y\right)\right] = 0 \tag{10}$$

Substituting equations 8, 9, and 10 into equation 7 yields equation 11.

$$E\left[C_2\left(\delta_x,\delta_y\right)\right] = \sum_{x=1}^{M}\sum_{y=1}^{N}E\left[i\left(x,y\right)i\left(\delta_x+x-\alpha,\delta_y+y-\beta\right)\right] \tag{11}$$

This is merely the autocorrelation of the image with lags $(\delta_x-\alpha,\delta_y-\beta)$, and is maximized when $\delta_x=\alpha$ and $\delta_y=\beta$. Therefore, under the assumption that the noise is uncorrelated and zero mean, the MMSE shift estimate is the set of parameters that maximizes the 2 dimensional cross correlation of the two data frames. Furthermore the expected value of the cross correlation of $d_1(x,y)$ and $d_2(\delta_x+x,\delta_y+y)$ equals the autocorrelation of the underlying true scene with lags $(\delta_x-\alpha,\delta_y-\beta)$, which means that the expected value of the estimator is the true shifts.

Probability Based Image Modeling

Digital images can be modeled as a random process, where each pixel is a random variable. For a particular scene all pixels are governed by the same general probability distribution form, but each pixel distribution has unique parameters. These parameters are a function of the underlying noiseless true scene. More specifically, for each pixel, the mean of the distribution is the value of the corresponding true scene pixel. Modeling an image as a joint distribution function provides a way to compute the total probability of collecting a particular data frame from a true scene. Assuming that the pixels of an image are independent random variables, the probability of receiving a particular data frame is the product of all the individual probabilities of the pixels within that data frame.

Two common distributions often used in this modeling approach are the Poisson distribution and the Gaussian distribution. Photon arrival is a Poisson random process.

Therefore in situations where most of the noise is a result of the light collection process, equation 12, which defines the total probability of receiving data frame *D* from true scene *i* with Poisson noise and random translational shifts of *(α,β)*, is a reasonable distribution choice.

$$P\left(d = D \mid \alpha, \beta\right) = \prod_x \prod_y \frac{i\left(x, y\right)^{D\left(x-\alpha, y-\beta\right)} e^{-i\left(x,y\right)}}{D\left(x-\alpha, y-\beta\right)!} \tag{12}$$

At light levels where the Poisson distribution becomes nearly symmetric (i.e. more than 100 photons per pixel) the shape of the Poisson distribution becomes similar to that of the Gaussian distribution. Therefore the Gaussian distribution is often used to approximate Poisson noise. If light levels are somewhat constant throughout the image then a single variance parameter can be used to approximate the individual pixel standard deviations. Equation 13 shows the total probability of receiving data frame *D* from true scene *i*, where σ_n is the standard deviation of the noise and *(α,β)* are the random translational shifts.

$$P\left(d = D \mid \alpha, \beta\right) = \prod_x \prod_y \frac{1}{\sqrt{2\pi}\sigma_n} e^{-\frac{1}{2\sigma_n^2}\left(D\left(x-\alpha, y-\beta\right)-i\left(x,y\right)\right)^2} \tag{13}$$

Since the likelihood of the Gaussian distribution is maximized at the mean, minimizing the MSE maximizes the likelihood of the Gaussian model. Therefore the estimator in equation 11 is the maximum likelihood estimator in the presence of Gaussian noise. Since the expected value of the data is the true scene, one of the data frames can be used

as a true scene estimate in the Gaussian model, and the other data frame can be used as the data with a shift relative to the scene estimate. Therefore, since the cross correlator estimates the true shifts (on average) for zero mean noise such as this, it minimizes the cost function in equation 4 and is therefore an unbiased shift estimator in the presence of Gaussian noise.

III. Multiframe Shift Estimator Algorithm

Existing multiframe shift estimation algorithms are effectively only a series of 2 frame shift estimators. For a set of K data frames, a single reference data frame is selected and the remaining K-1 data frames are registered to it. The performance of this method is highly dependent on the reference data frame chosen, and therefore can yield poor results if the reference data frame is unusually corrupted by noise. Furthermore, since each data frame is uniquely corrupted, any reference data frame chosen will skew performance by the noise present in that frame. This approach to multiframe shift estimation fails to use all available information. A better approach is one that exploits the fact that shift estimates will vary depending on which frame is used as a reference data frame.

Initial Approach To Solution

In order to compute meaningful shift estimates, all estimators must choose an origin to provide a common reference for the shift values. 2 frame based multiframe estimators force the reference data frame to be located at the origin and yield estimates (for the remaining data frames) that are relative to the reference data frame. Multiple sets of shift estimates, where each set contains shift estimates for the entire data set using a different reference data frame, can be compared by aligning the sets to the same reference point. This can be accomplished by normalizing each set to be zero mean in both shifting

dimensions. Figure 1 shows several examples of this for 10 frames of data, using the first and second frames as the references with the cross correlation shift estimator.

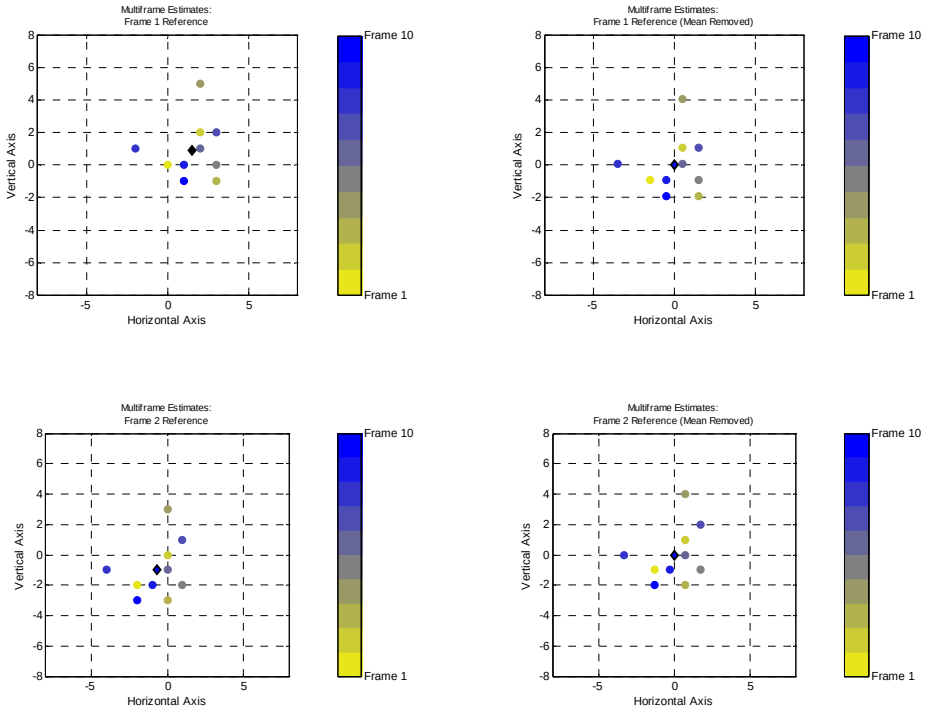

Figure 1. Examples of 2 frame based multiframe shift estimation using the cross correlation shift estimator where K equals 10. The black diamond indicates the mean of the scatter.

If this technique is repeated for the same data set, using a different reference data frame each time, two K by K matrices are populated (one for the horizontal shift estimates and the other for the vertical shift estimates) where K is the total number of data frames. Continuing from the example in figure 1, figure 2 demonstrates the resulting 10 sets of shift estimates (with respective means removed) plotted together. Here the small differences in shift estimates (due to using different reference data frames) are visible.

Since each data frame is uniquely corrupted by noise, rendering a slightly different set of relative shift estimates when used as the reference data frame, the shift estimates form clusters surrounding the true shift values.

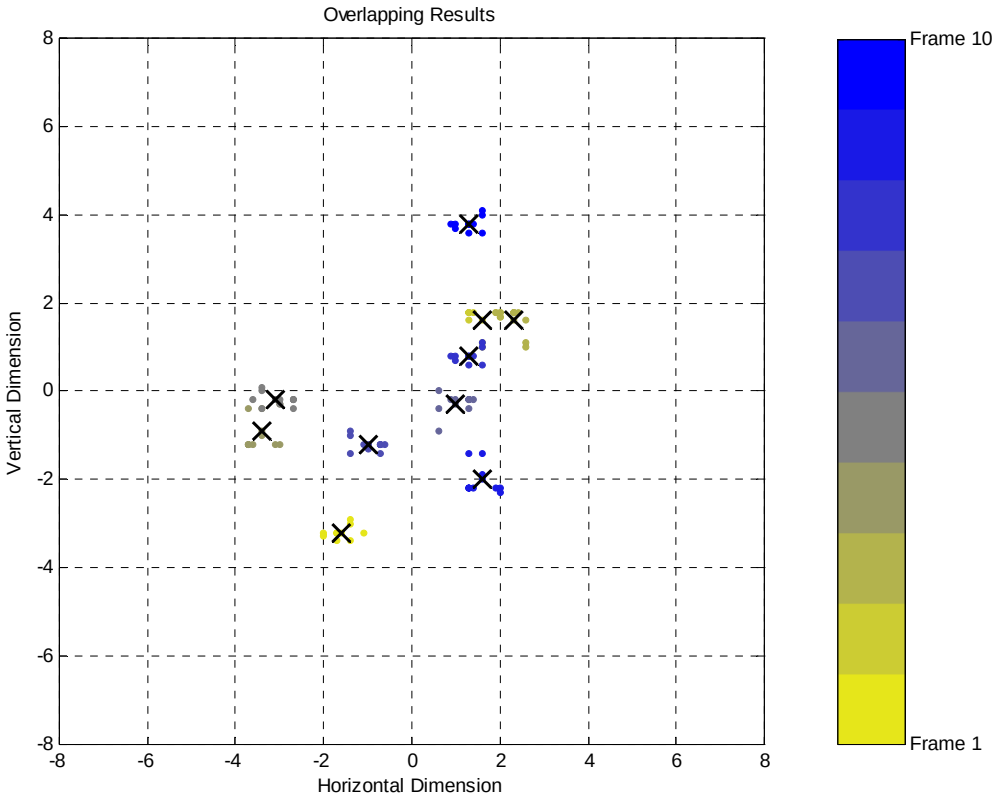

Figure 2. Example of K 2 frame based multiframe shift estimations where a different reference data frame is used for each estimation, and where K equals 10. Bold x characters indicate the mean of each cluster of shift estimates.

As the number of data frames used in this technique increases, the clusters become more densely populated. The new shift estimates are the means of the clusters, and have a mean of zero in both shifting dimensions.

This new multiframe shift estimator is practically implemented by using an unbiased 2 frame shift estimator to estimate shift differences between all combinations of 2 data frames. As described before, this can also be thought of as using a 2 frame based multiframe shift estimator K^2 times for a data set of K frames, using a different reference data frame each time. Each set of shift estimates (estimates for the entire data set using a particular reference data frame) is normalized to have a mean at the origin. Graphically this yields K clusters of K estimates each. The mean of each cluster is stored as the final estimate.

Mathematical Model

The mathematical model for this new multiframe shift estimator begins with an unbiased 2 frame shift estimator capable of estimating vertical and horizontal translational shifts between the 2 frames. This estimator is referred to as the *initial estimator*, and is modeled as *M(i,j)=(vertical estimate,horizontal estimate)* or $M_1(i,j)$*=(vertical estimate)* and $M_2(i,j)$*=(horizontal estimate)*. Inputs *(i,j)* are the indices of the two data frames being registered, where *i* is the (index number of the) reference data frame and *j* is the (index number of the) data frame being registered to it. The vertical and horizontal estimate outputs are each real numbers with a unit of one pixel (i.e. *vertical shift = 1.2* indicates a vertical shift difference of 1.2 pixels) and a maximum absolute value of one half the total number of pixels per dimension (i.e. for data with dimensions 256 by 256 a vertical or horizontal shift estimate of 350 is not possible, since

a shift of that size would indicate that there is no overlapping content between the two data frames being registered). Output sign convention is slightly different than the standard 2 dimensional sign convention, and is described in Figure 3.

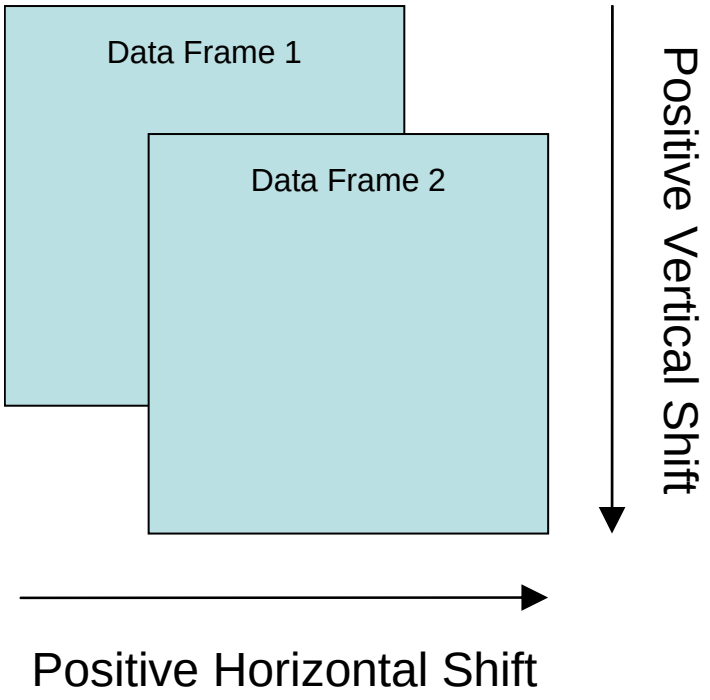

Figure 3. Sign convention of the new multiframe shift estimator model.

The initial estimator performs shift estimates for all combinations of 2 data frames in the set of K data frames. This is efficiently represented in equation 14 where shift estimates for each dimension populate a K by K matrix. Each row represents the result of a 2 frame based multiframe shift estimation.

$$\begin{bmatrix} M_1(1,1) & M_1(1,2) & \cdots & M_1(1,K) \\ M_1(2,1) & M_1(2,2) & & M_1(2,K) \\ \vdots & & \ddots & \vdots \\ M_1(K,1) & M_1(K,2) & \cdots & M_1(K,K) \end{bmatrix} \\ \begin{bmatrix} M_2(1,1) & M_2(1,2) & \cdots & M_2(1,K) \\ M_2(2,1) & M_2(2,2) & & M_2(2,K) \\ \vdots & & \ddots & \vdots \\ M_2(K,1) & M_2(K,2) & \cdots & M_2(K,K) \end{bmatrix} \tag{14}$$

The mean of each row is subtracted from all members of its row, as shown in equation 15.

$$\begin{bmatrix} \left(M_{\dim}(1,1) - \frac{1}{K}\sum_{i=1}^{K} M_{\dim}(1,i)\right) & \cdots & \left(M_{\dim}(1,K) - \frac{1}{K}\sum_{i=1}^{K} M_{\dim}(1,i)\right) \\ \vdots & \ddots & \vdots \\ \left(M_{\dim}(K,1) - \frac{1}{K}\sum_{i=1}^{K} M_{\dim}(K,i)\right) & \cdots & \left(M_{\dim}(K,K) - \frac{1}{K}\sum_{i=1}^{K} M_{\dim}(K,i)\right) \end{bmatrix} \tag{15}$$

The mean of each cluster is acquired by computing the mean of each column of equation 15. This results in 2 vectors of length K, representing the final shift estimates for all K data frames, where the mean of the shift estimates for both dimensions is zero, and where each vector contains the shift estimates for a particular dimension. Equation 16 shows this step and condenses the operation into a more compact equation.

$\tilde{M}_{\dim,k}\left(dataframe_1,...,dataframe_K\right)$ represents the new multiframe shift estimator, where *dim* is the dimension of operation, *k* refers to the data frame whose shift is being estimated, and $dataframe_1,\ldots,dataframe_K$ are the estimator inputs.

$$
\begin{aligned}
&\tilde{M}_{\dim,k}\left(dataframe_1,...,dataframe_K\right) \\
&\quad=\left(\frac{1}{K}\sum_{i=1}^{K}\left(M_{\dim}\left(i,k\right)-\frac{1}{K}\sum_{j=1}^{K}M_{\dim}\left(i,j\right)\right)\right) \\
&\quad=\frac{1}{K^2}\sum_{i=1}^{K}\left(\left(K\cdot M_{\dim}\left(i,k\right)\right)-\sum_{j=1}^{K}M_{\dim}\left(i,j\right)\right) \\
&\quad=\frac{1}{K^2}\begin{bmatrix}\left(K\cdot M_{\dim}\left(1,k\right)\right)-\sum_{j=1}^{K}M_{\dim}\left(1,j\right)\\ +\left(K\cdot M_{\dim}\left(2,k\right)\right)-\sum_{j=1}^{K}M_{\dim}\left(2,j\right)\\ +\cdots\\ +\left(K\cdot M_{\dim}\left(K,k\right)\right)-\sum_{j=1}^{K}M_{\dim}\left(K,j\right)\end{bmatrix} \\
&\quad=\frac{1}{K^2}\left[\left(K\sum_{i=1}^{K}M_{\dim}\left(i,k\right)\right)-\left(\sum_{i=1}^{K}\sum_{j=1}^{K}M_{\dim}\left(i,j\right)\right)\right]
\end{aligned}
\tag{16}
$$

Further simplifications can be made an important property is true of the initial estimator. This property (equation 17) states that, for shift difference estimation between 2 data frames, switching the reference data frame results only in a sign change of the estimate. If this property is true then the matrices (populated by the initial estimator) in equation 14 gain the property such that each is equal to the negative of its own transpose. If equation 17 is true then equation 18 is also true. It states that the shift estimate of a data frame with itself yields zero in both dimensions. This means that the diagonals of the matrices in equation 14 equal zero.

$$
M\left(i,j\right)=-M\left(j,i\right) \tag{17}
$$

$$M(i,i) = 0 \tag{18}$$

Finally, if equation 17 is true then a corollary property becomes true. Equation 19 states that the sum of all values in each matrix is zero.

$$\sum_{i=1}^{K}\sum_{j=1}^{K} M_{\dim}(i,j) = 0 \tag{19}$$

Substituting this equality into equation 16 simplifies the expression for the final shift estimates to be as described in equation 20.

$$\begin{aligned}
&\tilde{M}_{\dim,k}\left(dataframe_1,...,dataframe_K\right) \\
&\quad = \frac{1}{K^2}\left[\left(K\sum_{i=1}^{K} M_{\dim}(i,k)\right) - \left(\sum_{i=1}^{K}\sum_{j=1}^{K} M_{\dim}(i,j)\right)\right] \\
&\quad = \frac{1}{K^2}\left[\left(K\sum_{i=1}^{K} M_{\dim}(i,k)\right) - 0\right] \\
&\quad = \frac{1}{K}\sum_{i=1}^{K} M_{\dim}(i,k)
\end{aligned} \tag{20}$$

This means the initial shift estimates do not need to be normalized prior to averaging the columns.

This new multiframe shift estimator is independent of any underlying random shifting model. If prior information is known about the type of random shifting occurring, it can be factored into the choice of initial estimator for this new multiframe shift estimator.

Speed Of Implementation

Implementation of the new multiframe shift estimator involves selecting an unbiased 2 frame initial shift estimator, computing the shift differences between all combinations of 2 data frames, and averaging all results with respect to dimensionality and frame index. If equation 17 holds then significant speed improvements can be made. It has been shown that the cross correlation operation is an unbiased 2 frame shift estimator and it satisfies equation 17. Based on its merits, the cross correlation shift estimator is an excellent candidate for the initial estimator.

Since equation 17 is true of the cross correlator, implementation becomes much faster. Rather than performing K^2 estimations with the initial estimator, only the upper triangular portion of the two initial shift estimate matrices must be filled. The symmetry in the matrices reduces the number of estimations from K^2 to $(K^2-K)/2$ because the diagonal of each matrix equals zero, and because only the upper (or lower) triangle of the matrices needs to be populated since the opposing triangle equals its negative transpose. After populating the initial shift estimate matrices, final estimates are computed by averaging the columns.

Unbiased Property

For this thesis bias is defined as the mean-square error (MSE) of the estimator without noise. This value is computed by centering the mean of the shift estimates on the mean of the true shift values, calculating the distance squared between each true shift value and it's corresponding shift estimate (for each dimension), summing these squared distances, and dividing the sums (one for each dimension) by the number of data frames used. Equation 21 models this metric as $MSE_{\dim,K}$ where K is the number of data frames used, and where $trueshift_{\dim,i}$ is comprised of 2 vectors of length K containing the true shift values for each dimension and data frame.

$$MSE_{\dim,K} = \frac{\sum_{i=1}^{K}\left(\left(trueshift_{\dim,i} - \frac{\sum_{j=1}^{K} trueshift_{\dim,j}}{K}\right) - \left(\tilde{M}_{\dim,i} - \frac{\sum_{j=1}^{K}\tilde{M}_{\dim,j}}{K}\right)\right)^2}{K} \quad (21)$$

Equation 20 shows that the multiframe estimate is an average of unbiased estimates of the shifts. If the initial estimator is unbiased then the new multiframe shift estimator is also unbiased. The cross correlation estimator is unbiased. Therefore, if the cross correlation estimator is used as the initial estimator, the new multiframe estimator is unbiased. Furthermore, as K gets larger, the MSE of the estimates compared to the true

shifts approaches zero. This property can be demonstrated experimentally by simulating data, estimating the random shifts, comparing the truth to the estimates to acquire MSE information for each experiment, and averaging the MSE information over all the experiments to determine if any bias is present in the estimator.

The first step in simulating data is to select a true scene. This matrix of positive real integers is the uncorrupted Nyquist sampled digital representation of a scene. It is then filtered by the average atmospheric transfer function and optical transfer function using fast Fourier techniques. The result is randomly shifted according to the 2 dimensional Gaussian distribution with a mean at the origin and with variances of 2 pixels for both dimensions. Fast Fourier techniques are used to perform circular shifts and the result is cropped to smaller dimensions to eliminate unrealistic border effects. For this set of experiments a simulated point source is used because it has high contrast (higher contrast tends to yield better shift estimates). The true scene has dimensions of 64 pixels by 64 pixels and the final simulated data frame has dimensions of 32 pixels by 32 pixels. Figure 4 shows the true scene cropped to 32 pixels by 32 pixels but without any corruption. Using this method of simulation, 500 unique sets of data were generated for each K between 3 and 20, totaling 103,500 uniquely shifted data frames. The new multiframe shift estimator, using the cross correlation initial estimator, generated shift estimates for each of the 500 data sets per K. The estimated shifts were compared to the true shift values to compute the MSE metric, and the 500 MSE metrics per K were averaged. Figure 5 shows the average MSE (per dimension) as a function of the number of data frames used.

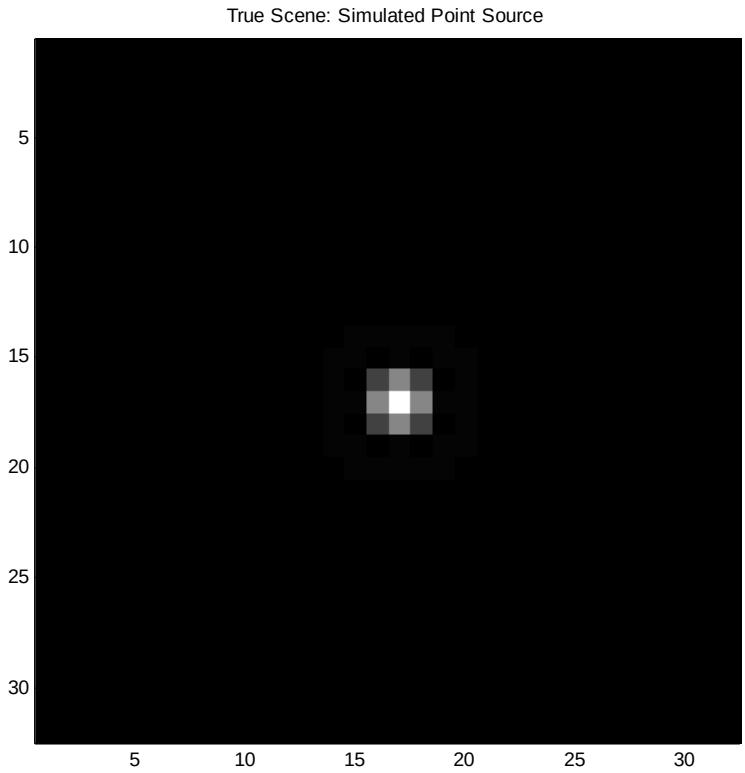

Figure 4. Simulated point source used to generate data and experimentally measure any bias in new multiframe shift estimator.

It also shows the best fitting curve of the form $MSE(K) = a*e^{-b*K} + c*e^{-d*K}$ for each data set to demonstrate convergence toward zero. This functional form is used because it closely follows the trend of the data and is a simple form to mathematically manipulate. As the independent variable increases the MSE converges to zero. Therefore, with no noise present, the estimator is unbiased as *K* grows large.

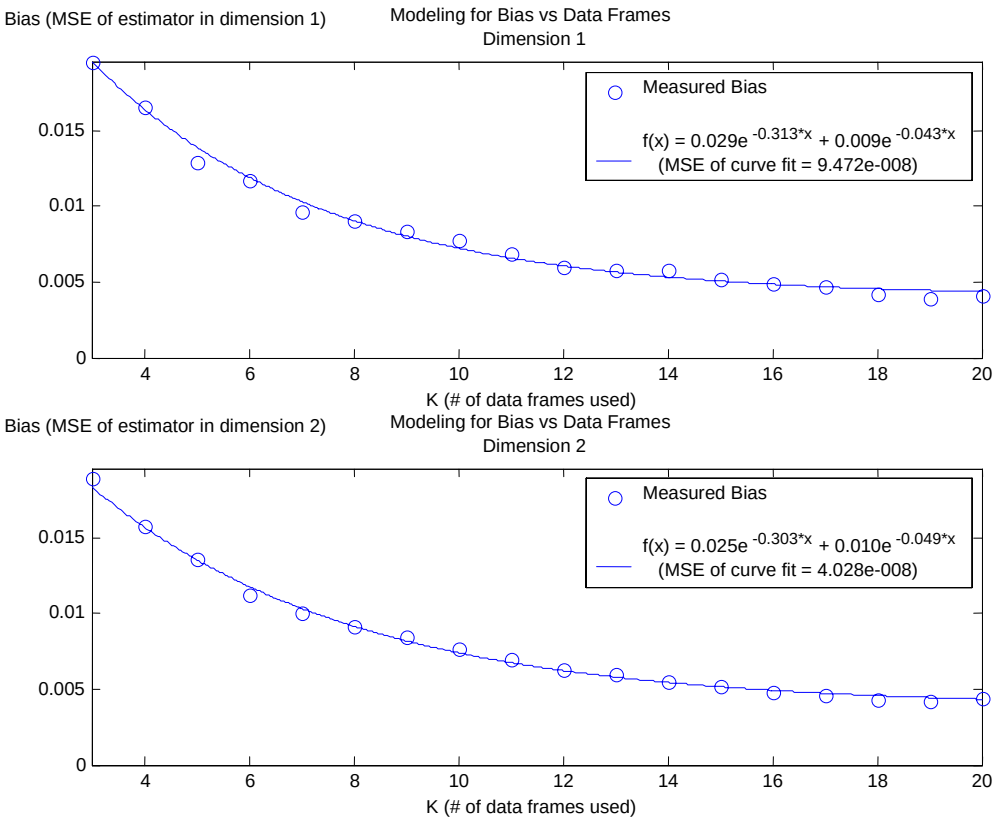

Figure 5. Simulated point source used to generate data and experimentally measure any bias in new multiframe shift estimator.

IV. Comparison With Existing Algorithms

Lower Bound Of Performance

A well documented and broadly respected performance limit in the field of estimation theory is the total mean-square error of the estimator. This is the lowest possible total mean-square error that any unbiased estimator can achieve for a given distribution model. It can be used to show that the lower bound on shift error improves with adding frames, and it can infer the relative performance improvement obtainable. Van Trees derives it by modifying the nonrandom parameter boundary derivation such that prior information is accounted for. Equation 22 shows the total information matrix J_T where J_D is the information matrix based on the data and derived in the nonrandom parameter case, and J_P is the prior information. J_D is also known as *Fisher's information matrix* (*FIM)* which is used to compute the Cramer Rao lower bound (CRLB) of variance.

$$\mathbf{J}_T = \mathbf{J}_D + \mathbf{J}_P \tag{22}$$

Van Trees defines the elements of J_P as

$$J_{P_{ij}} = -E\left[\frac{\partial^2 \ln p_{\mathbf{a}}(\mathbf{A})}{\partial A_i \partial A_j}\right] \tag{23}$$

and states that the "diagonal elements in the inverse of the total information matrix are lower bounds on the corresponding mean-square errors" (9:84) as shown in equation 24, where J_T^{ii} is the i^{th} diagonal element of the inverse of $\mathbf{J}_T$.

$$E\left[a_{\varepsilon_i}^2\right] \geq J_T^{ii} \tag{24}$$

In the multiframe shift estimation problem the lower performance bound can be determined as a function of the number of data frames used, the parameters of the random shifting model, and the gradient information of the true scene. Since there are two dimensions of random translational movement it is necessary to use the multiple parameter model of $\mathbf{J}_D$. The elements of $\mathbf{J}_D$ are defined in equation 25 (9:79).

$$J_{D_{ij}} = -E\left[\frac{\partial^2 \ln\left(p_{\mathbf{r}|\mathbf{a}}\left(\mathbf{R}\mid\mathbf{A}\right)\right)}{\partial A_i \partial A_j}\right] \tag{25}$$

It is assumed that any necessary conditions for equation 24 to be true are met (9:79-80).

The random translational shifts of a set of data can be modeled as the joint likelihood function $P(\underline{\alpha},\underline{\beta}|D)$ where $\underline{\alpha}$ and $\underline{\beta}$ represent the horizontal and vertical random translational shifts, respectively, and D is the true scene. This expression can be rewritten by applying Bayes theorem, as shown in equation 26.

$$P\left(\underline{\alpha},\underline{\beta}\mid D\right)=\frac{P\left(D\mid\underline{\alpha},\underline{\beta}\right)P\left(\underline{\alpha},\underline{\beta}\right)}{P\left(D\right)} \tag{26}$$

The natural log operator further simplifies this function, as shown in equation 27.

$$\ln\left(\frac{P\left(D\mid\underline{\alpha},\underline{\beta}\right)P\left(\underline{\alpha},\underline{\beta}\right)}{P\left(D\right)}\right)=\ln\left(P\left(D\mid\underline{\alpha},\underline{\beta}\right)\right)+\ln\left(P\left(\underline{\alpha},\underline{\beta}\right)\right)-\ln\left(P\left(D\right)\right) \tag{27}$$

In the first term the translational shifts are not treated as random variables but rather parameters of the true scene, which is modeled as a random matrix. More specifically, the first term models the true scene as the average value of random noise. The FIM is computed from this term, which means that the lower bound of the variance is image dependent. The Poisson distribution becomes nearly symmetric when its parameter becomes large, and begins to take on a shape similar to the Gaussian distribution. If there is sufficient light in a particular pixel (i.e. more than 100 photons) then the Gaussian distribution can be used to approximate the noise. Therefore the likelihood function of the first term is modeled as follows:

$$P\left(D\mid\underline{\alpha},\underline{\beta}\right)=\prod_{k=1}^{K}\frac{1}{\sqrt{2\pi}\sigma_n}\exp\left(\frac{\sum_{m,n}\left(D_k\left(m,n\right)-f\left(m-\alpha_k,n-\beta_k\right)\right)^2}{-2\sigma_n^2}\right) \tag{28}$$

σ_n is the standard deviation of the noise, D_k is data frame k, and $f(m-\alpha_k,n-\beta_k)$ is the true scene shifted by the true shift values of data frame k. The log likelihood expression is

$$\ln\left(P\left(D\,|\,\underline{\alpha},\underline{\beta}\right)\right) = -K\ln\left(\sqrt{2\pi}\sigma_n\right) - \frac{1}{2\sigma_n^2}\sum_{k=1}^{K}\left[\sum_{m,n}\left(D_k(m,n) - f\left(m-\alpha_k, n-\beta_k\right)\right)^2\right] \tag{29}$$

where the first term is just a constant. The partial derivative operator is also known as the *gradient*, and must be broken into 3 steps in order to fully account for all combinations of the shift parameters. In all three cases, the constant term disappears and the gradient of the second term is all that remains, resulting in the following three equations:

$$\frac{\partial^2}{\partial\alpha_i\partial\alpha_j}\ln\left(P\left(D\,|\,\underline{\alpha},\underline{\beta}\right)\right) = \frac{1}{\sigma_n^2}\sum_{k=1}^{K}\left[\sum_{m,n}\left(\left(D_k-\tilde{f}\right)\left(\frac{\partial^2\tilde{f}}{\partial\alpha_i\partial\alpha_j}\right) - \left(\frac{\partial\tilde{f}}{\partial\alpha_i}\right)\left(\frac{\partial\tilde{f}}{\partial\alpha_j}\right)\right)\right] \tag{30}$$

$$\frac{\partial^2}{\partial\beta_i\partial\beta_j}\ln\left(P\left(D\,|\,\underline{\alpha},\underline{\beta}\right)\right) = \frac{1}{\sigma_n^2}\sum_{k=1}^{K}\left[\sum_{m,n}\left(\left(D_k-\tilde{f}\right)\left(\frac{\partial^2\tilde{f}}{\partial\beta_i\partial\beta_j}\right) - \left(\frac{\partial\tilde{f}}{\partial\beta_i}\right)\left(\frac{\partial\tilde{f}}{\partial\beta_j}\right)\right)\right] \tag{31}$$

$$\frac{\partial^2}{\partial\alpha_i\partial\beta_j}\ln\left(P\left(D\,|\,\underline{\alpha},\underline{\beta}\right)\right) = \frac{1}{\sigma_n^2}\sum_{k=1}^{K}\left[\sum_{m,n}\left(\left(D_k-\tilde{f}\right)\left(\frac{\partial^2\tilde{f}}{\partial\alpha_i\partial\beta_j}\right)-\left(\frac{\partial\tilde{f}}{\partial\alpha_i}\right)\left(\frac{\partial\tilde{f}}{\partial\beta_j}\right)\right)\right] \tag{32}$$

For conciseness of notation the substitutions $\tilde{f} = f\left(m-\alpha_k, n-\beta_k\right)$ and $D_k = D_k\left(m,n\right)$ are made. Since the shift values are treated as parameters, and since they are different for each data frame, the gradient operations are expressed in a more general form. The expression greatly simplifies when the expectation operator is applied to the log likelihood function. This is because the only random element of the function is D_k, which has a mean of the true scene. Therefore the expected value of $\left(D_k-\tilde{f}\right)$ is zero, leaving only the gradient information of the true scene.

$$\begin{aligned}
&-E\left[\frac{\partial^2}{\partial\alpha_i\partial\alpha_j}\ln\left(P\left(D\,|\,\underline{\alpha},\underline{\beta}\right)\right)\right] \\
&\quad= -E\left[\frac{1}{\sigma_n^2}\sum_{k=1}^{K}\left[\sum_{m,n}\left(\left(D_k-\tilde{f}\right)\left(\frac{\partial^2\tilde{f}}{\partial\alpha_i\partial\alpha_j}\right)-\left(\frac{\partial\tilde{f}}{\partial\alpha_i}\right)\left(\frac{\partial\tilde{f}}{\partial\alpha_j}\right)\right)\right]\right] \\
&\quad= \frac{1}{\sigma_n^2}\sum_{k=1}^{K}\left[\sum_{m,n}\left(-E\left[\left(D_k-\tilde{f}\right)\left(\frac{\partial^2\tilde{f}}{\partial\alpha_i\partial\alpha_j}\right)\right]+E\left[\left(\frac{\partial\tilde{f}}{\partial\alpha_i}\right)\left(\frac{\partial\tilde{f}}{\partial\alpha_j}\right)\right]\right)\right] \\
&\quad= \frac{K}{\sigma_n^2}\sum_{m,n}\left[\left(\frac{\partial\tilde{f}}{\partial\alpha_i}\right)\left(\frac{\partial\tilde{f}}{\partial\alpha_j}\right)\right]
\end{aligned} \tag{33}$$

$$-E\left[\frac{\partial^2}{\partial\beta_i\partial\beta_j}\ln\left(P\left(D\,|\,\underline{\alpha},\underline{\beta}\right)\right)\right] = \frac{K}{\sigma_n^2}\sum_{m,n}\left[\left(\frac{\partial\tilde{f}}{\partial\beta_i}\right)\left(\frac{\partial\tilde{f}}{\partial\beta_j}\right)\right] \tag{34}$$

This information is more clearly and completely represented in matrix form in equation 35 where the substitutions $\frac{\partial \tilde{f}}{\partial \alpha_i} = \tilde{f}_{\alpha_i}$ and $\frac{\partial \tilde{f}}{\partial \beta_i} = \tilde{f}_{\beta_i}$ are made for more notation brevity. In this form all possible combinations of two gradients are accounted for (i.e. gradient of true scene with respect to α_i shift times gradient of true scene with respect to β_j shift).

$$\begin{aligned}\mathbf{J}_D &= -E\left[\frac{\partial^2}{\partial\left(\underline{\alpha},\underline{\beta}\right)^2}\ln\left(P\left(D\,|\,\underline{\alpha},\underline{\beta}\right)\right)\right] \\ &= \frac{K}{\sigma_n^2}\begin{bmatrix} \sum_{m,n}\tilde{f}_{\alpha_1}\tilde{f}_{\alpha_1} & \sum_{m,n}\tilde{f}_{\alpha_1}\tilde{f}_{\alpha_2} & \cdots & \sum_{m,n}\tilde{f}_{\alpha_1}\tilde{f}_{\alpha_K} & \sum_{m,n}\tilde{f}_{\alpha_1}\tilde{f}_{\beta_1} & \cdots & \sum_{m,n}\tilde{f}_{\alpha_1}\tilde{f}_{\beta_K} \\ \sum_{m,n}\tilde{f}_{\alpha_2}\tilde{f}_{\alpha_1} & \sum_{m,n}\tilde{f}_{\alpha_2}\tilde{f}_{\alpha_2} & & & & & \vdots \\ \vdots & & \ddots & & & & \vdots \\ \sum_{m,n}\tilde{f}_{\alpha_K}\tilde{f}_{\alpha_1} & & & \sum_{m,n}\tilde{f}_{\alpha_K}\tilde{f}_{\alpha_K} & & & \vdots \\ \sum_{m,n}\tilde{f}_{\beta_1}\tilde{f}_{\alpha_1} & & & & \sum_{m,n}\tilde{f}_{\beta_1}\tilde{f}_{\beta_1} & & \vdots \\ \vdots & & & & & \ddots & \vdots \\ \sum_{m,n}\tilde{f}_{\beta_K}\tilde{f}_{\alpha_1} & \cdots & \cdots & \cdots & \cdots & \cdots & \sum_{m,n}\tilde{f}_{\beta_K}\tilde{f}_{\beta_K} \end{bmatrix}\end{aligned} \tag{35}$$

The FIM for Gaussian noise is originally derived by Robinson and Milanfar (8:1186-1187).

The second term in equation 27 accounts for any prior knowledge of the translational shifts. Unlike $\mathbf{J}_D$, here the translational shifts are treated as random vectors which can be statistically modeled. For this derivation a Gaussian distribution is used to

describe the translational shifts. Equation 36 expresses the random translational shifts in a general 2 dimensional Gaussian form.

$$P\left(\underline{\alpha},\underline{\beta}\right)=\prod_{k=1}^{K}\frac{1}{2\pi\sigma_{\alpha}\sigma_{\beta}}\exp\left(\frac{-\alpha_k^2}{2\sigma_{\alpha}^2}\right)\exp\left(\frac{-\beta_k^2}{2\sigma_{\beta}^2}\right) \tag{36}$$

This model assumes that the shifts occur with a mean of zero. Furthermore, this model allows for different variances to govern each dimension of shifting. The natural log operator simplifies the expression to become the sum of $2K+1$ terms where all but one contain a random variable, as shown in equation 37.

$$\begin{aligned}\ln\left(P\left(\underline{\alpha},\underline{\beta}\right)\right) &= -K\ln\left(2\pi\sigma_{\alpha}\sigma_{\beta}\right)+\sum_{k=1}^{K}\left[\frac{-\alpha_k^2}{2\sigma_{\alpha}^2}+\frac{-\beta_k^2}{2\sigma_{\beta}^2}\right]\\ &= -K\ln\left(2\pi\sigma_{\alpha}\sigma_{\beta}\right)-\frac{1}{2}\sum_{k=1}^{K}\left[\frac{\alpha_k^2}{\sigma_{\alpha}^2}\right]-\frac{1}{2}\sum_{k=1}^{K}\left[\frac{\beta_k^2}{\sigma_{\beta}^2}\right]\end{aligned} \tag{37}$$

The partial derivative operators distribute to each term, immediately eliminating the constant term. Once again the results are dependent on the reference of the derivatives. Equations 38-42 show the results of all possible combinations of derivative operators.

$$\frac{\partial^2}{\partial\alpha_i^2}\ln\left(P\left(\underline{\alpha},\underline{\beta}\right)\right)=\frac{-1}{\sigma_{\alpha}^2} \tag{38}$$

$$\frac{\partial^2}{\partial \beta_i^2} \ln\left(P\left(\underline{\alpha}, \underline{\beta}\right)\right) = \frac{-1}{\sigma_\beta^2} \tag{39}$$

$$\frac{\partial^2}{\partial \alpha_i \partial \alpha_j} \ln\left(P\left(\underline{\alpha}, \underline{\beta}\right)\right) = 0 \qquad \left(i \neq j\right) \tag{40}$$

$$\frac{\partial^2}{\partial \beta_i \partial \beta_j} \ln\left(P\left(\underline{\alpha}, \underline{\beta}\right)\right) = 0 \qquad \left(i \neq j\right) \tag{41}$$

$$\frac{\partial^2}{\partial \alpha_i \partial \beta_j} \ln\left(P\left(\underline{\alpha}, \underline{\beta}\right)\right) = 0 \tag{42}$$

Many results are conveniently canceled out, leaving only 2*K* non-zero results. These terms pass through the expectation operator and finally result in the following two very simple expressions:

$$-E\left[\frac{\partial^2}{\partial \alpha_i^2} \ln\left(P\left(\underline{\alpha}, \underline{\beta}\right)\right)\right] = \frac{1}{\sigma_\alpha^2} \tag{43}$$

$$-E\left[\frac{\partial^2}{\partial \beta_i^2} \ln\left(P\left(\underline{\alpha}, \underline{\beta}\right)\right)\right] = \frac{1}{\sigma_\beta^2} \tag{44}$$

Using the same matrix notation form as before, $\mathbf{J}_\mathrm{P}$ is

$$\mathbf{J}_P = -E\left[\frac{\partial^2}{\partial\left(\underline{\alpha},\underline{\beta}\right)^2}\ln\left(P\left(\underline{\alpha},\underline{\beta}\right)\right)\right] = \begin{bmatrix} \frac{1}{\sigma_\alpha^2} & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & \ddots & & & & \vdots \\ \vdots & & \frac{1}{\sigma_\alpha^2} & & & \vdots \\ \vdots & & & \frac{1}{\sigma_\beta^2} & & \vdots \\ \vdots & & & & \ddots & 0 \\ 0 & \cdots & \cdots & \cdots & 0 & \frac{1}{\sigma_\beta^2} \end{bmatrix} \tag{45}$$

The third term in equation 27 is the easiest to solve of the three. Since the data does not contain shift parameters, the partial derivative operators with respect to the shift parameters cancel out this term completely.

Given that image noise is Gaussian with a mean of the true scene, and given that the random translational shifts occur according to a Gaussian model with a mean of zero, the final expression for the total mean-square error of the multiframe shift estimation scenario is

$$\mathbf{J}_T^{-1} = \begin{bmatrix} \frac{1}{\sigma_\alpha^2} + \frac{K}{\sigma_n^2}\sum_{x,y}\tilde{f}_{\alpha_1}^2 & \cdots & \frac{K}{\sigma_n^2}\sum_{x,y}\tilde{f}_{\alpha_1}\tilde{f}_{\alpha_K} & \frac{K}{\sigma_n^2}\sum_{x,y}\tilde{f}_{\alpha_1}\tilde{f}_{\beta_1} & \cdots & \frac{K}{\sigma_n^2}\sum_{x,y}\tilde{f}_{\alpha_1}\tilde{f}_{\beta_K} \\ \vdots & \ddots & & & & \vdots \\ \frac{K}{\sigma_n^2}\sum_{x,y}\tilde{f}_{\alpha_K}\tilde{f}_{\alpha_1} & & \frac{1}{\sigma_\alpha^2} + \frac{K}{\sigma_n^2}\sum_{x,y}\tilde{f}_{\alpha_K}^2 & & & \frac{K}{\sigma_n^2}\sum_{x,y}\tilde{f}_{\alpha_K}\tilde{f}_{\beta_K} \\ \frac{K}{\sigma_n^2}\sum_{x,y}\tilde{f}_{\beta_1}\tilde{f}_{\alpha_1} & & & \frac{1}{\sigma_\beta^2} + \frac{K}{\sigma_n^2}\sum_{x,y}\tilde{f}_{\beta_1}^2 & & \frac{K}{\sigma_n^2}\sum_{x,y}\tilde{f}_{\beta_1}\tilde{f}_{\beta_K} \\ \vdots & & & & \ddots & \vdots \\ \frac{K}{\sigma_n^2}\sum_{x,y}\tilde{f}_{\beta_K}\tilde{f}_{\alpha_1} & \cdots & \frac{K}{\sigma_n^2}\sum_{x,y}\tilde{f}_{\beta_K}\tilde{f}_{\alpha_K} & \frac{K}{\sigma_n^2}\sum_{x,y}\tilde{f}_{\beta_K}\tilde{f}_{\beta_1} & \cdots & \frac{1}{\sigma_\beta^2} + \frac{K}{\sigma_n^2}\sum_{x,y}\tilde{f}_{\beta_K}^2 \end{bmatrix}^{-1} \quad (46)$$

It can be shown that $\tilde{f}_{\alpha_i} = \tilde{f}_x$ and $\tilde{f}_{\beta_i} = \tilde{f}_y$ for all *i*. In other words, since all $\underline{\alpha}$ values are in the same dimension, taking the gradient of an image with respect to any $\underline{\alpha}$ value is the same as taking a gradient with respect to the dimension of $\underline{\alpha}$. The same is also true of $\underline{\beta}$. This property simplifies the final result to be as shown in equation 47, where the diagonal elements of the inverse of matrix $\mathbf{J}_T$ are the estimation performance bounds for all random shifts. This performance boundary is estimator independent and therefore represents the best possible performance that any estimator can achieve.

$$\mathbf{J}_T^{-1} = \begin{bmatrix} \frac{1}{\sigma_\alpha^2} + \frac{K}{\sigma_n^2}\sum_{x,y}\tilde{f}_x^2 & \cdots & \frac{K}{\sigma_n^2}\sum_{x,y}\tilde{f}_x^2 & \frac{K}{\sigma_n^2}\sum_{x,y}\tilde{f}_x\tilde{f}_y & \cdots & \frac{K}{\sigma_n^2}\sum_{x,y}\tilde{f}_x\tilde{f}_y \\ \vdots & \ddots & & & & \vdots \\ \frac{K}{\sigma_n^2}\sum_{x,y}\tilde{f}_x^2 & & \frac{1}{\sigma_\alpha^2} + \frac{K}{\sigma_n^2}\sum_{x,y}\tilde{f}_x^2 & & & \frac{K}{\sigma_n^2}\sum_{x,y}\tilde{f}_y\tilde{f}_x \\ \frac{K}{\sigma_n^2}\sum_{x,y}\tilde{f}_y\tilde{f}_x & & & \frac{1}{\sigma_\beta^2} + \frac{K}{\sigma_n^2}\sum_{x,y}\tilde{f}_y^2 & & \frac{K}{\sigma_n^2}\sum_{x,y}\tilde{f}_y\tilde{f}_y \\ \vdots & & & & \ddots & \vdots \\ \frac{K}{\sigma_n^2}\sum_{x,y}\tilde{f}_y\tilde{f}_x & \cdots & \frac{K}{\sigma_n^2}\sum_{x,y}\tilde{f}_y\tilde{f}_x & \frac{K}{\sigma_n^2}\sum_{x,y}\tilde{f}_y\tilde{f}_y & \cdots & \frac{1}{\sigma_\beta^2} + \frac{K}{\sigma_n^2}\sum_{x,y}\tilde{f}_y^2 \end{bmatrix}^{-1} \quad (47)$$

If the variance of the shifts is large then the prior information effectively cancels out, leaving $\mathbf{J}_T$ equal to the FIM, which yields the Cramer Rao lower bound (CRLB) of variance. However since the FIM is based only on the data, it is a lower bound than the actual performance bound (which could only be computed if all random elements of the imaging process were included). In the case of the new multiframe shift estimator the CRLB is useful since the estimator does not assume a particular (non-uniform) shifting model. This means that the estimator assumes the shifting variance is infinite (or rather that the shifting distribution is uniform).

Estimator Variance

Equation 47 provides a reference for estimator variance comparison. Experimentally acquiring the variance of the estimator involves selecting a true scene with gradient information that is greater than zero, and selecting an imaging system model that samples the scene well enough to prevent aliasing. In order to observe differences in the estimator results it is necessary to vary the noise of each frame in a set of data, without changing the random shifts of each frame. Figure 6 shows the results of this approach where a set of *K* data frames are uniquely corrupted by Gaussian noise (with a standard deviation of 3.5 photons) 500 times and the random translational shifts are estimated each time by the new multiframe shift estimator. The same true scene (figure 4) and imaging system simulator as the estimator bias experiment is used for this experiment. The average variance of the 500 shift estimates per data frame is retained, and those *K* values are averaged to become the empirically discovered estimator variance

for *K* data frames. *K* is varied from 3 to 20 data frames with the expectation that the variance of the estimator will decrease as the number of data frames used in the estimation increases. Finally, since only one set of random shifts is used in this experiment, and since the initial estimator performance can be influenced by individual sets of data, this entire experiment is run 50 times, changing the random translational shifts each time.

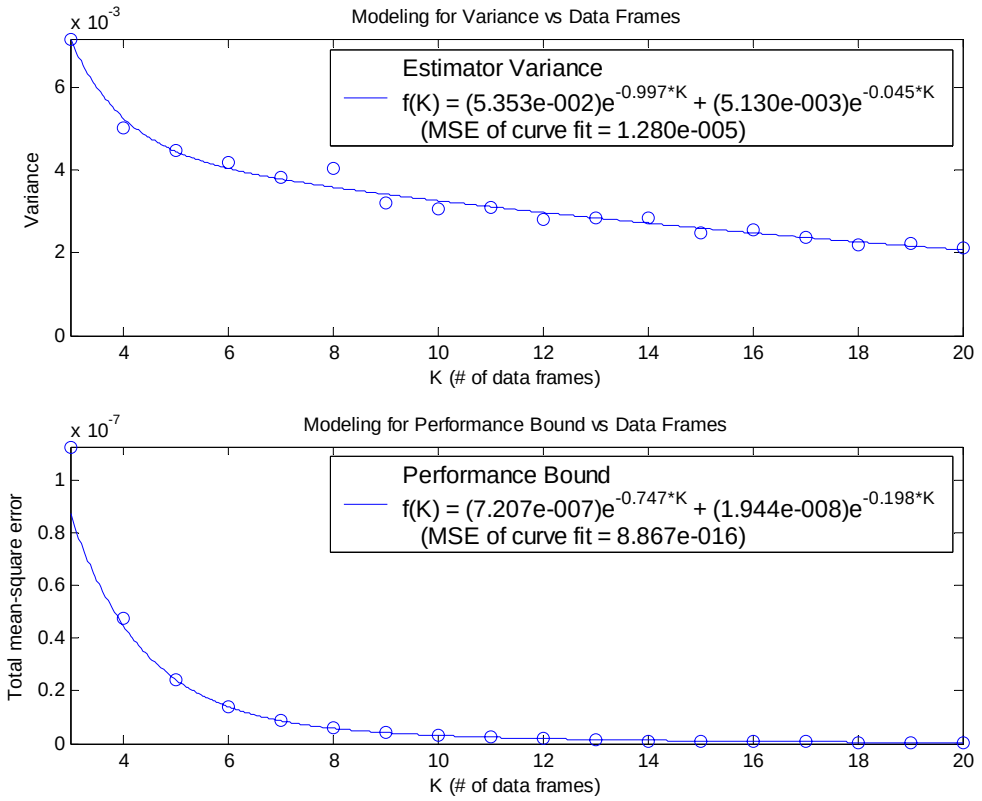

Figure 6. Experimentally determined variance of the new multiframe shift estimator, compared to the total mean-square error lower bound of infinite shifting variance, for 3 to 20 data frames, for estimation of the horizontal shift on frame 1 of the data set. This bound is effectively the CRLB of estimation variance.

In figure 6 both the estimator performance and the bound are modeled by a double exponential functional form, which retains the general trend of the data, and allows for simple analysis since both are of the same form. The gap between the performance and the bound (in figure 6) is a combined result of the type of image used, the failure of the cross correlation initial estimator to account for any prior shifting information, and the fact that the bound does not include scene estimation in its model.

Analysis Of Results

This performance bound offers a useful tool for finding points of diminishing returns of performance from adding frames. For a given type of data this performance bound is computed and fit with a curve model. The percent of improvement due to adding additional frames, relative to performance at a particular starting *K* value, can be quickly computed and offers the analyst practical insight into the potential improvement that can be gained by using more data frames in the estimation. This same technique can be applied to the estimator performance. Figure 7 shows several examples of this tool, using the estimator performance and performance bound shown in figure 6. Figure 8 shows the cumulative version of figure 7, where the improvement approaches 100% as K grows large.

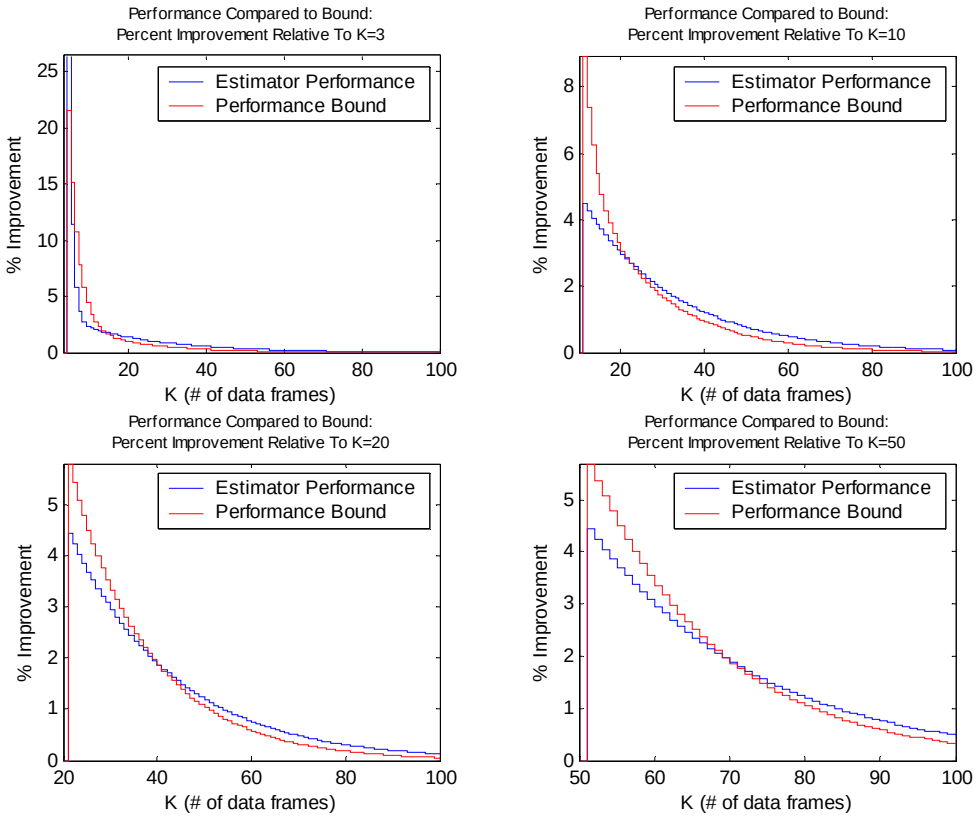

Figure 7. Percent improvement of the estimator performance and performance bound due to using more data frames in the estimation. This is based on the estimator performance and performance bound in figure 6. For instance, adding 10 data frames when starting with 3 will improve the potential performance bound much more significantly than adding 10 frames when starting with 50.

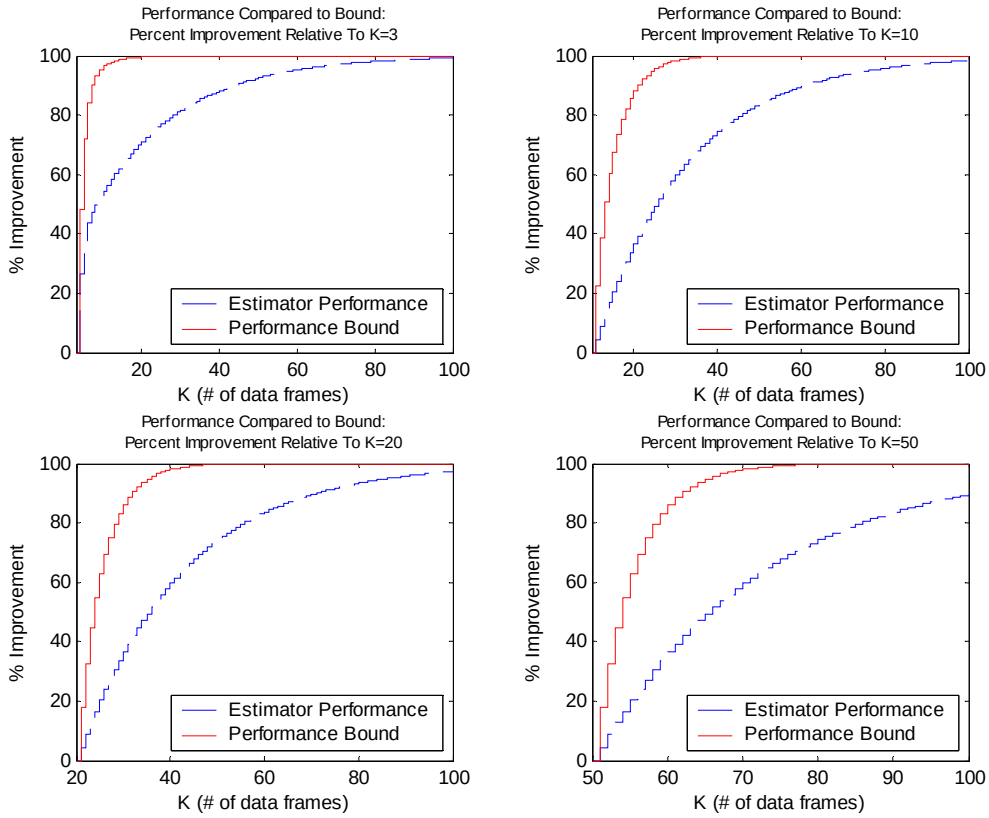

Figure 8. Cumulative percent improvement of the estimator performance and performance bound due to using more data frames in the estimation. This is based on the estimator performance and performance bound in figure 6, and is the cumulative result of the curves in figure 7.

Point sources have higher gradients than any other realistic images of the same dimensions and amount of light. According to equation 47 this means that point source images yield the lowest possible performance bounds. Since figure 4 is a point source, then the performance bound in figure 6 is the lowest possible bound for 32 pixel by 32 pixel images with the same total amount of light, making this bound image independent.

In realistic situations the true scene is not known. Therefore it is necessary to use an estimate of the true scene to compute the performance bound for a particular set of

data. Common practice is to use either the frame average of a registered set of data or an individual data frame depending on the severity of noise corruption.

Estimator Comparison Based On Likelihood Of Scene Estimate

It has been shown that a data frame can be modeled as a joint random function, where the total likelihood of receiving a particular data frame is the product of all the individual probabilities of the pixels for the amount of light received, respectively. Equation 12 models a particular data frame according to the Poisson distribution. Equation 48 shows how the natural log operator simplifies this expression without altering its monotonicity.

$$\ln\left(P\left(D\,|\,\alpha,\beta\right)\right) = \sum_{x}\sum_{y}\left[D\left(x-\alpha,y-\beta\right)\ln\left(i\left(x,y\right)\right)-i\left(x,y\right)-\ln\left(D\left(x-\alpha,y-\beta\right)!\right)\right] \tag{48}$$

The joint probability modeling approach provides a common frame of reference to compare scene estimates obtained from a single data set. It is therefore reasonable to argue superiority of one scene estimate over another on the grounds of higher likelihood of occurrence. For example, for two different scene estimations generated from the same data set, the scene estimation with a higher likelihood is the more probable estimate.

Since frame registration is a crucial part in scene estimation, and since translational shift estimation is the fundamental element of registration, this modeling approach provides a way to compare shift estimation performance of different shift estimators using real data. Frame averaging is a simple and effective scene estimation method. It involves translational shift estimation and removal (registration) followed by averaging across the registered data set. Different translational shift estimators will ultimately yield different scene estimates. Therefore, multiple shift estimators can be compared by using the frame averaging technique. If one shift estimator ultimately yields a scene estimate with a higher likelihood than another, then that shift estimator performed better.

Figure 9 shows the average results of 500 such comparisons for 3 to 30 frames. For each comparison and each number of frames a unique set of data is created with Poisson noise and Gaussian random shifting (variance of 2 pixels in both dimensions). The data set is registered using the new multiframe shift estimator (with cross correlation initial estimator) and the cross correlation multiframe shift estimator. The registered data sets are then averaged across the number of frames to estimate the true scene. Finally, the scene estimates are shifted such that they are centered on the true scene in order to maximize the likelihood functions that will be generated and to ensure that neither scene estimate is unfairly biased. The true scene is the image with the highest likelihood of occurrence, therefore providing an upper bound of likelihood. In this experiment the simulated 32 pixel by 32 pixel point source image of figure 4 is used.

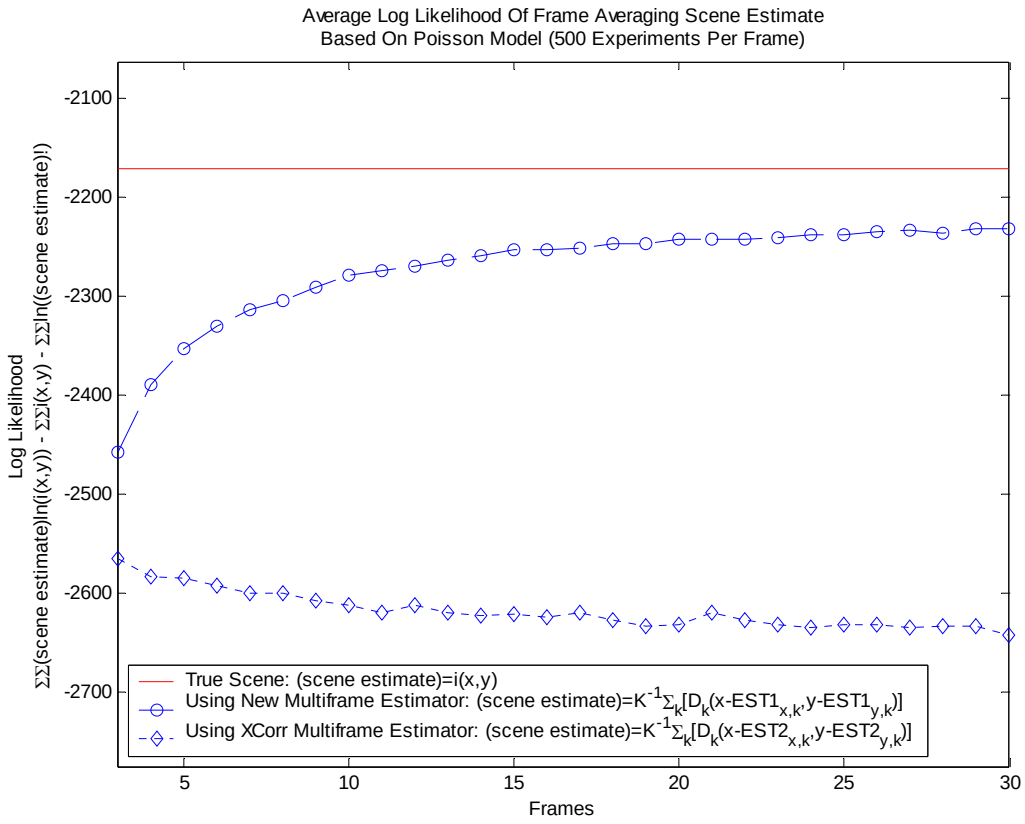

Figure 9. Average experimental log likelihood of the frame averaged scene estimate as a function of the number of frames.

It is plainly visible from this experiment that the log likelihood of the scene estimation resulting from the new multiframe shift estimator is substantially better than that of the cross correlation multiframe shift estimator. Furthermore, it is apparent that as the number of frames increases, the log likelihood of the scene estimation resulting from the new multiframe shift estimator increases, eventually converging to some value less than or equal to the log likelihood of the true scene.

Estimator Comparison Based On Likelihood Of Data Set

If the true scene is not known, such as in a realistic data collection situation, then a slightly different comparison method must be used. Just as an individual image can be modeled as a joint likelihood function, a set of data frames can also be modeled as a joint likelihood function. Assuming that all data frames within a given set are independent of one another, and using the model already established for a single image with Poisson noise, the total probability of receiving a particular data set is

$$P\left(D \mid \underline{\alpha}, \underline{\beta}\right) = \prod_{k}^{K} \prod_{x} \prod_{y} \frac{i\left(x, y\right)^{D_k\left(x-\alpha_k, y-\beta_k\right)} e^{-i\left(x,y\right)}}{D_k\left(x-\alpha_k, y-\beta_k\right)!} \tag{49}$$

This model is very similar to equation 12, except that it computes the product of all the probabilities of each data frame occurring. The log likelihood version of this model is very similar to equation 48, except that it adds all the individual log likelihoods of each data frame, as shown in equation 50.

$$\begin{aligned} &\ln\left(P\left(D \mid \underline{\alpha}, \underline{\beta}\right)\right) \\ &= \sum_{k}^{K} \sum_{x} \sum_{y} \left[D_k\left(x-\alpha_k, y-\beta_k\right) \ln\left(i\left(x, y\right)\right) - i\left(x, y\right) - \ln\left(D_k\left(x-\alpha_k, y-\beta_k\right)!\right) \right] \end{aligned} \tag{50}$$

For a particular shift estimator, the average of the registered frames is used as the true scene. The log likelihood of the registered data set occurring is computed as a final

performance metric, representative of how well the shift estimator performs. Accurate shift estimates yield higher individual data frame likelihoods. Therefore higher overall data set likelihoods indicate better shift estimation. Figure 10 shows experimental results comparing the new multiframe shift estimator and the cross correlation multiframe shift estimator. The same data sets from the previous experiment are used. Since this is an experiment and the true scene is actually known, an upper performance bound can be calculated. For a data set of K frames the highest possible metric that can be acquired K times the log likelihood of the known true scene. This boundary is shown in figure 10.

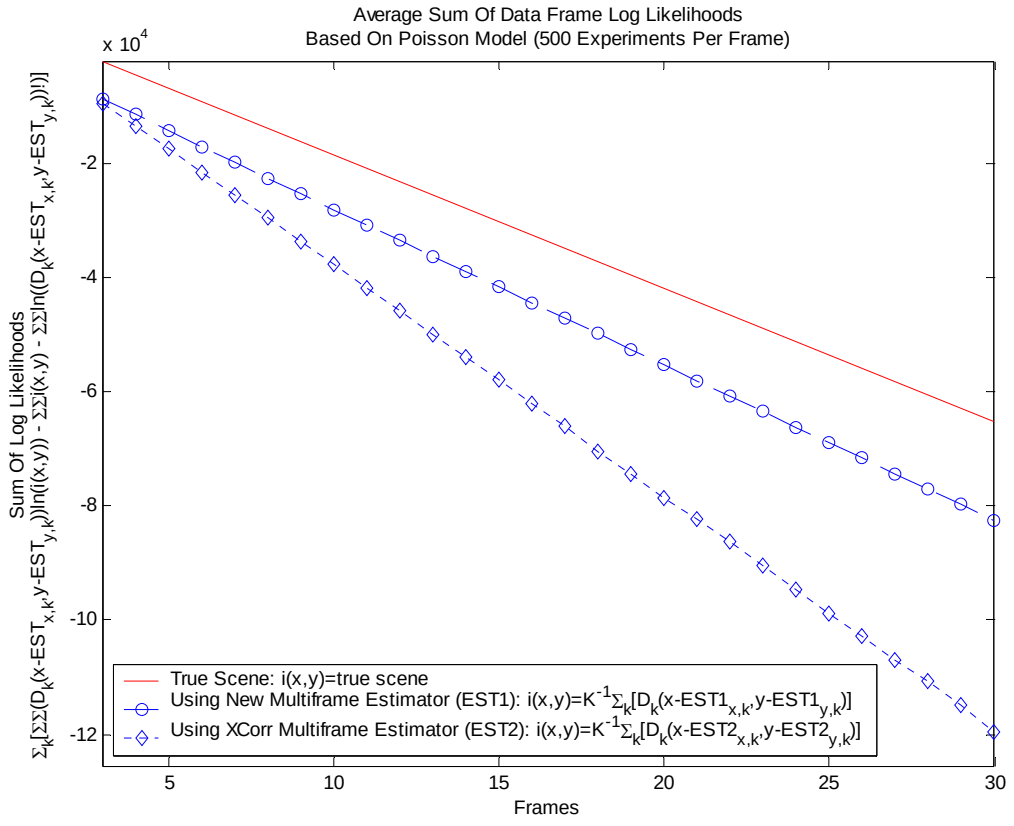

Figure 10. Average log likelihood of a particular data set, using the frame averaged scene estimate as the true scene, as a function of number of frames.

It is clear in figure 10 that the new multiframe shift estimator (using cross correlation initial estimator) performs significantly better than the cross correlation multiframe shift estimator.

Figure 11 illustrates the log likelihood results of a real data set of black and white photographs of a coin, located 12 inches from a 20 watt halogen lamp light source. This data is from a handheld digital camera, and the random shifting (resulting from ordinary hand instabilities) occurs about the center of the coin. An example data frame is shown in figure 12. For this experiment data sets ranging in size from 3 to 50 frames are collected. For each collection the new multiframe shift estimator (with cross correlation initial estimator) and the cross correlation multiframe shift estimator are used to estimate all translational shifts. The total data set log likelihood for each collection is calculated (using frame averaging of the respective registered data sets as the scene estimate). Figure 11 clearly shows that, for this data set, the new multiframe shift estimator performs better than the cross correlation multiframe shift estimator. Furthermore, it shows that as the number of data frames is increased, the the new multiframe shift estimator performs increasingly better compared to the cross correlation multiframe shift estimator.

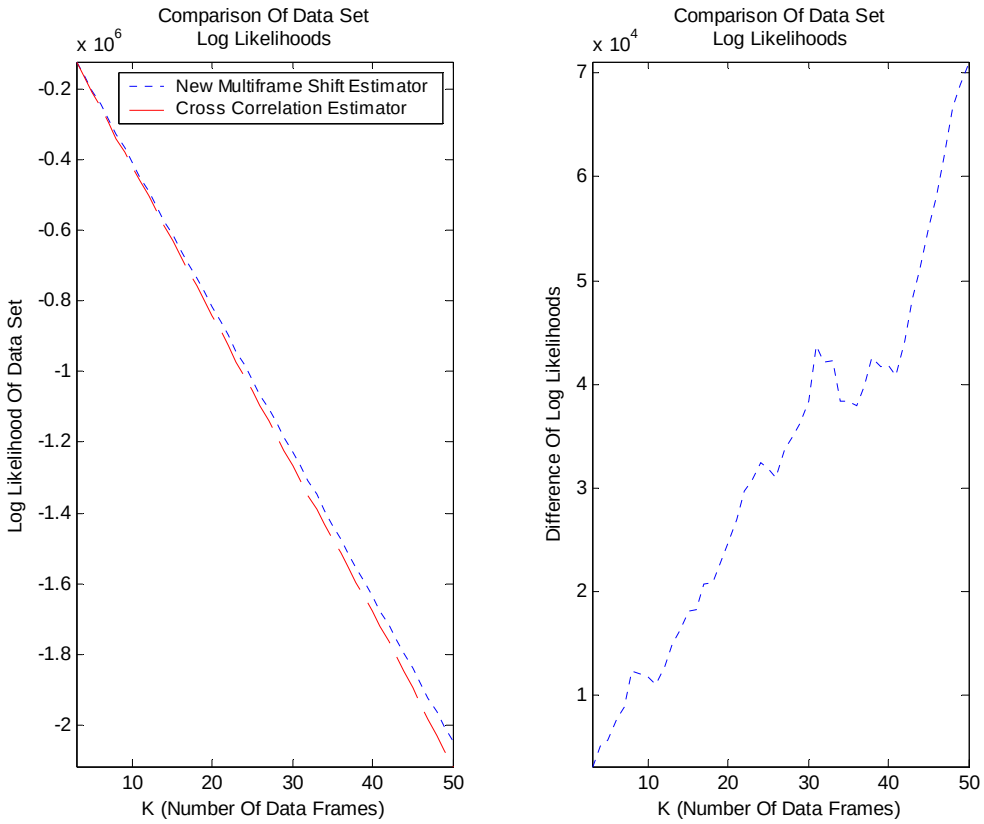

Figure 11. Comparison of data set log likelihoods for 3 to 50 frames of real data. The right graph shows the log likelihood difference between the two estimators, such that a higher positive difference indicates a higher likelihood of the new multiframe shift estimator.

Figure 12. Sample data frame from real data set. The data set contains black and white photographs of a nickel on a dark gray background. The 20 Watt halogen lamp light source was located 12 inches from the coin, and the camera was located approximately 24 inches from the coin.

V. Conclusion

Random translational shifting is a common problem in data collected from imaging systems mounted on a moving or vibrating platform. This shifting degrades the quality of information in the data, and complicates the scene estimation process. However, much of this corruption can be removed if reliable and accurate translational shift estimates are obtained.

It has been shown that there exist estimators able to determine the translational shift difference between 2 frames of the same data set. More specifically, it has been shown that the maximum likelihood estimator in the presence of Gaussian noise is the cross correlation shift estimator. While there exists a broad variety of 2 frame shift estimation techniques, multiframe shift estimation theory is relatively undeveloped. Most current multiframe shift estimation methods are merely a series of 2 frame based shift estimations. This approach can suffer from skewed performance due to its image dependence, thereby motivating the development of a multiframe shift estimation approach that overcomes any image dependence.

The new multiframe shift estimator presented in this report uses existing 2 frame based estimators to determine shift differences among a set of data frames, and applies prior knowledge of the shifting to remove any performance skewing due to particular image dependencies. This results in estimates that are more accurate than the corresponding 2 frame based estimates of the shifts. This new approach to multiframe shift estimation effectively uses more of the information in the data than the 2 frame

approach. As a result it performs better as more data frames are included in the estimation process. Furthermore, this new estimator does not add any bias to the initial estimator that drives it. Therefore, if the initial estimator is unbiased (meaning that the expected value of the mean-square error is zero) then the corresponding new multiframe estimator is also unbiased.

It has been shown that images and data sets can be modeled as joint random functions where the true scene is the mean of the noise. This modeling approach allows performance of different shift estimators to be compared on the basis of total likelihood. In all experiments performed, the new shift estimator yielded results with a higher likelihood than the cross correlation estimator.

The lower bound of performance is comprised of several parts. The first part is a result of the data, where the translational shifts are treated as non-random parameters. The boundary that results from this part is the Cramer Rao lower bound of variance, and has been derived for the 2 frame estimation case by Robinson and Milanfar. The second part of the performance bound accounts for the translational shifts as random parameters, and utilizes any prior information about the shifting. The final result is a lower bound on the total mean-square error of a particular true scene, independent of any estimator. The multiframe derivation of this shows that the performance bound asymptotically approaches zero as the number of frames used in the estimation increases. This boundary equips the analyst with a helpful tool for determining how many frames to include in the estimation process.

Potential further research in this area includes modeling with aliasing, optimization for real time implementation, characterization of shifting without a zero mean, and performance comparison using initial estimators other than the cross correlator.

This thesis offers the field of image processing a new approach to multiframe shift estimation. Theoretical proof and experimental results show that this new approach is valid and useful for data analysis. The performance boundary derivations in this thesis provide a frame of reference for all multiframe translational shift estimation problems. Current research in estimation theory for the field of image processing has moved forward as a result of this thesis.

Bibliography

1. Frischolz, Robert W. and Klaus P. Spinnler. "Class of Algorithms For Realtime Subpixel Registration," *Proceedings of SPIE*, 1989: 50-59 (December 1993).

2. Gottesfeld Brown, Lisa. "A Survey of Image Registration Techniques," *ACM Computing Surveys*, 24: 325-376 (December 1992).

3. Jain, Anil K. *Fundamentals of Digital Image Processing*. Upper Saddle River, New Jersey: Prentice-Hall, Inc., 1989.

4. Kuglin,C.D. and D.C. Hines. "The Phase Correlation Image Alignment Method," *Proc. IEEE 1975 Int. Conf. Cybernetics and Society*, 163-165 (September 1975).

5. Leon-Garcia, Alberto. *Probability and Random Processes for Electrical Engineers (Second Edition)*. Reading, Massachusetts: Addison-Wesley Publishing Company, 1994.

6. "Optics." *Encyclopedia Britannica Premium*. 13 February 2006 http://www.britannica.com/eb/article-37988

7. Pham, Tuan Q., Marijn Bezuijen, Lucas J. van Vliet, Klamer Schutte, and Cris L. Luengo Hendriks. "Performance of Optimal Registration Estimators," *Visual Information Processing XIV (Proc. SPIE Defense and Security Symposium, Orlando, Florida, USA, Mar.29-Apr.1)*, 5817: 133-144, 2005.

8. Robinson, Dirk and Peyman Milanfar. "Fundamental Performance Limits in Image Registration," *IEEE Transactions on Image Processing*, 13: 1185-1199 (September 2004).

9. Van Trees, Harry L. *Detection, Estimation, And Modulation Theory (Part I)*. New York: John Wiley & Sons, Inc., 2001.

Vita

2Lt Stephen A. Bruckart graduated from American Christian Academy High School in Boulder Creek, California. He completed his undergraduate studies at Cedarville University in Cedarville, Ohio, earning a Bachelor of Science degree in Electrical Engineering in May of 2004. Upon graduation, he was commissioned through AFROTC detachment 643 of Wright State University of Ohio where he was recognized as a Distinguished Graduate.

Stephen's first assignment was the Air Force Institute of Technology (AFIT), under a joint educational effort between the National Air and Space Intelligence Center (NASIC) and AFIT, where he entered the Graduate School of Engineering and Management for a Master's degree in electrical engineering. Upon graduation in March of 2006 he will be assigned to NASIC, Wright Patterson AFB, Ohio.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From – To)*
03-23-2006	**Master's Thesis**	Aug 2004 – Mar 2006

4. TITLE AND SUBTITLE

Multiframe Shift Estimation

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Bruckart, Stephen, A., 2DLt, USAF

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES(S) AND ADDRESS(S)
Air Force Institute of Technology
Graduate School of Engineering and Management (AFIT/EN)
2950 Hobson Way
WPAFB OH 45433-7765

8. PERFORMING ORGANIZATION REPORT NUMBER

AFIT/GE/ENG/06-08

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
AFRL/SNJ (AFMC)
Attn: Dr. Edward A. Watson
2241 Avionics Circle, Bldg 620
WPAFB OH 45433-7765 DSN: 785-7859 x4030

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED

13. SUPPLEMENTARY NOTES

14. ABSTRACT

The purpose of this research was to develop a fundamental framework for a new approach to multiframe translational shift estimation in image processing. This thesis sought to create a new multiframe shift estimator, to theoretically prove and experimentally test key properties of it, and to quantify its performance according to several metrics. The new estimator was modeled successfully and was proven to be an unbiased estimator under certain common image noise conditions. Furthermore its performance was shown to be superior to the cross correlation shift estimator, a robust estimator widely used in similar image processing cases, according to several criteria.

This research effort led to the derivation of a lower bound of estimation performance for the multiframe case. This valuable data analysis tool extends current boundary derivations to include prior information about the random shifting, thereby providing a more precise performance boundary.

15. SUBJECT TERMS
Image Registration, Translational Shift Estimation, Multiframe Estimation,
Performance Bound, Variance Analysis, Cramer Rao Lower Bound

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
REPORT	**ABSTRACT**	**c. THIS PAGE**			Stephen C. Cain, AFIT/ENG
U	**U**	**U**	**UU**	63	**19b. TELEPHONE NUMBER** *(Include area code)* (937) 255-3636, ext 4716; e-mail: stephen.cain@afit.edu

Standard Form 298 (Rev: 8-98)
Prescribed by ANSI Std. Z39-18