MULTI-REFERENCE FRAME IMAGE REGISTRATION
FOR ROTATION, TRANSLATION, AND SCALE

THESIS

Christopher Costello, Second Lieutenant, USAF

AFIT/GE/ENG/08-05

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

AFIT/GE/ENG/08-05

MULTI-REFERENCE FRAME IMAGE REGISTRATION
FOR ROTATION, TRANSLATION, AND SCALE

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 Electrical Engineering

Christopher Costello, B.S.E.E.

Second Lieutenant, USAF

March 2008

AFIT/GE/ENG/08-05

Multi-Reference Frame Image Registration
for Rotation, Translation, and Scale

Christopher Costello, B.S.E.E.

Second Lieutenant, USAF

Approved:

/signed/	29 Feb 2008
Dr. Richard Martin (Chairman)	date
/signed/	29 Feb 2008
Dr. Steven Gustafson (Member)	date
/signed/	29 Feb 2008
Dr. Stephen Cain (Member)	date

AFIT/GE/ENG/08-05

Abstract

This thesis investigates applications of multi-reference frame image registration for image sets with various translation, rotation, and scale combinations. It focuses on registration accuracy improvement over traditional pairwise registration, and also compares the quality of scene estimation from frame averaging. Three experiments are developed which use cross-correlation to estimate translation, the Radon transform to estimate translation and rotation, and the Fourier-Mellin transform to estimate translation, rotation, and scale. Results from applying multi-reference frame registration in these experiments show distinct improvements in both registration accuracy and quality of frame averaging compared to single-reference frame registration. Furthermore, it is shown that the new registration technique is equivalent to the optimal Gauss-Markov estimator of the relative shifts given all pairwise shifts.

Acknowledgements

First off, I must thank Dr. Martin for his enduring patience through the last six months and for his guidance and knowledge without which this thesis would not be possible. I would also like to thank Dr. Cain and Dr. Gustafson for their time and insights as committee members. I thank all those who helped me on my path to graduation and to my wife for her love and unwavering support throughout this process.

Christopher Costello

Table of Contents

List of Figures

ix

List of Tables

List of Abbreviations

Multi-Reference Frame Image Registration
for Rotation, Translation, and Scale

I. Introduction

Multiple images of a target or scene captured from a single sensor, are generally distorted from one image to the next. The severity and type of distortion directly depend on the sensor, target, and information acquired. The term "image" commonly refers to capturing information of a specific target or scene using the visible range of the electromagnetic spectrum or light. However, other forms of information could be gathered and simply converted to a visual representation. Virtually anything with varying magnitude can be mapped to a range of colors, such as infrared radiation for thermal imaging or sound waves for ultrasound imaging.

The term "image" should be more loosely defined as any visual representation of information. The information gathered could also have several dimensions of visual significance, as in hyperspectral imaging. Hyperspectral imaging is a fairly new field of imaging science, but as the name suggests, it contains many dimensions of information, and there could be several hundred dimensions of data. The distortions for a hyperspectral image could be completely different than the distortions in an image of visible light.

To maintain generality, this thesis refers to an image as any two-dimensional (2-D) visual representation of information taken from a single sensor and analyzes only linear geometric distortions, specifically translation, rotation, and scale and combinations thereof. These distortions mostly stem from a simple principle: unless the sensor and target are completely stationary, the act of repeatedly capturing images introduces relative random distortion.

1.1 Geometric Image Distortion

For images captured at a relatively fast rate, the distortions are most likely caused by spatial changes. Even if the images are captured quickly, there could still be temporal changes that cause relative distortion between images.

Relative translative distortion is present in any set of images that are captured from non-stationary sensors or targets. A familiar example involves the simple act of a person photographing multiple pictures of the same object. Random shaking of the hand causes small differences in where the camera is pointing. The images collected from the photographs also have rotational distortions because as the photographer pushes the button to take a picture the camera tilts and twists slightly, introducing random relative rotation. Any non-stationary sensor could introduce translative or rotational distortion when capturing multiple images.

For images with significant differences in scaling, the sensor could be moving relative to the target or vice versa. For example, an airborne sensor may introduce relative scaling. If a sensor is mounted on a airborne platform that is flying towards a target, each successive capture of the target is at time when the platform is actually closer to the target. This discrepancy in distance from one capture to next results in relative scaling. The airborne sensor undoubtedly has translational and rotational distortion to some degree. It is difficult to completely stabilize a sensor, especially if it is mounted to a mobile platform.

Distorted images are not only caused by sensors or even targets; distortion may be independent of the sensor. The image may also be distorted by some form of post processing or intentional manipulation, such as encryption or digital watermarking [1]. The foregoing is not an exhaustive account of the causes of distortion: the focus here is not the cause of geometric distortion but rather correcting the effects of distortion.

1.1.1 Mathematical Models. Each distortion or combinations of distortions can be modeled mathematically. Let $f(x, y)$ be a 2-D image. A 2-D image $f_\tau(x, y)$

(a) $f(x, y)$ (b) $f_\tau(x, y)$
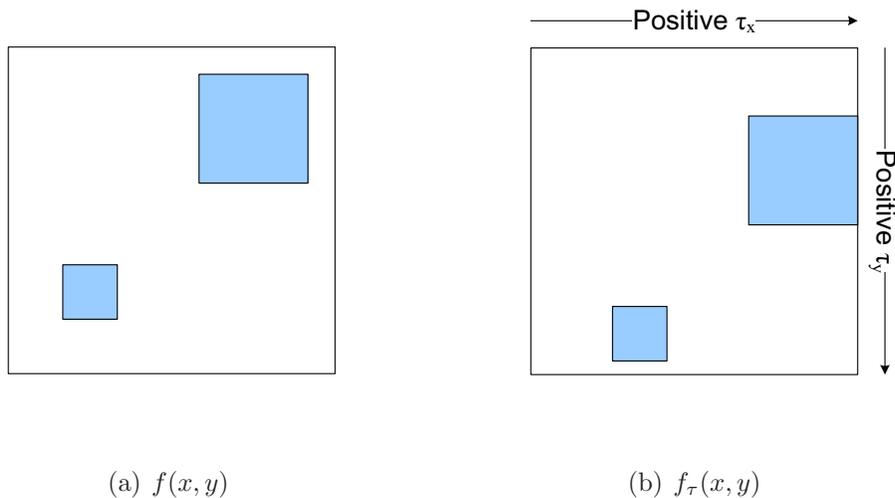
Figure 1.1:   Convention for the model of an image translated by $[\tau_x, \tau_y]$.

translated by the vector $[\tau_x, \tau_y]$ is

$$f_\tau(x, y) = f(x - \tau_x, y - \tau_y), \tag{1.1}$$

where τ_x is the horizontal translation and τ_y is the vertical translation, as in Figure 1.1. We assume that an image can only be translated half as much as its width in any dimension otherwise, the translated image would have more new information than information that is in common with the original image $f(x, y)$. A 2-D image $f_\phi(x, y)$ rotated counter-clockwise about its origin by an angle ϕ is

$$f_\phi(x, y) = f(x \cos \phi + y \sin \phi, y \cos \phi - x \sin \phi), \tag{1.2}$$

where $\phi \in [0, 2\pi]$, as in Figure 1.2. Finally, an image $f_\beta(x, y)$ scaled by a positive factor of β is

$$f_\beta(x, y) = f(\beta x, \beta y). \tag{1.3}$$

If $\beta > 1$ then $f_\beta(x, y)$ is smaller than $f(x, y)$. Conversely, if $\beta < 1$ then $f_\beta(x, y)$ is larger than $f(x, y)$.

(a) $f(x, y)$ (b) $f_\phi(x, y)$
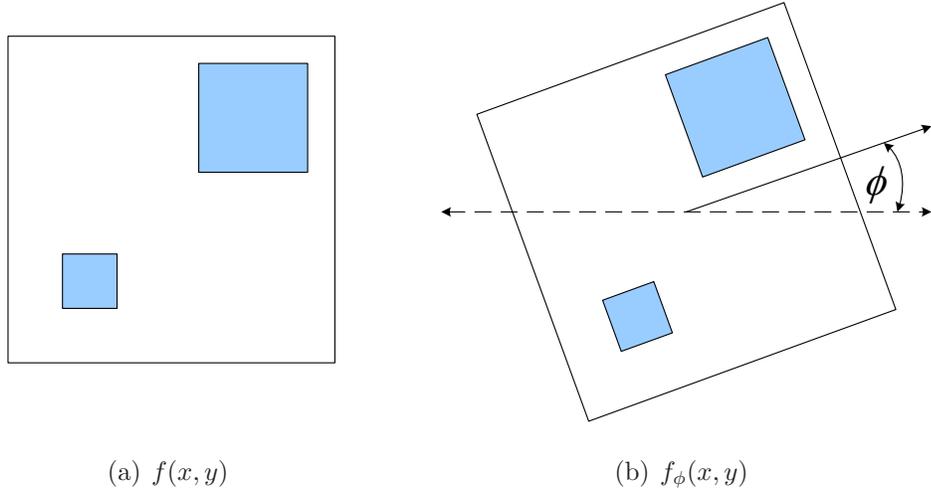
Figure 1.2:   Convention for the model of an image rotated by ϕ.

Mathematical models of combinations of distortions are more complicated because the order of distortions can change the model, e.g., there is a difference in the model for an image that is first translated and then rotated versus an image that is first rotated and then translated. An image translated by the vector $[\tau_x, \tau_y]$ and then rotated by ϕ is

$$f_{\tau,\phi}(x, y) = f((x - \tau_x)\cos\phi + (y - \tau_y)\sin\phi, -(x - \tau_x)\sin\phi + (y - \tau_y)\cos\phi), \quad (1.4)$$

where as an image rotated by ϕ and then translation by the vector $[\tau_x, \tau_y]$ is

$$f_{\phi,\tau}(x, y) = f(x\cos\phi + y\sin\phi - \tau_x, -x\sin\phi + y\cos\phi - \tau_y). \quad (1.5)$$

These two models result in different images when the same $[\tau_x, \tau_y]$ and ϕ are used, as in Figure 1.3. The order of distortions is denoted by the order of the indices of f. For example, an image translated and then rotated is denoted $f_{\tau,\phi}(x, y)$, and an image rotated and then translated image is denoted $f_{\phi,\tau}(x, y)$.

4

(a) $f_{\phi,\tau}(x,y)$

(b) $f_{\tau,\phi}(x,y)$

(c) $f_{\phi,\tau}(x,y)$ & $f_{\tau,\phi}(x,y)$
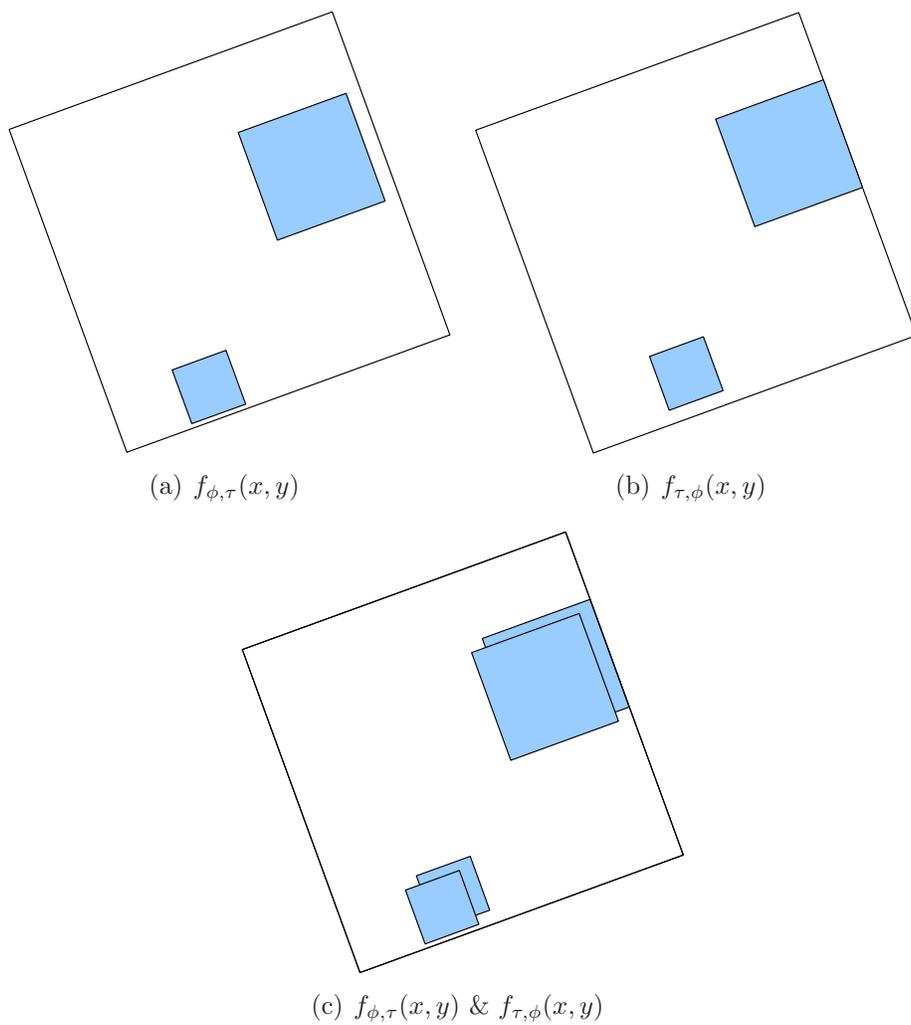
Figure 1.3:   Illustration showing the difference between models for an image that is translated and then rotated, $f_{\tau,\phi}(x,y)$, versus an image that is rotated and then translated, $f_{\phi,\tau}(x,y)$. Each model has the same values for $[\tau_x, \tau_y]$ and ϕ and are overlayed to highlight the differences.

There are six combinations of translation, rotation, and scale. The models for all six combinations are

$$f_{\beta,\tau,\phi}(x,y) = f((\beta x - \tau_x)\cos\phi + (\beta y - \tau_y)\sin\phi, -(\beta x - \tau_x)\sin\phi + (\beta y - \tau_y)\cos\phi)$$
$$(1.6)$$

$$f_{\beta,\phi,\tau}(x,y) = f(\beta x \cos\phi + \beta y \sin\phi - \tau_x, -\beta x \sin\phi + \beta y \cos\phi - \tau_y) \tag{1.7}$$

$$f_{\tau,\beta,\phi}(x,y) = f(\beta(x - \tau_x)\cos\phi + \beta(y - \tau_y)\sin\phi, -\beta(x - \tau_x)\sin\phi + \beta(y - \tau_y)\cos\phi)$$
$$(1.8)$$

$$f_{\tau,\phi,\beta}(x,y) = f(\beta[(x - \tau_x)\cos\phi + (y - \tau_y)\sin\phi], \beta[-(x - \tau_x)\sin\phi + (y - \tau_y)\cos\phi])$$
$$(1.9)$$

$$f_{\phi,\beta,\tau}(x,y) = f(\beta[x\cos\phi + y\sin\phi] - \tau_x, \beta[-x\sin\phi + y\cos\phi] - \tau_y) \tag{1.10}$$

$$f_{\phi,\tau,\beta}(x,y) = f(\beta[x\cos\phi + y\sin\phi - \tau_x], \beta[-x\sin\phi + y\cos\phi - \tau_y]) \tag{1.11}$$

Note that $f_{\tau,\beta,\phi}(x,y) = f_{\tau,\phi,\beta}(x,y)$. Aside from the obvious mathematical differences the models are essentially the same they each model an image that has been translated, rotated, and scaled.

1.2 Image Registration and Applications

It is useful to compare a data set of multiple images of the same target in order to gather more information than a single image can provide. The relative distortions between images could inhibit proper comparison or processing of the images and, in turn, adversely affect analysis and conclusions. Image registration determines parameters of the geometric distortion(s) relating two images and then removes them.

Frame averaging is a common technique that uses registration to average a set of commonly aligned images; it is a primary focus of this thesis. If the distortions are not removed, then the average of the image could produce a worse composite image. The goal of frame averaging is to generate an image with more detail. Image registration can also be used to stabilize video. It is difficult to extract valuable

information from video that is unstable or shaky, and human perception is sensitive to such instabilities. Image matching is another technique that uses image registration to determine the presence of a known object or to determine if a set of images contain the same object. If the object in a set of images is distorted, it could be difficult to determine if each image contains the same object. Image matching is commonly used in digital watermarking to determine if a pair of images contain a known watermark, but often the watermark is distorted and, therefore, image registration is needed to properly match the images.

Traditionally, image registration is performed using a single reference image to which all other images are compared. This procedure results in registration estimates that are inherently biased to the reference image. Also, there is information in the data that is not used. A registration technique is needed that is unbiased and utilizes all available information. Multi-reference frame image registration has exactly this function [2].

1.3 Research Objective

The objective of this thesis is to further investigate the work of Bruckart [2] in the area of multi-reference frame image registration by expanding its applications to include not only translation but also rotation and scale. This research focuses on multi-reference frame registration accuracy improvements over traditional registration and its implications, while also comparing the quality of scene estimation from frame averaging. The results will demonstrate the advantages of multi-reference frame registration and its possible applications. To accomplish this objective, several experiments are developed that use combinations of geometric distortions and that apply different registration algorithms to estimate distortions. The aim of this research is to compare the performance of the multi-reference and single reference frame registrations methods and the quality of their respective frame averaging processes.

1.4 Resources

All coding of registration algorithms and post-processing of results are performed using MATLAB® version 7.4.0287 (R2007a) on a dual processor Intel Zeon 3.6 Ghz with 3 GB of memory using Windows XP Service Pack 2. All simulations are performed using MATLAB® version 7.1.0.183 (R14) Service Pack 3 on a Linux High Performance Computing (HPC) cluster, which has 64 nodes (64 bit) and 128 Opteron 248 2.2 GHz CPUs with 4 GB of memory per CPU.

II. Review of Registration Techniques

A majority of the literature on image registration is based on a single reference frame approach where one frame, the reference image, is compared or matched to another frame, the secondary image. Image registration for a set of frames is performed by repeating this process for different secondary images, using the same reference image, hence the name "single reference frame method". When only one reference image is used, registration estimates are biased to the reference image. If this image is heavily corrupted by noise or other interference, the results of the registration are inaccurate. However, under most circumstances the single reference frame approach works well and is computationally less taxing than a multi-reference frame approach.

2.1   Single Reference Frame Image Registration

One common and effective single reference frame technique for image registration is based on the 2-D cross-correlation [3]. The cross-correlation technique requires complicated calculations when the geometric transforms between the two images include rotation and scaling. Another technique uses moments and moment invariants for image matching. Moments are sensitive to noise, moment invariants have limited discriminating power, and high-order moments require extensive computations [4]. Image matching using symmetric phase-only matched filters (SPOMF) have been shown to be computationally efficient, to be robust against noise, and to have high discriminating power (sharpness in correlation peak), but they are sensitive to variation in rotation and scale [5] [6].

Ideally, techniques are needed that are invariant to some of the geometric transforms relating two images, thus reducing complexity and dimensionality. An invariant technique, if applied to a distorted image, results in an output that is equivalent to the undistorted input image, assuming that the technique is invariant for the specific distortions of the image. One such technique is the Fourier-Mellin transform (FMT) [1] [7]. The FMT is invariant to translation, rotation, and scale, making it

9

an ideal image matching technique. Another such technique is the Radon Transform (RT) [8] [9] [10] [11]. This thesis focuses on registration using cross-correlation, FMT, and RT.

2.1.1 Cross-Correlation. Cross-correlation is commonly used in many signal processing applications such as image registration and pattern recognition. It measures the similarity in shape of two signals. The un-normalized cross-correlation of two discrete vectors $g(n)$ and $h(n)$ both of size N is

$$c(m)\{g,h\} = \begin{cases} \sum\limits_{n=1}^{m} g(n)h^*(n+(N-m)), & 1 \leq m \leq N \\ \sum\limits_{n=1}^{2N-m} g(n+(m-N))h^*(n), & N+1 \leq m \leq 2N-1. \end{cases} \qquad (2.1)$$

If one vector is shorter than the other, then the shortest is zero padded to the length of the other vector. Cross-correlation computes a sum of the product of the two vectors for all possible combinations of overlap. The location of the maximum of the cross-correlation can be used to estimate the translation between two vectors.

If the cross-correlation is used for estimating translation between two discrete vectors and a reasonable estimation for the maximum possible translation in either direction, $\hat{\tau}_{max}$, is known, then the cross-correlation can be modified to possibly increase accuracy. Using one vector as a reference vector, the non-reference vector is windowed by removing samples from the ends, where the amount removed depends on the estimated maximum possible translation. For example, if the two vectors are 128 samples and the estimated maximum shift is $\hat{\tau}_{max} = 10$ samples, then 10 samples are removed from each end of the non-reference vector, resulting in a reference vector with 128 samples and another vector of 108 samples.

Here equation (2.1) does not apply because the shorter vector should not be zero padded and partial overlaps should not be calculated. Not calculating partial overlap has the added benefit of reducing the number of computations required to

perform the correlation. The windowed cross-correlation of two discrete vectors $g(n)$ and $h(p)$ of size N and P, respectively, is

$$c_{win}(m)\{g, h\} = \sum_{n=1}^{N} g(n)h^*(n + m - 1), \quad 1 \le m \le 2\hat{\tau}_{max} + 1, \qquad (2.2)$$

where $N = P - 2\hat{\tau}_{max}$ and $N \ge 4\hat{\tau}_{max}$. The computations of the windowed cross-correlation is reduced to $2\hat{\tau}_{max} + 1$ compared to the $2N - 1$ computations of equation (2.1).

To possibly even further improve the performance of cross-correlation for estimating translation between vectors, the vectors can be normalized. There are many ways to normalize vectors, but the focus focus here is on normalizing by the power of each vector. The vector g normalized by its power is

$$\breve{g} = \frac{g}{\sqrt{\langle g, g \rangle}}, \qquad (2.3)$$

where $\langle g, g \rangle$ is the inner product of g with itself. Applying normalization to equation (2.2), results in the normalized, windowed cross-correlation

$$c_{win}^{norm}(m)\{g, h\} = \sum_{n=1}^{N} \breve{g}(n)\breve{h}^*(n + m - 1), \quad 1 \le m \le 2\hat{\tau}_{max} + 1. \qquad (2.4)$$

The term normalized, windowed cross-correlation is misleading because it is not the cross-correlation that is normalized but rather the vectors being correlated.

Cross-covariance is similar to cross-correlation except that cross-covariance uses mean-removed vectors. The un-normalized cross-covariance of two discrete vectors

$g(n)$ and $h(n)$ both of size N is

$$v(m)\{g,h\} = \begin{cases} \displaystyle\sum_{n=1}^{m}(g(n)-\mu_g)(h^*(n+(N-m))-\mu_h) & 1 \le m \le N \\ \displaystyle\sum_{n=1}^{2N-m}(g(n+(m-N))-\mu_g)(h^*(n)-\mu_h) & N+1 \le m \le 2N-1, \end{cases}$$

$$(2.5)$$

where μ_g and μ_h are the means of $g(n)$ and $h(n)$, respectively. The terms cross-correlation and cross-covariance are often used interchangeably despite their distinct differences. The same normalization and windowing used for cross-correlation can be applied to the cross-covariance to possibly achieve better performance.

Until now, all cross-correlations discussed here have used vectors, but cross-correlation can also apply to 2-D functions, such as images. Equations (2.1), (2.2), and (2.4) can be expanded to the 2-D case, but would require drastically more computations. To resolve this computational complexity issue, the spectral implementation of the cross-correlation can be used. Let g and h be digital images and let G^F and H^F be their discrete Fourier transforms (DFT), respectively. The spectral implementation of the cross-correlation is the inverse DFT of the cross-power spectrum of g and h:

$$C\{g,h\} = \mathfrak{F}^{-1}\{G^F(H^F)^*\}, \tag{2.6}$$

where $\mathfrak{F}^{-1}$ is the inverse DFT and $G^F(H^F)^*$ is the cross-power spectrum of g and h. The DFT and inverse DFT are fast implementations of the Fourier transform (FT) and, therefore, are an ideal choice for 2-D cross-correlation.

2.1.2 Phase Correlation.   One special case of the spectral implementation of the cross-correlation is called phase correlation [3]. As its name implies, phase correlation measures the correlation between two images using the phase of the FT. To extract the phase from g and h, the cross-power spectrum is normalized by its magnitude

$$p\{g,h\} = \frac{G^F(H^F)^*}{|G^F(H^F)^*|}. \tag{2.7}$$

Phase correlation is the inverse FT of the normalized cross-power spectrum

$$P\{g,h\} = \mathfrak{F}^{-1}\{p\}, \tag{2.8}$$

which is equivalent to the normalized spectral implementation of the cross-correlation

$$C^{norm}\{g,h\} = \mathfrak{F}^{-1}\left\{\frac{G^F(H^F)^*}{|G^F(H^F)^*|}\right\} = P\{g,h\}. \tag{2.9}$$

Relative translative movement between g and h can also be estimated using the location of the maximum of the phase correlation.

2.1.3 Radon Transform. The RT recently received attention for use in a wide range of applications, one of which is image registration. There are several definitions of the RT, all of which are related. The Radon transform $H^R(r,\theta)$ for a 2-D continuous function $h(x,y)$ is found by computing line integrals along h, where the lines are defined by their perpendicular distance from the origin, r, and the angle that r makes with the horizontal axis, θ, as in Figure 2.1. The RT of a two-dimensional function $h(x,y)$ is

$$\mathfrak{R}\{h(x,y)\} = \int\int h(x,y)\delta(r - x\cos\theta + y\sin\theta)dxdy = H^R(r,\theta). \tag{2.10}$$

For discrete applications such as digital image registration, the line integrals in equation (2.10) are replaced by projections, and interpolation is required because the projections do not always pass through the centers of pixels. There are many algorithms to compute the discrete RT. Here, the discrete RT is computed using the MATLAB® command `radon`, as in Figure 2.2. An example of the discrete RT is shown in Figure 2.3.

The power of the RT for image registration comes from its unique properties with respect to rotation, translation, and scale. The 2-D image $f_\phi(x,y)$, as shown in

13
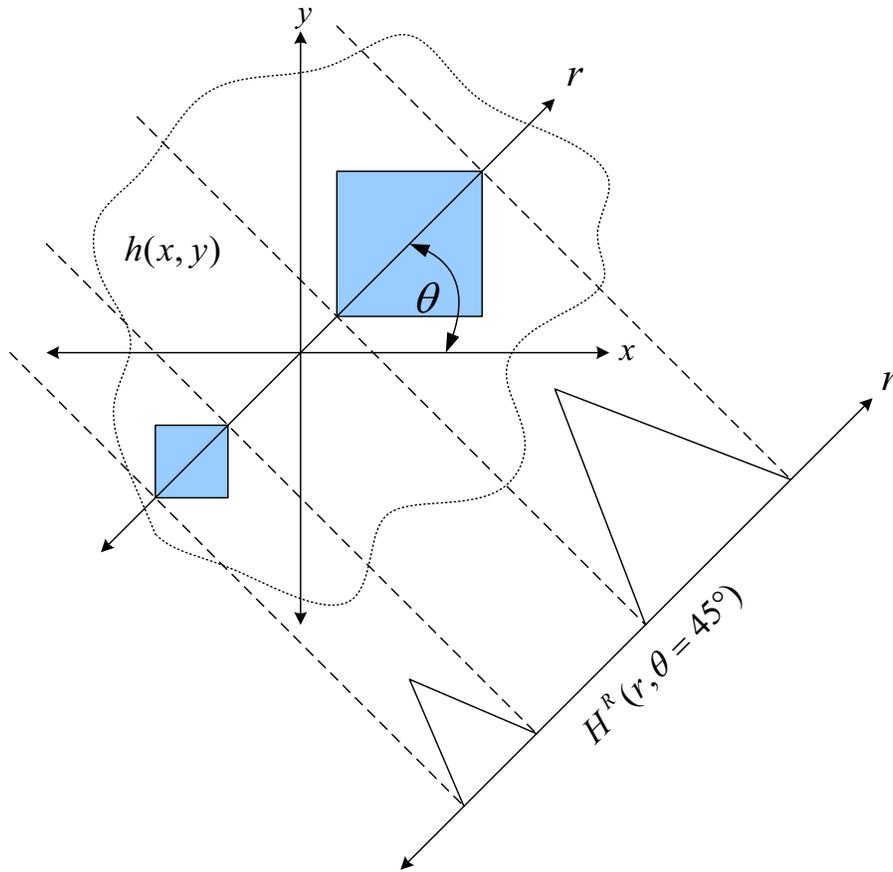
Figure 2.1: Illustration of the continuous RT. The line integrals across h, shown as dotted lines, are defined by their distance along r and the angle r makes with the x-axis, θ. The origin for the RT is the same as the origin of the the function $h(x, y)$. The result of the RT for $\theta = 45°$ is also shown. This figure is adapted from [8].
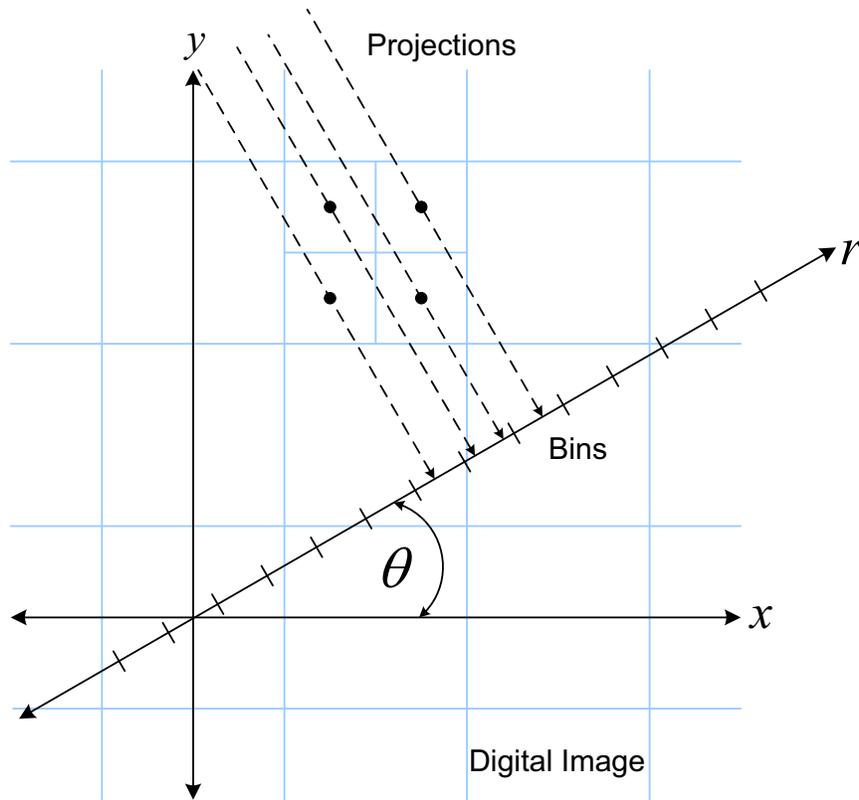
Figure 2.2:   The discrete RT of an image is the sum of RTs of each individual pixel. The `radon` command divides each pixel in the image into four subpixels, each with the same value as the original pixel, and projects each subpixel separately. Each subpixels contribution is proportionally split into the two nearest bins according to the distance between the projected location and the bin centers. If the subpixel projection intersects the center of a bin, the bin has the full value of the subpixel. If the subpixel projection intersects the border between two bins, the subpixel value is divided evenly between bins. The origin of the RT and the image axes is the center pixel of the image. This figure and description is adapted from the MATLAB® help document for the `radon` command.

(a) $f(x, y)$

(b) $\Re\{f(x, y)\}$

(c) $F^R(r, \theta = 45°)$
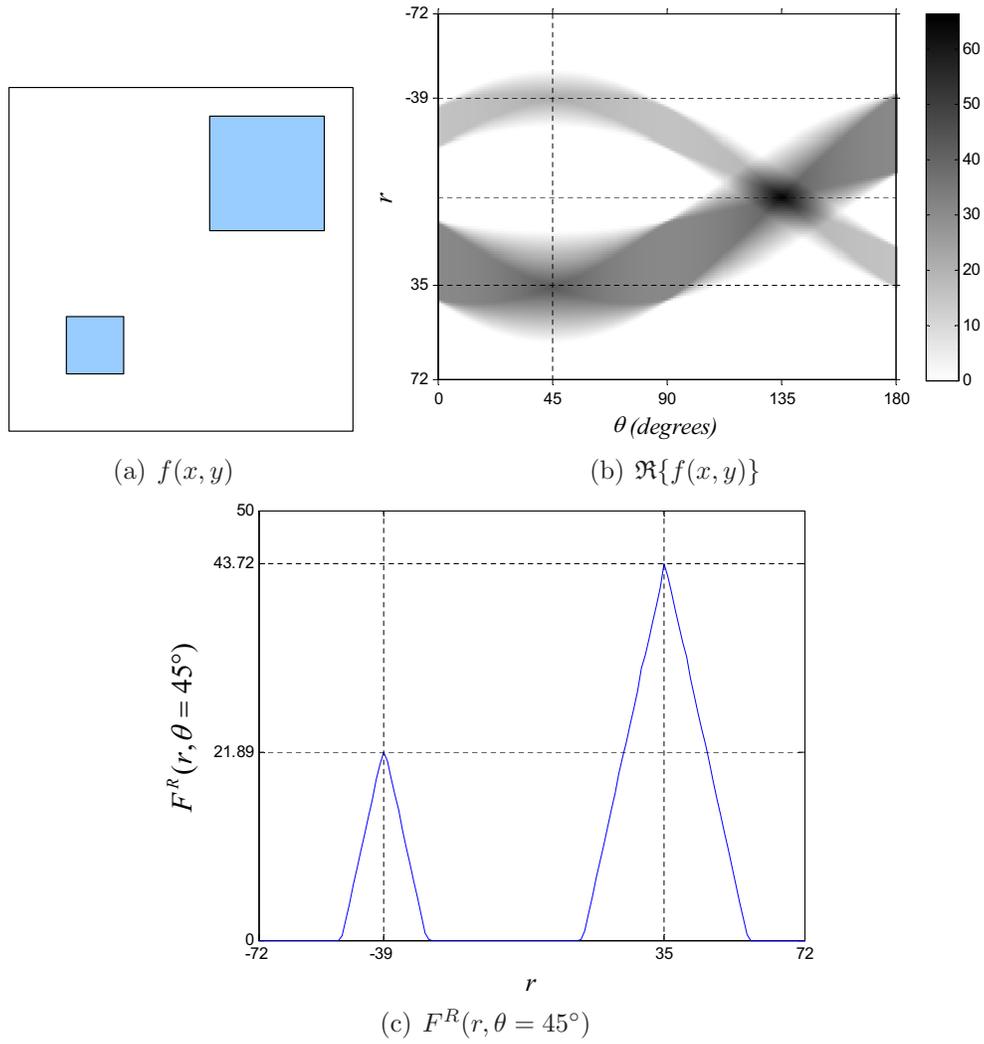
Figure 2.3: The RT of the digital image $f(x, y)$. Also shown is the discrete RT for $\theta = 45°$. Note the similarity to Figure 2.1.

equation (1.2), can be represented by its RT as

$$\Re\{f_\phi(x,y)\} = F^R(r, \theta - \phi) = F_\phi^R(r, \theta). \tag{2.11}$$

This equation shows that a rotation of an image corresponds to a translation of the angular variable of the RT [8], as in Figure 2.4. The image $f_\tau(x, y)$, as in equation (1.1), can be represented by its RT as

$$\Re\{f_\tau(x,y)\} = F^R(r - \tau_x \cos\theta + \tau_y \sin\theta, \theta) = F_\tau^R(r, \theta). \tag{2.12}$$

Therefore, translating an image corresponds to a translation of the spatial variable of the RT by an amount that depends on both the image translation, $[\tau_x, \tau_y]$, and the angular variable, θ, as in Figure 2.5.

However, for a given θ the RT of a translated image is translated along the spatial variable. Certain choices of θ lead to direct estimation of $[\tau_x, \tau_y]$, as in Figure 2.6. The RT of $f_\tau(x, y)$ evaluated for $\theta = 0$ is

$$
\begin{aligned}
F_\tau^R(r, \theta = 0) &= F^R(r - \tau_x \cos(0) + \tau_y \sin(0), 0) \\
&= F^R(r - \tau_x, 0).
\end{aligned}
\tag{2.13}
$$

This choice of θ eliminates the effect of vertical translation. Therefore, the vector $\Re\{f_\tau(x, y)\}$ for $\theta = 0$ is translated along the spatial domain by τ_x relative to the vector $\Re\{f(x, y)\}$ for $\theta = 0$.

The RT of $f_\tau(x, y)$ evaluated for $\theta = 90°$ is

$$
\begin{aligned}
F_\tau^R(r, \theta = 90°) &= F^R(r - \tau_x \cos(90°) + \tau_y \sin(90°), 90°) \\
&= F^R(r + \tau_y, 90°).
\end{aligned}
\tag{2.14}
$$

This choice of θ eliminates the effect of horizontal translation. Therefore, the vector $\Re\{f_\tau(x, y)\}$ for $\theta = 90°$ is translated along the spatial domain by τ_y relative to the vector $\Re\{f(x, y)\}$ for $\theta = 90°$. The translations τ_x and τ_y can be directly ex-

(a) $\Re\{f(x,y)\}$

(b) $\Re\{f_\phi(x,y)\}$
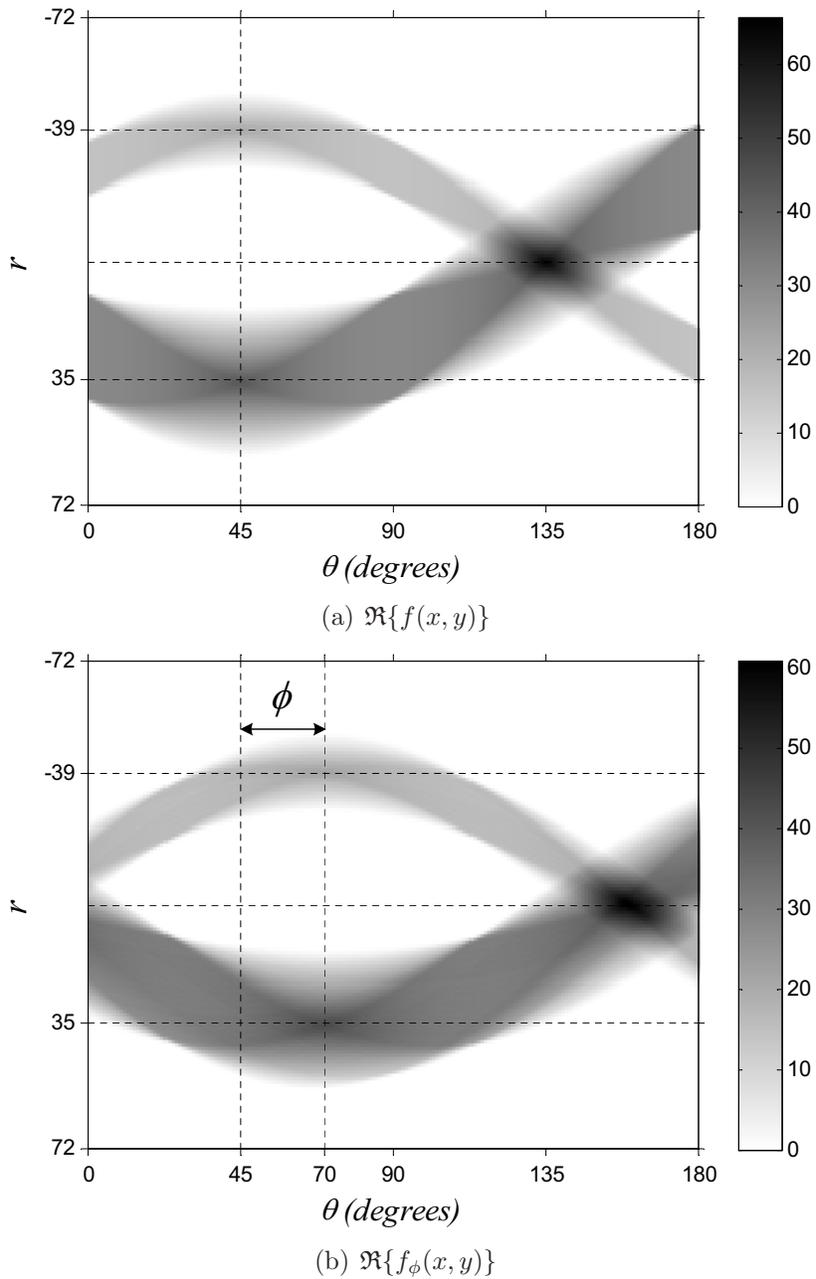
Figure 2.4:   The RT of the digital image $f(x,y)$ and the rotated digital image $f_\phi(x,y)$ displayed together for visual comparison. The rotated image is by 25 degrees. Note that the RT of the rotated image is by shifted $\phi = 25°$ relative to the original image.

tracted using the location of the maximum of the cross-correlation of $\Re\{f(x,y)\}$ and $\Re\{f_\tau(x,y)\}$ for $\theta = 0°$ and $90°$, respectively.

Lastly, the image $f_\beta(x,y)$ can be represented by its RT as

$$\Re\{f_\beta(x,y)\} = \frac{1}{\beta} F^R(\beta r, \theta) = F_\beta^R(r,\theta). \tag{2.15}$$

Scaling an image corresponds to scaling the spatial variable of the RT and scaling the intensity values [8]. The combination of these properties make the RT a powerful technique for image registration.

2.1.4 The Fourier-Mellin Transform.   The FMT is an application of the shifting property of the FT combined with a log-polar coordinate mapping. The shifting property of the FT states that a spatial shift, τ_o, corresponds to a linear phase change in the spectral domain [12]

$$\mathfrak{F}\{f(x-\tau_o)\} = F(e^{jw})e^{-jw\tau_o}. \tag{2.16}$$

The magnitude of the FT of the $f(x-\tau_o)$ is

$$|\mathfrak{F}\{f(x-\tau_o)\}| = |F(e^{jw})e^{-jw\tau_o}| = F(e^{jw}) = \mathfrak{F}\{f(x)\}. \tag{2.17}$$

Therefore, the magnitude of the FT is translation invariant.

This same property can be applied to both rotation and scaling if log-polar coordinate mapping is used. Using a point (x,y) in the real-valued Cartesian coordinate plane, the log-polar mapping is

$$x = e^\mu \cos(\theta), \tag{2.18}$$

$$y = e^\mu \sin(\theta), \tag{2.19}$$
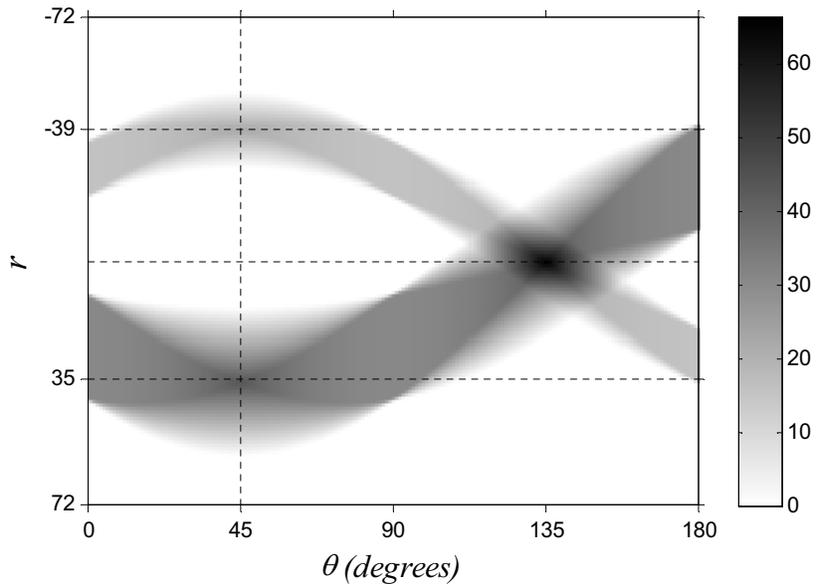
(a) $\Re\{f(x, y)\}$
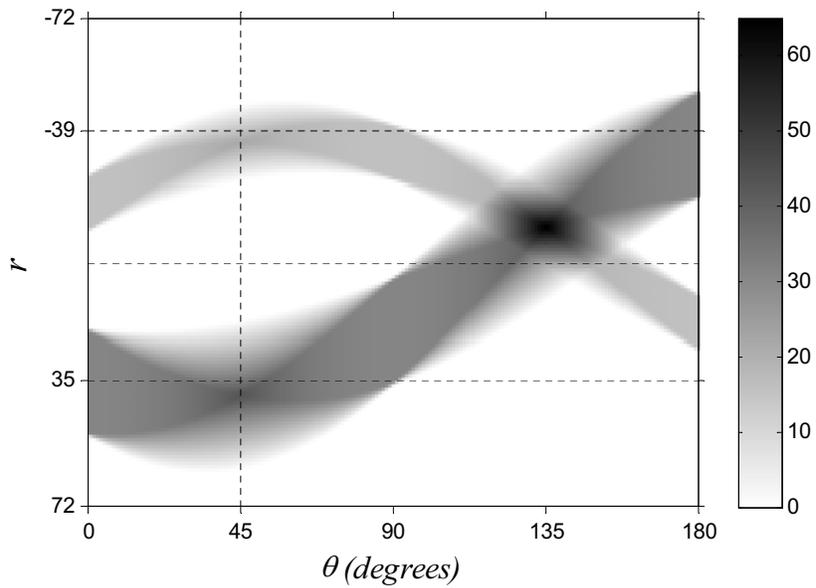
(b) $\Re\{f_\tau(x, y)\}$

Figure 2.5:   The RT of the digital image $f(x, y)$ and the same image translated by $\tau_x = 10$ pixels and $\tau_y = 5$ pixels, $f_\tau(x, y)$, displayed together for visual comparison. The RT of the translated image is warped along the r axis relative to the RT of the original image.
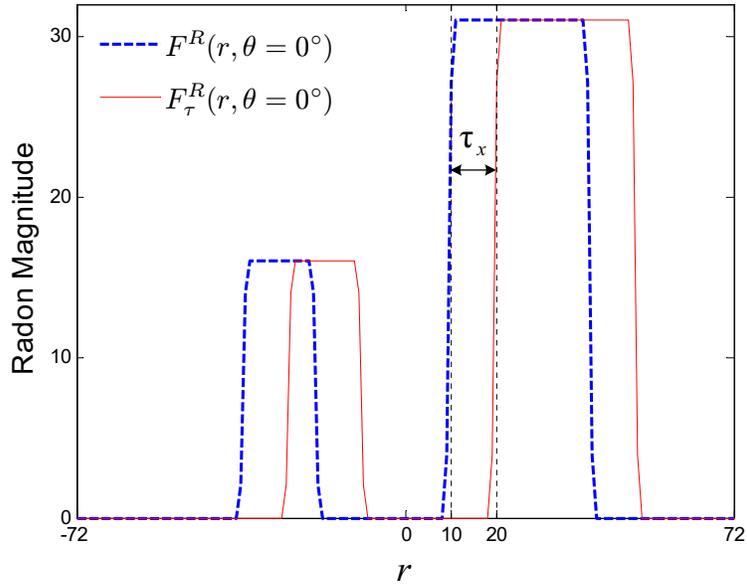
(a) $\Re\{f_\tau(x,y)\}$ for $\theta = 0°$
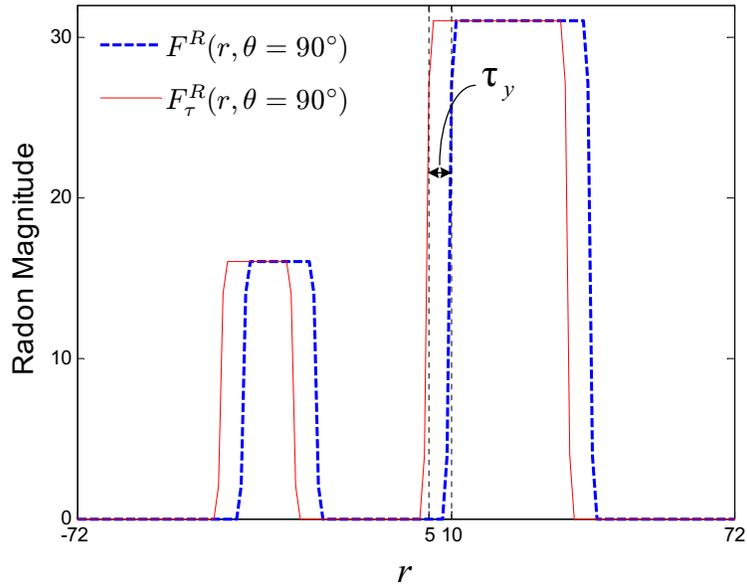
(b) $\Re\{f_\tau(x,y)\}$ for $\theta = 90°$

Figure 2.6: The RT the original image $f(x,y)$ and the translated image $f_\tau(x,y)$ for specific values of θ, showing that τ_x and τ_y can be estimated from their RTs.

21

where $e^\mu \in \Re$ is the scaling factor and $\theta \in [0, 2\pi)$ is the rotation angle. For every μ and θ there is a unique point in the Cartesian plane. A change in rotation or scale of an image corresponds to a translation of θ or μ, respectively. Therefore, the magnitude of the FT of the log-polar mapping is rotation and scaling invariant [1] in accordance with equation (2.17).

The FMT is the FT of the log-polar mapping of the magnitude of the FT of a function, as shown in Figure 2.7. The FMT of a 2-D function $h(x, y)$ is

$$\mathcal{M}\{h(x, y)\} = \mathfrak{F}\{H^F(\mu, \theta)\} = H^M(u, v), \tag{2.20}$$

where $H^F(\mu, \theta)$ is the log-polar mapping of the magnitude of the FT of $h(x, y)$. The magnitude of the FT of $h(x, y)$ is not represented directly in equation (2.20) but is implicit. The FMT is invariant to translation because it uses the magnitude of the FT of $h(x, y)$, which is translation invariant as described above.

The magnitude of the FMT is invariant to both rotation and scale. The FMT of the image $f_{\tau, \beta, \phi}(x, y)$, (see equation (1.8)), which is translated by the vector $[\tau_x, \tau_y]$, scaled by β, and rotated by ϕ, is

$$\mathcal{M}\{f_{\tau, \beta, \phi}(x, y)\} = F^M(u, v)e^{-ju\ln\beta}e^{-jv\phi} = F^M_{\tau, \beta, \phi}(u, v) \tag{2.21}$$

and its magnitude is

$$|F^M_{\tau, \beta, \phi}(u, v)| = |F^M(u, v)e^{-ju\ln\beta}e^{-jv\phi}| = F^M(u, v). \tag{2.22}$$

Therefore, rotation and scaling of an image result in linear phase shifts in the FMT. Notice that translation does not appear in the FMT and that neither rotation nor scale appear in the magnitude of the FMT, thus showing that the FMT is invariant to translation, rotation, and scale.

$$h(x, y)$$

$$\boxed{\begin{array}{c} \text{Magnitude of FT} \\ \left| \Im\{h(x, y)\} \right| \end{array}}$$

$$H^F(x, y)$$

$$\boxed{\begin{array}{c} \text{Log-Polar Mapping} \\ LP\{H^F(x, y)\} \end{array}}$$

$$H^F(\mu, \theta)$$

$$\boxed{\begin{array}{c} \text{FT of Log-Polar Mapping} \\ \Im\{H^F(\mu, \theta)\} \end{array}}$$

$$H^M(u, v)$$
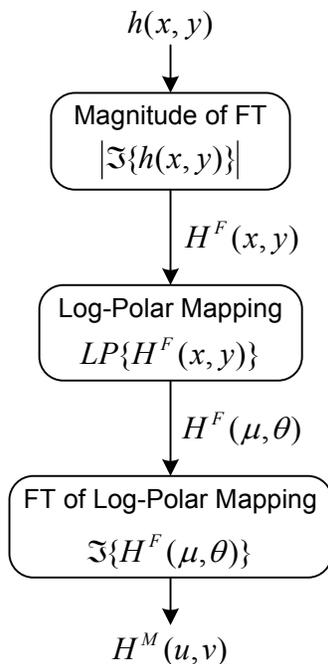
Figure 2.7:   Flow chart for computing the FMT for the 2-D function $h(x, y)$.

2.2 *Multi-Reference Frame Image Registration*

Multi-reference frame image registration aims to minimize or eliminate image biasing and to be more robust in noisy environments than traditional single reference frame registration. A multi-reference frame image registration technique for estimating translational shifts uses the location of the maximum of the two-dimensional cross-correlation as the *initial estimator*. The *initial estimator* is the single reference frame registration technique used to estimate the relative distortion between all combinations of two images [2].

An initial reference image is chosen, and all images in the set of N images are compared to the reference image to obtain relative distortion estimates. Each reference image results in a different set of distortion estimations, as in Figure 2.8. This procedure is repeated for all images in the set using a different reference image each time. To ensure comparable distortion estimates, each collection of relative estimates from a particular reference image are normalized to zero mean. Final distortion estimates are made by averaging the relative distortion estimates with respect

(a) Frame 1 Reference

(b) Frame 1 Reference Mean Removed

(c) Frame 2 Reference

(d) Frame 2 Reference Mean Removed
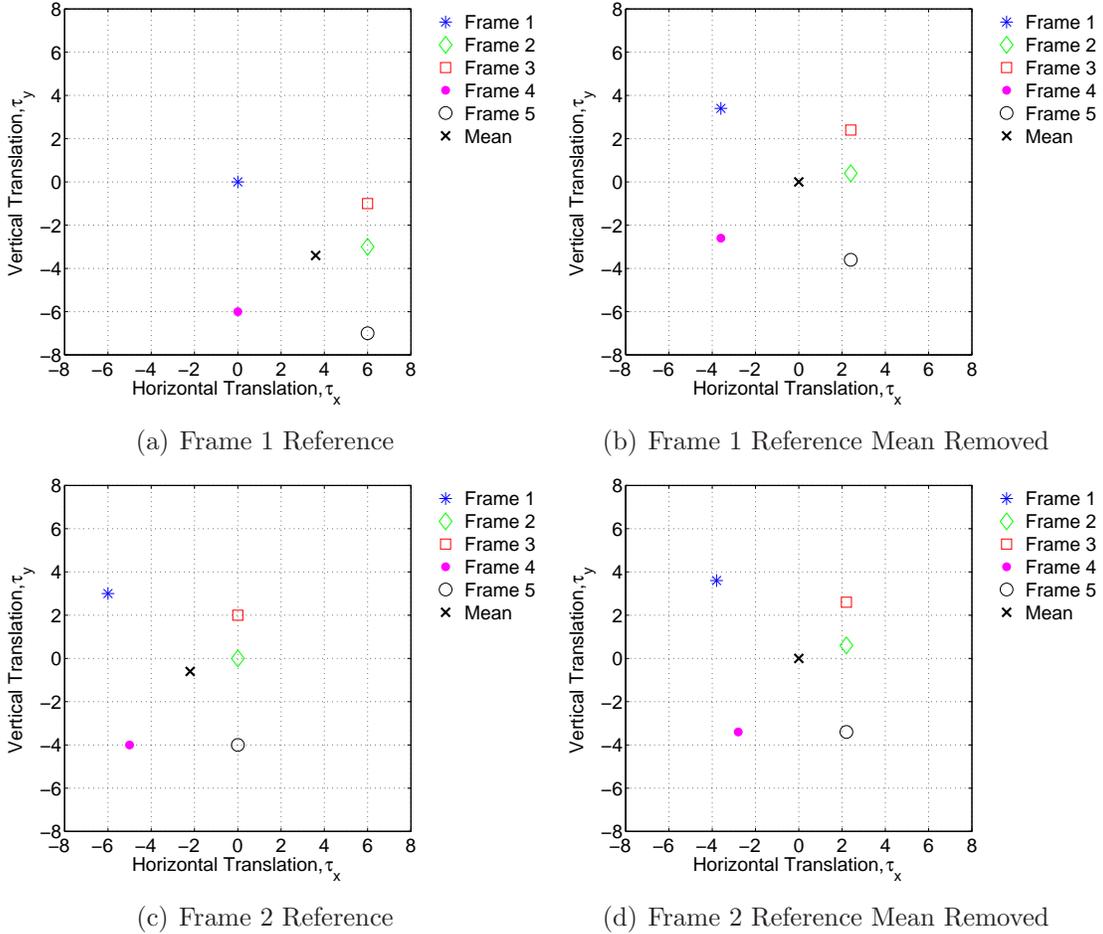
Figure 2.8:   The result of distortion estimation using different reference frames. The estimations are made from 5 frames that have random translative distortion.

to dimensionality and reference frame, as in Figure 2.9. This averaging requires N^2 estimations, which implies significantly more calculations then single reference frame registration [2].

The process is applied with a model using a $N \times N$ matrix with each element of the matrix corresponding to one distortion estimation. Let $\hat{\Omega}\{i, j\}$ be the estimation of some distortion between image i and image j. The matrix of multi-reference frame
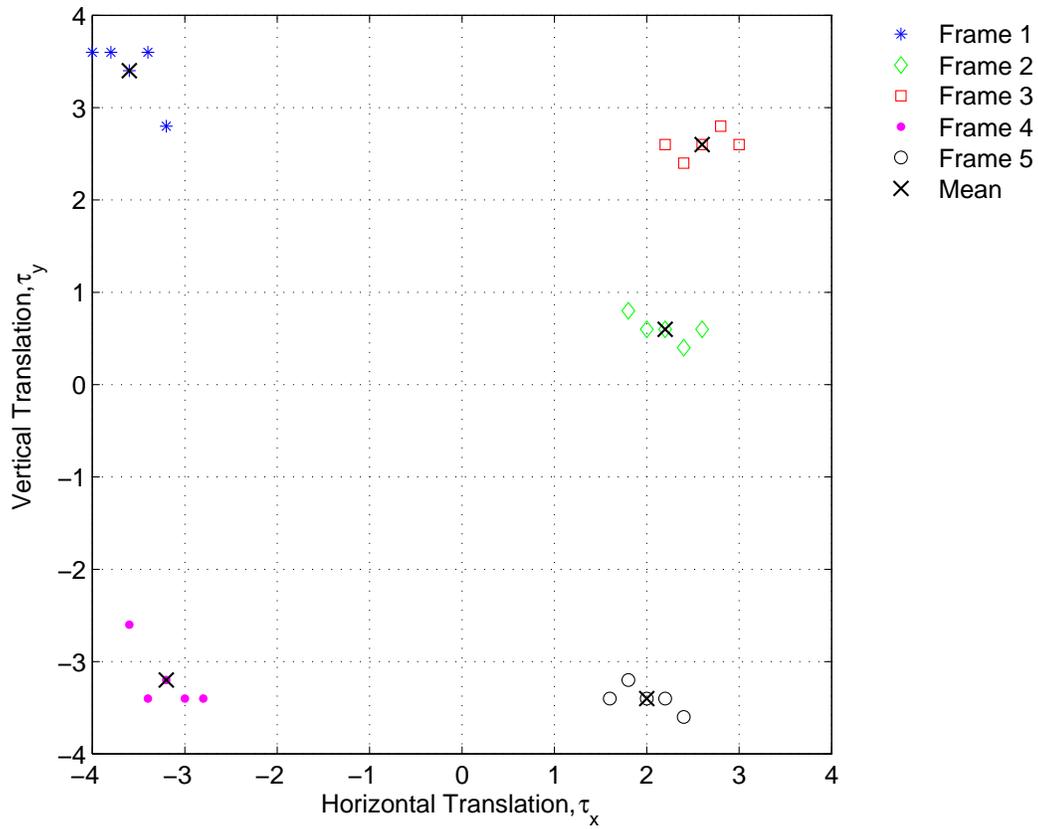
Figure 2.9:   The overlapped results of translation estimation using each image as a reference. This figure highlights the differences in estimation when using different reference frames. Multi-frame registration combines all data to compute the distortion estimation by averaging the cluster of estimations. The means of the clusters are the final distortion estimates for a given frame.

estimations is

$$\hat{\boldsymbol{\Omega}} = \begin{bmatrix} \hat{\Omega}\{1,1\} & \hat{\Omega}\{1,2\} & \cdots & \hat{\Omega}\{1,N\} \\ \hat{\Omega}\{2,1\} & \hat{\Omega}\{2,2\} & & \hat{\Omega}\{2,N\} \\ \vdots & & \ddots & \vdots \\ \hat{\Omega}\{N,1\} & \hat{\Omega}\{N,2\} & \cdots & \hat{\Omega}\{N,N\} \end{bmatrix}. \tag{2.23}$$

Simplifications can be made using symmetry. Specifically, the distortion between an image and itself is zero or one,

$$\hat{\Omega}\{i,i\} = 0 \text{ or } 1. \tag{2.24}$$

The distortion is zero for translation or rotation, and one for scaling. Also, switching the reference frame with the frame to be registered results in the same distortion estimate only negated or inverted,

$$\hat{\Omega}\{i,j\} = -\hat{\Omega}\{j,i\} \text{ or } \frac{1}{\hat{\Omega}\{j,i\}}. \tag{2.25}$$

The estimate is negated for translation or rotation, and inverted for scaling. These simplifications reduce the number of estimations to $(N^2 - N)/2$. and reduce equation (2.23) to

$$
\hat{\boldsymbol{\Omega}} = \begin{bmatrix} 0 & \hat{\Omega}\{1,2\} & \cdots & \hat{\Omega}\{1,N\} \\ -\hat{\Omega}\{1,2\} & 0 & & \hat{\Omega}\{2,N\} \\ \vdots & & \ddots & \vdots \\ -\hat{\Omega}\{1,N\} & -\hat{\Omega}\{2,N\} & \cdots & 0 \end{bmatrix}
$$

or

$$
= \begin{bmatrix} 1 & \hat{\Omega}\{1,2\} & \cdots & \hat{\Omega}\{1,N\} \\ \frac{1}{\hat{\Omega}\{1,2\}} & 1 & & \hat{\Omega}\{2,N\} \\ \vdots & & \ddots & \vdots \\ \frac{1}{\hat{\Omega}\{1,N\}} & \frac{1}{\hat{\Omega}\{2,N\}} & \cdots & 1 \end{bmatrix}. \tag{2.26}
$$

26

Another consequence of these simplifications is that the initial distortion estimates do not need to be normalized because the sum of all values in each matrix is zero. The final vector of distortion estimates is calculated by averaging the columns or rows of the matrix of multi-reference frame estimations [2]:

$$\hat{\Omega}_M(m) = \frac{\sum\limits_{n=1}^{N} \hat{\Omega}(m,n)}{N}, \ \forall m \in \{1, 2, \ldots, N\}. \tag{2.27}$$

This multi-reference frame registration method is independent of the distortion model and can account for prior knowledge of the distortion by using a different initial estimator. Multi-frame registration is an unbiased estimator if the initial estimator used is unbiased. It is shown that the multi-reference frame registration for shift estimation (using cross-correlation as the initial estimator) performs better than single reference frame cross-correlation [2].

2.2.1 Optimal Gauss-Markov Estimator. The new multi-reference frame estimation technique is equivalent to the optimal Gauss-Markov estimator for distortion estimation, as in [13]. In [13], the Gauss-Markov estimator is used for time difference of arrival estimation in radio navigation given all pairwise estimations of delay. The Gauss-Markov estimate for the vector, D, of $N-1$ true distortions relative to the first image is

$$\hat{D} = [A^T P_E^{-1} A]^{-1} A^T D_M, \tag{2.28}$$

where

$$D\{1, k\} = [\Omega\{1, k\}]^T, \ \text{for } k \in \{2, \ldots, N\}, \tag{2.29}$$

D_M is the vector of $(N^2 - N)/2$ distortion estimates for all N images taken two at a time or

$$D_M\{i, j\} = [\hat{\Omega}\{1,2\}, \hat{\Omega}\{1,3\}, \cdots, \hat{\Omega}\{1,N\}, \hat{\Omega}\{2,3\}, \cdots, \hat{\Omega}\{N-1, N\}]^T, \tag{2.30}$$

P_E is the normalized error covariance of the distortion estimates which for independent and identically distributed estimates is equal to the identity matrix of size $(N^2 - N)/2$ or

$$A = \delta\{j, k\} - \delta\{i, k\}, \tag{2.31}$$

and δ is the Kronecker delta function

$$\delta\{i, j\} = \left\{ \begin{array}{ll} 1, & \text{for } i = j \\ 0, & \text{for } i \neq j \end{array} \right\}. \tag{2.32}$$

Using the multi-frame estimates, $\hat{\Omega}_M$, it can be shown that the Gauss-Markov estimator is equal to the multi-frame estimates with respect to the first image

$$\hat{D} = \hat{\Omega}_M(1) - \left[\begin{array}{c} \hat{\Omega}_M(2) \\ \vdots \\ \hat{\Omega}_M(N) \end{array} \right]. \tag{2.33}$$

Appendix A provides an example of the calculations that lead to the result in equation (2.33).

III. Experimental Research Methodology

The basic methodology for the experiments developed in this thesis involves single and multi-reference frame registration using a set of N images. Here, single reference frame registration is also called the single frame method, and multi-reference frame registration is also called the multi-frame method. The set of N images are multiple images of a true object taken at different times and/or perspectives, which results in relative geometric distortions of translation, rotation, and/or scale. The final results of registration are estimates of the relative distortions and an estimate of the true object or scene for both the single and multi-frame methods. The goal is to compare the quality of the scene estimate and the performance of the methods used for image registration.

Comparisons of the single and multi-frame methods are made for different combinations of geometric distortions using appropriate registration algorithms to estimate and remove the distortions. The first experiment, focuses on data that has relative translation and uses cross-correlation to register the data as in [2]. The second experiment, uses data distorted translationally and rotationally and registered using the RT and cross-correlation. The last experiment, uses images translated, rotated and scaled, and registered using the FMT and phase correlation. For each experiment a simulated data set is used for comparing the performance of single and multi-frame registration.

3.1   Simulated Experiment Setup

Simulated data is used to compare the single and multi-frame methods. Initially, an over-sampled image (1024 by 1024 pixels) is converted to a grayscale image, $I(x,y)$, with magnitude values from 0 to 255. The grayscale image is randomly distorted N times using a mathematical model, found in section 1.1.1, to simulate geometric distortions from capturing multiple images of the same object. This procedure results in N images that are randomly distorted relative to a true image, $I_\Omega(x,y,n)$, and a vector of the true distortion, $\Omega(n)$. The grayscale undistorted image is the true

image. Next, Gaussian white noise is added to all images. Interpolation is required to simulate rotation and scale. Also, to correct/remove rotation or scale interpolation is required, although this need is not unique to simulated data. The same interpolation techniques are used for the single and multi-frame methods.

Thus the simulated data consists of N distorted, noisy images and one undistorted, noiseless image (the true image). An unrealistic limitation of simulated data is that translation can only be an integer number of pixels unless computationally intense interpolation is used. In reality, multiple images of the same target has noninteger pixel shifts between frames. To account for sub-pixel shifts, the simulated data is downsampled. Downsampling also has the benefit of reducing the computational time needed for registration. Downsampling yields the final simulated data, $f_\Omega(x, y, n) + n(x, y, n)$, consisting of N noisy images that are randomly and realistically distorted between frames and one undistorted, noiseless image, $f(x, y)$, (the true image), all of which are 256 by 256 pixels.

The simulated data is registered using single and multi-frame methods. The results of registration are distortion estimates, $\hat{\Omega}(n)$, and a scene estimate, $\hat{f}(x, y)$. The root mean square error (RMSE) of the distortion estimates is computed and the signal-to-noise ratio (SNR) of the scene estimate is computed. The SNR and the RMSE provide a comparison of the performance of the single and mutli-frame registration methods. Figure 3.1 provides a flow of the methodology used for the simulated experiments.

3.1.1 Metrics. Appropriate metrics are required to compare the multi-frame and single frame methods for the simulated experiments. The SNR is a popular image quality metric; it is the ratio of signal power to noise power. The SNR of the true

image, $f(x, y)$, and its estimate, $\hat{f}(x, y)$, is

$$SNR(f, \hat{f}) = \frac{\sum_{n=1}^{N} \sum_{m=1}^{M} \left[f(n, m) - \frac{1}{RS} \sum_{r=1}^{R} \sum_{s=1}^{S} f(r, s) \right]^2}{\sum_{n=1}^{N} \sum_{m=1}^{M} \left[\hat{f}(n, m) - f(n, m) \right]^2}. \qquad (3.1)$$

The scene estimate, $\hat{f}(x, y)$, is the result of averaging the N images that are commonly aligned using the distortion estimates, $\hat{\Omega}(n)$. However, the scene estimates and the true image are not commonly aligned and, therefore, the scene estimates are registered and aligned to the true image before calculating the SNR.

The RMSE is a standard metric for measuring registration accuracy. For the simulated data, the geometric distortions are artificially applied to the images, and, therefore, the true values are known. The distortion estimates from the two methods can be compared with the known, true distortion values when using simulated data.

The distortion estimates from the two methods cannot be directly compared to the true distortion values because the true values and estimated values have a different reference point from which they are measured. Therefore, the differences of true and estimated values are used to compute the error. A difference matrix for the true distortion, $\Omega(n)$, is computed, where each element corresponds to a difference between distortion values,

$$\Delta\boldsymbol{\Omega} = \begin{bmatrix} \Omega(1) - \Omega(1) & \cdots & \Omega(1) - \Omega(N) \\ \vdots & \ddots & \vdots \\ \Omega(N) - \Omega(1) & \cdots & \Omega(N) - \Omega(N) \end{bmatrix}, \qquad (3.2)$$

and similarly for the estimates, $\Delta\hat{\boldsymbol{\Omega}}$. These matrices are compared using the RMSE. The RMSE between the known distortion differences, $\Delta\boldsymbol{\Omega}$, and the estimated distor-

tion differences, $\Delta\hat{\mathbf{\Omega}}$, is

$$RMSE(\Delta\mathbf{\Omega}, \Delta\hat{\mathbf{\Omega}}) = \sqrt{\frac{\sum_{n=1}^{N}\sum_{m=1}^{N}(\Delta\mathbf{\Omega}(n,m) - \Delta\hat{\mathbf{\Omega}}(n,m))^2}{N^2}}. \tag{3.3}$$

However, the distortion differences have zeros along the diagonal, which creates a bias in the RMSE. To remove this bias, the zeros along the diagonal of the difference matrix are not used in the calculation of the RMSE leaving only $N(N-1)$ calculations and resulting in a new RMSE

$$RMSE(\Delta\mathbf{\Omega}, \Delta\hat{\mathbf{\Omega}}) = \sqrt{\frac{\sum_{n=1}^{N}\sum_{\substack{m=1 \\ m\neq n}}^{N}(\Delta\mathbf{\Omega}(n,m) - \Delta\hat{\mathbf{\Omega}}(n,m))^2}{N(N-1)}}. \tag{3.4}$$

Also, the error is calculated using $(N-1)^2$ more calculations then needed because the vector of N estimates is used to create a matrix of N^2 estimates . Therefore, the actual RMSE is calculated using $\sqrt{\frac{N(N-1)}{(N-1)^2}}$RMSE or $\sqrt{\frac{N}{(N-1)}}$RMSE. The RMSE of the distortion estimates is

$$RMSE(\Delta\mathbf{\Omega}, \Delta\hat{\mathbf{\Omega}}) = \frac{\sqrt{\sum_{n=1}^{N}\sum_{\substack{m=1 \\ m\neq n}}^{N}(\Delta\mathbf{\Omega}(n,m) - \Delta\hat{\mathbf{\Omega}}(n,m))^2}}{(N-1)}. \tag{3.5}$$

3.2 Experiment for Translative Distortion

This section further develops the work of Bruckart [2] in the area of multi-frame image registration for translation only.

The methodology for the experiment for translation applies the spectral implementation of the 2-D cross correlation, equation (2.6). The model $f_\tau(x, y)$, as in equation (1.1), is used to simulate the translative distortion. The translation $[\tau_x, \tau_y]$
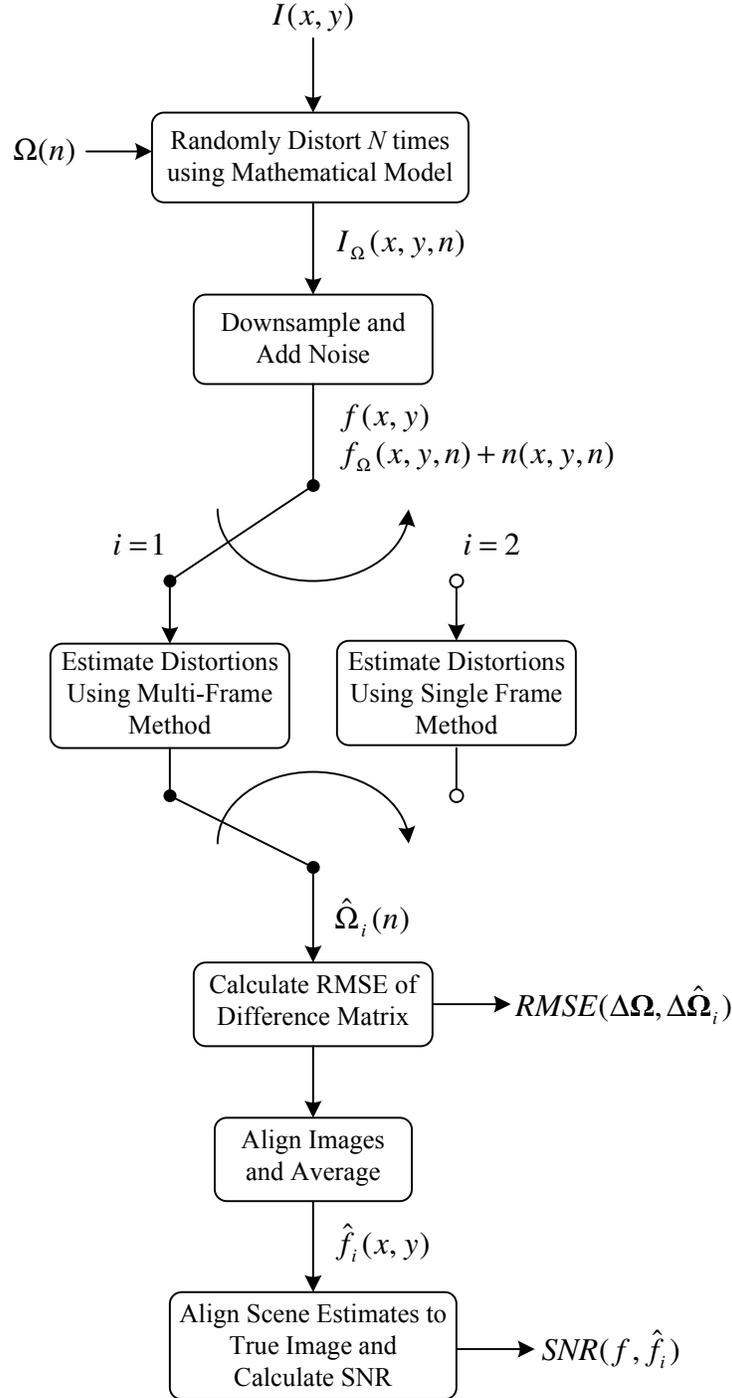
Figure 3.1: Process flow in the simulated experiments for analyzing the performance of single and multi-frame image registration methods. The chart outlines the process of using an oversampled gray scale image, $I(x, y)$, to create a simulated data set of randomly distorted images and to register them using single and multi-frame registration methods.
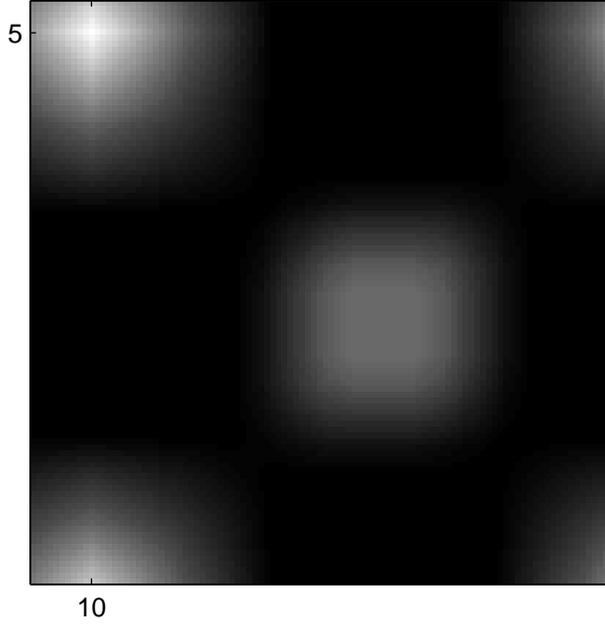
Figure 3.2:   The 2-D cross correlation of the image $f(x,y)$, Figure 2.3(a), and the same image translated by $\tau_x = 10$ pixels and $\tau_y = 5$ pixels, $f_\tau(x,y)$. The location of the maximum is indicated by the labelled markers, which correspond to the estimates $\hat{\tau}_x$ and $\hat{\tau}_y$, in accordance with equation (3.6).

is estimated from the location of the maximum of the cross-correlation of $f(x,y)$ and $f_\tau(x,y)$

$$[\hat{\tau}_x, \hat{\tau}_y] = \arg\max_{x,y} C\{f(x,y), f_\tau(x,y)\}, \tag{3.6}$$

as in Figure 3.2. After the estimates for translation are computed, the distortion can be removed. Once the shifts are removed the images are all commonly aligned and can be averaged together, which reduces the noise. This process is image averaging. The registration is performed using both single and multi-frame methods, resulting in two reconstructed images, or scene estimates, of the true object. The goal is to compare the quality of the two scene estimates and the performance of the multi-frame and single frame methods.

3.3   Experiment for Translative and Rotational Distortion

The methodology for the experiment for translation and rotation combines normalized, windowed 1-D cross correlation in the spatial domain and properties of the

RT. As stated in section 2.1.3, translations in an image correspond to a translation in the spatial parameter of the RT as a function of angle parameter, as in equation (2.12). Also, rotation in an image corresponds to a translation in the angle parameter of the RT, as in equation (2.11). An effective registration method is developed using a combination of these properties for images that are rotated and translated.

Several papers use the RT or a combination of the RT and other methods for image registration [8] [9] [10] [11] [14]. However, few of those papers explicitly and properly deal with registering images that are rotated *and* translated. The few that do so use images that contain a white shape on a black background, have different textures, or contain lines or borders, all of which are cases where the RT excels in discriminating translation and/or rotation. When realistic images are used these methods do not work accurately for a wide range of images. Therefore, it is necessary to develop a novel approach for rotation and translation estimation using the RT for realistic images.

A model for rotated and translated images is needed to develop the new method. There is a difference in the model for an image that is first translated and then rotated versus an image that is first rotated and then translated, as indicated in section 1.1.1. The 2-D image $f_{\tau,\phi}(x, y)$, equation (1.4), is represented by its RT as

$$\Re\{f_{\tau,\phi}(x, y)\} = F^R(r - \tau_x \cos(\theta - \phi) + \tau_y \sin(\theta - \phi), \theta - \phi) = F^R_{\tau,\phi}(r, \theta). \qquad (3.7)$$

The 2-D image $f_{\phi,\tau}(x, y)$, equation (1.5), can be represented by its RT as

$$\Re\{f_{\phi,\tau}(x, y)\} = F^R(r - \tau'_x \cos(\theta - \phi) + \tau'_y \sin(\theta - \phi), \theta - \phi) = F^R_{\phi,\tau}(r, \theta), \qquad (3.8)$$

where

$$\tau_x = \tau'_x \cos \phi + \tau'_y \sin \phi, \qquad (3.9)$$

$$\tau_y = -\tau'_x \sin \phi + \tau'_y \cos \phi, \qquad (3.10)$$

and

$$\tau_x' = \tau_x \cos \phi - \tau_y \sin \phi, \tag{3.11}$$

$$\tau_y' = \tau_x \sin \phi + \tau_y \cos \phi. \tag{3.12}$$

The vector $[\tau_x, \tau_y]$ is the rotational transformation of the vector $[\tau_x', \tau_y']$. The order of translation and rotation does not matter for registering the images, but if the parameters ϕ and $[\tau_x, \tau_y]$ are to be estimated, then the order is important. For simplicity, equation (3.7) is used as the model.

This model is developed as an iterative algorithm to estimate the relative translation and rotation between a pair of images. To estimate τ_x, let $\theta - \phi = 0°$ and

$$
\begin{aligned}
F_{\tau,\phi}^R(r, \theta = \phi) &= F^R(r - \tau_x \cos(0) + \tau_y \sin(0), 0) \\
&= F^R(r - \tau_x, 0).
\end{aligned}
\tag{3.13}
$$

Therefore, if ϕ is known then the vector $\Re\{f_{\tau,\phi}(x, y)\}$ for $\theta = \phi$ is translated along the spatial domain by τ_x relative to the vector $\Re\{f(x, y)\}$ for $\theta = 0$. The translation is estimated using the maximum of the 1-D normalized, windowed cross-covariance between the pair of vectors

$$\hat{\tau}_x(\tilde{\phi}) = \arg \max_m v_{win}^{norm}(m)\{F^R(r, 0), F_{\tau,\phi}^R(r, \theta = \tilde{\phi})\}. \tag{3.14}$$

To estimate τ_y, let $\theta - \phi = 90°$ and

$$
\begin{aligned}
F_{\tau,\phi}^R(r, \theta = 90° + \phi) &= F^R(r - \tau_x \cos(90°) + \tau_y \sin(90°), 90°) \\
&= F^R(r + \tau_y, 90°).
\end{aligned}
\tag{3.15}
$$

Therefore, if ϕ is known, then the vector $\Re\{f_{\tau,\phi}(x, y)\}$ for $\theta = 90° + \phi$ is translated along the spatial domain by τ_y relative to the vector $\Re\{f(x, y)\}$ for $\theta = 90°$. The translation is again estimated using the maximum of the 1-D normalized, windowed

cross-covariance between the pair of vectors

$$\hat{\tau}_y(\tilde{\phi}) = \arg\max_m v_{win}^{norm}(m)\{F^R(r, 90°), F_{\tau,\phi}^R(r, \theta = 90° + \tilde{\phi})\}. \qquad (3.16)$$

Since ϕ is unknown, the previous estimations cannot be computed directly. However, with an iterative approach the rotation can be accurately estimated. Performing the previous steps for a range of rotation values results in a set of maxima of 1-D normalized, windowed cross-covariances that correspond to the range of rotation values, as in Figures 3.3 and 3.4. The location of the highest maxima gives the estimate for the rotation

$$\hat{\phi} = \arg\max_{\tilde{\phi}} M(\tilde{\phi}), \qquad (3.17)$$

where

$$M(\tilde{\phi}) = \max_m v_{win}^{norm}(m)\{F^R(r, 0), F_{\tau,\phi}^R(r, \theta = \tilde{\phi})\} \qquad (3.18)$$

and $\tilde{\phi}$ ranges from $-\hat{\phi}_{max}$ to $\hat{\phi}_{max}$ incremented by $\Delta\tilde{\phi}$, as in Figure 3.5. Using the estimate of ϕ from equation (3.17), the parameters τ_x and τ_y can then be accurately estimated as described in equations (3.14) and (3.16).

The 1-D normalized, windowed cross-covariance is used to estimate the distortion parameters, because other correlation techniques are ineffective for a wide range of digital images when correlating using the RT, primarily due to the fact that the RT of most digital images varies drastically in magnitude. Normalizing eliminates the effect of varying magnitude and windowing reduces the number of computations and can increase the accuracy of correlation. Another important design consideration when using the RT is the need for circular windowing of the images. Digital images have rectangular boundaries, and the boundaries create a fixed pattern in the RT, Figure 3.6. Circular windowing eliminates the fixed pattern caused by the boundaries of the image, as in Figure 3.7.

This iterative algorithm requires some criteria. First, the range of rotation values must be wide enough to account for the maximum possible rotational difference

(a) $\tilde{\phi} = 0°$

(b) Aligned for Maximum Correlation

(c) $\tilde{\phi} = 5°$

(d) Aligned for Maximum Correlation
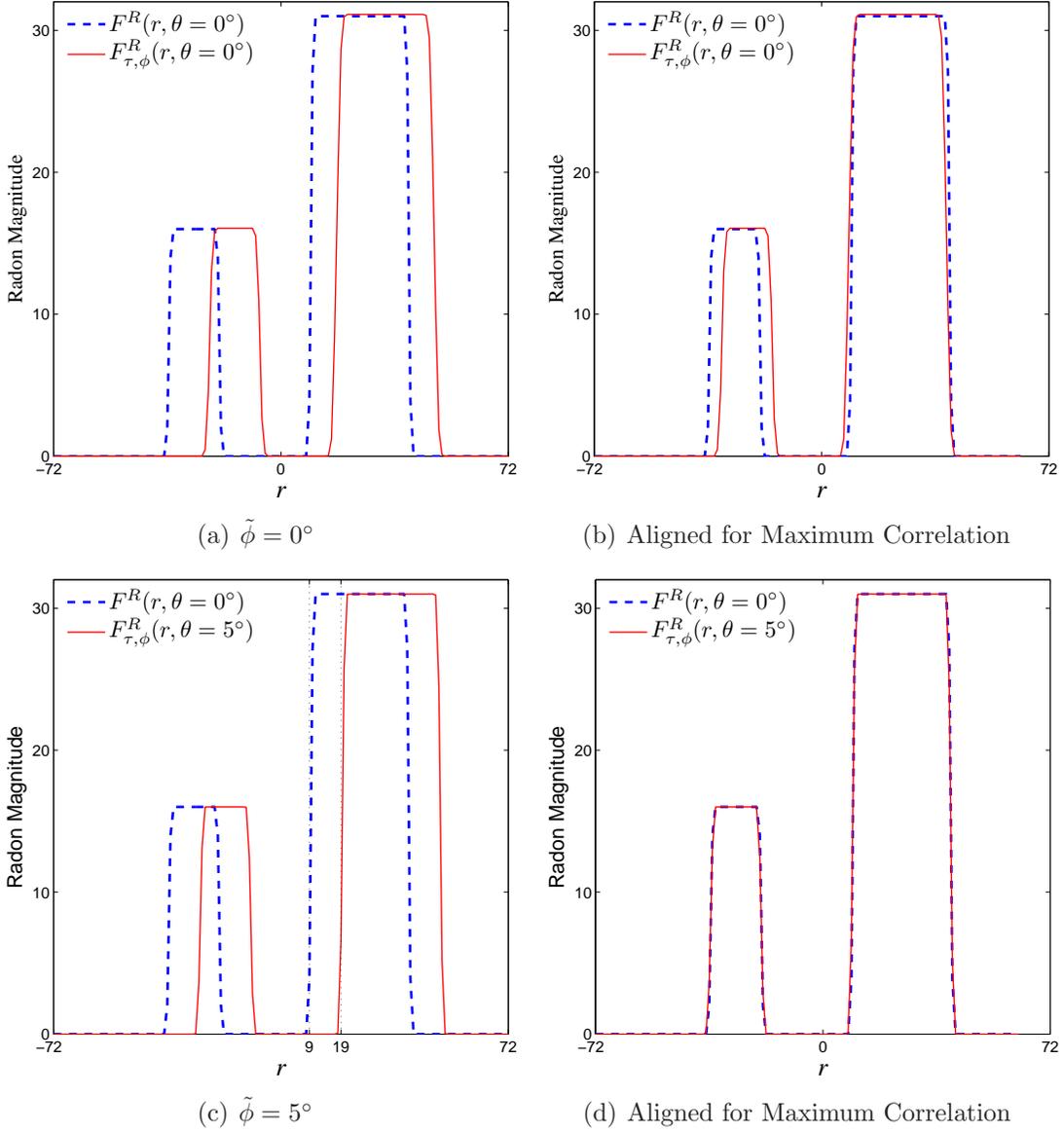
Figure 3.3: The RTs $F^R(r, 0°)$ and $F^R_{\tau,\phi}(r, \theta = \tilde{\phi})$ for two values of $\tilde{\phi}$, showing that τ_x can be estimated using the RT. The image $f(x, y)$, Figure 2.3(a), is translated by $\tau_x = 10$ pixels and $\tau_y = 5$ pixels then rotated by $\phi = 5°$. When $\tilde{\phi} = \phi$ the estimation of τ_x is easily seen because $\mathfrak{R}\{f_{\tau,\phi}(x, y)\}$ is translated by τ_x relative to $\mathfrak{R}\{f(x, y)\}$ in accordance with equation (3.13).

(a) $\tilde{\phi} = 10°$

(b) Aligned for Maximum Correlation

(c) $\tilde{\phi} = 15°$

(d) Aligned for Maximum Correlation
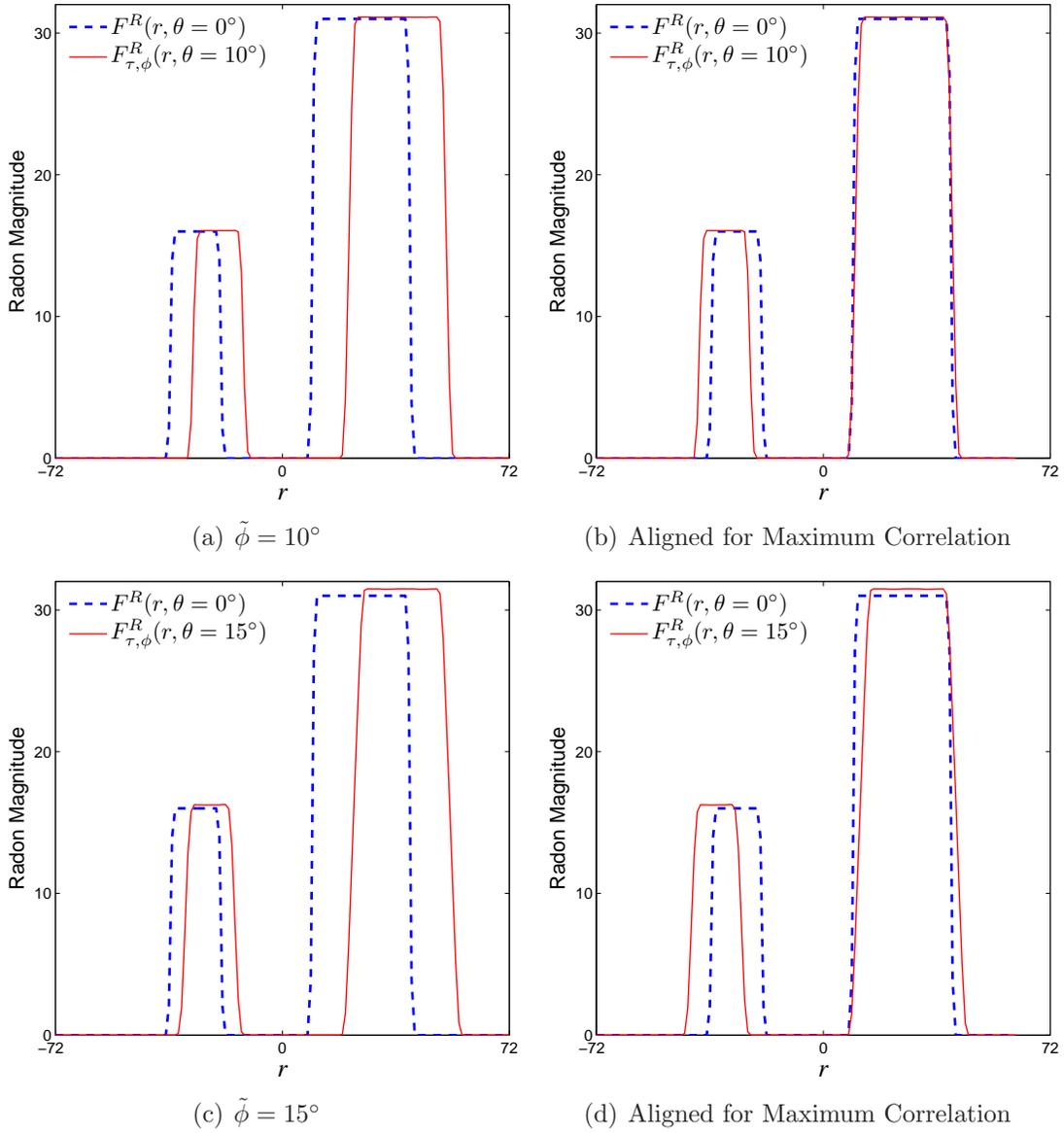
Figure 3.4: The RTs $F^R(r, 0°)$ and $F^R_{\tau,\phi}(r, \theta = \tilde{\phi})$ for another two values of $\tilde{\phi}$, showing that if $\tilde{\phi} \neq \phi$ then the RTs do not correlate well. The image $f(x, y)$, Figure 2.3(a), is translated by $\tau_x = 10$ pixels and $\tau_y = 5$ pixels then rotated by $\phi = 5°$, illustrating that the RT algorithm can accurately estimate the translation because the RTs correlate best when $\tilde{\phi} = \phi$.
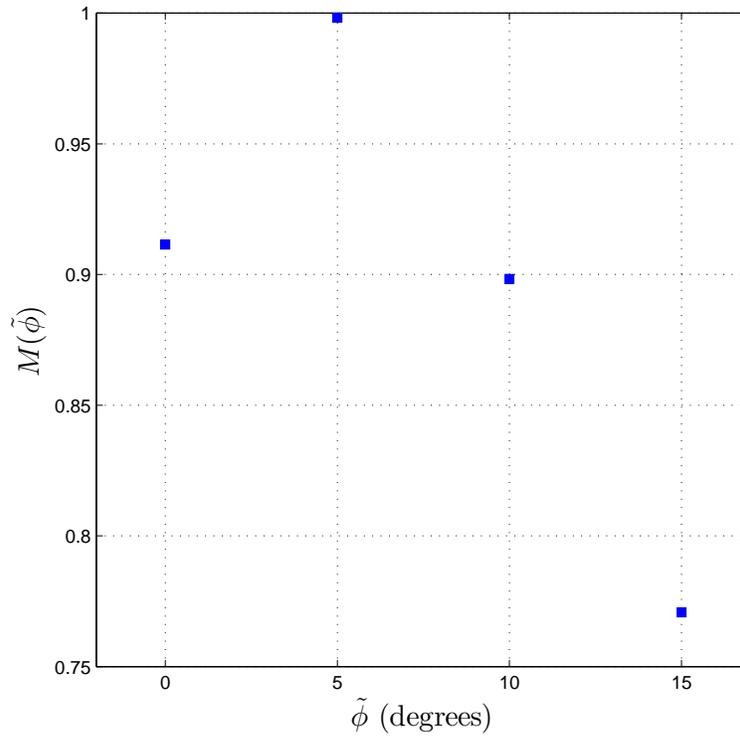
Figure 3.5: Maximum of the 1-D normalized, windowed cross-covariance between $F^R(r, 0°)$ and $F^R_{\tau,\phi}(r, \theta = \tilde{\phi})$ for a range of $\tilde{\phi}$; see equation (3.18). The estimate of ϕ is the value of $\tilde{\phi}$ at the maximum of $M(\tilde{\phi})$; see equation (3.17). The estimated rotation, $\hat{\phi}$, is 5 degrees, which equals the true ϕ.

(a) $b(x, y)$

(b) $\Re\{b(x, y)\}$

(c) $b_\phi(x, y)$
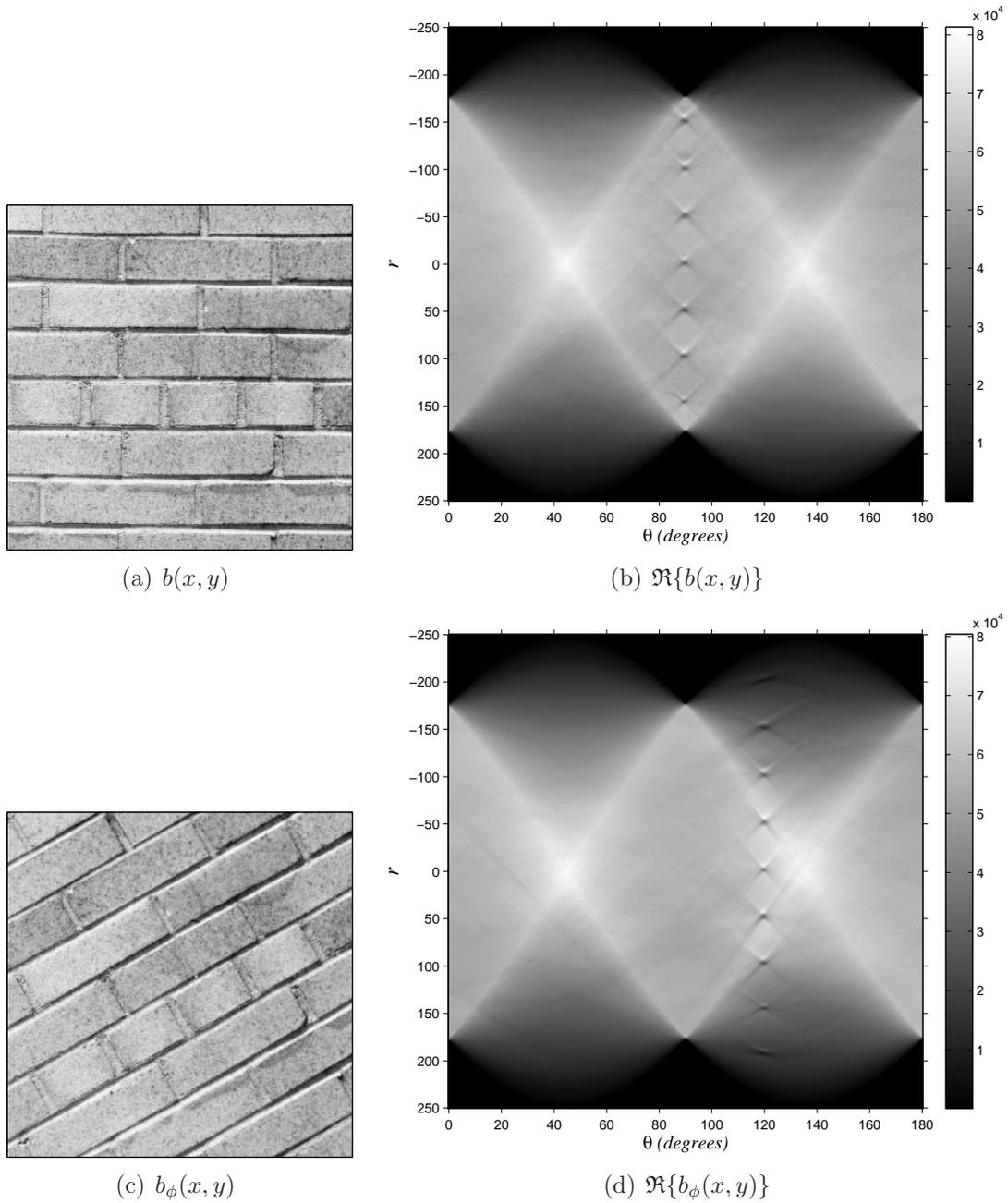
(d) $\Re\{b_\phi(x, y)\}$

Figure 3.6:   The RT of an image of a brick wall, $b(x, y)$, and the same image rotated by 30 degrees, $b_\phi(x, y)$. Note that the RTs both have a similar diamond-shaped pattern, which masks the translation along the θ-axis due to the rotation.
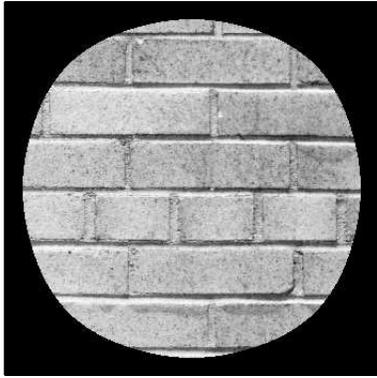
(a) Windowed $b(x, y)$
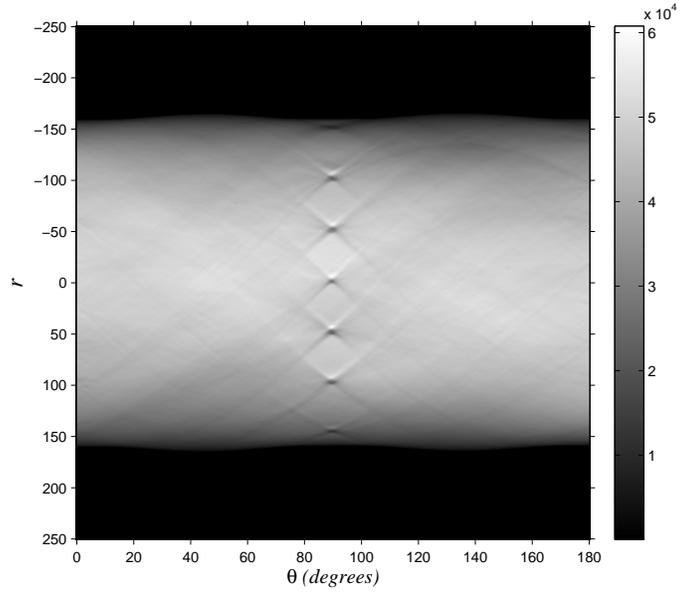
(b) Windowed $\Re\{b(x, y)\}$
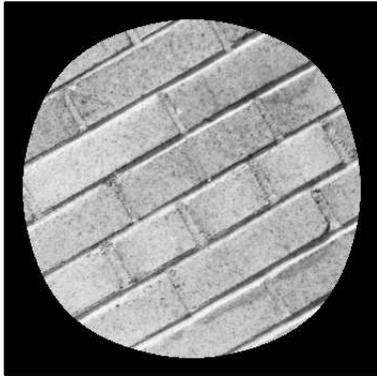
(c) Windowed $b_\phi(x, y)$
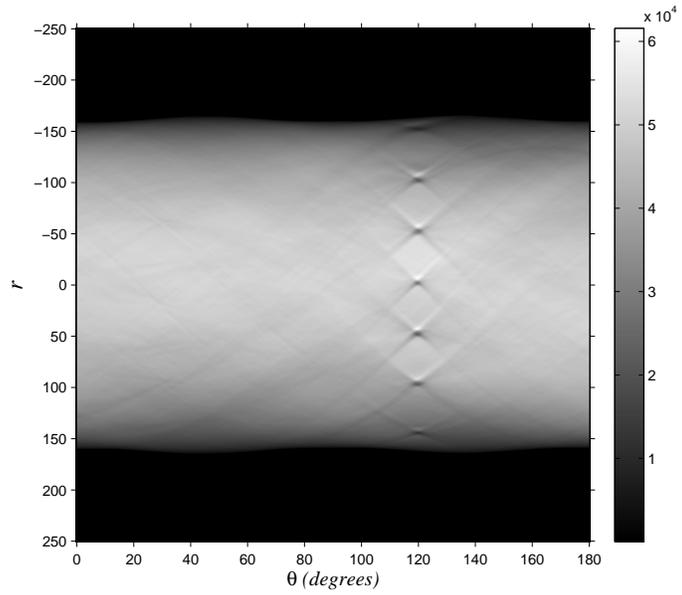
(d) Windowed $\Re\{b_\phi(x, y)\}$

Figure 3.7:   The RT of a windowed image of a brick wall and the same windowed image rotated by 30 degrees. With windowing the diamond pattern is removed because the edges of the image no longer contribute to the magnitude of the RT, so the translation due to rotation is more obvious.

between any two images in the set. Second, the 1-D normalized, windowed cross-covariance requires that the assumed maximum translation for each dimension be large enough to account for the maximum possible translational difference between any two images in the set. Thus, registration always fails for rotation and translation above the maximum assumed value.

After the estimates for translation and rotation are computed, the distortions can be removed, thus aligning the images. The aligned images are averaged to create a scene estimate. There is a scene estimate for both single and multi-frame registration. The goal is to compare the quality of the two scene estimates and the performance of the registration methods.

3.4 Experiment for Translation, Rotation, and Scale Distortion

The methodology for the experiment for translation, rotation, and scale combines the FMT and phase correlation. As discussed in section 2.1.4, a scale and/or rotation corresponds to linear shifts in the phase of the FMT, as in equation (2.21). Also, translations in a image correspond to linear shifts in the phase of the FT, as in equation (2.16). An effective registration method has been developed by Adam Wilmer using these properties and is used for registering images that are translated, rotated, and scaled.

The model $f_{\tau,\beta,\phi}(x,y)$, as in equation (1.8), is used to simulate the distortion for this experiment. The algorithm computes the FMT of two images and uses phase correlation to estimate the relative rotation and scale, which is accomplished by first computing the normalized cross-power spectrum

$$
\begin{aligned}
p\{F^F(\mu,\theta), F^F_{\tau,\beta,\phi}(\mu,\theta)\} &= \frac{\mathfrak{F}\{F^F(\mu,\theta)\}\mathfrak{F}\{F^F_{\tau,\beta,\phi}(\mu,\theta)\}^*}{|\mathfrak{F}\{F^F(\mu,\theta)\}\mathfrak{F}\{F^F_{\tau,\beta,\phi}(\mu,\theta)\}^*|} \\
&= \frac{F^M(u,v)F^M_{\tau,\beta,\phi}(u,v)^*}{|F^M(u,v)F^M_{\tau,\beta,\phi}(u,v)^*|},
\end{aligned}
\tag{3.19}
$$

and, using equation (2.21), the normalized cross-power spectrum is

$$
\begin{aligned}
p\{F^F(\mu,\theta), F^F_{\tau,\beta,\phi}(\mu,\theta)\} &= \frac{F^M(u,v)F^M(u,v)^*e^{j(u\ln\beta+v\phi)}}{|F^M(u,v)F^M(u,v)^*e^{j(u\ln\beta+v\phi)}|} \\
&= \frac{F^M(u,v)F^M(u,v)^*e^{j(u\ln\beta+v\phi)}}{|F^M(u,v)F^M(u,v)^*|} \\
&= e^{j(u\ln\beta+v\phi)}.
\end{aligned}
\tag{3.20}
$$

The scale and rotation are estimated from the maximum of phase correlation

$$
\begin{aligned}
[\ln\hat{\beta},\hat{\phi}] &= \arg\max_{\mu,\theta} P\{F^F(\mu,\theta), F^F_{\tau,\beta,\phi}(\mu,\theta)\} \\
&= \arg\max_{\mu,\theta} \mathfrak{F}^{-1}\{p\{F^F(\mu,\theta), F^F_{\tau,\beta,\phi}(\mu,\theta)\}\} \\
&= \arg\max_{\mu,\theta} \mathfrak{F}^{-1}\{e^{j(u\ln\beta+v\phi)}\} \\
&= \arg\max_{\mu,\theta} \delta(\mu+\ln\beta, \theta+\phi),
\end{aligned}
\tag{3.21}
$$

as in Figure 3.8. The phase correlation of the FMT in Figure 3.8 does not result in a delta function because the log-polar mapping for discrete functions requires interpolation.

The log-polar mapping can have a resolution different from the original images. The resolution of the log-polar mapping determines the logarithmic spacing of the translation due to scaling and the proportional spacing of the translation due to rotation. For scaling, if the log-polar mapping has a resolution of 256 pixels for the scale axis, the pixels correspond to logarithmically spaced values from $\log(1)$ to $\log(128)$. For this algorithm, the logarithmic spacing only extends to half the resolution of the scale axis because beyond half the resolution the scale is less than one. For rotation, if the log-polar mapping has a resolution of 512 pixels for the angular axis then the translation of a pixel corresponds to $\frac{360°}{512} \approx 0.703°$ per pixel.

The relative rotation and scale are removed from the image that is reduced in scale. The smaller scaled image is used to preserve the size of the images after each is adjusted for rotation and scale. The images now only have relative translative distortion. The translation is estimated using phase correlation of the images after the
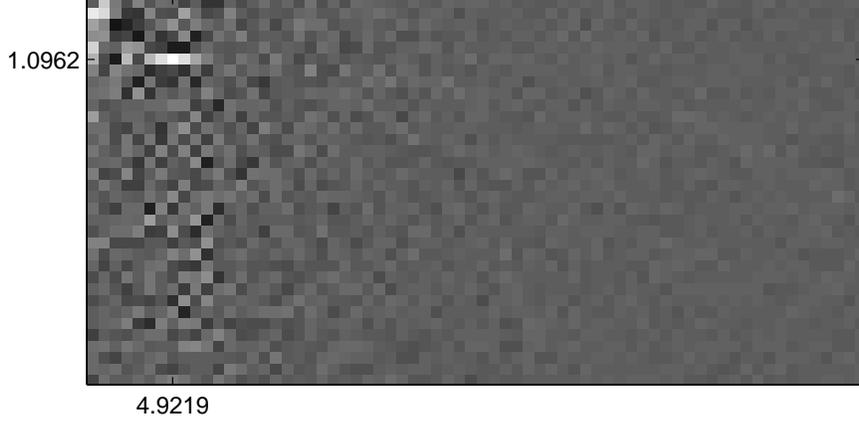
Figure 3.8:   Phase correlation of the FMT of $f(x,y)$, as in Figure 2.3(a), and the FMT of the same image translated by $\tau_x = 10$ pixels and $\tau_y = 5$ pixels, scaled by $\beta = 1.1$, and rotated by $\phi = 5°$, $f_{\tau,\beta,\phi}(x,y)$. The location of the maximum is indicated by the labelled markers, which correspond to the estimates $\hat{\beta}$ and $\hat{\phi}$, in accordance with equation (3.21). The pixels of the vertical axis correspond to logarithmically spaced values of scale and the labelled marker for scale reflects this spacing. The pixels of the horizontal axis correspond to equal proportions of $360°$.

rotation and scale have been removed from the smaller scaled image. The cross-power spectrum of $f(x,y)$ and $f_\tau(x,y)$ is

$$
\begin{aligned}
p\{f(x,y), f_\tau(x,y)\} &= \frac{F^F(u,v)F_\tau^F(u,v)^*}{|F^F(u,v)F_\tau^F(u,v)^*|} \\
&= \frac{F^F(u,v)F^F(u,v)^*e^{j(u\tau_x+v\tau_y)}}{|F^F(u,v)F^F(u,v)^*e^{j(u\tau_x+v\tau_y)}|} \\
&= e^{j(u\tau_x+v\tau_y)}
\end{aligned}
\tag{3.22}
$$

The relative translation is also estimated using the maximum of phase correlation

$$
\begin{aligned}
[\hat{\tau}_x, \hat{\tau}_y] &= \arg\max_{x,y} P\{f(x,y), f_\tau(x,y)\} \\
&= \arg\max_{x,y} \mathfrak{F}^{-1}\{p\{f(x,y), f_\tau(x,y)\}\} \\
&= \arg\max_{x,y} \mathfrak{F}^{-1}\{e^{j(u\tau_x+v\tau_y)}\} \\
&= \arg\max_{x,y} \delta(x + \tau_x, y + \tau_y),
\end{aligned}
\tag{3.23}
$$

as in Figure 3.9. The translative distortions are then removed, aligning the images. The aligned images are averaged to create a scene estimate. There is a scene estimate
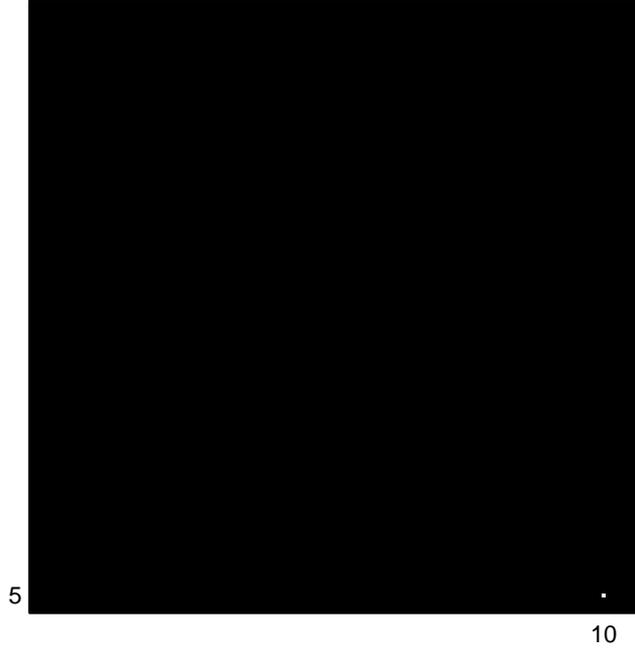
Figure 3.9: Phase correlation of the FMT of $f(x, y)$ and the FMT of $f_{\tau, \beta, \phi}$ after it has been adjusted for rotation and scale by the estimates labelled on Figure 3.8. The location of the maximum is indicated by the labelled markers, which correspond to the estimates $\hat{\tau}_x$ and $\hat{\tau}_y$, in accordance with equation (3.23).

for both single and multi-frame registration. The goal is to compare the quality of the two scene estimates and the performance of the registration methods.

IV. Results and Analysis

For each experiment, statistics on the accuracy and quality of the registration methods are generated using multiple simulated data sets. Each simulated data set is defined by the model used to generate distortions, the maximum value of those distortions, the variance of the Gaussian white noise, and the number of simulations performed to generate the statistics. Table 4.1 summarizes the parameters for each experiment. The random translational and rotational distortion are generated using uniform, zero-mean random generators from $-\tau_{max}$ to τ_{max} and $-\phi_{max}$ to ϕ_{max}, respectively. For the experiment using the RT, the rotation estimates depend on the increments of rotation values, $\Delta\tilde{\phi}$, as in equation (3.18). The random scaling is generated using uniform random generators from 1 to some maximum scale, β_{max}. The same true image is used for each experiment to provide comparable results. The true image is from NASA and is a satellite image of crops in Kansas, United States [15]. Each experiment produces relatively similar results, therefore, the majority of the analysis is applied to all experiments.

The cross-correlation experiment has a large noise variance to demonstrate the robustness of cross-correlation against noise. The added noise is random and unlikely to correlate. Also, with only translative distortion the experiment is relatively simple. The decreased noise variance for the FMT experiment is because the algorithm is not designed to handle noisy data. The algorithm is only designed as a demonstrative

	2-D Cross-Correlation Experiment	Radon Transform Experiment	Fourier-Mellin Transform Experiment
Mathematical Model	$f_\tau(x,y)$	$f_{\tau,\phi}(x,y)$	$f_{\tau,\phi,\beta}(x,y)$
Noise Variance	60	20	10
# of Simulations	1000	10000	10000
Max Translation, $2\tau_{\mathbf{max}}$	25 pixels (10%)	16 pixels (6%)	16 pixels (6%)
Max Rotation, $2\phi_{\mathbf{max}}$		16°	16°
Rotation Increment, $\Delta\phi$		1°	
Max Scale, $\beta_{\mathbf{max}}$			130%

Table 4.1:   Summary of the simulation parameters used for each experiment.

tool for the applications of the FMT and, therefore, performs poorly with noisy data. This result does not indicate that the FMT is a poor technique for noisy data, but rather that only this particular implementation of the FMT is a poor technique for noisy data.

The results show an obvious improvement in both registration accuracy and quality of the scene estimate for the multi-frame method; see Figures 4.1- 4.12. This result is expected because an unbiased registration method, on average, performs better than a biased method. More interesting is the non-linear improvement in registration accuracy for the multi-reference frame estimator with respect to using more frames. The registration accuracy and quality statistics are computed for different frame set sizes as indicated by the independent axis of the figures. The goal is to analyze how the methods perform when more frames are added. However, the statistics for different frame set sizes are not independent, which is intentional. For one simulation, the metrics are computed for 3 frames then the same 3 frames are used with another frame added, the metrics re-computed, and so on. The idea is that more data should produce better results.

It is known that the single frame method is not more accurate with more frames, which is supported by the results. This effect is a byproduct of the bias of the single frame method, which highlights a key benefit of the multi-frame approach. The multi-frame method averages all the available frames, resulting in better accuracy because there is no bias. The non-linear improvement of the multi-frame method is not only because it is unbiased, but also because with more frames it averages all combinations of frames, which provides a boost in performance when more frames are used. With N frames the single frame method uses only $N-1$ estimations. The multi-frame method averages over $(N^2 - N)/2$ estimations. This averaging over more estimations results in less error and gives the multi-frame method the capability of subpixel estimation.

The multi-frame method produces a better scene estimate as a result of the improved registration accuracy. The gain in image quality is less drastic for the cross-

correlation experiment because both methods are accurate to less than one pixel. It may seem exceptional that the image quality of the scene estimates from the single frame method improve as more frames are added, since the registration accuracy is constant, (see Figures 4.2 and 4.3). However, the scene estimate image quality for the single frame method still improves with more frames because more images are used in averaging, which decreases the effects of the additive noise. Also, averaging reduces the effect of a registration error, therefore, even without noise the single frame method generates an improved scene estimate as more frames are added.

The results from the experiments clearly show the benefits of a multi-frame registration technique. The primary disadvantage is that the multi-frame technique requires considerably more computational time due to the increased number of estimations required, as in Figure 4.13, which is especially true for large data sets. If a real-time application is required this method might prove too costly, but with current and future computer performance advances this disadvantage is less of a factor.
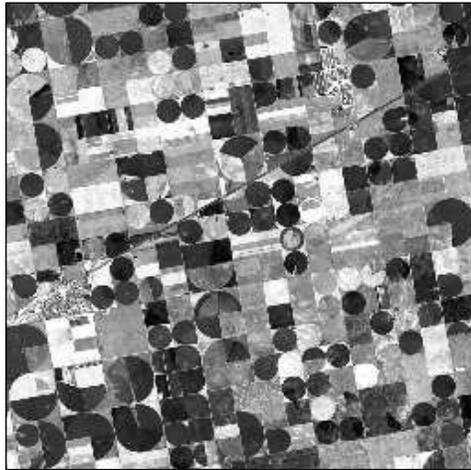
(a) True Image   (b) Sample Noisy Frame
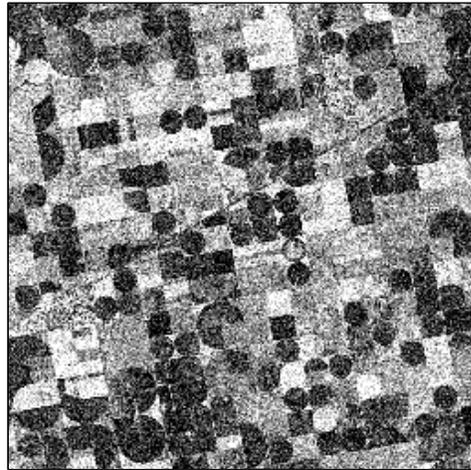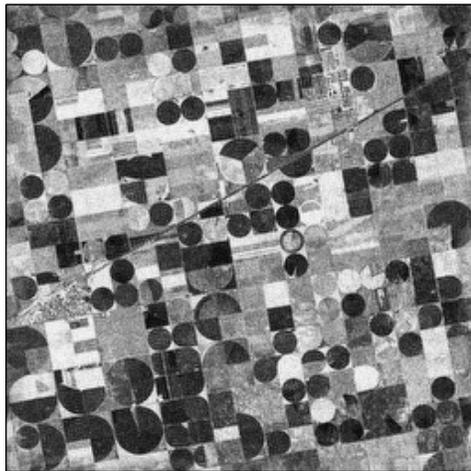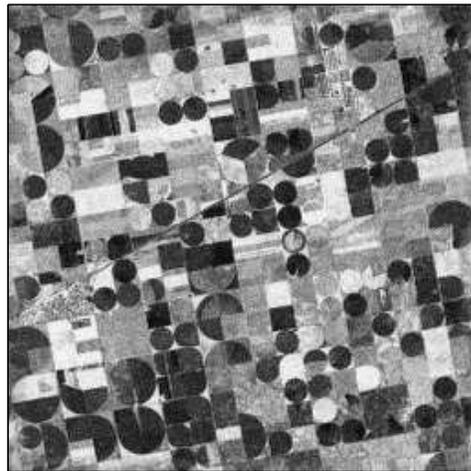
(c) Multi-Frame Estimate   (d) Single Reference Frame Estimate

Figure 4.1:   Scene estimates of the multi-frame and single frame registration methods from the cross-correlation experiment. Scene estimates are made using 24 frames of data. Also shown are the true image and a sample frame from the noisy data.
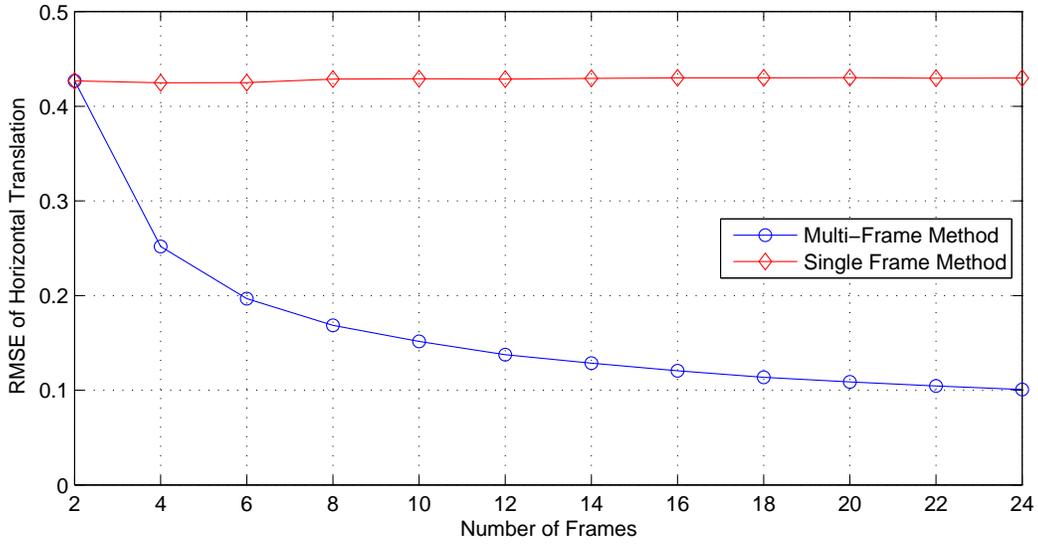
(a) RMSE of Horizontal Shift Estimates
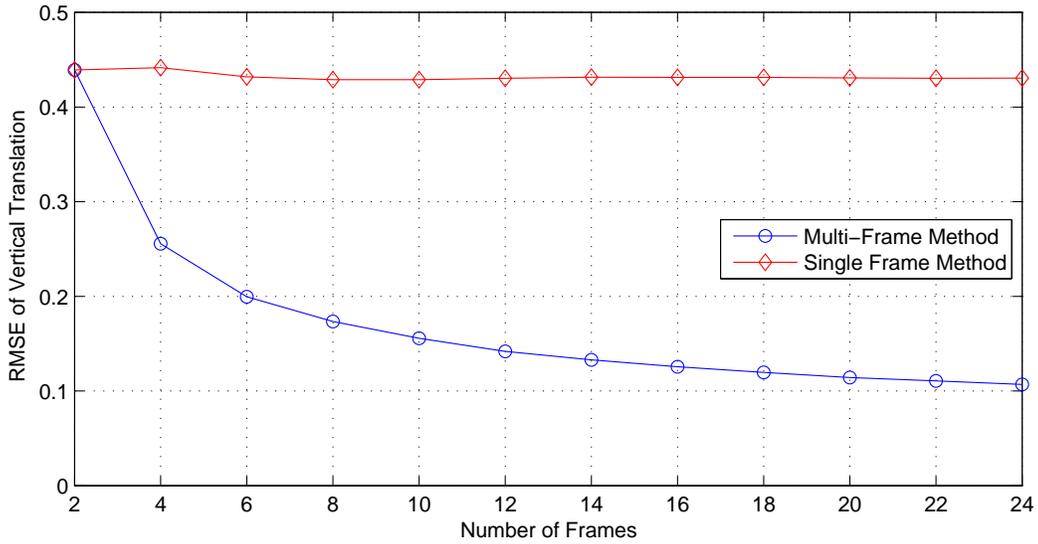
(b) RMSE of Vertical Shift Estimates

Figure 4.2:   Registration error of translation estimates (in pixels) from the cross-correlation experiment using the aerial image of a crop field.
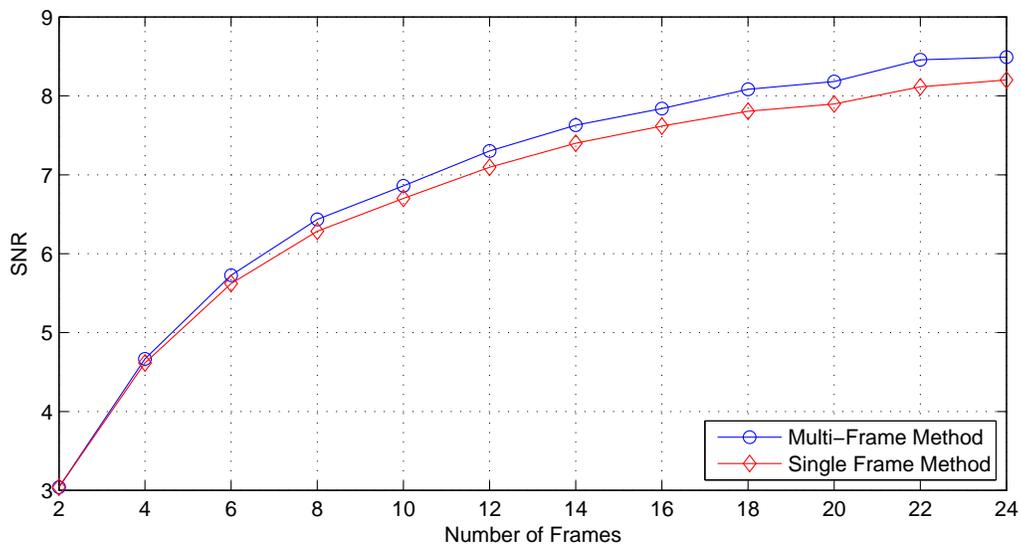
Figure 4.3: Quality of the scene estimates from the cross-correlation experiment using the aerial image of the crop field. The SNR is computed using a linear scale.
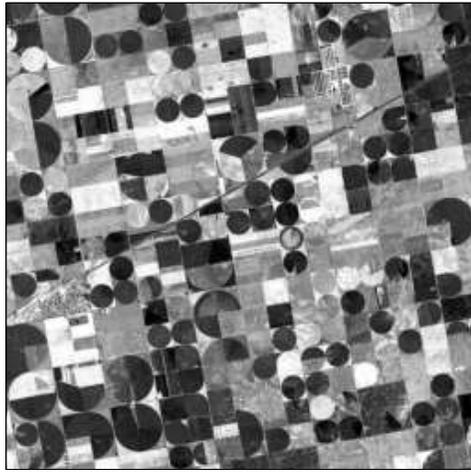
(a) True Image
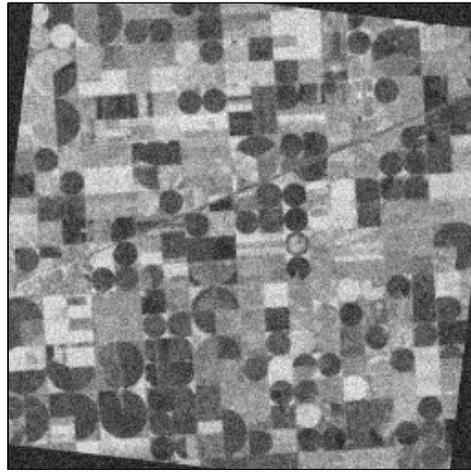
(b) Sample Noisy Frame
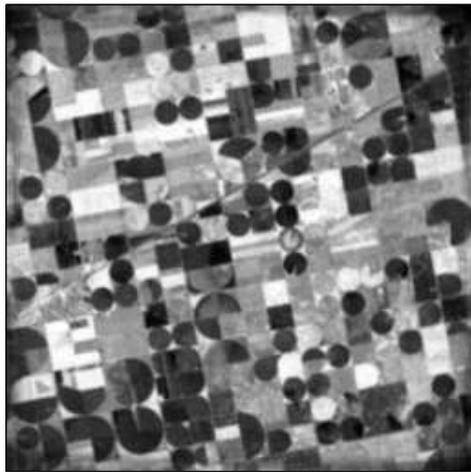
(c) Multi-Frame Estimate
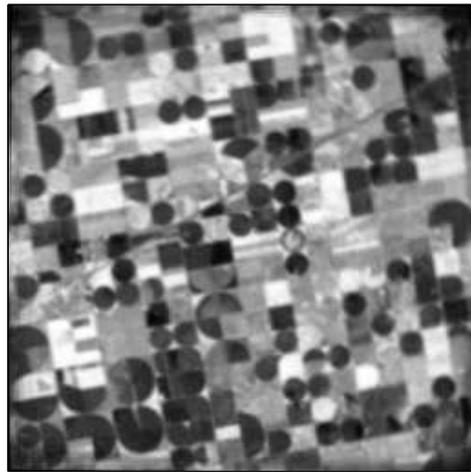
(d) Single Reference Frame Estimate

Figure 4.4: Scene estimates of the multi-frame and single frame registration methods from the RT experiment. Scene estimates are made using 24 frames of data. Also shown are the true image and a sample frame from the noisy data.
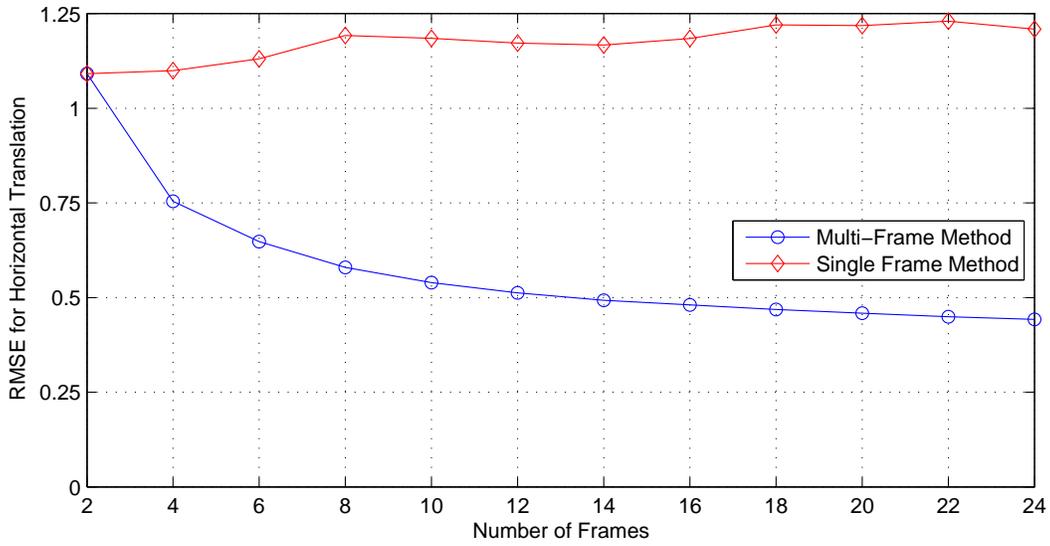
(a) RMSE of Horizontal Shift Estimates
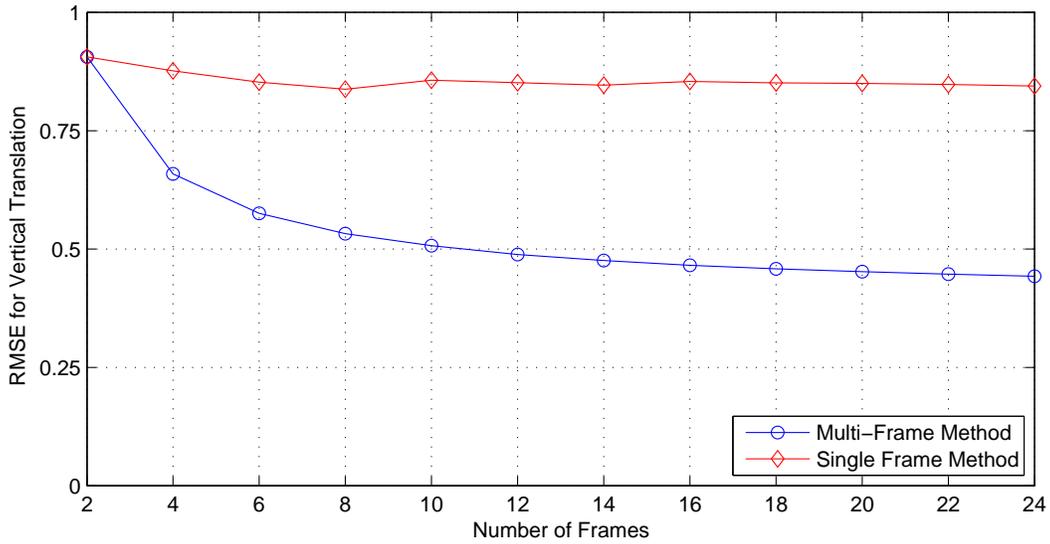
(b) RMSE of Vertical Shift Estimates

Figure 4.5: Registration error of translation estimates (in pixels) from the RT experiment using the aerial image of a crop field.
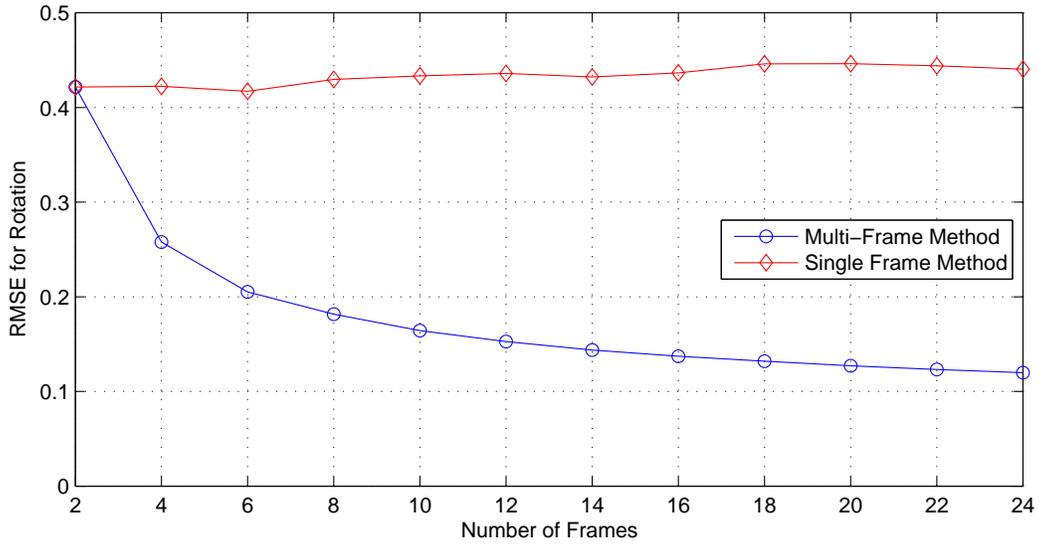
Figure 4.6: Registration error of rotation estimates (in degrees) from the RT experiment using the aerial image of a crop field.
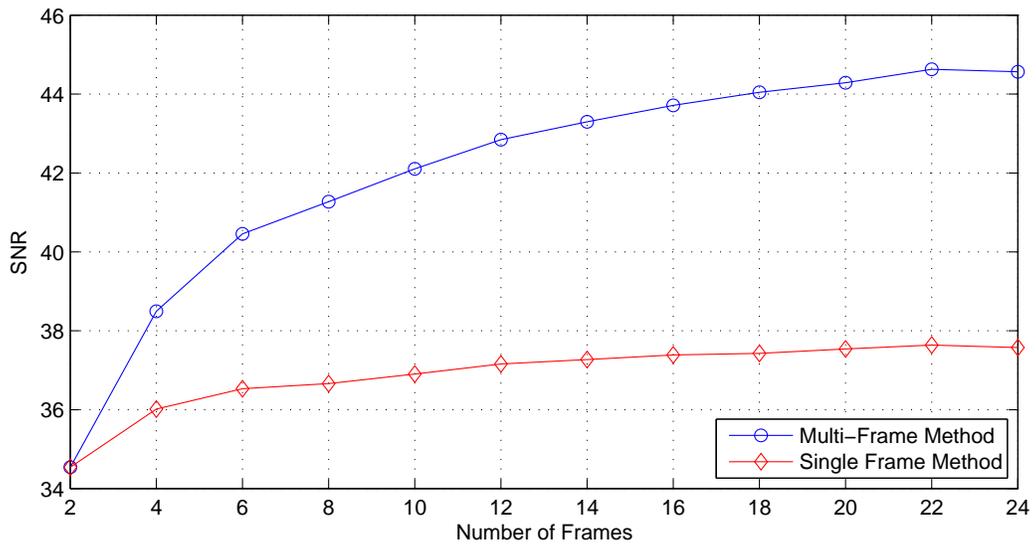

Figure 4.7: Quality of the scene estimates from the RT experiment using the aerial image of a crop field. The SNR is computed using a linear scale.
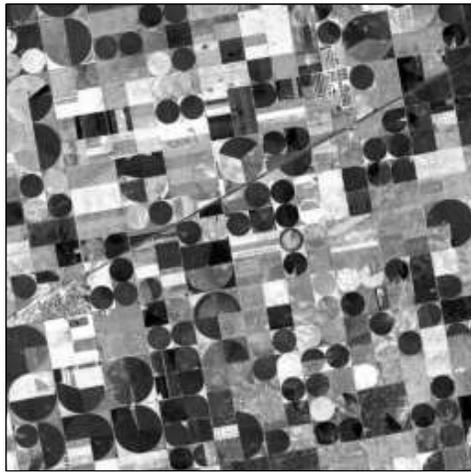
(a) True Image
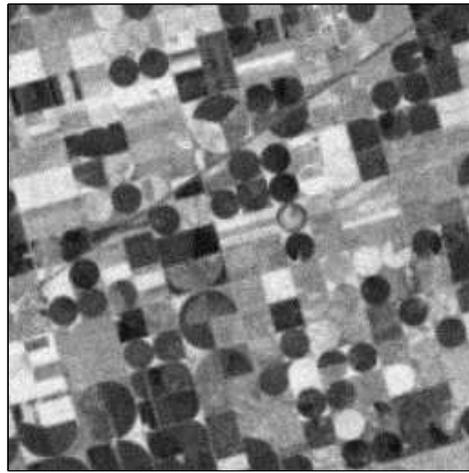
(b) Sample Noisy Frame
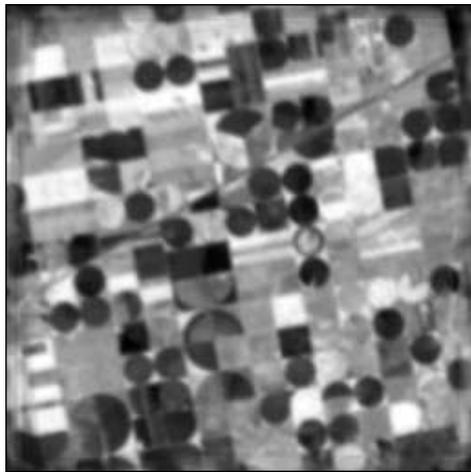
(c) Multi-Frame Estimate
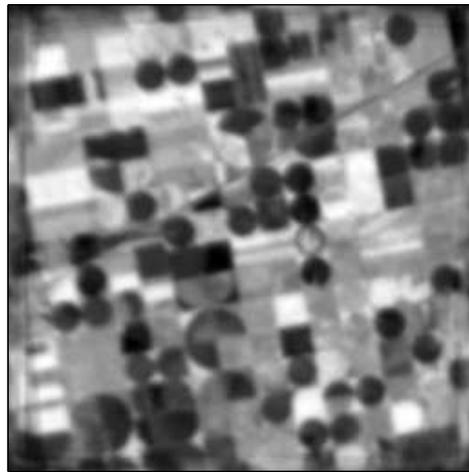
(d) Single Reference Frame Estimate

Figure 4.8:   Scene estimates of the multi-frame and single frame registration methods from the FMT experiment. Scene estimates are made using 24 frames of data. Also shown are the true image and a sample frame from the noisy data. The sample noisy frame visually appears to be higher quality than the scene estimates. However, the SNR of the individual noisy frame is approximately 20, which is significantly lower than the SNR for the single and multi-frame scene estimates of approximately 40.5 and 33.5, respectively. Also, the SNR of the sample noisy frame corresponds to the trend shown in Figure 4.12.
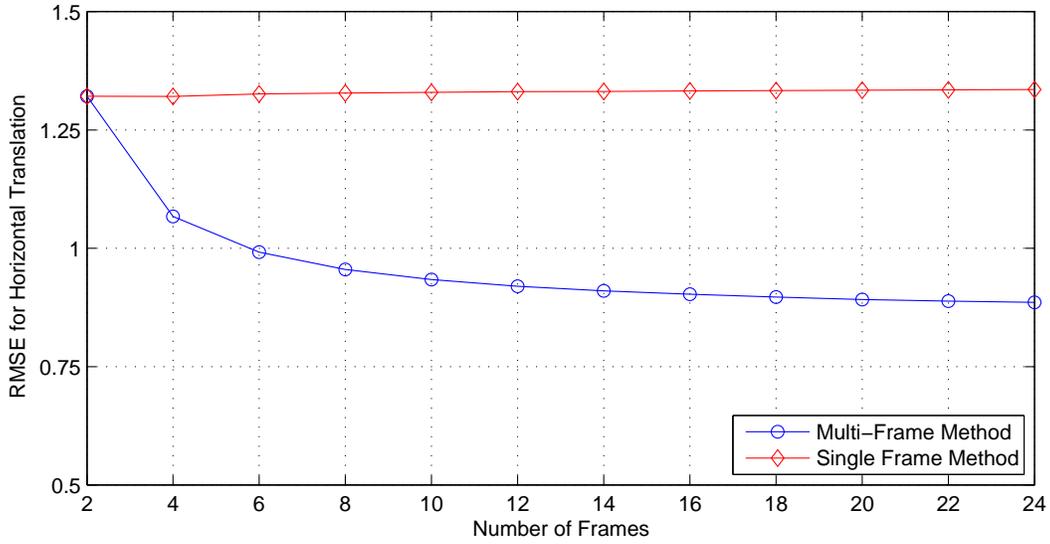
(a) RMSE of Horizontal Shift Estimates
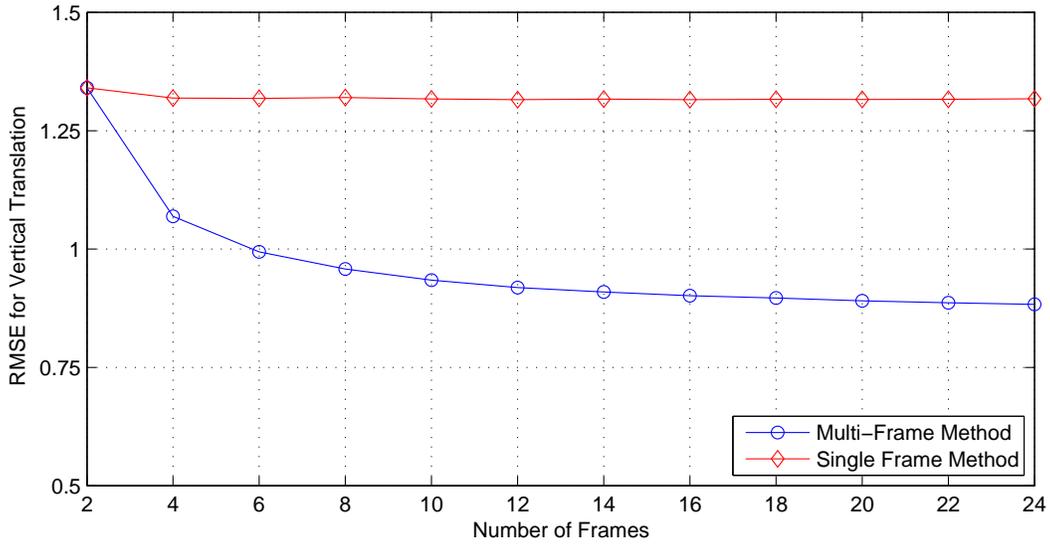
(b) RMSE of Vertical Shift Estimates

Figure 4.9: Registration error of translation estimates (in pixels) from the FMT experiment using the aerial image of a crop field.
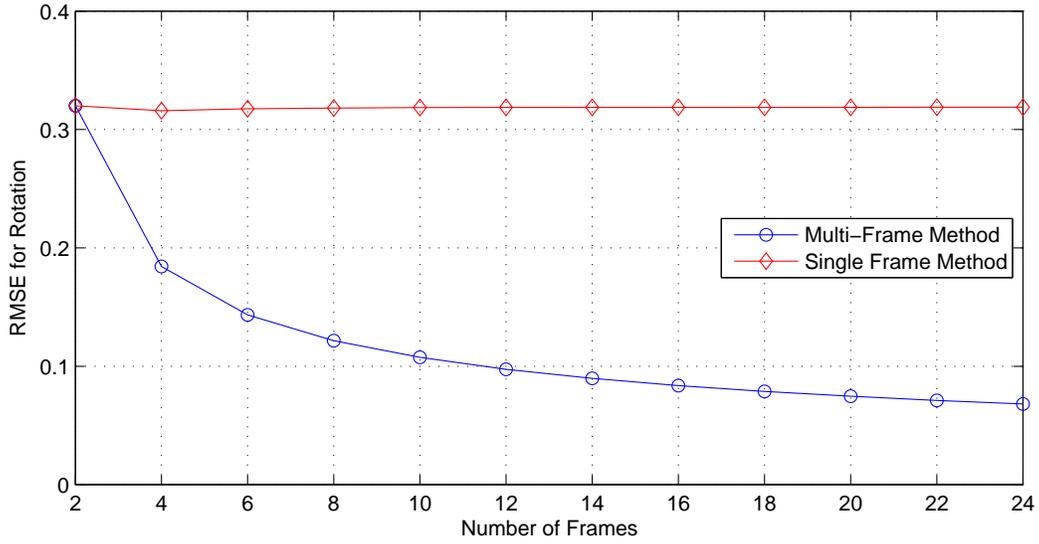
Figure 4.10:   Registration error of rotation estimates (in degrees) from the FMT experiment using the aerial image of a crop field.
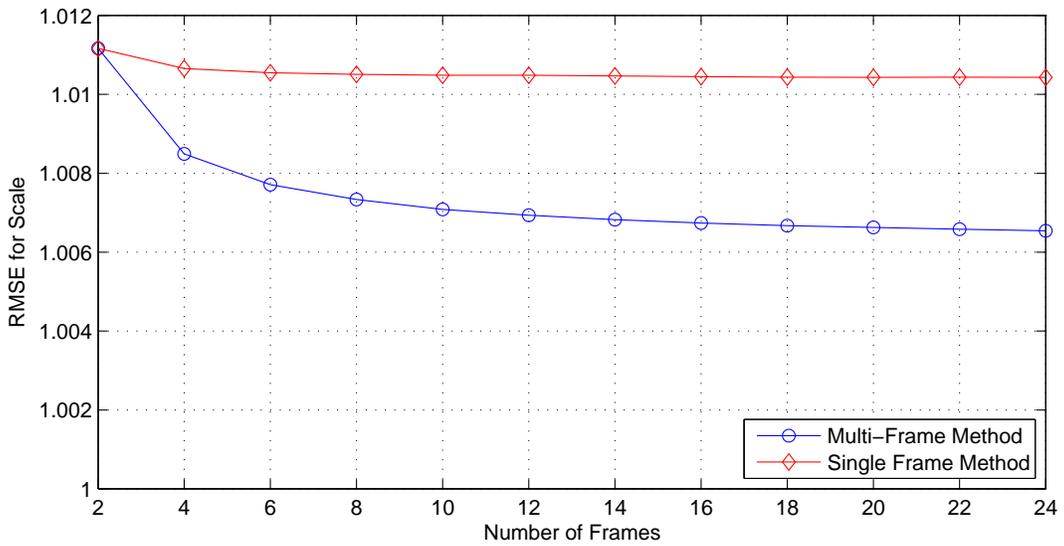

Figure 4.11:   Registration error of scale estimates from the FMT experiment using the aerial image of a crop field. Since scale estimates are an estimate of the relative size ratio, the lower limit for scale estimation is 1.
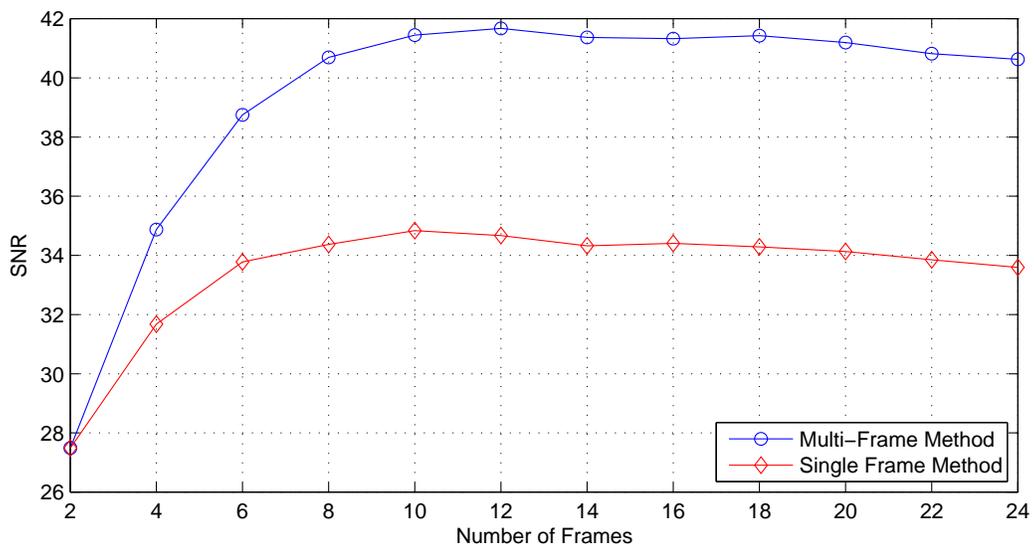
Figure 4.12: Quality of the scene estimates from the FMT experiment using the aerial image of the crop field. The SNR is computed using a linear scale. The curves flatten because the registration error outweighs the effect of the image averaging, especially since low noise is added to the images.
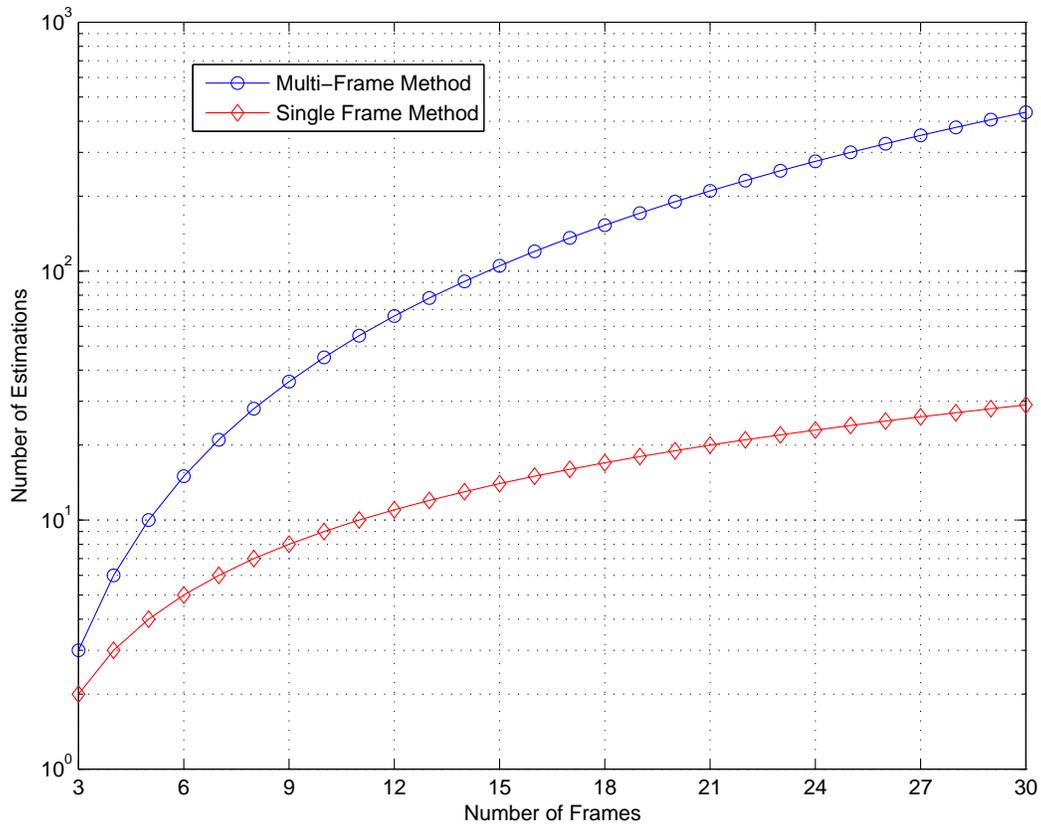
Figure 4.13: Comparison of the number of estimations required for both registration methods, emphasizing the main disadvantage of the multi-frame method. Considerably more estimations are required for large data sets.

V. Conclusions

Image distortions such as translation, scale, and rotation potentially obstruct the ability to gather useful information from a set of images. Three experiments are developed to simulate systems that introduce different combinations of distortion and noise. The first uses 2-D cross correlation to estimate translation. The second experiment employs a new registration technique that uses the RT to estimate translation and rotation. The final experiment uses the FMT and phase correlation to estimate translation, rotation, and scale. In each experiment, single and multi-frame image registration are used to estimate distortions and perform frame averaging. A distinct and noticeable improvement in both registration accuracy and quality of frame averaging is observed and quantified. The new multi-frame registration method is shown to be reliable even for the registration algorithms that perform poorly. Also, the unbiased estimation and utilization all available data for the multi-frame method proved it to be superior over traditional registration.

The results provide further justification that the new method is superior and maintains its superiority through a wide range of applications. It is shown that the multi-frame method displays increased performance with increased number of images but with decreased processing speed; thus it is ideal for post processing applications. Multi-frame registration could still be used in pseudo-real-time applications, because a window of data could be processed to exercise the benefits of the new method. Furthermore, it is shown that the new registration technique is equivalent to the optimal Gauss-Markov estimator for relative distortion in images.

5.1 Future Research

Future research on multi-reference frame registration is needed to realize its full potential. Research in the area of the real-time applications using measured (non-simulated) data is needed. Multi-frame registration could be applied to registration algorithms that handle more complex distortions, such as images from different viewpoints and/or multiple sensors. For example, image fusion combines images from

different viewpoints and/or sensors to create a larger 2-D image or 3-D representation of the scene, and misregistration results in poor construction of the composite image. Multi-frame registration could improve this process. The multi-frame method could also be applied to feature detection or feature matching algorithms. These applications sometimes require accurate detection of multiple features and are sensitive to image degradations and different imaging conditions. Another area that may prove fruitful is analyzing systems that have unusual noise patterns or corruption. These systems could benefit greatly from the unbiased estimation of the multi-frame method. It may be interesting to investigate systems with non-uniform distributions of distortion. In particular, it may be possible to account for knowledge of the distribution and improve registration performance.

Appendix A. Calculations for Optimal Gauss-Markov Estimation

The new multi-frame estimation method is an optimal Gauss-Markov estimator for $N-1$ relative distortion in N images given the $(N^2-N)/2$ pairwise distortion estimates; see section 2.2.1. Here calculations are performed for a specific case, and the results for this case can be generalized to any N.

For $N=4$, the vector of true distortions with respect of the first image is

$$D\{1,k\} = [\Omega\{1,2\}, \Omega\{1,3\}, \Omega\{1,4\}]^T, \tag{A.1}$$

and the vector of $(N^2-N)/2$ distortion estimates for all N images taken two at a time is

$$D_M\{i,j\} = [\hat{\Omega}\{1,2\}, \hat{\Omega}\{1,3\}, \hat{\Omega}\{1,4\}, \hat{\Omega}\{2,3\}, \hat{\Omega}\{2,4\}, \hat{\Omega}\{3,4\}]^T. \tag{A.2}$$

The matrices A and P_E are

$$A = \begin{bmatrix} \delta\{2,2\} - \delta\{1,2\} & \delta\{2,3\} - \delta\{1,3\} & \delta\{2,4\} - \delta\{1,4\} \\ \delta\{3,2\} - \delta\{1,2\} & \delta\{3,3\} - \delta\{1,3\} & \delta\{3,4\} - \delta\{1,4\} \\ \delta\{4,2\} - \delta\{1,2\} & \delta\{4,3\} - \delta\{1,3\} & \delta\{4,4\} - \delta\{1,4\} \\ \delta\{3,2\} - \delta\{2,2\} & \delta\{3,3\} - \delta\{2,3\} & \delta\{3,4\} - \delta\{2,4\} \\ \delta\{4,2\} - \delta\{2,2\} & \delta\{4,3\} - \delta\{2,3\} & \delta\{4,4\} - \delta\{2,4\} \\ \delta\{4,2\} - \delta\{3,2\} & \delta\{4,3\} - \delta\{3,3\} & \delta\{4,4\} - \delta\{3,4\} \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ -1 & -1 & 0 \\ -1 & 0 & 1 \\ 0 & -1 & 1 \end{bmatrix} \tag{A.3}$$

and

$$P_E = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}. \tag{A.4}$$

The optimal Gauss-Markov estimator for D is

$$\begin{aligned} \hat{D} &= [A^T P_E^{-1} A]^{-1} A^T D_M \\ &= \begin{bmatrix} \frac{1}{2}\hat{\Omega}\{1,2\} + \frac{1}{4}\hat{\Omega}\{1,3\} + \frac{1}{4}\hat{\Omega}\{1,4\} - \frac{1}{4}\hat{\Omega}\{2,3\} - \frac{1}{4}\hat{\Omega}\{2,4\} + 0\hat{\Omega}\{3,4\} \\ \frac{1}{4}\hat{\Omega}\{1,2\} + \frac{1}{2}\hat{\Omega}\{1,3\} + \frac{1}{4}\hat{\Omega}\{1,4\} + \frac{1}{4}\hat{\Omega}\{2,3\} - 0\hat{\Omega}\{2,4\} - \frac{1}{4}\hat{\Omega}\{3,4\} \\ \frac{1}{4}\hat{\Omega}\{1,2\} + \frac{1}{4}\hat{\Omega}\{1,3\} + \frac{1}{2}\hat{\Omega}\{1,4\} + 0\hat{\Omega}\{2,3\} - \frac{1}{4}\hat{\Omega}\{2,4\} + \frac{1}{4}\hat{\Omega}\{3,4\} \end{bmatrix}. \end{aligned} \tag{A.5}$$

The same result is achieved using the multi-frame estimations. The matrix of N^2 distortions estimates for all N images taken two at a time is

$$\hat{\Omega} = \begin{bmatrix} 0 & \hat{\Omega}\{1,2\} & \hat{\Omega}\{1,3\} & \hat{\Omega}\{1,4\} \\ -\hat{\Omega}\{1,2\} & 0 & \hat{\Omega}\{2,3\} & \hat{\Omega}\{2,4\} \\ -\hat{\Omega}\{1,3\} & -\hat{\Omega}\{2,3\} & 0 & \hat{\Omega}\{3,4\} \\ -\hat{\Omega}\{1,4\} & -\hat{\Omega}\{2,4\} & -\hat{\Omega}\{3,4\} & 0 \end{bmatrix}. \tag{A.6}$$

The multi-frame estimates are calculated by averaging along the rows, as in equation (2.27),

$$\hat{\Omega}_M = \begin{bmatrix} \frac{1}{4}\hat{\Omega}\{1,2\} + \frac{1}{4}\hat{\Omega}\{1,3\} + \frac{1}{4}\hat{\Omega}\{1,4\} \\ -\frac{1}{4}\hat{\Omega}\{1,2\} + \frac{1}{4}\hat{\Omega}\{2,3\} + \frac{1}{4}\hat{\Omega}\{2,4\} \\ -\frac{1}{4}\hat{\Omega}\{1,3\} - \frac{1}{4}\hat{\Omega}\{2,3\} + \frac{1}{4}\hat{\Omega}\{3,4\} \\ -\frac{1}{4}\hat{\Omega}\{1,4\} - \frac{1}{4}\hat{\Omega}\{2,4\} - \frac{1}{4}\hat{\Omega}\{4,4\} \end{bmatrix}. \tag{A.7}$$

The $N - 1$ distortion estimates may now be computed relative to the first image by subtracting the first image estimates, $\hat{\Omega}_M(1)$, from all other multi-frame estimates:

$$
\hat{D} = \hat{\Omega}_M(1) - \begin{bmatrix} \hat{\Omega}_M(2) \\ \hat{\Omega}_M(3) \\ \hat{\Omega}_M(4) \end{bmatrix}
$$

$$
= \begin{bmatrix} \frac{1}{2}\hat{\Omega}\{1,2\} + \frac{1}{4}\hat{\Omega}\{1,3\} + \frac{1}{4}\hat{\Omega}\{1,4\} - \frac{1}{4}\hat{\Omega}\{2,3\} - \frac{1}{4}\hat{\Omega}\{2,4\} + 0\hat{\Omega}\{3,4\} \\ \frac{1}{4}\hat{\Omega}\{1,2\} + \frac{1}{2}\hat{\Omega}\{1,3\} + \frac{1}{4}\hat{\Omega}\{1,4\} + \frac{1}{4}\hat{\Omega}\{2,3\} - 0\hat{\Omega}\{2,4\} - \frac{1}{4}\hat{\Omega}\{3,4\} \\ \frac{1}{4}\hat{\Omega}\{1,2\} + \frac{1}{4}\hat{\Omega}\{1,3\} + \frac{1}{2}\hat{\Omega}\{1,4\} + 0\hat{\Omega}\{2,3\} - \frac{1}{4}\hat{\Omega}\{2,4\} + \frac{1}{4}\hat{\Omega}\{3,4\} \end{bmatrix} .
$$

$$(A.8)$$

Bibliography

1. O'Ruanaidh, J. and T. Pun. "Rotation, Scale, and Translation Invariant Digital Image Watermarking". *International Conference on Image Processing*, volume 1, 536–539. Oct 1997.

2. Bruckart, S. A. *Multiframe Shift Estimation*. Master's thesis, AFIT, 2006.

3. Manduchi, R. and G. Mian. "Accuracy analysis for correlation-based image registration algorithms". *International Symposium on Circuits and Systems*, volume 1, 834–837. May 1993.

4. Labunets, E., V. Labunets, K. Egiazarian, and J. Astola. "Fast Spectral Algorithms of Invariants Calculation". *International Conference on Image Analysis and Processing*, volume 1, 203–208. Sept 1999.

5. Miller, M., T. Nguyen, and C. Yang. "Symmetric Phase-Only Matched Filter (SPOMF) for Frequency-Domain Software GNSS". *IEEE/ION Position, Location, and Navigation Symposium*, 187–197. April 2006.

6. Teuner, A., O. Pichler, J. E. Santos Conde, and B. J. Hosticka. "Orientation- and scale-invariance recoginition of textures in multi-object scenes". *International Conference on Image Processing*, volume 3, 174–177. Oct 1997.

7. Chen, Q., Michel Defrise, and F. Deconinck. "Symmetric Phase-Only Mathced Filtering of Fourier-Mellin Transforms for Image Registration and Recognition". *IEEE Transactions on Patter Analysis and Machine Intelligence*, 16(12):1156–1168, December 1994.

8. Peyrin, F. and R. Goutte. "Image invariant via the Radon Transform". *International Conference on Image Processing and its Applications*, 458–461. April 1992.

9. Li, L., Z. Qu, Q. Zeng, and F. Meng. "A Novel Approach to Image Rototranslation Estimation". *International Conference on Automation and Logisitics*, 2612–2616. August 2007.

10. Cai, L. and S. Du. "Rotation, scale and translation invariant image watermarking using Radon transform and Fourier transform". *Emerging Technologies: Frontiers of Mobile and Wireless Communication*, volume 3, 281–284. 2004.

11. Li, J., Q. Pan, H. Zhang, and P. Cui. "Image recognition using Radon Transform". *Intelligent Transportation Systems*, volume 1, 741–744. August 2003.

12. Oppenheim, A., R. Schafer, and J. Buck. *Discrete-Time Signal Processing*. Prentice Hall, 2nd edition, 1999.

13. Hahn, W. and S Tretter. "Optimum Processing for Delay-Vector Estimation in Passive Signal Arrays". *IEEE Transactions on Information Theory*, 608–614. September 1973.

14. Jafari-Khouzani, K. and H. Soltanian-Zadeh. "Rotation-invariant multiresolution texture analysis using Radon and wavelet transforms". *IEEE Transactions on Image Processing*, volume 14, 783–795. June 2005.

15. "New Images: Crop Circles in Kansas". Earth Observatory Newsroom, NASA. `http://earthobservatory.nasa.gov/Newsroom/NewImages/images.php3?img_id=17006`.

Vita

2Lt Christopher S. Costello graduated from Bossier High School, in Bossier City, Louisiana in May 2001. In May of 2006, he graduated Magna Cum Laude from Virginia Tech in Blacksburg, Virginia, earning a Bachelor of Science degree in Electrical Engineering with a minor in Leadership. Upon graduation, he was commissioned into the USAF through AFROTC Detachment 875 of Virginia Tech.

Chris's first assignment was the Air Force Institute of Technology (AFIT), where he entered the Graduate School of Engineering and Management for a Mater's degree in electrical engineering. Upon graduation in March of 2008 he will be assigned to the 705th Combat Training Squadron at Kirtland, AFB, New Mexico.

Permanent address: 2950 Hobson Way
Air Force Institute of Technology
Wright-Patterson AFB, OH 45433

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704–0188

1. REPORT DATE *(DD–MM–YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From — To)*
27–03–2008	Master's Thesis	Sept 2006 — Mar 2008

4. TITLE AND SUBTITLE

Multi-Reference Frame Image Registration
for Rotation, Translation, and Scale

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Christopher S. Costello, 2Lt, USAF

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Air Force Institute of Technology
Graduate School of Engineering and Management
2950 Hobson Way
WPAFB OH 45433-7765

8. PERFORMING ORGANIZATION REPORT NUMBER

AFIT/GE/ENG/08-05

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Randall L. Bostick
NASIC/DEMI
4180 Watson Way
WPAFB OH 45433–5648
DSN: 986-0186
email: Randall.Bostick2@wpafb.af.mil

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

This thesis investigates applications of multi-reference frame image registration for image sets with various translation, rotation, and scale combinations. It focuses on registration accuracy improvement over traditional pairwise registration, and also compares the quality of scene estimation from frame averaging. Three experiments are developed which use cross-correlation to estimate translation, the Radon transform to estimate translation and rotation, and the Fourier-Mellin transform to estimate translation, rotation, and scale. Results from applying multi-reference frame registration in these experiments show distinct improvements in both registration accuracy and quality of frame averaging compared to single-reference frame registration. Furthermore, it is shown that the new registration technique is equivalent to the optimal Gauss-Markov estimator of the relative shifts given all pairwise shifts.

15. SUBJECT TERMS

image registration; Radon transform; Fourier-Mellin transform; translation estimation; rotation estimation; scale estimation; multi-frame registration; multi-reference frame registration

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Dr. Richard K. Martin, richard.martin@afit.edu
U	U	U	UU	81	**19b. TELEPHONE NUMBER** *(include area code)* (937) 255–3636, ext 4625

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