8-2017

Computational Imaging Approach to Recovery of Target Coordinates Using Orbital Sensor Data

Michael D. Vaughan
University of Tennessee, Knoxville, mvaugha3@vols.utk.edu

To the Graduate Council:

I am submitting herewith a dissertation written by Michael D. Vaughan entitled "Computational Imaging Approach to Recovery of Target Coordinates Using Orbital Sensor Data." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Computer Engineering.

Mongi A. Abidi, Major Professor

We have read this dissertation and recommend its acceptance:

Seddik M. Djouadi, Suzanne Lenhart, Hairong Qi

Accepted for the Council:
Dixie L. Thompson

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

Computational Imaging Approach to Recovery of Target Coordinates Using Orbital Sensor Data

A Dissertation Presented for the
Doctor of Philosophy
Degree
The University of Tennessee, Knoxville

Michael D. Vaughan
August 2017

Dedication

To

My Father

Whose strength and drive has been an inspiration to me my entire life

My Mother

Whose calm and enduring love provided me with a firm and stable foundation

My Grandfather

William W. Vaughan, who first inspired me to pursue a technical field and whose dedication to his work has motivated me to continue during the difficult times

Acknowledgements

I would like to express my gratitude to Dr. Mongi Abidi for his support and direction during the course of my graduate career. His knowledge of the field and patience in overseeing my work have allowed me to greatly expand my capabilities. I would also like to thank Drs. Seddik Djouadi, Suzanne Lenhart, and Hairong Qi for serving on my committee and for the instruction I received in their courses.

I am also grateful for the fellow students with whom I collaborated on this work. For Timothy Ragland and Jacob D'Avy, my closest lab mates with whom I began this journey, their commiseration helped me to keep up my spirits in the early years. For Harishwaran Hariharan, who was always willing to lend a listening ear to new ideas. For Chung-Hao Chen, with whom I greatly enjoyed collaborating on several works. For Steven Clukey, who helped me through the final stretch of my doctorate. I am indebted to you all.

I would also like to express appreciation for the various sponsors of my work, as well as to collaborating labs who provided ideas and sample data that inspired some of the solutions presented.

Abstract

This dissertation addresses the components necessary for simulation of an image-based recovery of the position of a target using orbital image sensors. Each component is considered in detail, focusing on the effect that design choices and system parameters have on the accuracy of the position estimate. Changes in sensor resolution, varying amounts of blur, differences in image noise level, selection of algorithms used for each component, and lag introduced by excessive processing time all contribute to the accuracy of the result regarding recovery of target coordinates using orbital sensor data.

Using physical targets and sensors in this scenario would be cost-prohibitive in the exploratory setting posed, therefore a simulated target path is generated using Bezier curves which approximate representative paths followed by the targets of interest. Orbital trajectories for the sensors are designed on an elliptical model representative of the motion of physical orbital sensors. Images from each sensor are simulated based on the position and orientation of the sensor, the position of the target, and the imaging parameters selected for the experiment (resolution, noise level, blur level, etc.). Post-processing of the simulated imagery seeks to reduce noise and blur and increase resolution. The only information available for calculating the target position by a fully implemented system are the sensor position and orientation vectors and the images from each sensor. From these data we develop a reliable method of recovering the target position and analyze the impact on near-realtime processing. We also discuss the influence of adjustments to system components on overall capabilities and address the potential system size, weight, and power requirements from realistic implementation approaches.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

3D *three-dimensional*

Bi3 *bicubic interpolation*

BP *biomodality priori*

BSR *blind super resolution*, the name for a method developed by Šroubek et al. which simultaneously performs blind recovery of the blur kernel from a set of images, calculates the inter-frame motion, and reconstructs a higher resolution image

CCD *charge-coupled device*, an imaging device consisting of a capacitive photoactive region which accumulates an electric charge proportional to incident light intensity

CMOS *complementary metal–oxide–semiconductor*, an imaging device where each pixel consists of a photodetector and an active amplifier

CSD *cubic spline and deblur*

DOF *degrees of freedom*

FOV *field-of-view*

FRSR *Farsiu et al Robust SR*

Gm *transconductance*, the electrical characteristic relating the current through the device output to the change in voltage at its input

GUI *graphical user interface*

HR *high resolution*

HSI *hue, saturation, intensity*, an alternative color space representation for images, the most familiar being the RGB color space

IBP *iterated back projection*

IR *infrared*

ISO *International Organization for Standardization*, as used in this document the term refers to the organization’s film speed standards or, for the application to digital sensors, the sensitivity of the sensor to light, often correlated to the potential for an increased noise floor as a tradeoff for increased low light sensitivity

LR *low resolution*

MRF *Markov random fields*

ML *maximum likelihood*

MTF *modulation transfer function*

NC *normalized convolution*

NXC *normalized cross correlation*

PCA *principal component analysis*

PG *Papoulis-Gerchberg*

POCS *projection onto convex sets*, a set theoretic approach to SR

PSF *point-spread function*, a representation of the spreading that occurs of a point source of light as it passes through the focusing optics, most often represented at the focal plane of an imaging system. Akin to a 2D analog of the impulse response of a 1D system

RGB *red, green, blue*, one of the color spaces used to represent color images

RSR *robust super resolution*

SIFT *scale-invariant feature transform*

SNR *signal-to-noise ratio*

SR *super resolution*, a class of image processing methods which enhance the spatial resolution of a set of LR image observations of a HR continuous scene

SRM *statistical region merging*

SSIM *structural similarity index*, a reference-based image quality metric

SURF *speeded-up robust features*

TV *total variation*

ZRSR *Zomet et al Robust SR*

1 Introduction

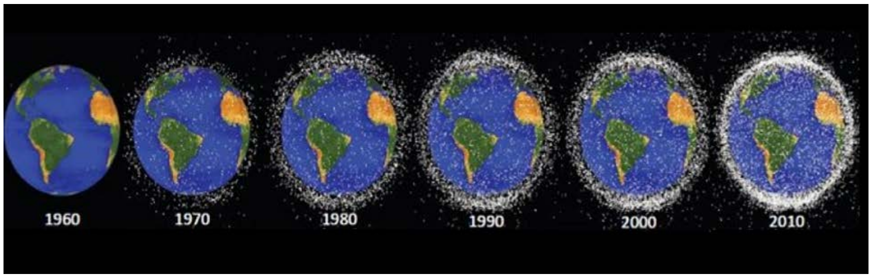
The purpose of this research is to contribute to the development of an approach for recovery of the 3D world-centric coordinates of an aerial target in motion. The targets of interest exhibit predictable spectral signatures suitable for identifying them from low resolution images with a noisy background. The sensor images are processed to enable location of the targets in the image despite significant degradations in the images. Knowledge about the position and orientation of the sensors is also used to calculate a reliable estimate of the target's position. The motivation of this work and associated research objectives are detailed in Section 1.1. Then in Section 1.2 we discuss the completed work and present our contributions.

1.1 Motivation

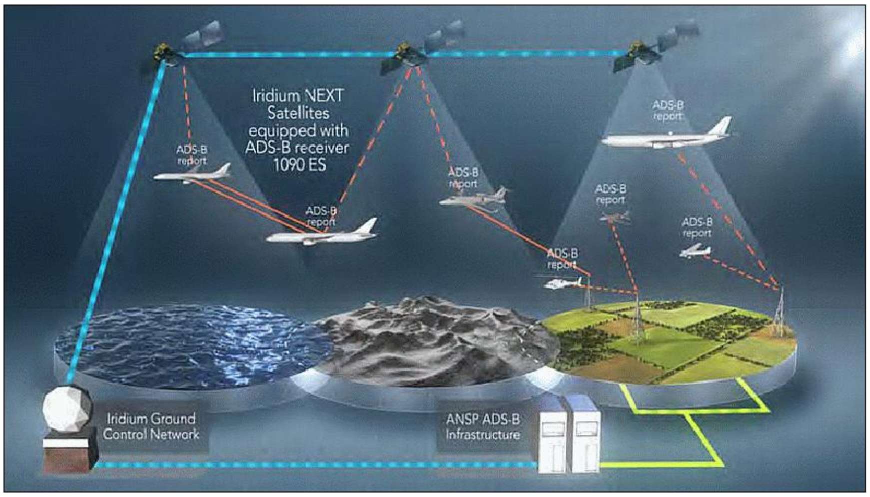
The application that motivates this dissertation is based on the needs of one of the sponsors who supported my research. The areas of contribution are my own research interests performed prior, during, and after the sponsor's need was met and adapted for the structure of the application presented. Each contribution was initially developed in response to the needs of various sponsors and designed as generalized solutions and adapted or applied to the needs of the sponsor. The work presented as existing was developed either by other researchers with whom I collaborated or whose work was competitive in the state of the art for the functional block to complete the system as designed. Therefore, the individual components presented have much broader application than the specific case presented (some possible applications are shown in Figure 1.1), and results shown are tailored for this case. Conclusions drawn may not be directly applicable to a significantly different application, but effort has been made to detail the range of capability offered which should be instructive for decisions in other use cases.

Many factors influence the accuracy of 3D positional recovery of a target from image-based sensors. Conditions of the atmosphere degrade the path by which light travels from the target to the sensors. The quality of the optics can impede the recovery process. The resolution and other sensor characteristics may cause errors to be introduced into the estimation process. An analysis of the contribution to and scope of the error from these sources enables proper planning and design of improvements to the system components responsible for the greatest contribution of error to maximize system capability and reliability while minimizing cost.

There are limitations to the capabilities a given hardware implementation can provide. With respect to the optics, for instance, it is impossible to image beyond the resolution limit established by the Rayleigh criterion. Higher resolution imaging sensors are always in demand, but due to shot noise there is a limit in how small the physical dimensions a given pixel can be before the noise overtakes the signal, and often making larger sensors to accommodate higher resolution sensors with reasonably sized pixels is infeasible due to equipment size constraints. Where improvements to the hardware design and implementation cannot overcome the physical limits restricting them, more advanced image processing and computer vision algorithms may be applied in order to remove the deleterious artifacts in the images or utilize fusion methods developed to take advantage of the system setup. Noise from the sensor and blur and distortion from the lens may be diminished or removed with proper calibration. Turbulent effects of the atmosphere may be modeled and accounted for. Low resolution of the sensor and restrictions due to the diffraction limit may be overcome by proper application of super resolution.



(a)



(b)

(c)

Figure 1.1: *Motivating applications and improvement on the state of the art described in this dissertation. Accuracte knowledge of the location of (a) orbital debris, (b) airplanes, and other targets of interest is vital for making intelligent, timely decisions. Existing systems using radar (c) are expensive and inaccurate in many situations. The LANZA radar, for instance, costs ~$30 million and has an accuracy limit depicted above at 250 nautical miles (NM). Utilizing orbital sensors to detect and calculate the position of moving targets can improve the accuracy and timeliness of measurements at a lower cost and can provide additional information not possible with current systems.*

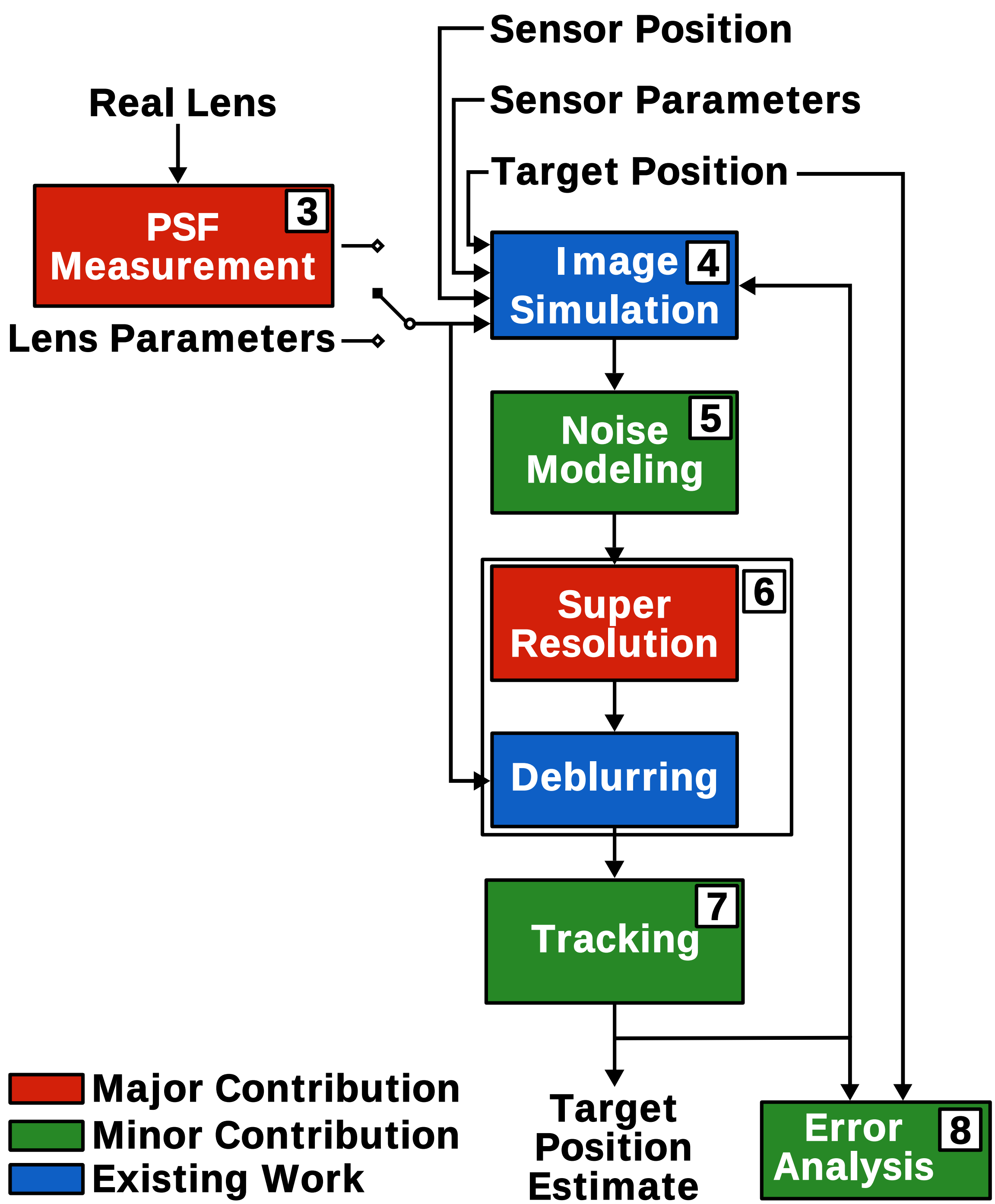


Figure 1.2: *Pipeline of contributions. The ideal target image is generated from known information of the sensor's position, resolution, and field-of-view (FOV) combined with the position of the target from the simulated path. We simulate the effects which cause a disturbance to the trajectory of the target, sensor, and infrared (IR) light as it travels between the two locations, and the lens aberration step factors in the effect a real lens has on image acquisition. We then apply our image segmentation methods to separate candidate signal information from the background noise. Next we separate the target from these candidate regions based on appropriate IR signatures. Finally, we estimate the 3D pose of the target in the frame and analyze the error with respect to the known position of the target.*

1.2 Contributions

The 3D position recovery process pipeline is presented in Figure 1.2 as the structure for the work performed. The process pipeline consists of seven key steps: point-spread function (PSF) measurement, image simulation, noise modeling/mitigation, super resolution, deblurring/deconvolution, tracking, and error analysis. The image simulation step depends on simulated target path positions. The sensor position is determined by the orbital parameters and the orientation is adjustable based on an update algorithm. The lens parameters are adjustable, and are represented using the PSF of either the measurement of an existing lens or one simulated using a lens prescription or using a generic blur kernel. The sensor parameters are also adjustable and include noise level/distribution and sensor resolution and size. The PSF measurement approach was developed in order to create an accurate spatially-variant representation of a physical lens.

We want to handle multiple sensors designed to image in the visual and IR bands, we include a step to handle noise modeling, analysis, and mitigation. The targets of interest in the images are faint, often near the noise floor. As such we must exercise caution in our detection and removal of the noise. We also must handle the low resolution of the sensors with respect to the size of the target, often on the order of a single pixel. By employing multiple unique measurements of the scene we can enhance the spatial resolution of the images in order to enhance our ability to recover the position of the targets within the FOV. We employ deblurring/deconvolution of these images utilizing our known lens blur in order to further improve our 3D position recovery capabilities. Once we have acquired and restored the images of the scene, we employ our target tracking step to recover the coordinates of the target in 3D space.

1.2.1 Major Contributions

1.2.1.1 PSF Measurement

Nearly all modern imaging systems make use of optical lenses in order to focus the image of the scene while the aperture is set to allow more light from the scene. However, physical optics introduce geometric distortion and spatially variant blur of the light from the scene recorded by the sensor. The blur introduced is modeled by the PSF. If we accurately measure the PSF of a lens, we are able to reduce or eliminate the negative influence of the lens optics on image quality. Our method makes use of a grid of pointlike sources which image to less than a pixel with respect to the sensor used in the imaging system. The blur effects then will be the components resulting from the lens blur. In this way we are able to directly measure the spatially variant PSF of the lens and use these measurements in later corrective steps.

1.2.1.2 Super Resolution

The light entering a digital optical system is recorded by a sensor with limited resolution. The resolution available is often insufficient for the intended use and obtaining a higher resolution sensor may be prohibitive due to cost or technology limitations. Classical super resolution (SR) involves integration of a series of unique images of the scene which can be thought of as compression of the time domain in order to enhance resolution in the spatial domain. We acquire images of the scene at a high framerate with respect to motion of the targets within the frame and apply a sliding window combination of the frames in order to create a higher resolution reconstruction of the scene at approximately the same framerate with minimal delay.

A second approach we developed involves applying the same algorithms on a set of images acquired with a modified imaging setup. One of the most difficult parts of SR is obtaining good data of a scene with rapid movement. We created a multi-aperture system so that the requisite number of images required for applying SR could be obtained simultaneously.

1.2.2 Minor Contributions

1.2.2.1 Noise Mitigation

The targets in the scene comprise a small region of the image, and the contrast of the targets is such that the targets may easily be hidden near the noise floor. The noise distribution for a sensor can be measured and the noise in the image estimated. We take advantage of this a priori knowledge of the distribution type and utilize analysis of the images themselves. Care must be taken as overly aggressive mitigation techniques have the potential to obscure what little signal we have available in each individual image. The approach we have taken with respect to handling the noise in the images breaks into two areas: (1) modeling, where we analyze the image to obtain a model of the noise type and intensity; and (2) removal or mitigation, where we apply techniques to correct for the noise present in the images. We only model the noise when we are able to apply a simple thresholding to obtain candidate target regions (usually for low noise levels). By applying a time-domain correlation of the result we can reduce the number of false positives. Where we attempt to remove the noise from the image before further processing, we carefully analyze the statistics of the image and apply current state-of-the-art removal techniques [Liu2012][Liu2013][Luisier2011] [Zanella2009] in order to improve our position estimate in preparation for subsequent processing steps.

1.2.2.2 Tracking

Our tracking step encompasses the image-based aspects of target localization with respect to target detection or separation from other image elements and 3D recovery of the target's position in world coordinates from available sensor imagery. Imagery available at this stage has had all preprocessing and correction steps performed so the tracking step comprises all higher-level image understanding and decision making. First, we detect the presence of candidate locations of a target of interest within the image. Our application employs hyperspectral signatures of a target of interest and we conducted tests using a variety of spectral bands based on the signatures available. Next, we must localize the target with respect to a global coordinate system, eliminating false positives from the candidate locations. The target detection step is performed for each sensor for which the target is visible, and once we have isolated the target in image coordinates, we utilize the sensor position and orientation information to estimate the target position in global coordinates. This is possible with a minimum of two sensors and may be framed as a stereo recovery. However, we have also implemented an approach which allows information from additional sensors to be included in order to improve the estimation accuracy of the position recovery.

1.2.2.3 Error Analysis

The error analysis step is not explicitly in the feedback loop of the system. Data for an entire experimental run is collected and analyzed after completion though the incremental information is displayed live. This is due to the fact that we will not know the ground truth information live and cannot base any of our system calculations or updates on this information. We again make use of the ground truth target positions and compare the position estimate from our raw and fusion approaches to the true target position at each timestep. We represent the XYZ position data as a difference from the ground truth XYZ position of the target, so that the data representing the ground truth in our comparison is zero for each time step, and the data for the raw estimates and the fused results represents the Euclidean distance of the estimate from the true target location. Represented as a distance from a perfect recovery, an indicator of the quality of a processing method would be closer to zero while an increasing distance away would indicate a less accurate method. We use this final set of data in our evaluation of the quality of our method and as an indicator of the effect of varying parameters such as the resolution, blur level, baseline, etc. and are thus able to perform sensitivity analysis on these parameters.

1.2.3 Existing Work

1.2.3.1 Image Simulation

Image generation requires four sets of information at each time step: (1) the target position, (2) the sensor position and orientation, (3) the lens parameters, and (4) the sensor parameters. The position of the targets we wish to track includes additional objects surrounding the target which serve to obscure the true position of the target from detection. Some paths are derived from measurements of actual flight paths of representative targets and some tracks were generated using Bezier curves matching the profile type of the flight paths of true target trajectories. The sensor orbits are set prior to each experiment, but are adjustable in order to explore the impact of the number of sensors, baseline between sensors, and height of the sensor relative to the target path with respect to the accuracy of target position recovery. Thus the position of the sensor is adjustable prior to an experimental run, but not adjustable in real time. Likewise the lens and sensor parameters are adjustable prior to a given experiment in order to investigate the influence of differing types and combinations of each on the accuracy of position recovery. We encode the information about the influence of the lens in the non-parametric PSF and can explore both spatially variant and global PSFs. We measure several types of physical lenses and also include standard Gaussian models and other types of blur. The sensor model includes adjustable parameters of the sensor size and resolution, as well as an adjustable level and distribution of noise. Finally, we use Maxwell's Demons method to model the atmospheric degradation given the significant distance from target to sensor in the experiments performed.

1.2.3.2 Sensor Orientation Update

Once we have recovered the position of the target with the imagery from the sensors, we must update the sensor's orientation to keep the targets in view, ideally in the center of the FOV. In order to avoid ambiguity in the reorientation, specifically to avoid gimbal lock and to maintain calculation of a smooth path, we have adapted the use of quaternions for representing the current orientation and for calculation of the incremental orientation update vectors. We have developed two approaches for calculating the update vectors. The first of these assumes a simple calibrated projection calculation of the relationship between the x, y motion of the target in pixel coordinates and the necessary change in angle of the rotational degrees of freedom of the sensor. This amounts to a scaling of the motion seen as projected to the image plane and directs the update to happen relative to this plane with respect to the current orientation of the sensor. The second method is predictive and is based on the current estimate of the target position and the velocity of the target based on the positional estimates of the target between the current and previous time steps. The velocity of the target changes slowly between adjacent timesteps, so utilizing this estimate in order to predict the target position in order to reorient the sensor for the subsequent timestep gives a reasonably accurate result.

1.2.3.3 Deconvolution

The problem of restoring an image from a blurred observation is termed deconvolution. We know the spatially variant PSF of the lens system source of intrinsic blur in the acquired images from our prior PSF measurement of the lens, therefore the type of deconvolution algorithms we will employ are called non-blind (as opposed to blind approaches for which the blur kernel is not known). We apply existing methods to recover the original scene by utilizing the knowledge of an image of the scene and the PSF of the lens which corrupted the image. While deconvolution does not improve the spatial resolution in terms of pixels, it can significantly improve the effective resolution and enhance contrast between closely-spaced features.

1.3 Document Organization

The remainder of this document is arranged as follows:

- **Chapter 2** presents the state of the art of the literature for the areas addressed in this dissertation.
- **Chapter 3** discusses the method developed to measure the PSF of a physical lens. A description of the approach taken to measure the geometric distortion field is also included.
- **Chapter 4** contains a description of our image simulation approach and our application of competitive approaches to each component in the process. We also present our approach to updating the sensor orientation.
- **Chapter 5** briefly introduces the problem of noise in images and our approach to modeling and mitigation of the noise.
- **Chapter 6** introduces a description of our contribution in analyzing the capabilities of super resolution approaches to the enhancement of our image sets. We also discuss our application of deconvolution using either the measured or synthetic PSF.
- **Chapter 7** addresses our tracking work, covering the areas of segmentation of the image, target detection in the presence of heavy noise, target separation from similar false positives, and our contribution in 3D position recovery of the world coordinates.
- **Chapter 8** covers our comparison of experimental results comprising an analysis of the sensitivity of our position estimate to the various input parameters which leads to an understanding of the scope of capabilities of the overall system.
- **Chapter 9** concludes with a summary of tasks completed for this dissertation.

2 State of the Art

In this chapter we discuss the general problems of PSF measurement, synthetic image generation, noise modeling and reduction, super resolution, deconvolution, tracking, and error analysis. For synthetic image generation, a description of the approach and methods from the literature for the main components will be discussed. Our approach includes some elements less commonly found in conventional imaging but with the exception of one component, all areas should be considered compatible. Also for sensitivity analysis, the approach taken will be outlined and comparable approaches taken by contemporary papers in the literature will be compared. For PSF measurement, noise modeling and reduction, super resolution, deconvolution, and tracking, multiple current approaches will be described and the methods selected for comparison in order to determine the most appropriate method for application to our scenario. In addition, a new method for PSF measurement will be demonstrated.

2.1 PSF Measurement

Accurate knowledge of the PSF is essential in applying deconvolution which enhances the data by reducing blur, thus increasing the effective resolution. PSF estimation techniques generally utilize a calibration pattern and are known as non-blind, taking the form of targets relying on random noise [Levy1999], [Delbracio2012], [Brauers2010], sharp edges [Tang2013], [Claxton2007], or other means [Hu2012], [Schuler2012]. Blind methods use various single-image or multi-image approaches [Sroubek2008], [Hu2012], [Schuler2012] that generally use statistical analysis or blurred edge detection. Some of these methods require a parametric fitting of the PSF [Aguet2008], [Howell1996], [Zheng2006], [Claxton2007], while others do not impose this constraint [Liu2008], [Sroubek2008], [Levy1999], [Delbracio2012], [Brauers2010], [Hu2012].

Delbracio et al [Delbracio2012] indicate several requirements for a calibration pattern-based PSF estimation technique: namely, that a PSF estimation method should be non-blind; that kernel estimation must be two-dimensional, local, and subpixel; and that optical distortion, nonuniform illumination, and nonlinear sensor response must be accounted for in the method. Furthermore, no regularization should be used to enforce radial symmetry constraints, parametric models of the PSF, or other kernel regularization constraints. They claim an adequately chosen noise pattern (they use Bernoulli distribution) eliminates the need for regularization, which dampens the high-frequency content of the PSF. Due to the relative generality of the method and careful consideration of relevant factors, their method will be used as a comparison despite the fact that it is an estimation method, rather than a direct measurement method.

Direct measurement of the PSF [Liu2008], [SVI] refers to methods that utilize a point source or point-like source of light with specific properties. Delbracio et al [Delbracio2012] and Gunturk and Li [Gunturk2013] state that the ideal target for measuring the PSF is a perfect pinhole image simulating a 2D impulse. They then dismiss the approach as being unfeasible due to the low signal to noise ratio (SNR) as the ideal spot size is infinitesimal. Some methods [Tang2013], [Aguet2008] incorporate a fitting of the PSF. This direct measurement yields a sampled version of the continuous PSF, which contains all intrinsic distortions of the aperture, optics, and sensor array. Extrinsic factors that affect the degradation of the captured scene include motion blur (including motion of the system relative to the scene and the motion of subjects within the scene), atmospheric effects, lighting variations, and object distance. These factors will depend on conditions during acquisition and cannot be included in our measurement.

There exists a standard for measurement of the modulation transfer function (MTF) of an imaging system [ISO12233], but as yet no agreed-upon standard exists for a digital optical system regarding measurement of its PSF. Direct measurement of the PSF necessitates construction of a point-like source, and in the case of our method and the method of Shih et al [Shih2012] and Navas-Moya et al [NavasMoya2013], a grid of point-like sources are used for measurement of the spatially variant PSF.

Liu and Chen [Liu2008] propose a PSF measurement method that uses an interferometer to directly measure the high resolution PSF of the lens. The lens is removed from its paired sensor and since the exact alignment of the optical axis relative to a given pixel is not known, an average is taken of a 10x10 grid across the area of a pixel from the original sensor and all 100 images of the PSF are averaged to account for the unknown alignment. This version of the PSF is then downsampled to match the pixel size of the sensor array that was originally paired with the lens. First, removal of the lens from the optical system and sensor that the lens was intended to be used with carries the potential for damage to the sensor, and reassembly of the system requires specialized skill. Further, measurement of the PSF using a device (the interferometer) that contains other lenses and sensors each with their own PSF and pixel response carries with it the introduction of spurious influences on the measured result. Finally, Liu and Chen compare their measured PSF applied to an image obtained with a different lens/sensor combination of the same model. Shih et al [Shih2012] note the problem with assuming all physical lenses are representative of the lens prescription used during their manufacture; manufacturing tolerances introduce a significant difference in the PSF of individual lenses.

We propose a direct PSF measurement method that uses a pinhole grid pattern and employ a super resolution approach to reach a subpixel measurement as noted in [Liu2008], [Delbracio2012], which is then downsampled to the pixel size of the paired sensor. We are able to use the optical system to measure its own PSF and the device is still able to function. Our method is nondestructive to the equipment used, and re-measurement of the PSF may be performed at any time with limited interruption to the use of the system. This feature is of particular importance for systems in environments where pressure and temperature changes and vibration are prevalent which leads to small changes in the alignment of the sensor and lens elements.

2.2 Synthetic Image Generation

Knowledge of the performance of image processing and computer vision algorithms is critical in the selection process for a given application. Proper testing of an algorithm's performance involves accurate knowledge of the expected outcome, often referred to as ground truth. Synthetic test images offer the ability to easily test an algorithm's performance by generating features in the image for the algorithm to detect, classify, remove, or enhance. The images are generally created to have information which is complex enough to be representative of the type of imagery that will be encountered while still being simple enough to determine the success of the method.

Synthetic test images have been used in the development of iris recognition [Zuo2006], biomedical microwave imaging [Kundu2010], vehicular safety [Cosel2009], distortion field estimation [Tian2001], stereo depth estimation [Chen2011], edge detection [Benjelloun2007], image-to-text classification [Wang2006], and satellite image segmentation [Marcal2010] algorithms, among others. As the formation of the image is specific to the construction of the imaging system and environment being imaged, the approach taken will vary by application.

2.2.1 Blur Types

Many authors [Lucas1996][Smith1997][Kuhl2006][Aguet2008][Kuthirummal2011][Tang2013] have attempted to represent the blur found in images using parametric models. However, Shih et al [Shih2012] state that the assumptions made which lead to these simplified forms are not valid for real lens systems and as such are not useful in general when complex degradations are present in the image, which is the case in varying degrees for all real data not synthetically generated. Shih et al further note that for fixed focal-length lenses, such as the ones we consider in our work, the PSF is a 6-dimensional function of the light wavelength, λ, image plane coordinates, x,y, lens aperture, a, lens to object distance, d_{obj}, and back focal distance, d_{bfd}. They and others [Delbracio2012][Liu2008][Joshi2008] advocate the use of the nonparametric PSF as a representation of the effects of the optical system. The use of parametric models for the PSF are necessary when using an ill-posed blind PSF estimation approach in order to simplify the solution space but the accuracy of the estimation is affected by this approach [Delbracio2012].

When dealing with real imaging systems, we can measure the PSF or if we know the lens design we can simulate the PSF using raytracing to obtain a nonparametric blur kernel [Delbracio2012]. When using a nonparametric blur kernel, it is encoded in a matrix similar to how digital images are represented; the blur then is applied using a convolution operation. This section contains a description of the types of blur models which can generate a nonparametric representation of the blur.

2.2.1.1 Pillbox Function

The PSF of an ideal thin lens is represented by a uniform-intensity disk, called the pillbox function [Tang2013][Kuthirummal2011]. The model for this function is of a single parameter, *r*, the radius of the disk as seen in (2.1). There are several authors who advocate the use of the pillbox model in fitting the measured PSF [Zheng2006][Smith1997][Tang2013], but the most useful application is in modeling defocus blur. The form of the pillbox model is given in [Kuthirummal2011] as

$$H(r,b) = \frac{4}{\pi b^2} \prod \left(\frac{r}{b}\right) \tag{2.1}$$

Where r is the distance of the pixel from the centroid of the PSF,

$\prod(\mathrm{x})$ is the rectangle function, and

b is the diameter of the PSF, given by

$$b = \frac{a}{v} \tag{2.2}$$

Where a is the diameter of the system aperture and

v is the distance from the lens to the sensor (back-focal distance).

Smith [Smith1997] states that the pillbox is similar enough to the Gaussian and the exponential to be considered the same with respect to resolution preserved in the image. Smith also states that the source of a pillbox PSF for an imaging system when the lens is improperly focused. Tang and Kutulakos [Tang2013] use the pillbox model to represent the PSF of an ideal thin-lens system and compare the performance of reconstruction algorithms using a real-lens PSF (simulated with 3rd-order Seidel aberrations). Kuthirummal et al [Kuthirummal2011] also describe the pillbox as the form of an ideal model for the PSF. Their findings indicate that more information about the original image is contained in images observed with the real-lens PSF than with that of the simple model. Finally, the method introduced by Zheng and Hellwich [Zheng2006] allows the algorithm to select which model of blur fits the data best, which includes the pillbox function as an option.

2.2.1.2 Gaussian

The Gaussian function is 2-dimensional and assumed to be rotationally symmetric, for purposes of modeling the PSF of an imaging system. As such the representation of this model of the PSF is as described in (2.3).

$$H(r,b) = \frac{2}{\pi b^2} exp\left(-\frac{2r^2}{b^2}\right) \tag{2.3}$$

Where r and b are given as described above.

The Gaussian model is the most popular simple parametric model used to describe the PSF of a lens, partly due to its simple form and common use in other areas such as noise modeling when the exact distribution is unknown or difficult to calculate. This is likely influenced by the central limit theorem as indicated by [Zheng2006], who also include the Gaussian in their model-fitting selection structure. Many authors [Dore2004][Zheng2006][Capel2004][Smith1997]

[Kuhl2006][Kuthirummal2011] make use of the Gaussian model in their work. Smith [Smith1997] indicates that the Gaussian form results from the random combination of errors, such as from viewing stars through a turbulent atmosphere. He also notes that since the Gaussian decays rapidly it is a simple matter to truncate the function without introducing side effects. Kuhl et al [Kuhl2006] state that the assumption of a Gaussian model for the PSF is reasonable when a variety of aberrations are present due to the lens. Finally, Kuthirummal et al [Kuthirummal2011] make use of the Gaussian model in their proposal of an integrated PSF (IPSF) constructed by varying the distance of the lens to the sensor during the integration time of the sensor. Gunturk and Li address experimental verification of a Gaussian model of a sensor's pixel-level PSF [Gunturk2012]

2.2.1.3 Sinc

Of the simple parametric models used, the sinc function is the most interesting as it more accurately represents the nature of the PSF to have components which continue in diminishing periodic groupings. One of the drawbacks is the fact that the PSF will inherently contain no negative components, while the sinc function does.

$$sinc(x) = \frac{sin\, x}{x} \tag{2.4}$$

A simple solution to this is to take the absolute value of the function, or to only consider the positive components.

2.2.1.4 Jinc

The jinc function, also called the besinc or sombrero function, is more commonly used in image processing than the sinc function [Hendee1997], and is defined as:

$$jinc(x) = \frac{J_1(x)}{x} \tag{2.5}$$

where $J_1(x)$ is a Bessel function of the first kind. However, since

$$\lim_{x \to 0} jinc(x) = \frac{1}{2} \tag{2.6}$$

the function is often [Siegman1986] multiplied by a factor of 2 in order to achieve $jinc(0) = 1$.

Smith [Smith1997] claims that these functions are seldom used in image processing because images do not have their information encoded in the frequency domain. He notes that the long "tail" of the function can cause some problems and must be truncated to be of any practical use in computation or filter design. Aguet et al [Aguet2008] indicate that for their system, the use of the jinc function in modeling the PSF simplified the computational complexity while preserving the accuracy of their solution.

Kuhl et al [Kuhl2006] state that the jinc function used exactly describes the diffraction pattern due to a circular aperture, using the modified form in (2.7), but state that the Gaussian is a reasonable approximation to the PSF when lens effects are considered.

$$A(r) = \left[\frac{J_1(r)}{r}\right]^2 \tag{2.7}$$

Where $r = \sqrt{x^2 + y^2}$ and x and y are the Cartesian image plane coordinates.

2.2.1.5 Exponential

For modeling of the lens effects of a camera operating in the visual spectrum, we have not encountered systems which follow this form. However, Smith [Smith1997] states that the exponential is representative of PSFs commonly found in imaging systems. He further notes that this form is generated when electrons and x-rays strike a phosphor layer and are converted into light, and that the model is used in radiation detectors, night vision amplifiers, and CRT displays.

It seems for some of these systems that this may be an arguably valid model, as Lucas and Roche [Lucas1996] also make use of this form in astronomical imaging. We have not used systems of this type in our work but this parametric model is included for completeness.

$$H(r,g) = e^{-\left(\frac{-2g\left(\sqrt[2]{x^2+y^2}-r\right)}{ln(2)}\right)} \tag{2.8}$$

Where r is the radius of the PSF at full-width half maximum (FWHM),

g is the slope of the PSF at FWHM, and

x and y are the pixel coordinates, centered at 0,0.

2.2.1.6 Defocus Blur

Defocus blur refers to the blur seen in an image not due to the intrinsic blur, but rather due to the camera's focus setting relative to the distance of an object in a scene. As there can be a wide range of distances for various objects within the field of view, this function can be quite complex. A common compromise in designing lenses with no adjustable focus range is to set the focus equal to the hyperfocal distance [Piper1901].

$$H = \frac{f^2}{Nc} + f \tag{2.9}$$

Where H is the hyperfocal distance,

f is the focal length of the lens,

N is the f-number, and

c is the circle of confusion.

By setting the focus of a fixed-focus lens to the hyperfocal distance, the maximum depth of field is achieved and all objects in the scene from half the hyperfocal distance to infinity are in acceptable focus.

Although the pillbox model is generally accepted for representing defocus blur [Zheng2006][Smith1997][Kuthirummal2011], Kuhl et al [Kuhl2006] use the Gaussian model, which is also noted as acceptable by Kuthirummal et al [Kuthirummal2011] when the intrinsic blur is significant enough to not be ignored. For scenes where the object depths vary significantly, a depth map should be constructed in order to handle each region appropriately [Gunturk2012].

Many authors [Lucas1996][Kuhl2006][Kuthirummal2011][Delbracio2012] assume spatial invariance when applying blur when constructing synthetic images, although Kuthirummal et al [Kuthirummal2011] claim their approach to PSF measurement is invariant to both shift and depth. This approach is acceptable when demonstrating a deconvolution method's effectiveness on a given type of blur. However, images from real lens systems will typically exhibit a strong spatial variance in the PSF. This means that when the nonparametric PSF of an imaging system is represented, a single PSF measurement will not be sufficient but rather an array of PSF measurements will be necessary to accurately represent the local anisotropic blur kernel [Delbracio2012]. When using real images it is important to keep this distinctive attribute in mind when handling deblurring. In demonstrating our ability to recover 3D coordinates from representative imagery applying a stationary blur kernel is sufficient.

2.2.2 Noise Types

Noise in digital images can be classified with respect to its source or its distribution. Sources of noise are typically classified as fixed-pattern noise (FPN), banding noise (horizontal or vertical), or random noise. FPN arises from the construction of the device but can be influenced by environmental conditions. Presence of banding noise comes from the readout and amplification circuitry of the device. Random noise comes from a variety of physical sources on the device but is heavily influenced by environmental conditions.

Noise can also be classified as white or colored noise. White noise varies with time while colored noise is more stable and can generally be removed through a calibration process. This arises from their different sources within the device. White noise comes from the light-collecting pixel region while colored noise comes from the readout and amplification circuitry.

The noise types include shot noise, generation/recombination noise, popcorn noise, thermal noise, kT/C noise, and flicker noise. Shot noise has a Poisson distribution and describes the expected arrival event of a photon on a given pixel [Lundberg2002]. Generation/recombination noise occurs due to the trapping centers in the bulk of the silicon making up the device [Hooge1994]. Popcorn noise is caused by discrete modulation of the channel current [Lundberg2002]. Thermal noise is caused by the voltage fluctuation due to Brownian motion of electrons in a resistive medium [Lundberg2002]. kT/C noise describes thermal noise in the presence of a filter capacitor [Lundberg2002]. Flicker noise comes from oxide traps resulting from imperfections in the construction of the transistor material [Jayaraman1989].

2.2.3 Motion Model

Fusion methods employ data from multiple sensors to gain a more comprehensive and stable representation of the observed scene. Image fusion methods accomplish this from processing multiple images of the scene to achieve a goal that would be less accurate or reliable using only a single image. The motion of objects in the scene or egomotion of the sensor or sensors gives rise to the image registration problem, meaning that data contained in two or more images must be correlated in order to combine or otherwise perform higher order analysis. The expected motion of the objects in the scene, relative rigidity or deformability of the objects, and path and change in orientation of the sensors drives the selection of a motion model to which points in the image are matched and parameters are calculated to describe the change of these points from one image frame to the next. This extracted information is then used in subsequent processing steps.

Table 2.1: *Comparison of global 2D planar transformations included in each motion model. For a (3x3) matrix, the number of DOF indicate the number of adjustable parameters available for computing the transformation. In the fourth column, the given entries indicate the image content preserved during the transformation. The plus (+) symbol indicates that each attribute below the given row is also included as each row is a subset of the row below, thus a translation operation preserves orientation, Euclidean distance, angles, etc. whereas the Euclidean motion model does not preserve orientation.*

Name	Degrees of Freedom (DOF)	#DOF	Preserves
Translation	Translation	2	Orientation+
Euclidean	Translation, Rotation	3	Euclidean distance+
Similarity	Translation, Rotation, Scale	4	Angles, curvatures+
Affine	Translation, Rotation, Scale, Aspect, Shear	6	Straight line parallelism+
Projective	Translation, Rotation, Scale, Aspect, Shear, Keystone	8	Straight lines

Table 2.2: *Matrix form of the 2D planar transformations. For the motion models shown in Table 2.1, the transformations may be applied serially or composited into a single transformation matrix. As keystone (projective) transformations are considered, the homogenous coordinates are used, if this transformation type was not considered then (2x2) matrices would be sufficient to represent the remaining transformations.*

Translation	Rotation	Scale	Aspect	Skew	Keystone
$\begin{pmatrix}1 & 0 & t_x\\ 0 & 1 & t_y\\ 0 & 0 & 1\end{pmatrix}$	$\begin{pmatrix}\cos\theta & -\sin\theta & 0\\ \sin\theta & \cos\theta & 0\\ 0 & 0 & 1\end{pmatrix}$	$\begin{pmatrix}s & 0 & 0\\ 0 & s & 0\\ 0 & 0 & 1\end{pmatrix}$	$\begin{pmatrix}s & 0 & 0\\ 0 & \frac{1}{s} & 0\\ 0 & 0 & 1\end{pmatrix}$	$\begin{pmatrix}1 & h_x & 0\\ h_y & 1 & 0\\ 0 & 0 & 1\end{pmatrix}$	$\begin{pmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ k_x & k_y & 1\end{pmatrix}$

Motion estimation models may be classified as global or local. Global methods use all extracted control points in the image for calculating the model parameters of a transformation. Local methods use a patch-based approach and may warrant application of an appropriate segmentation method. Images captured with significant lens distortion are particularly well-suited for application of local motion estimation correction as an object appearing in different areas of the image will be subjected to differing portions of the geometric distortion field.

These motion models utilize all of the basic planar transformations. While more available DOF allows for more flexible calculation of the relationship between frames, this comes at increased computational cost. Further, if a higher-order motion estimation is selected for an image sequence for which the additional DOF are not present then extraneous fitting of those parameters may occur, especially in images with high blur and noise components. The mathematical form of the DOF described in Table 2.1 is shown in Table 2.2. Note that all are modifications to the identity matrix with the on-diagonal elements representing scaling along one of the dimensions.

These matrices may be composited into a single transformation matrix by successive multiplication of the component transformations. In order to apply these motion estimation models effectively, the images of a scene are assumed to have been taken either from the same camera or from a multi-camera system with very narrow baseline with respect to the distance to the scene. If there is significant angular distance between the cameras or position of a single camera between image acquisitions then a 3D model of the image acquisition is needed to accurately relate the inter-frame motion rather than the 2D planar methods described above. Figure 2.1 gives a graphical representation of the matrix transformations shown in Table 2.2, noting that the keystone and projective transformations are equivalent.

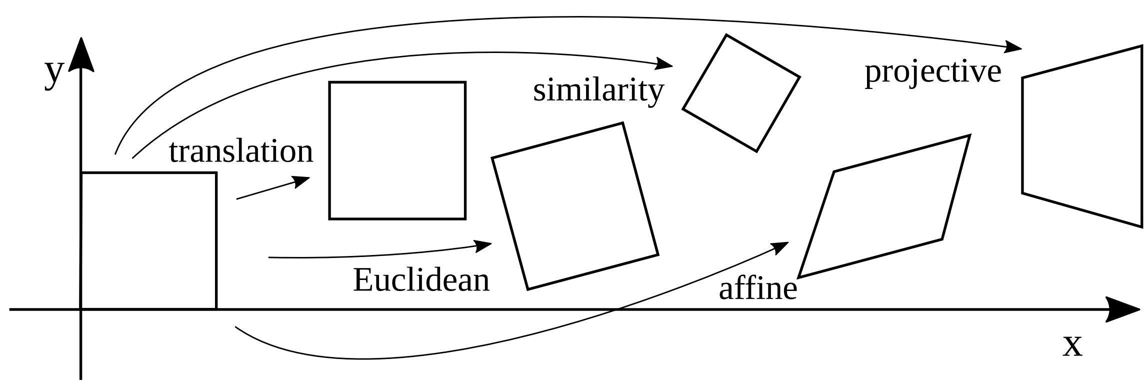


Figure 2.1: *Basic set of 2D planar transformations. For image sequences in which the egomotion of the camera or the disparity in a multi-camera system is small, this set of 2D planar transformations is sufficient to represent all motion assuming rigid objects in the scene or simultaneous acquisition. Reproduced from [Szeliski2006]*

If the objects in a scene are static, then global motion models work well but will generally fail in the presence of occlusion or due to transparency or other uniform regions for which control points are difficult to match accurately [Babu2011]. Quite often, the objects in a scene move independently of one another and with respect to the egomotion of the sensor(s). Application of local motion estimation methods can be useful for many applications. Delbracio et al [Delbracio2012] use local perspective transformation on image patches to correct for complex lens distortion.

Application of an image segmentation method to extract information relevant to the desired subject can be one way to not only reduce the size of the data but also to remove extraneous information in the scene and to calculate motion only on the relevant information. If the object of interest is rigid a valuable approach can be to segment the available frames to create a data subset with the relevant image data corresponding to the object of interest rather than the whole scene (e.g. enhancing license plate data for passing to an optical character recognition algorithm).

Most estimation methods use a two-frame approach, where one frame from the image series is selected as a reference for all other frames. While still an open problem, simultaneous multiframe motion estimation should provide needed robustness [Babu2011]. Szeliski calls this bundle adjustment when referring to calculation of the camera positions simultaneously to achieve a globally consistent solution [Szeliski2006].

2.3 Noise Modeling and Reduction

Many potential sources contribute to the noise type and level present in digital images. Understanding these sources and the methods for reducing their impact pre- and post-acquisition can improve the SNR of the image and the accuracy of algorithms which use the images as input. It is better practice to minimize the potential for noise during the acquisition process, but often due to environmental conditions, cost, or other factors it is necessary to remove or mitigate the influence of noise post-acquisition.

2.3.1 Pre-Acquisition Approach

Environmental conditions can have a significant impact on noise levels in an image. Two common sources are the contribution of increased temperature on the noise level which influence many hardware sources of noise, and radiation which is more likely to occur in installations within a nuclear facility or in orbital applications where cosmic rays can impact the image acquisition process, often manifesting as a “salt and pepper” type of noise. Designing proper shielding into the acquisition system to handle radiation sources serves to minimize the influence from this source. Thermal effects can be reduced by active cooling of the system. Another environmental influence is the amount of light available in a scene during capture time. In some situations the light level can be increased but in situations where this is not possible the shot noise can be severe. Hoshino and Nishimura [Hoshino2007] discuss methods for reduction of noise from the sensor.

Some noise types are persistent across image capture sequences, arising from the system itself and not the light entering the system or due to environmental effects. A “dark image” may be acquired which encodes this type of noise and may be stored on the device and removed from each acquisition so that the result obtained is only influenced by factors uniquely experienced at the time of image capture.

2.3.2 Post-Acquisition Mitigation Approach

Noise removal methods estimate statistical information of input images in order to reduce the noise level and increase the SNR. A side effect of an aggressive application of these methods results in over-smoothing of the image. This is due to the high-frequency content in an image containing both fine details and noise. Some methods [He2010] approach the problem with the goal of preserving edges in the image but often introduce noticeable smoothing. Luisier et al [Luisier2011] propose a method for noise reduction in images corrupted by a mixture of Poisson and Gaussian noise.

2.3.3 Hybrid Approach

An alternative approach which we apply is to modify the image acquisition process in order to mitigate the effect of random noise even in unfavorably high noise situations. This process works by acquiring multiple, identical images of the scene and performing statistical analysis to determine the appropriate value for each pixel. The number of images needed to reach an acceptable SNR will vary based on the device characteristics, available light, and noise level encountered at acquisition time. Utilization of a noise level estimation [Liu2013] is helpful for determining an appropriate number of images to acquire.

2.4 Super Resolution

Images acquired from a digital image acquisition device are comprised of an array of numbers which correspond to the light incident at individual pixel locations on the sensor. In order to obtain a more accurate representation of the scene of interest, higher pixel resolution is desired. Though hardware solutions to this desire exist there are side effects which limit the ever-denser packing of ever-smaller pixels into an imaging sensor. The methods which seek to resolve the resolution enhancement task can be categorized into three different types, as shown in Figure 2.2.

Interpolation methods aim to create a HR image by upsampling a LR image [Siu2012]. Often, polynomial-based approaches are employed, such as the bilinear or bicubic approaches. The benefit of this class of resolution enhancement methods is the computational simplicity, though there are more advanced methods which seek to preserve edges in images rather than treating all image content uniformly. Siu and Hung [Siu2012] detail the classification of this set of approaches rather well. Yang et al [Yang2014] describe upscaling (they use single-image super-resolution, or SISR) as generating high-quality HR images from LR input by exploiting certain image priors. They classify the methods into four groups: prediction models, which operate without training data; edge-based methods, which learn priors from edge features in calculating HR images; statistical methods, which employ analysis of the statistical distribution of HR/LR image pairs; and patch-based methods, which employ a set of LR/HR paired image patches to upscale a new LR image based on past observations.

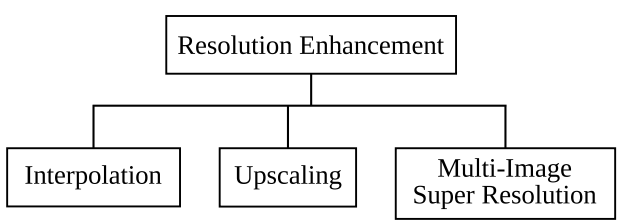


Figure 2.2: *Classification of resolution enhancement methods. Interpolation methods use simple mathematical models to calculate the value of new data points from existing known sample values. Upscaling methods, also known as single-image SR (SISR), often use a database of LR/HR image patches which are used to upscale a single LR image into a HR representation. Multi-image SR, on the other hand, employs multiple LR observations of a continuous scene and combines these images into a HR representation from actual data of the scene of interest. Our work focuses on this class of methods, utilizing multiple observations to calculate a HR output. We detail the types of methods developed which fall into this area in a later chapter.*

Instead of relying on simple polynomial models to interpolate image points or utilizing content from a priori image observations, multi-image SR methods combine multiple independent LR observations of a scene to calculate a HR representation. In choosing to employ methods from this class we can combine actual information from the scene to arrive at an improved understanding of the features of the scene. In order to utilize these independent samples, the relationship between them must be determined, termed image alignment or registration.

In our comparison we utilize the work of Šroubek et al [Sroubek2008], Irani and Peleg [Irani1991], Zomet et al [Zomet2001], Pham et al [Pham2006], and Farsiu et al [Farsiu2004]. Iterated Back Projection (IBP) [Irani1991] and Robust SR (RSR) [Zomet2001] are classical SR methods while Blind SR (BSR) [Sroubek2008], Normalized Convolution (NC) [Pham2006], and Fast RSR (FRSR) [Farsiu2004] are more recent methods. IBP is based on a regularized error minimization based on back projection and includes a noise prior. RSR [Zomet2001] is similar to IBP [Irani1991] but uses the median of the noise prior as opposed to the average used in IBP.

NC [Pham2006] uses a windowing function applied to local linear structures in the image to create subspaces sampled by the multiple low-resolution images available. NC is also a regularization-based method, using L_1 minimization of the cost function. BSR [Sroubek2008] is regularization-based, but uses an alternating minimization scheme to iteratively estimate the stationary blur kernel and high-resolution image.

We use SR in two places in our work: to enhance the accuracy of the PSF measurement step and in determining the influence this application has on improving target localization. In each case we desire to use a method that will work with our data and give an accurate result. Siu and Hung [Siu2012] give an overview of both interpolation algorithms and SR algorithms. For our application, we include the use of multi-frame static SR algorithms as described by their classification but apply the taxonomy of MISR methods proposed by Nasrollahi and Moeslund [Nasrollahi2014].

2.5 Deconvolution

In our evaluation of popular deconvolution methods to use in validating our method, we considered three methods: Wiener deconvolution [Gonzalez2008], Lucy-Richardson deconvolution [Richardson1972], [Lucy1974], and Regularized deconvolution [Steve2008], [Kotera2013]. Image deconvolution is an ill-posed inverse problem [Bertero2005] and as such no perfect solution exists; however, numerous methods exist by which the quality of the image can be improved. We test the methods noted above to determine which gives the best results using data obtained with our imaging systems.

Wiener deconvolution is a frequency-domain method that seeks to minimize noise effects in frequencies with low SNR [Gonzalez2008]. The method is developed to find a solution that provides a minimum mean-square error (MMSE) estimate of the original image f. In the implementation we use, the image to be deblurred, the measured PSF of the system, and the estimated SNR of the image are supplied to the algorithm. An assumption of Wiener deconvolution that may not apply in all situations is that the image is modeled as a random process, and that 2nd-order statistics and noise variance are known or can be estimated [Steve2008].

Lucy-Richardson deconvolution makes use of the expectation-maximization algorithm in order to maximize the likelihood of the restored image [Biggs1997]. Shepp and Vardi [Shepp1982] have shown that if the algorithm converges, it will converge to the maximum likelihood solution.

Regularization-based methods of deconvolution do not make the same assumptions as those made in Wiener deconvolution or attempt calculation of an EM-based solution. A side-effect of regularization-based methods is an over-smoothed solution, which while diminishing the influence of undesired noise in the result also limits image sharpness and crisp edges [Steve2008]. Regularizing terms are added to the minimization function in order to penalize selected aspects of unacceptable solutions. Methods exist that seek to enhance edges [Yang2009], diminish noise by penalizing high frequencies, etc. In the implementation of regularized deconvolution that we apply, a high-pass filter is employed in order to minimize the effects of the noise present in the captured images used for testing.

Kotera, Šroubek and Milanfar [Kotera2013] apply their alternating minimization approach in the nonblind deconvolution method presented and adapt their approach to better correct natural images by using sparse priors ($L^p, p < 1$) that better fit the gradient distribution of natural images.

2.6 Tracking

A number of applications tackle a similar problem with different goals and approaches. Visual tracking (surveillance) [Hedayati2015][Bouchrika2014][Leal-Taixe2012], visual odometry [Kazik2012], structure from motion (SfM) [Vaca-Castano2012][Lim2012], and visual motion capture (animation) [Leal-Taixe2012] methods all utilize the imagery from the visual spectrum in order to capture and represent information from the real world in a useful way. As these methods tackle similar problems using similar classes of data capture devices there is much overlap in the types of problems encountered and therefore in the application of solutions to these problems in the other areas. The basic format of any of these methods is to select hardware (lens type, sensor resolution, number of capture devices and geometry, etc.) capture data from a target of interest and extract useful information about the target using descriptive features and utilizing other information about the capture scenario such as the camera calibration (including both intrinsic and extrinsic information such as position and orientation information about each of the capture devices). At this point, the information about the target at each timestep from each device can be integrated in order to determine path information and 3D position information. Leal-Taixé et al call the second two steps detection and data association [Leal-Taixe2012].

Elhayek et al [Elhayek2012] utilize multiple cameras with unsynchronized frames in order to enhance the temporal resolution of their markerless motion capture method. Bouchrika et al [Bouchrika2014] use multiple views and a markerless capture method focused on the target's walking pattern to construct a recognition system based on the unique gait of the target. Lim et al [Lim2012] use a single camera to recover 6D information about the ego-motion of their camera while Kazik et al [Kazik2012] use multiple cameras to recover the same information from a target of interest. Leal-Taixé et al [Leal-Taixe2012] combine both the detection and data association steps into a single optimization. Soto et al [Soto2009] develop a 10-camera network of independent nodes which reach a consensus about the state of the target from their individual estimates.

Tracking, as we use the term, is an accurate knowledge of the historical, current, and expected location of a point of interest in real space. Tracking used in this sense is important in a variety of fields, from eye tracking [Fuhl2015] for human-machine interfaces designed for handicapped individuals to autonomous navigation [Chen2006] for mobile robot platforms and the expanding field of self-driving cars to exercise activity monitors [Vepakomma2015] used to recognize the wearer's behavior and automatically infer the calories burned to cell network tracking of handsets to coordinate selection of a service tower [Yassin2015]. Our interest lies in camera-based tracking, specifically for multi-camera systems since we desire to know with some degree of accuracy the precise location of the target. The theme of multi-sensor tracking centers on locating a subject of interest traveling in 3D space. The subject will have some unique attributes which distinguish it from its surroundings. The method used will also either include or omit some form of path planning or other prediction of the anticipated future location of the target. Matching images of a target from multiple cameras in an NP-hard problem and multiple methods have been developed to find a solution [Leal-Taixe2012][Soto2009][Zou2009].

2.6.1 Feature Descriptors

A number of feature descriptors have been proposed for uniquely identifying a target from its surroundings, including the scale-invariant feature transform (SIFT) [Lowe1999], speeded-up robust features (SURF) [Knopp2010], corners [Harris1988][Tomasi2004], and histogram of oriented gradients (HOG) [Dalal2005] descriptors. Some have developed other approaches for extracting features about targets including geometric information such as the height of a person [Bouchrika2014] and modeling of the color distribution of a given target [Orwell1999].

While these descriptors are very useful when the target is close to the sensor and relatively high resolution, color, and texture information about the target can be gathered, in our application we are limited to relatively sparse information about the target in the terms just described. However, we are able to distinguish the desired target from the background as there is little expectation of a cluttered field of targets and we mostly contend with noise and other traditional imaging artifacts. Thus our initial approach is a spatio-temporal correlation of the brightest point or brightest blob in the image. At times the noise is high enough to bury the target beneath the noise floor but by applying path prediction we are able to recover from these instances. We also consider extension into spectral signature recognition of the target since there is a distinct spectrum for which we are able to more accurately determine the location of the target within the image from each sensor.

2.6.2 3D Recovery Methods

The first step in determining the location of a target using a multi-camera system is to locate the target in image space. This is done by determining some way of locating the target based on unique information recorded about it in each image, as described above. The next step is to combine the location information in image space of each sensor with some information about the relative position and orientation of the sensors with respect to one another. This is due to the issue of tracking with a single sensor and the resolution of this problem with the addition of a sensor with an alternate perspective of the target, as shown in Figure 2.3. The camera on the left is able to see the target but has no sense of depth. In general, even information about the size of the object can only marginally aid in determining its distance to the sensor with even an assumed perfect knowledge of the camera matrix. Adding a second camera enables us to dispense with estimates based on size as only the centroid of the target is necessary for extracting the necessary information.

Moving from one camera to two allows calculation of target depth since unique projections of the 3D scene into 2D space are available to make the calculation. Introducing additional cameras to the system will allow fallback when communication with a sensor is lost, help to resolve occlusion in a crowded scene, and may increase the accuracy of the system. Analysis of the system developed for a given application will be necessary to determine the benefit of each additional camera due to the inherent cost of equipment and processing time versus the incremental benefit in features due to redundancy and potential for improved accuracy.

Two methods in particular will be discussed in greater detail: generalized stereo as presented in Figure 2.3, and an adaptation of this approach developed to allow for an arbitrary number of sensors. Rather than allowing the target to

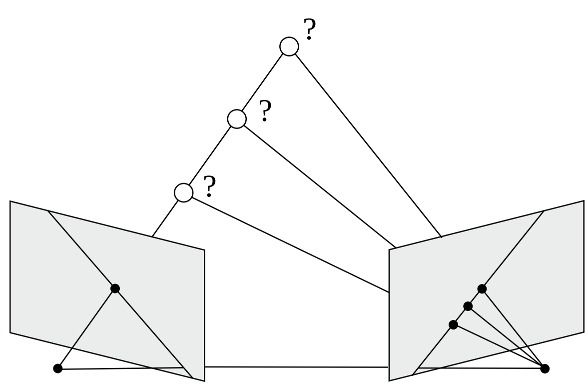


Figure 2.3: *Illustration of stereo matching ambiguity. The point in the left image has multiple possible physical locations in the scene along a line which information in the right image resolves by satisfying the epipolar constraint. Information from additional images can further reduce the error in determining scene geometry to a point, but also provides the ability to retain sufficient information about the scene in the case that an object occludes the view of a sensor or communication with a given sensor is momentarily lost. Reproduced from [Zou2009].*

appear anywhere within the FOV, the sensors are oriented to keep the target of interest always at the center of the image. Then, the position and orientation of each sensor is used to construct a line segment in 3D space. A minimization of the intersection point of these lines is used to recover the target position. In this way, any number of sensors may be considered in the target position recovery. Analysis using synthetic path information allows development of improved models for target tracking and position recovery as well as sensor orientation update approaches in order to minimize the error in target position recovery.

2.6.3 Path Prediction Methods

Many methods for navigation and tracking exist and positional accuracy for the system varies based on the constraints of the environmental conditions, speed of travel, method employed, sensor accuracy and resolution, measurement noise, processing speed, etc. Application of path prediction methods can help to improve target position estimation accuracy, reduce accumulation of errors from the many potential sources, and improve response time. Potential also exists for reduction of equipment costs and processing time when the behavioral model of the target movement is well-defined and simply from the fact that multiple-camera tracking is more efficient [Pham2007].

Many approaches and relevant applications have been proposed in this area. Van Verth and Bishop utilize path prediction to address the concern of lag in player positon updates in networked games [VanVerth2008]. Hedayati et al [Hedayati2015] utilize a system that factors in the egomotion of the cameras. Elhayek et al [Elhayek2012] utilize unsynchronized cameras in order to improve the temporal resolution of target position estimates. Pham et al [Pham2007] extend the Gaussian mixture probability hypothesis density (GMPHD) filter method to a multi-camera setup in order to handle multi-object tracking with occlusion without the computational overhead of a multi-object state space.

$$x_k = x_{k+1} + w_k \tag{2.10}$$

Pham et al [Pham2007] present the model in (2.11) for a dynamic moving equation for 3D tracking where the state of an object $x_k = \{x_1, k, x_2, k, x_3, k\}$ is a 3D coordinate and w_k is the process noise. The measurement equation of the ith camera is

$$\begin{gathered}\begin{pmatrix} l_{1,k} \\ l_{2,k} \\ l_{3,k} \end{pmatrix} = \begin{pmatrix} a_{11}^i & a_{12}^i & a_{13}^i & a_{14}^i \\ a_{21}^i & a_{22}^i & a_{23}^i & a_{24}^i \\ a_{31}^i & a_{32}^i & a_{33}^i & a_{34}^i \end{pmatrix} \begin{pmatrix} x_{1,k} \\ x_{2,k} \\ x_{3,k} \\ 1 \end{pmatrix} \\ \begin{pmatrix} y_{1,k}^i \\ y_{2,k}^i \end{pmatrix} = \begin{pmatrix} {l_{1,k}}/{l_{3,k}} \\ {l_{2,k}}/{l_{3,k}} \end{pmatrix} + u_k, \end{gathered} \tag{2.11}$$

where u_k is the measurement noise and a_{mn}^i are the projection parameters from 3D coordinate space to the ith camera plane. Calibration of the cameras yields this matrix. For a small room (3m x 2m x 2m) simulation, they achieve a 5-7cm error using fixed cameras.

Approaches which employ a predictive model makes use of one of a set of possible assumptions about the motion of the target in order to extrapolate the target position. Formulations have been employed for constant or variable acceleration (kinematics) and some include the influence of momentum (kinetics) [VanVerth2008].

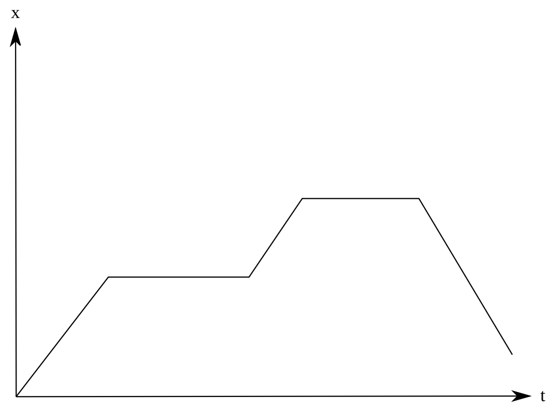


Figure 2.4: *Graph of a piecewise continuous path in one dimension of target movement as a function of time. Reproduced from [VanVerth2008].*

The simplest of the motion models considered is shown in Figure 2.4 and employs a piecewise constant velocity model, with zero acceleration. The path is generated by use of instantaneous measurements of position and velocity and then holding the velocity as constant until the next measurement. The next step towards a generalized solution involves a constant, but nonzero acceleration. Shown in Figure 2.5 is a depiction of a target whose motion is influenced by the constant acceleration of gravity.

Other sources of constant acceleration exist as well, such as from a car engine, airplane engine, atmospheric drag, etc. and a complete model from the expected motion of a given target should be incorporated to most accurately predict the target path. In our application scenario, we can expect relatively predictable parabolic motion and as such will include explanations up to this point and handle any deviations from our prediction with the noise term as we update at each time step. Should longer prediction times be desired or more complex forces influencing the target motion necessitate further generalization of the motion model (such as variable acceleration from sources such as control systems) our work can be further extended. In particular our sensor coverage area is wide enough that the target will not be capable of exhibiting unexpected changes beyond the bounds for which we are able to maintain reliable tracking.

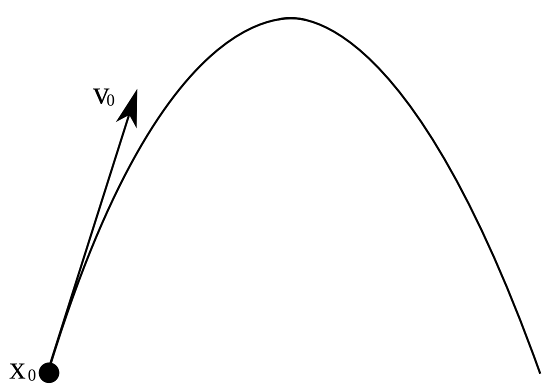


Figure 2.5: *Graph of a parabolic path of target as a function of time. Shown with initial velocity vector, source of constant acceleration is gravity. Reproduced from [VanVerth2008].*

2.7 Error Analysis

The value most important to our system is the final position estimate. Therefore, the metric of greatest importance is the error of this estimate with respect to the true target position. In our sensitivity analysis, we will investigate the adjustable input parameters' influence on the final accuracy of the target position estimate. The purpose of this influence will be to explore the flexibility of the system presented in order to better understand the limits and the factors which could make the system fail in its purpose, serving to drive the outer edge of our capabilities and influence direction for future efforts to improve the system to focus attention and advancement where it will have the greatest impact.

Sensitivity analysis is a crucial step to perform for any complex system [Zhang2010] and has seen effective use in medical [Wong2010][Zhang2007][Zhang2010], mathematical [Hadigheh2006], imaging [Desurmont2006] [Muhammed2010][Jalobeanu2000][Nanda1994][Vega2003][Yang2000][Molina2006][Yao2006] and other related fields [Vytyaz2008]. Unfortunately, as the system complexity increases, so often also does the interaction of the influence of the parameters [Pappenberger2006][Racu2006]. Further, the computational time will increase exponentially with each additional parameter and the range and granularity of sampling for each parameter [Kwisthout2012].

A closely related problem is parameter estimation, as it is known in the image enhancement field. Most image processing algorithms require selection of a set of parameters a priori, and the quality of the output varies widely with this selection. Our system is also bound by not only this requirement, but also the simulation parameters (resolution, baseline, etc.) which we can adjust to investigate their influence on the accuracy of recovery of the target position.

The factors we are most concerned with are accuracy of the method [Yao2006][Muhammed2010], computational complexity as it relates to processing time [Racu2005][Racu2006][Vytyaz2008][Kwisthout2012][Yu2015], and stability of the method [Yao2006][Hadigheh2006] to produce consistent results with varying parameter solutions [Yao2006]. Where possible, we will follow the Type I (Basis Invariancy) approach described by Hadigheh and Terlaky [Hadigheh2006] to find the range of parameters for which the accuracy of our reconstructed target position remains largely unperturbed. In some of our analysis, however we expect to find need for intelligent analysis of tradeoff while still defining theoretic limits and will seek to quantify or formulate these as appropriate [Vytyaz2008].

Desurmont et al [Desurmont2006] introduced a new evaluation metric as part of their work and produced a set of recommended parameters for particle filtering algorithms tailored to people tracking in video. Vytyaz et al [Vytyaz2008] derived a formula relating the oscillation frequency of a PLL to the sensitivity of the device to a given parameter set. Both results could be found in our recommendations due to the construction of the system and the nature of the components in our pipeline. For instance, we expect a more solid set or range of appropriate parameters for the processing algorithms, but anticipate needed tradeoff analysis from the resolution findings. Computer vision researchers always want more resolution, and if an increase in bandwidth, storage, and processing time did not also attend our analysis may deliver this result.

Our sensitivity analysis will include the simulation parameters to discover the influence of the sensor construction and the limits of available resolution and introduced blur, as well as to explore the effect of adjusting the baseline of the sensors and their height relative to the target. We will also analyze the effect of applying selected algorithms as part of our pipeline, more so for the effect that varying processing time will have on reconstruction accuracy due to system lag from overly-complex methods. In this way we can select an envelope of allowable processing time for a given acceptable frame lag and select suitable methods.

3 Point Spread Function Measurement

The methods for obtaining the point-spread function (PSF) of a system fall under two broad areas: PSF estimation and PSF measurement. PSF estimation methods utilize a calibration target in order to solve a known deconvolution process to recover the PSF, often by way of the Modulation Transfer Function (MTF). Other methods utilize assumptions such as sharp edge response and are known as blind methods. Generally PSF measurement techniques will attempt to directly measure the PSF by imaging a point source constructed for this purpose. It is into this category our method falls.

Many estimation methods do not account for the spatially variant and isoplanatic nature of the PSF as they perform measurements over too large of an area of the system's field of view (FOV). The PSF is constant over a small region [Hornberg2006], and a larger area must be used to properly capture the calibration target, thus violating this principle. We seek to solve the problem of determining the PSF of a digital optical system by using a direct measurement method. We solve this problem in a nondestructive way and such that the cost to perform initial and subsequent measurements is negligible relative to the cost of the sensor. Doing so removes the barriers to regular re-measurement of the system so that the most accurate, current state of the system can always be known.

Obtaining the best data from an optical system requires accurate knowledge of the PSF, whether that system is a high-precision optical metrological system or a casual consumer photograph. Research in this area has yielded a number of methods that vary in complexity regarding equipment requirements and data processing. We describe a method that allows for low cost in both required hardware and software processing time. Furthermore, the process is quick enough to allow for regular repeated measurements so that systems subject to environmental conditions such as vibration and temperature fluctuation may be accurately represented in their current state.

Digital optical systems have been adopted in modern society with near ubiquity. The quality of these systems has significantly improved over recent decades, but there still exists a need to correct the inherent aberrations introduced by both the lens optics and the sampling of the digital sensor. Figure 3.1 depicts the image acquisition process of a typical digital optical system. Many methods have been proposed for measurement of the PSF or related quality indicators [Liu2008], [Levy1999], [Park1984], [Delbracio2012], [Brauers2010], [Hu2012], [Tang2013], [Schuler2012], [Aguet2008], [Howell1996], [Zheng2006], [Claxton2007], [NavasMoya2013]. Each of these methods contain some feature which we seek to improve, such as destruction of the system [Liu2008], lack of measurement of the spatially-variant nature of the PSF [Liu2008], [Zheng2006] or optical transfer function [Levy1999], [Park1984], use of blind estimation [Levy1999], [Delbracio2012], [Brauers2010], [Hu2012], [Schuler2012], [Zheng2006], or restriction of the form of the PSF to mathematical models [Tang2013], [Aguet2008], [Howell1996], [Zheng2006], [Claxton2007].

Our method meets the requirements detailed by Delbracio et al [Delbracio2012] for a PSF measurement method while resolving the step of removing the lens for measurement in Liu and Chen's method [Liu2008] which introduces the ability to make repeated measurements. Further, we enhance Liu and Chen's approach [Liu2008] by utilizing the framework of Navas-Moya et al [NavasMoya2013] and applying Šroubek et al.'s method [Sroubek2008] to enable more accurate measurements by exploiting the benefits of super resolution (SR) in sampling the PSF directly. Any physically realizable system will have some error in its measurement capabilities, but by leveraging the computational power of modern systems we can recover the sampled data and better represent the real-world scene.

Applications of our method include but are not limited to: consumer mobile phone cameras, optical metrology, computer vision/robot vision, aerial reconnaissance, automated air- and land-based navigation, amateur astronomy, remote sensing, and consumer and professional level photography.

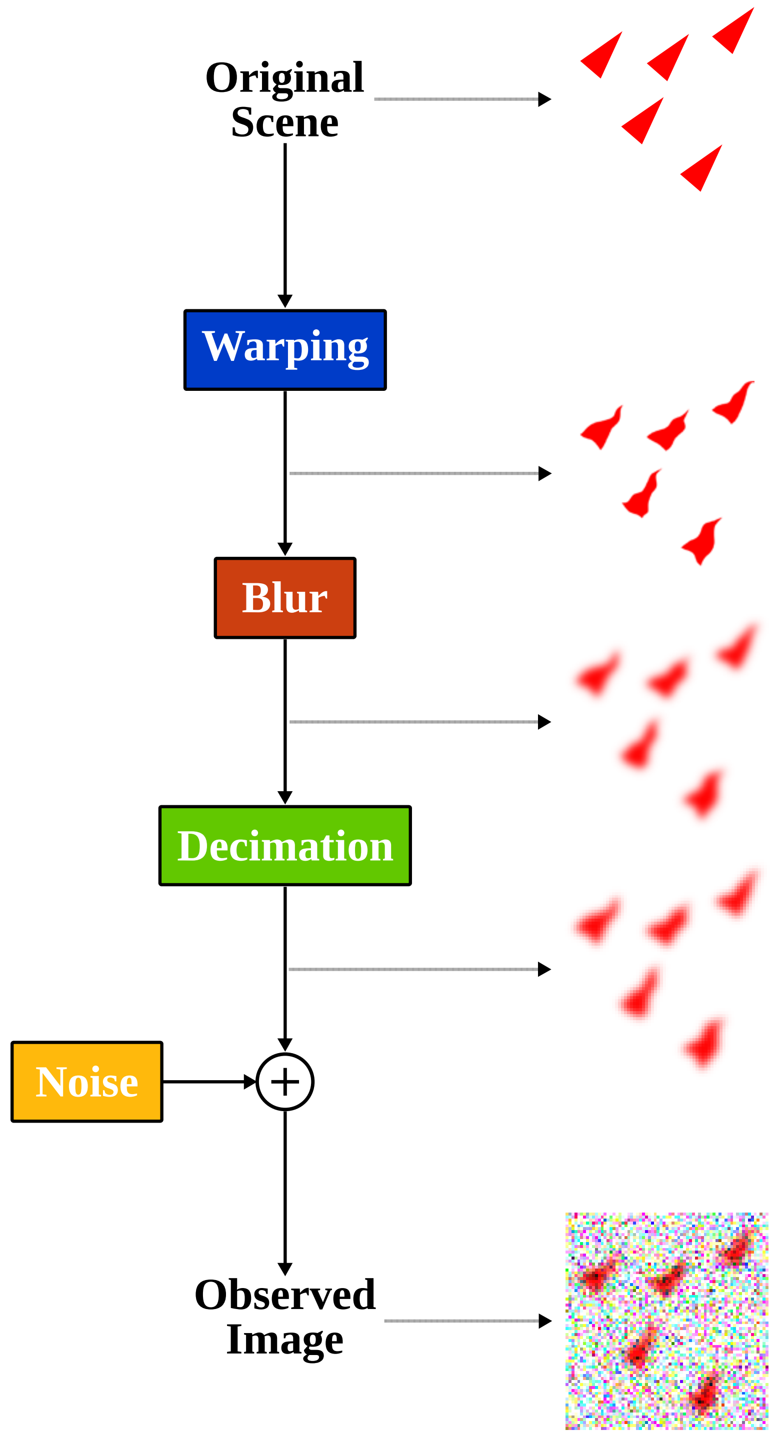


Figure 3.1: *Image acquisition models include the simple model used by Kotera et al [Kotera2013]:* $\boldsymbol{g} = \boldsymbol{o} * \boldsymbol{h} + \boldsymbol{n}$*,* $\boldsymbol{g}$ *is the observed image,* $\boldsymbol{o}$ *is the original scene,* $\boldsymbol{h}$ *is the system PSF, and* $\boldsymbol{n}$ *represents additive noise. The model used in [Sroubek2008] is more generalized:* $\boldsymbol{g}_{\boldsymbol{k}} = \boldsymbol{D}[\boldsymbol{h}_{\boldsymbol{k}} * \boldsymbol{W}_{\boldsymbol{k}}(\boldsymbol{o})] + \boldsymbol{n}_{\boldsymbol{k}}$*, including the geometric distortion of the lens,* $\boldsymbol{W}$*, and the decimation of the sensor,* $\boldsymbol{D}$*. The deconvolution operation is an ill-posed inverse operation and to successfully remove the blur effects of the lens requires accurate knowledge of the PSF.*

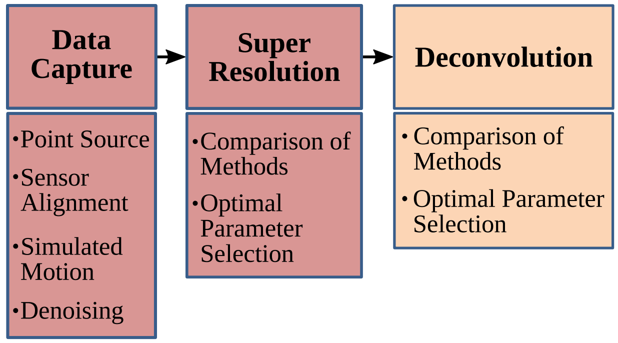


Figure 3.2: *Process pipeline for PSF measurement method. Data capture includes generation of a point source grid suitable for measurement of the PSF. The sensor must be well-aligned with the grid, which is 'moved' by shifting the pattern to adjacent pixels. The images are analyzed to mitigate the effects of noise and other degradation in our data. We apply SR to our captured data in order to enhance our measurements before downsampling to the original pixel size for use as part of our deconvolution to validate the measurements made.*

As shown in Figure 3.2, we align the sensor with a point-like grid prior to capturing data. Multiframe SR requires subpixel motion between input frames. In light of the low SNR of an infinitesimal point-source grid, we synthetically enhance the SNR by combining a sequence of frames with a static target [NavasMoya2013]. We then apply SR on these combined frame sets in order to obtain an enhanced representation of the PSF while remaining within an appropriate isoplanatic region of the lens [Hornberg2006].

3.1 Experimental Setup

In this section we detail the components of the experimental setup, demonstrating the design choices, current practices, and validation of our assumptions to ensure the data captured will yield a valid measurement of the PSF of the systems tested.

3.1.1 Generation of a Suitable Point Source

Levy et al [Levy1999] use a CRT monitor in estimating the MTF of a lens in the visible spectrum. While use of a monitor is presented as acceptable in the literature [Levy1999], [Delbracio2012], [Brauers2010], [NavasMoya2013], we want to validate the use of an LCD monitor for PSF measurement. There are two necessary criteria for such use: the light used must be incoherent, and the size of a single pixel used must image to less than a single pixel in the sensor's image plane.

Navas-Moya et al [NavasMoya2013] note that use of an LCD display has two major advantages over a printed target: there is no need for an additional light source, and the pattern may be easily changed—a feature that we exploit in our development. They further note the lack of critical need for alignment with random noise patterns and the problem of a lack of information regarding the spatially variant nature of PSF measurement obtained using random noise patterns.

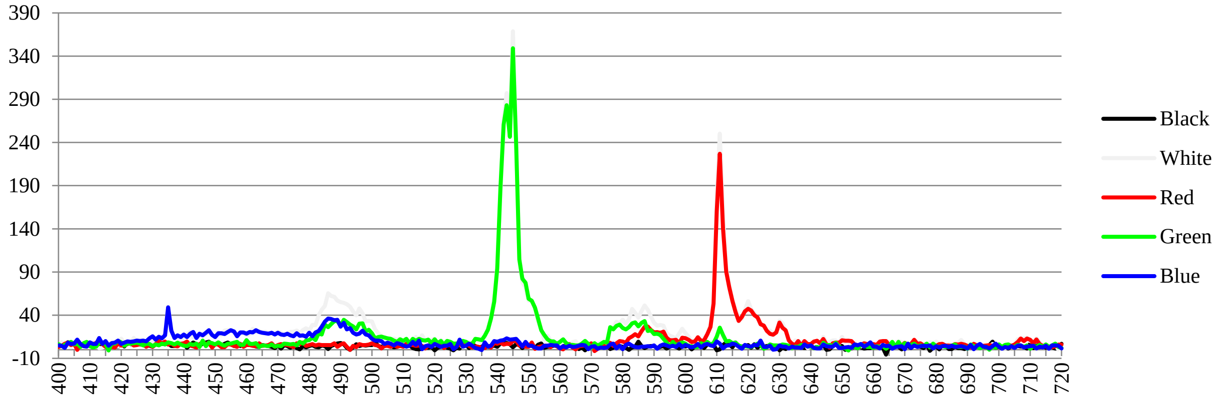


Figure 3.3: *Spectral Response of the Dell U2311Hb Monitor. The light emitted from the monitor is broad-spectrum, incoherent and the comparison of the red, green, and blue lines shows the bandpass effect of the individual color filters in each pixel. Coherent light is useful for measuring the PSF when working with holography but for general-purpose photography applications incoherent light is needed. Use of broad-spectrum light ensures that we obtain measurements that will be valid for imaging within the visual spectrum, for our chosen application. Using an IR light source would afford the capability of measuring the response of such a system within that spectral range.*

Figure 3.3 shows the result of our measurement of the light emitted by the LCD monitor used in our experiments. The light emitted has typical fluorescent spikes. Measurements were performed with all pixels off (but with the fluorescent backlights still on), labeled Black; with all pixels on, labeled White; and with each of the Red, Green, and Blue pixels only on, labeled respectively.

Due to the nature of the fluorescent light source [LCDSpectra] used in the monitor, we know that the light used for measurement is broad-spectrum within our desired visual range of the spectrum. As we need the size of an object pixel on the display to image to less than a single pixel on the sensor, we need to ensure that our fixed-focus prime lens is situated sufficiently far away from the display in order to meet this criterion.

3.1.2 Imaging Distance

It is impossible to construct a truly infinitesimal point source. Selecting the imaging distance is important to create a point-like source. There will be blurring due to aberrations of the lens and diffraction effects, causing the resultant measurement to be larger than one pixel. In order to determine an appropriate imaging distance from which to make our measurements, we consider three metrics, outlined as follows:

3.1.2.1 Mathematical Formulation

The angular resolution of a diffraction-limited system is [Born1999]:

$$\theta \ll \frac{\lambda}{D} \tag{3.1}$$

where θ is the angular size of the target,

λ is the wavelength of light, and

D is the diameter of the lens aperture.

Substituting for θ in order to obtain a form in terms of the distance from the target to the sensor, (3.1) becomes:

$$\theta = 2arctan\frac{s}{2d}, \tag{3.2}$$

where s is the physical size of the source and

d is the distance from the source to the sensor.

Substituting (3.2) into (3.1), we obtain:

$$2arctan\frac{s}{2d} \ll \frac{\lambda}{D}. \tag{3.3}$$

Solving for d, we obtain:

$$d \gg \frac{s}{2tan\frac{\lambda}{2D}} \tag{3.4}$$

For values of s = 265µm,

λ = 0.380µm, and

D = 755µm,

we obtain a requirement of d > 52.7cm for the Galaxy S4.

3.1.2.2 Shih Formulation

Shih et al. [Shih2012] propose a formulation related to the focal length of the lens:

$$d > \frac{fm}{s} \tag{3.5}$$

where f is the focal length of the lens,

m is the pixel pitch of the monitor, and

s is the pixel pitch of the sensor.

Using the same values above:

f = 1.85mm,

m =265µm, and

s =3.768µm,

we obtain a minimum imaging distance of d > 13cm for the Galaxy S4.

3.1.2.3 Field-of-View Restrictions and Batch Size

The maximum imaging distance is restricted by the angular FOV of the sensor relative to a given target size. However, the noise level does rise with increasing distance from image to sensor. In order to maintain an acceptable SNR in the presence of significant CMOS shot noise due to the necessarily low intensity of our imaging grid, we stack batches of images taken with no motion between frames. We perform this step in order to overcome the limitation of a non-user-configurable integration time of the sensor. The work of [Liu2012] was used to determine an appropriate minimum number of images to use. We use our alignment target (seen in Figure 3.7) to calculate the angular FOV of the sensor. From this image it is determined that the horizontal angular FOV of the Galaxy S4 is 72.7°, which is the direction in which we first run into the extent. This gives a maximum distance of 34.5cm for the Galaxy S4. These values were calculated using

$$FOV = 2arctan\left(\frac{D}{2d}\right) \tag{3.6}$$

where FOV is the field-of-view of the lens,

D is the amount of the monitor visible within the frame, and

d is the distance between the sensor and the monitor.

For the setup where the measurements were taken, D=41cm and d=28cm. Using the results obtained (FOV=72.7^o), we solve for the maximum distance.

$$d < \frac{D}{2tan\left(\frac{FOV}{2}\right)} \tag{3.7}$$

Where the variable labels are as previously described in (3.6). D=50.8cm is the monitor's horizontal extent, FOV is as calculated previously, 72.7^o, which gives the result of d=34.5cm as our maximum imaging distance.

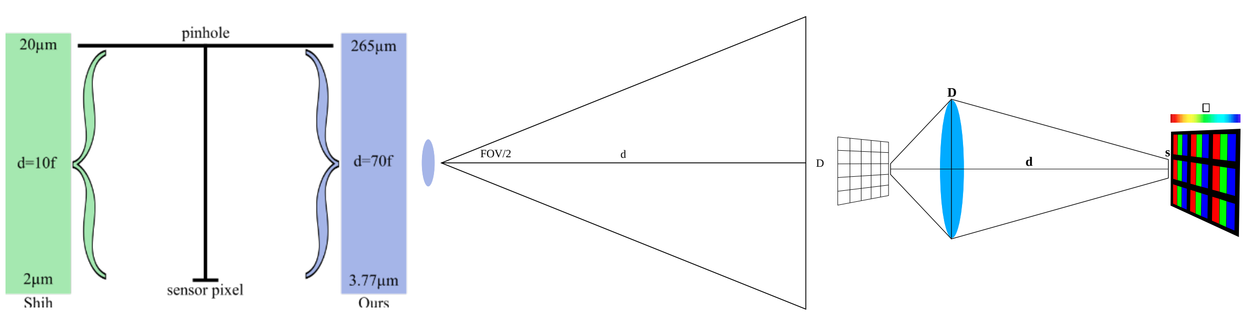
3.1.2.4 Pixel mapping method

Navas-Moya et al [NavasMoya2013] do not make any distance verification measurements, but they assume that the pixels in the object plane image to less than one pixel in their image space due to the fact that there are more pixels in the object plane than in the image plane. Because the pixel count in our monitor and sensor are the same, this would only be potentially possible at the maximum imaging distance. However, at this distance we began to encounter low SNR. Given that our other metrics agree with one another, the requirement noted by Navas-Moya et al [NavasMoya2013] is valid, but not required.

3.1.2.5 Minimum Imaging Distance Validation

We verify the minimum distance empirically. Imaging too closely results in a PSF diameter of 50 or more pixels at the optical center where the diameter of the PSF should be smallest. Noticing that the size of the PSF diminished as we moved further away, we plotted the change as seen in Figure 3.5.

When the size of the PSF stops diminishing, we have reached the minimum imaging distance. Measurements obtained further than this distance result in diminishing SNR and increase the minimum batch size and acquisition time per batch. As seen in Figure 3.5 below, this value appears to occur between 24cm and 33cm. We selected a distance of 29cm in order to allow a margin of error in our measurement distance. This range is less than the FOV limit and agrees with all but one of the distance metrics.



(a) (b) (c)

Figure 3.4: *Imaging distance prediction approaches. We want to increase the SNR of the measurements without overly lengthening the data acquisition time. We therefore select a distance close enough to the monitor to gather more light and measure across the FOV while being sufficiently far away to meet the optical constraints for measurement of a point source. (a) Illustration of the approach of Shih et al. (b) Illustration of the FOV restriction. (c) Illustration of the mathematical formulation.*

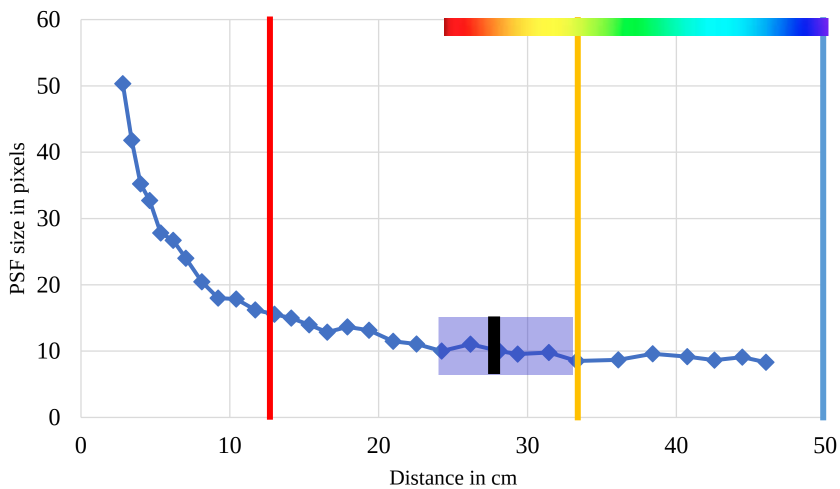


Figure 3.5: *Imaging distance validation. The diameter of the central PSF is overly large if too close to the monitor to measure the PSF accurately (as a grid of pointlike sources of light). As we move further from the monitor, we see the diameter reach a stable size as viewed from our optical system. Although it is possible to measure the pointlike sources on our grid from further away, doing so would decrease the SNR of each frame, necessitating additional frames per measurement and needlessly lengthening the data acquisition time. We therefore select a distance close to the monitor which is still sufficiently far away. The red vertical line represents the result using Shih's formulation. The orange vertical line represents the maximum imaging distance due to FOV restrictions. The blue vertical line represents the minimum imaging distance according to the mathematical formulation from the system optics. The blue shaded region represents the earliest region where the validation measurements settle and the black vertical bar within this region represents our selected imaging distance.*

A crucial part of our PSF measurement method is determining that we are capturing our measurements from a valid distance. Several criteria have been considered and an experimental validation of the minimum imaging distance was conducted, with low correlation between the results of each. Definitive analysis of the relevant factors contributing to the distance we wish to calculate would increase the method's validity, increase understanding of the principles of optics that apply, and save a significant amount of time when performing a measurement of each new device to be tested.

3.1.3 Induced Motion

Liu and Chen [Liu2008] propose a subpixel algorithm to average the effects of lens alignment with respect to a given pixel. The inherent subsampling of the high-resolution PSF of the lens by the paired sensor changes the representation of the PSF at the image plane depending on this alignment. In our method this alignment is not known, and over the life of the system a number of factors (temperature and pressure changes, system vibration) will likely cause slight changes in alignment. We match this part of their method by introducing a spatial averaging step accomplished using SR [Sroubek2008], [Milanfar2011].

	7	8	9	
	6	1	2	
	5	4	3	

Figure 3.6: *Display of the order in which samples were taken for subpixel-induced motion, beginning at the central pixel and progressing sequentially as indicated to form a 3x3 grid of measurements in the individual areas of the image. Liu and Chen [Liu2008] use an averaging step in order to account for alignment of the optical axis with a pixel on the sensor. We employ SR in order to account for these same effects without separating the lens from the sensor.*

For SR reconstruction-based algorithms, there must be subpixel motion between frames [Milanfar2011], and we want to avoid the introduction of in-frame motion blur while introducing between-frame motion [BenEzra2005]. Additionally, we do not want to violate the isoplanatic region assumption in measuring the PSF of the system [Hornberg2006]. There is a relationship between the SR factor and the minimum number of images required [Lin2004]. We induce subpixel motion at the image plane by creating synthetic motion at the object plane [NavasMoya2013]. To accomplish this, we shifted the point field ±1 pixel vertically and horizontally, creating a 3x3 sampling area, as shown in Figure 3.6. In this small area we take 9 measurements, which allows us the option to perform integer SR-factor reconstructions of 2X and 3X [Lin2004]. Larger sampling areas are possible, but it is necessary to investigate the tradeoff of additional accuracy versus acquisition and computational time while maintaining the measurements within an isoplanatic region.

3.1.4 Capture of Images

The image-capture process is automated using a MATLAB script on the processing device and, on the device under test, a program called IP Webcam [IPCamera2015], which gives network access and control of the camera module of the host device. Communicating with the device in this manner enables simplified alignment of the device under test and enables the operator to be sufficiently far away from the test apparatus so as not to induce vibration during data acquisition.

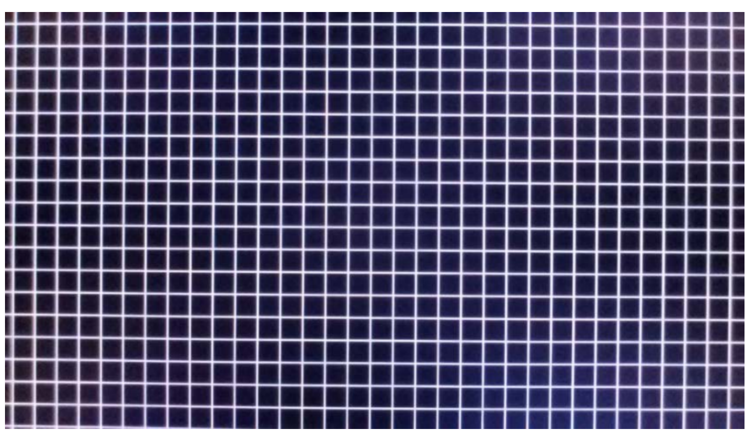


Figure 3.7: *Alignment target using the Galaxy S4 front camera, used to eliminate extraneous rotation about all three axes. When the camera is rotated about any axis with respect to the monitor, it is apparent in the image by a distortion of the scene. We manually adjust the sensor in a mount designed for this flexibility until the distortion is removed as seen above. The translation offsets are less problematic except in maintaining a minimum distance away from the monitor.*

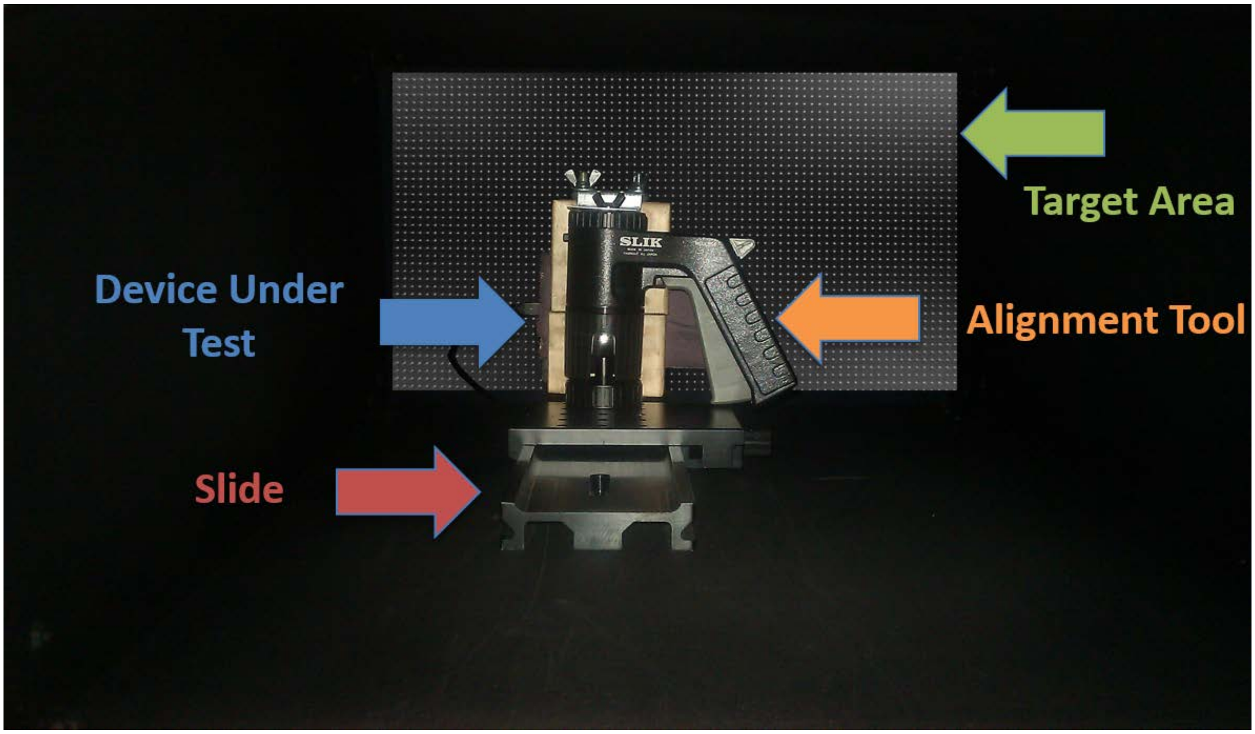


Figure 3.8: *Data acquisition setup. The alignment target or point grid is located at the back of the apparatus used to eliminate ambient light from disrupting the measurements of the target area. The device under test is mounted in the alignment tool and the video feed from the device is used to align the device with the target area. The measurement position is adjusted via the slide to which the alignment tool is attached.*

The sensor and target are first aligned using the video preview, which gives a handful of frames per second, ample for alignment using the target shown in Figure 3.7. The lenses have radially symmetric distortion fields, therefore when the viewed distortion of the feed from the sensor is symmetric across the FOV, we have properly aligned the target and sensor. The mount on which the sensor rests has four degrees of freedom to facilitate alignment, while the monitor used to display the target patterns contains a further two degrees of freedom to enable 6DOF alignment for our experiments.

After the manual alignment step is completed, the nine target images are displayed in succession, with the appropriate number of images taken per target, as discussed below. After the image acquisition portion is completed, we combine the corresponding images into a set with images stored for further processing. In order to determine an appropriate cutoff that would not overly extend the data acquisition time, we normalize and fit the data to a 6-dimensional polynomial to smooth the estimate of the incremental percentage improvement of each added image.

The number of images to combine per target was determined by applying a single-image, noise-level measurement method [Liu2012]. A set of 100 images of the pointlike grid pattern was taken without any induced motion between frames. These images were assembled into batches of images of increasing number from the smallest batch size of 1 image to a batch containing every image in the set of 100. It was assumed that an optimal batch size within this range existed and measurements for each batch size were taken. The results can be seen in Figure 3.9.

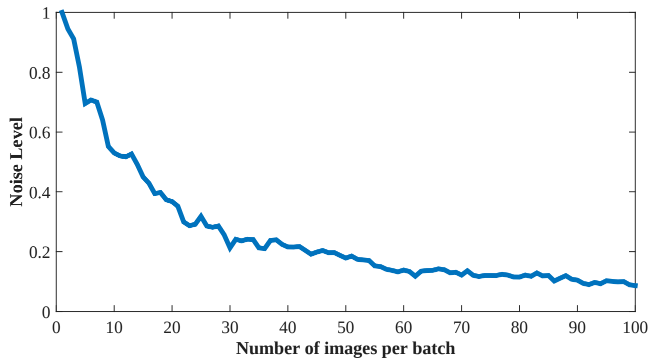


Figure 3.9: *Noise level attenuation to determine minimum batch size. The addition of multiple frames per measurement allows for a reduction in the noise which gives a more accurate result. We wish to find the minimum number which gives a satisfactory results so that our final result will be more accurate while not overly increasing the data acquisition time.*

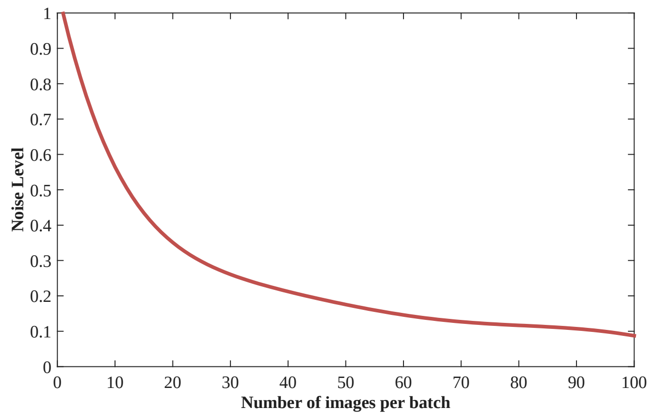


Figure 3.10: *Polynomial fit of the normalized noise measurement data. Since the original data in Figure 3.9 was not smooth, we need to fit the data to get an accurate estimate of the minimum number of input frames as this process requires the first derivative. A six-degree polynomial was used to fit the data as lower-order models were found to insufficiently match the data.*

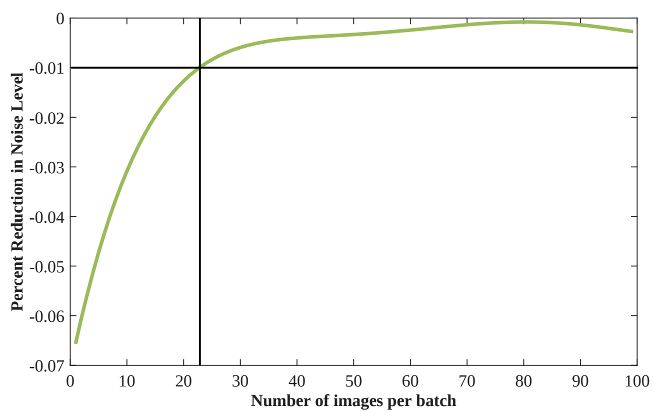


Figure 3.11: *Plot of the percentage reduction in noise level with annotation at the 1% level. We want to get as much data as we are able without overly extending the data acquisition time. An additional motivation for using 1% as the cutoff value is the error in our measurements which shows a diminishing contribution to the accuracy of the measurement followed by increased accuracy. By avoiding the effect introduced by too many measurements we avoid this problem.*

After plotting the data seen in Figure 3.9, the MATLAB built-in Basic Fitting Tool was used to create a smooth, approximately monotonic function to fit our data. It was found that the best fit that met our criteria came from a 6-dimensional polynomial of the form

$$y = a_1x^6 + a_2x^5 + a_3x^4 + a_4x^3 + a_5x^2 + a_6x + a_7 \quad (3.8)$$

The parameters were calculated to be $a_1 = 1.0517 \times 10^{-11}$,

$a_2 = -4.3842 \times 10^{-9}$,

$a_3 = 7.2604 \times 10^{-7}$,

$a_4 = -6.1413 \times 10^{-5}$,

$a_5 = 2.8499 \times 10^{-3}$,

$a_6 = -7.3475 \times 10^{-2}$,

$a_7 = 1.069$

and the norm of the residuals was $a_1 = 0.16315$.

The difference in the average image from the lowest-sized batches quickly diminishes and then levels out. An analysis of the percentage benefit of an additional image being added to the batch was set at a threshold of 1%, shown in Figure 3.11. The values for Figure 3.11 are calculated from the data points in Figure 3.10 using the following formula:

$$b_k = a_{k+1} - a_k \quad (3.9)$$

where b_k is the value at a given index in Figure 3.11,

a_{k+1} is the value at the subsequent index in Figure 3.10, and

a_k is the value at the same index as b_k.

This results in a value of 23 images per batch. A lower value than 1% may be selected should an alternate application necessitate a lower noise level. This curve will vary based on the sensor used, and as such the appropriate value of images per batch for the sensor should be determined before acquiring data.

3.2 Algorithms and Experimental Results

A more complete description of our work with and analysis of SR and deconvolution algorithms is presented in Chapter 0. The specific case detailed here comprises our findings as they relate to the type of data used in measurement of the PSF of a lens. Application to different types of data may result in findings unique to the application.

3.2.1 Choice of SR Method and Results

We compare one multi-frame interpolation method and five SR methods to find a suitable method to apply in our pipeline. We evaluate these methods in two groups, as indicated in our earlier discussion of the methods. The input to these methods are the averaged batch images obtained of the point-grid area. The resulting nine sample images with subpixel motion allows for integer SR factors of 2 and 3 [Gunturk2013]. For the experiments shown in this section, we set an SR factor of 2.

3.2.1.1 Classical SR Methods

The classical SR methods are grouped together based on the timeline of SR methods in [Nasrollahi2014]. As the RSR method [Zomet2001] is a derivative of IBP [Irani1991], it is included in this grouping despite being released much later.

3.2.1.1.1 Bicubic Interpolation [Zomet2001]

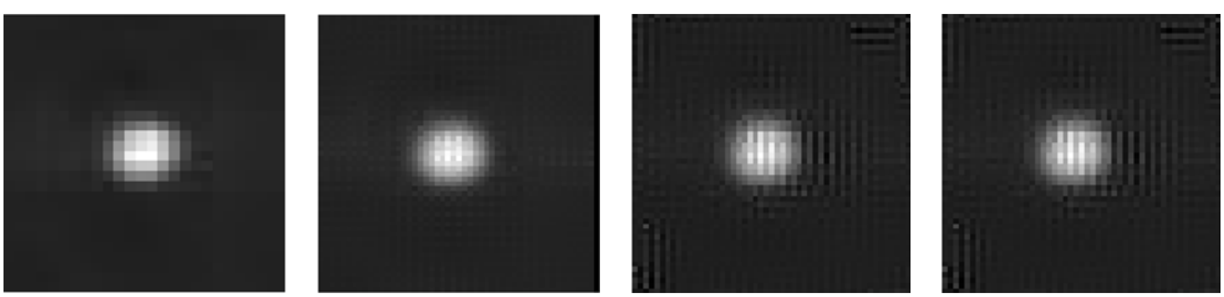
Bicubic interpolation utilizes all of the low-resolution input images as opposed to taking the more common single-frame approach. Our result is shown in Figure 3.12(b). The overshoot effect common to this method is apparent in the SR reconstruction. For natural images this may be acceptable, as it increases the acutance of the image. However, in our measurements of the PSF from the measured data, it is an undesirable artifact of the interpolation method.

3.2.1.1.2 Iterated Back Projection (IBP) [Irani1991]

This method requires a high-resolution initial estimate and projects how the high-resolution image would influence the low-resolution images based on the motion estimation of the input frames. Irani and Peleg [Irani1991] explain that the initial guess does not influence the speed or stability of the method's output, suggesting a simple average of the input images be used as the initial estimate. The results with regard to our measured data also exhibit some artifacts in the SR reconstruction. It is likely that the IBP method is sensitive to residual noise contained in the input frames.

3.2.1.1.3 Robust Super Resolution [Zomet2001]

Zomet et al.'s RSR [Zomet2001] has principally the same structure as Irani and Peleg's IBP (Peleg also being one of the authors). However, the average [Irani1991] was replaced with a median [Zomet2001] to estimate the reprojection error of the high-resolution image estimate, to avoid outliers and sidestep issues with noise sensitivity commonly seen in IBP. There is less sensitivity to noise using RSR as compared to IBP although the apparent ringing artifacts remain. Keren et al's method [Keren1988] was used to register the images before applying SR in these two methods. Keren et al's motion estimation method models the motion between frames as translation and a Euclidean rotation about the center of the image. An error function is constructed by expanding the Taylor series representation of the motion and minimizing relative to the x and y translations and θ rotation of the parameters estimated for each image in the sequence, using the first image in the sequence as a reference.



(a) (b) (c) (d)

Figure 3.12: *Comparison of input and result of classical SR methods. One input frame (a), bicubic interpolation using all nine input frames (b), IBP (c), and RSR (d) results are shown. The input frame is displayed using nearest neighbor (NN) interpolation to better demonstrate the improved quality in the SR result.*

3.2.1.2 Contemporary SR Methods

3.2.1.2.1 Normalized Convolution [Pham2006]

Normalized Convolution (NC) is a local signal modeling technique utilizing projection onto polynomial basis functions, equivalent to the Taylor series expansion for traditional NC. "First-order NC with three bases {1, x, y} can model edges, and second-order NC with six bases {1, x, y, x^2, xy, y^2} can further model ridges and blobs. Higher-order NC can fit more complex structures at a higher computational cost. However, NC with order greater than two is rarely used since the high-order bases are often fit to noise rather than the signal itself." First-order NC is used for SR fusion [Pham2006]. SR reconstruction using NC is both aesthetically pleasing, and in comparing the HR output with the LR input images it appears to be a reasonable result.

The method uses a least-squares estimation to determine the best-fit basis function to model the LR data in estimation of the HR image. The method is "a good interpolator for uncertain data [but] requires the signal uncertainty to be known in advance." This feature is billed in the form of Noise Robustness, which will first calculate the certainty for each LR input image prior to execution. Because the robust NC approach does not have a closed-form solution like least-squares NC, a two-pass approach uses the result of least-squares NC as the initial estimate for robust NC using an iterative weighted least-squares approach. A comparison of the results of this method with and without the two-pass step is shown in Figure 3.13.

3.2.1.2.2 Robust Super Resolution [Farsiu2004]

Farsiu et al.'s Robust SR method requires an estimate of the HR image (known as a seed image) and treats the supplied between-frame motion as an estimated prior. The authors note that if the data and noise do not conform to the assumed model, then the performance of the method degrades and the presence of outliers will disrupt the estimates made. The authors assume a white Gaussian noise model, this coupled with an admitted flaw in their assumed translational motion model that contributes to increased error. These issues could explain, in part, the results obtained.

When this method was applied in other preliminary experiments performed on synthetic images, it produced mostly acceptable results; however, applied to our measured PSF data, it did not produce acceptable results. A selection of priors was used to determine whether the results obtained were simply due to a poor choice in seed image, but none of the options attempted resulted in a more acceptable outcome.

(a) (b)

Figure 3.13: *Comparison of the least-squares one-pass NC reconstruction (left) with the two-pass robust NC result (right). No difference in output is seen despite the significant increase in processing time. Given the extended processing time required for the two-pass approach and the evidence that the one-pass approach is not sensitive to the residual level of noise in our measured LR samples of the PSF, we will use the one-pass result in our later comparison.*

3.2.1.2.3 Blind Super Resolution (BSR) [Sroubek2008]

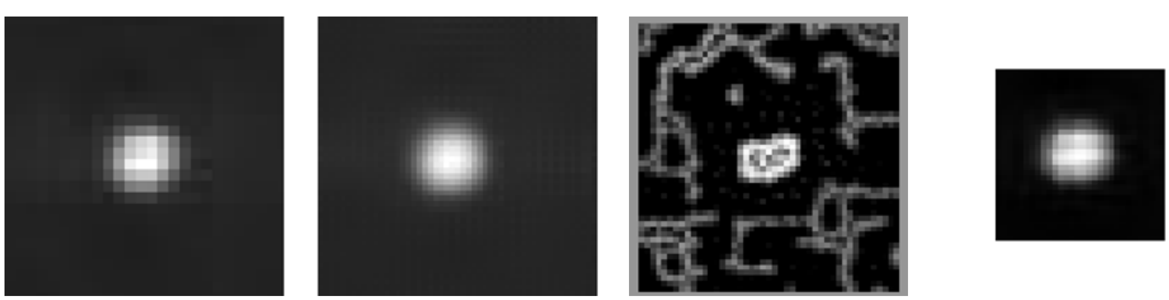
Šroubek et al.'s method combines a regularization-based SR approach with simultaneous estimation of the spatially invariant blur kernel, which are used in estimating the motion of the LR images. The method can be controlled by selecting a potential function used in minimizing the cost function. An alternating minimization scheme is employed where the algorithm performs a minimization first in the image domain, then in the orthogonal blur domain, and repeats until either the maximum number of iterations have occurred or a stopping criterion is met. The selection of the parameters was performed empirically after applying the method to a wide range of datasets. A best result would be obtained by intelligent or exhaustive search of the parameter space, but the high dimensionality of the space and length of time required per experimental run renders this approach prohibitive.

The BSR method generates visually pleasing results compared to the input images. Of note is the difference in the pixel dimensions of the output. Because a 25x25 kernel estimation size was used, the method returns a slightly undersized output, cropped about the edges proportional to the size of the kernel estimate matrix. A comparison of the results of the contemporary class of methods is shown in Figure 3.14.

3.2.1.3 Comparison of Methods

We omit the result obtained using [Farsiu2004] in our comparison. The images have been 2X NN upsized for easier viewing. The diameter and shape of the varying results is consistent, but the content varies significantly from method to method. IBP (Figure 3.15 (b),(f)) retains the most noise outside the extents of the PSF area. Bicubic interpolation and RSR (Figure 3.15 (a,f) and (c,h), respectively) have irregularities and noise in the PSF area compared to the NC and BSR results (Figure 3.15 (d,i) and (e,j), respectively). Aesthetically, the IBP reconstruction (second column from the left) retains the most noise outside the extents of the approximately 7x7 pixel PSF area. Meanwhile, the bicubic interpolation and ZRSR results on either side have irregularities and noise in the PSF area as compared to the NC and BSR results on the right side of Figure 3.15. A comparison to the nine input samples from which these results were calculated is needed.

A visual inspection of the nine input samples in Figure 3.16 based on contrast, relative illumination across the PSF, and shape leads us to conclude the NC result is the appropriate results to use in our method. We also include BSR [Sroubek2008] to determine whether our preliminary conclusions are justified.



(a) (b) (c) (d)

Figure 3.14: *Image summary of application of the modern SR methods. One input frame (a), the Normalized Convolution result (b), Farsiu et al.'s Robust SR (c), and Šroubek et al.'s BSR (d) results are shown. The input frame is displayed using nearest neighbor (NN) interpolation to better demonstrate the improved quality in the SR result.*

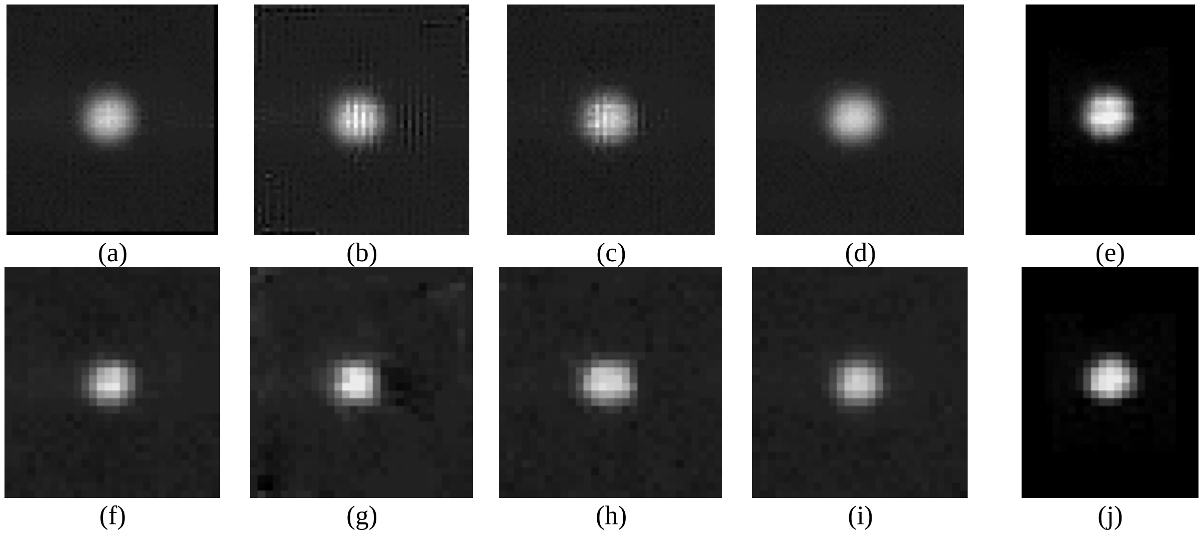


Figure 3.15: *(a-e) Comparison of SR reconstructions at the 2XSR resolution level (left to right - Bicubic interpolation, IBP, RSR, NC, BSR). As we desire to use the PSF measurements to correct for the effects of the lens rather than to give an accurate, high-resolution model of the shape we cannot use the SR results directly. (f-j) Comparison of SR reconstructions after downsampling to the original pixel size (left to right - Bicubic interpolation, IBP, RSR, NC, BSR). After downsampling the data shown in (a-e) we obtain a representation of the PSF at a sampling that may be used directly to correct the effects of the lens.*

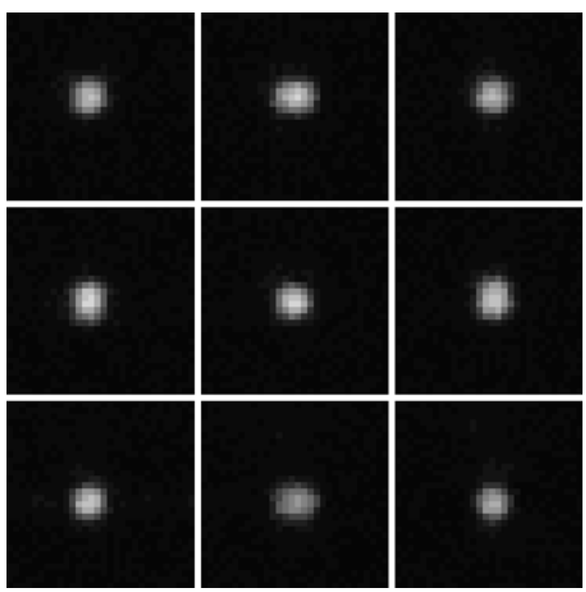


Figure 3.16: *Nine input measurements of the central PSF of the Galaxy S4. We inspect all of the input images when selecting a method so that we can be sure to have a method which most accurately represents the true underlying data. As we do not know how the lens is aligned relative to the sensor, as seen in the difference between measurements above, we must combine the data in order to obtain a more accurate representation.*

3.2.2 Choice of Deconvolution Method and Results

As a mathematical operation used in image restoration, deconvolution is widely used in image de-blurring and contrast enhancement techniques. Image degradation comes from different sources such as aperture diffraction, sensor noise, and lens blur. Thus, the objective of image restoration is to reduce or eliminate some or all of these effects. It is important to note that these forms of image degradation are independent from one another. This independence allows us to use deconvolution techniques.

The effects of diffraction are often less significant than the effects of blur and are most commonly handled in the deconvolution operation as part of the PSF estimation. To reduce or eliminate the effects of noise, there exist algorithms that can pre-filter the noise, thus making deconvolution results more robust, as the deconvolution process is inherently sensitive to noise because it is an inverse operation [Bertero2005]. Lens blur is a non-random spreading of the light that occurs before recording the scene at the sensor plane. Lens blur is intrinsic and thus can be modeled a priori. Wiener [Gonzalez2008], Lucy-Richardson [Richardson1972], [Lucy1974], Regularized Deconvolution [Steve2008], and Kotera at al’s [Kotera2013] methods are applied to sample image captures to demonstrate the improved quality.

Though our method was initially developed in response to issues the authors saw with Liu and Chen’s approach [Liu2008], we lack the equipment to directly compare to their method. The approach of Delbracio et al [Delbracio2012], though a calibrated target-based approach, presents a viable comparison as they also consider subpixel measurement of the PSF.

We calculated the centroid of each local measurement and generated a map of the matching area using kNN (with k=1) to generate the map. The value of each pixel corresponds to the index number in the data structure used to store the individual extracted PSF measurements. Our result is shown in Figure 3.19. Though not encountered in this measurement, some systems we tested had some “bad” pixel regions that reported a reading higher than the surrounding areas, which were filtered out during our extraction process.

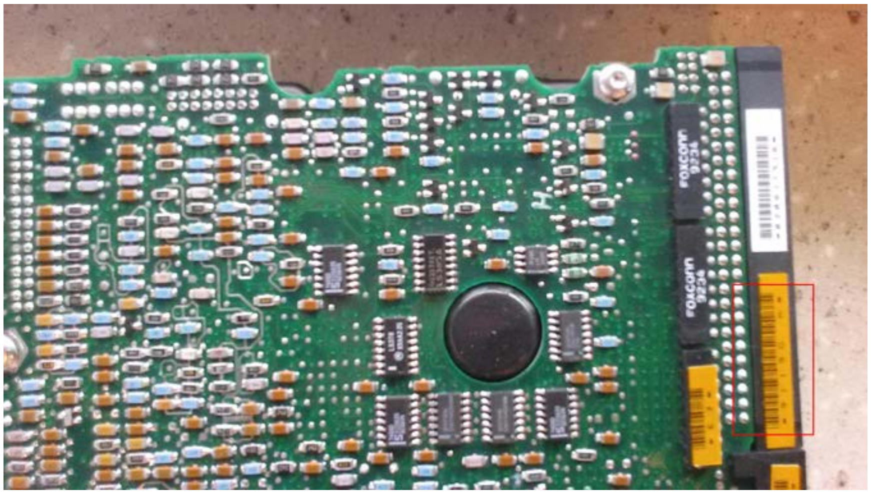


Figure 3.17: *Input image from the Galaxy S4. This circuit board was selected as it provides many hard edges, text, and parallel lines in the barcodes shown in the highlighted region. The highlighted region is displayed in Figure 3.18(a) to enhance visibility.*

(a) (b) (c) (d)

Figure 3.18: *Comparison of the results of deconvolution applied to the image shown in Figure 3.17. (a) Highlighted section of the image shown in Figure 3.17. (b) Result of the application of Wiener spatially variant deconvolution. The Weiner filter is sensitive to noise and parameters were carefully chosen to avoid this artifact. (c) Result of the application of Lucy-Richardson spatially variant deconvolution. At a comparable processing time to Weiner, the results are the best of the methods tested. Note in particular the improved clarity of the lines in the barcode. (d) Result of application of Kotera et al.'s spatially variant deconvolution. This method has comparable processing time to both Weiner and Lucy-Richardson, but results in an over-smoothing of the result.*

Our spatially variant deconvolution function iterates in the order indicated by the map, extracting the region from the image acquired, and applies the desired deconvolution method on the region, adding the masked result to the output image.

All experiments shown were performed on a system with an Intel Xeon E5507 Processor (Quad-core 2.26GHz) and 12GB DDR3 RAM.

More detail on the background of deconvolution in general and algorithms used in particular is given in Section 6.3.

3.2.2.1 Wiener Deconvolution [Gonzalez2008]

Wiener deconvolution takes just under three minutes to complete, and is shown in Figure 3.18(b). The results do not produce significant improvement in the clarity of features compared to the input and results in an overall darkening of the image.

3.2.2.2 Lucy-Richardson Deconvolution [Richardson1972], [Lucy1974]

Lucy-Richardson deconvolution completed in just under 3.5 minutes. Shown in Figure 3.18(c), the result is much clearer compared to that of Wiener. In our experimentation, we found that restricting the number of iterations of the algorithm to four yielded results that did not significantly diminish in contrast.

3.2.2.3 Regularized Deconvolution [Steve2008]

Regularized deconvolution took 12.4 minutes to complete, and introduces blur rather than removing it in addition to reducing contrast. This effect increases in proportion to the distance from the center of the image. For this reason, the result is not included in our comparison.

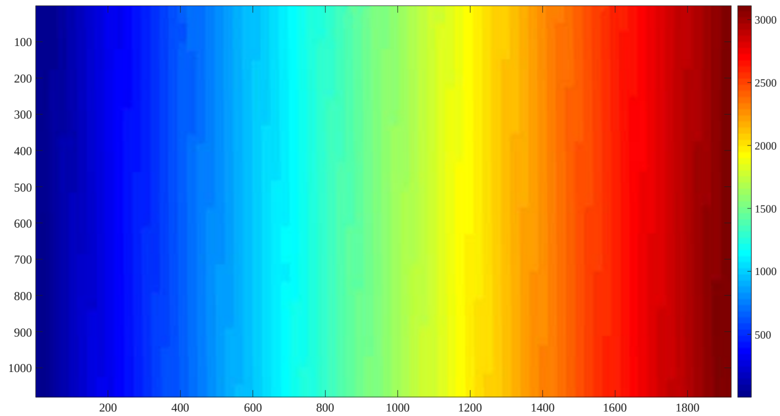


Figure 3.19: *The deconvolution mask mapping the PSFs of the Galaxy S4 system, colors correspond to indices of the PSF in data storage. PSFs are corrected for noise and bias and measurements too near the edge are discarded as incomplete.*

3.2.2.4 Kotera et al Nonblind Deconvolution [Kotera2013]

Shown in Figure 3.18(d), the method of Kotera et al [Kotera2013] took just over four minutes to complete. The method causes overly smooth regions in the result. Despite the reduction in noise in the image, the blur is not removed.

3.2.3 Verification and Application to Other Optical Systems

Though our method was initially developed in response to issues the authors saw with the work of Liu and Chen [Liu2008], we lack the equipment to directly compare our results with their method. Their use of specialty equipment was also a concern that we sought to address in the development of our method to enable simple direct PSF measurement using commodity hardware. The work of Delbracio et al [Delbracio2012], though a calibrated target-based approach, presents a viable comparison in that the authors also consider subpixel measurement of the PSF and employ a repeatable, nondestructive approach for determining the PSF of an optical system.

The overall shape and extents of the two representations shown in Figure 3.20 match well, though there are some differences in the edges and distribution of energy within the PSF area. The results of this measurement are available for more detailed inspection [IPOL1]. One explanation to account for the difference, other than the fundamental difference in the approach of the methods compared, is that Delbracio et al's approach [Delbracio2012] recommends a measurement area of 100x100 pixels in order to make an accurate estimate. Within this same pixel area, we have 16 measurements of the spatially varying PSF.

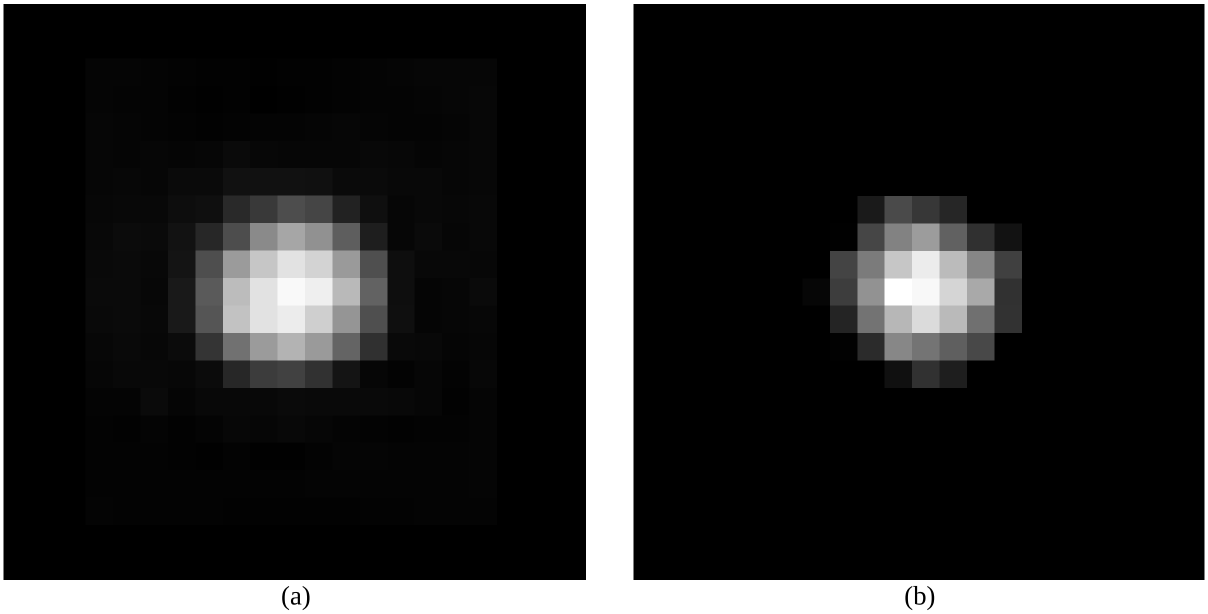


Figure 3.20: *Comparison of our result (a) at the optical center of the Galaxy S4 versus using Delbracio et al's method [Delbracio2012] (b). Though the size of the PSF obtained using the two methods is quite similar in diameter, the content is quite different, owing to the drastically different approaches of obtaining the respective results. Further, we are able to only make a coarse comparison to [Delbracio2012] due to constraints of their method and have selected the area around the central PSF as it changes the slowest from one sample to the next.*

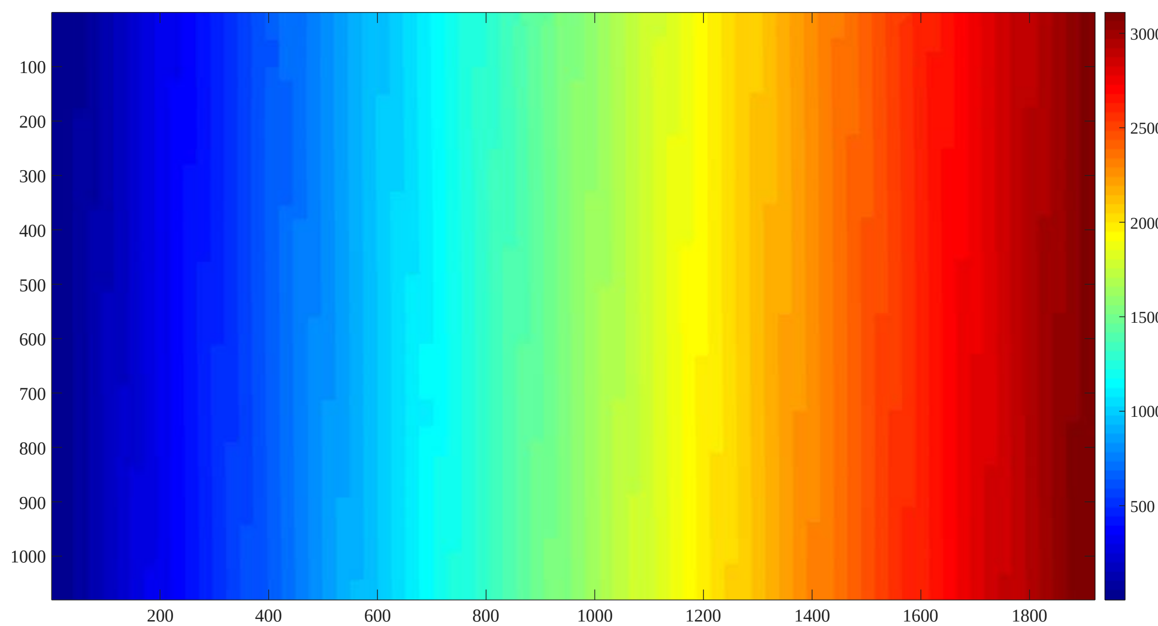
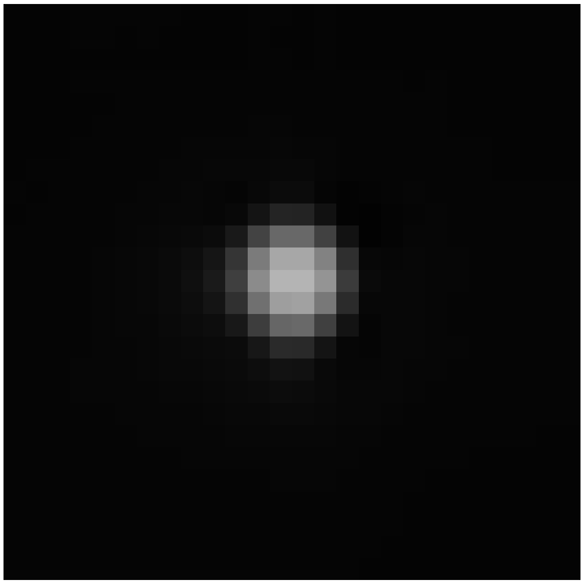
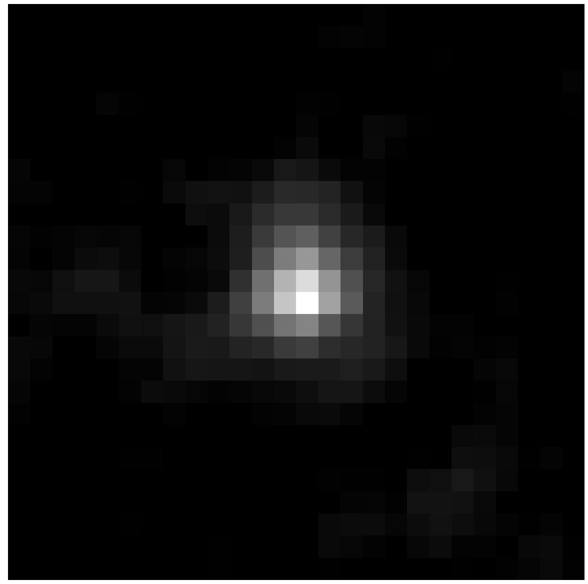


Figure 3.21: *The deconvolution mask mapping the PSFs of the Asus system, colors correspond to indices of the PSF in data storage. PSFs are corrected for noise and bias and measurements too near the edge are discarded as incomplete.*



(a)



(b)

Figure 3.22: *Comparison of our result (a) at the optical center of the Asus versus using Delbracio et al's method [Delbracio2012] (b). Though the size of the PSF obtained using the two methods is quite similar in diameter, the content is quite different, owing to the drastically different approaches of obtaining the respective results. Further, we are able to only make a coarse comparison to [Delbracio2012] due to constraints of their method and have selected the area around the central PSF as it changes the slowest from one sample to the next.*

3.2.3.1 Asus Transformer TF700T

3.2.3.1.1 PSF

The result of the automatic PSF extraction is applied to the data from the device under consideration and the result is shown in Figure 3.21.

3.2.3.1.2 Deconvolution

The Asus Transformer TF700T front-facing camera has similar lens and sensor characteristics as seen by the results compared to the Galaxy S4 measurement.

3.2.3.1.3 Delbracio Comparison

The results of the measurement using Delbracio et al's approach shown in Figure 3.22 are available for more detailed inspection [IPOL2].

3.2.3.2 GoPro Hero 3+ Black Edition

This system has a significant distortion field due to its wide-angle fisheye lens. We captured the requisite number of images using the companion application, which provided its own video preview, easing the alignment process.

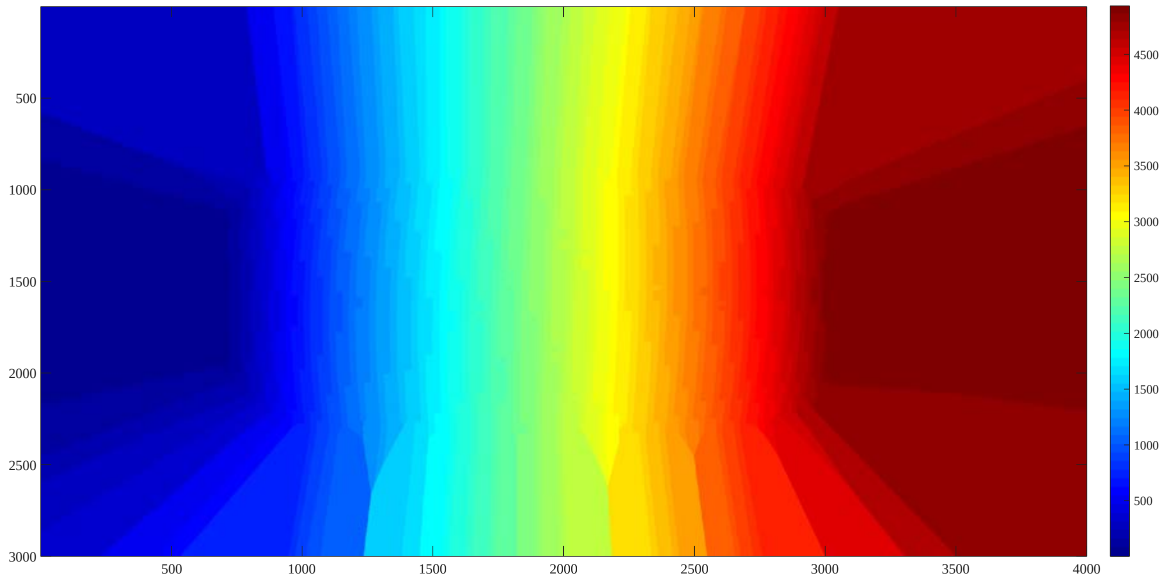


Figure 3.23: *Deconvolution map of the PSFs for the GoPro (colors correspond to indices of the PSF in data storage). PSFs are corrected for noise and bias. Note that the regions of the image towards the edge, away from the measurements at the center of the FOV, are still assigned to use the closest measurement in applying deconvolution. This results in reduced accuracy toward the edges, but these regions are more likely to be cropped in a fully corrected image.*

3.2.3.2.1 PSF

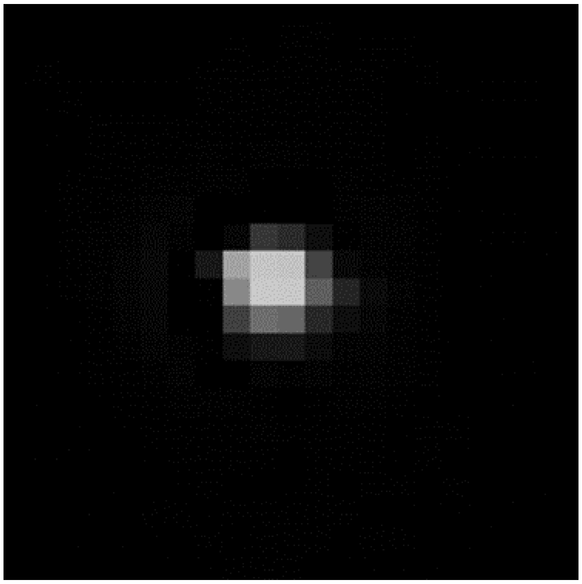
We performed experiments at a range of imaging distances and selected a distance of 30.5cm for comparison. If we used a greater distance, the noise in the measurements made the reconstructions unreliable. Meanwhile, measuring from a closer distance gave too high a sampling of the target, and individual samples of the PSF were not unique. As the sensor was placed at several distances closer to the target, the size of the PSF changed because the point-like source would now no longer image to less than one pixel.

3.2.3.2.2 Deconvolution

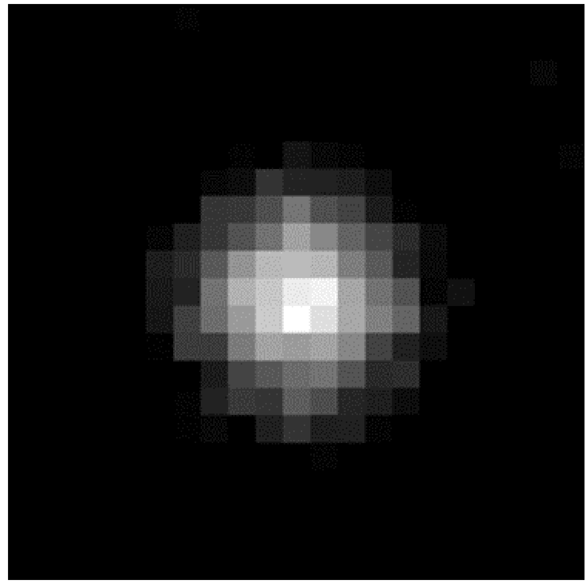
Because this system includes a lens with such a large FOV, imaging at the optimal distance did not allow measurements across the entire FOV. Given the method of assigning points in the deconvolution map to the nearest valid PSF measurement, we assigned each point in the images supplied for testing a valid PSF measurement, as seen in Figure 3.23; however, the accuracy of the deconvolution decreases toward the edges of the image.

3.2.3.2.3 Delbracio Comparison

The results from processing on this sensor correlated little to the result obtained by applying [Delbracio2012]. We noted during experimentation that the PSF for this lens changed quite rapidly based on the distance to target, and we surmise that the disparity seen in Figure 3.24 may be due to the requisite distance of several feet in order to abide by the 100x100 pixel requirement established by Delbracio et al [Delbracio2012] as compared to the 10cm to 35cm range measured using our method. Further information about our result after applying the work of [Delbracio2012] is available for more detailed inspection [IPOL3].



(a)



(b)

Figure 3.24: *Comparison of our result (a) at the optical center of the GoPro versus using Delbracio et al's method [Delbracio2012] (b). Delbracio et al take measurements of a single PSF over a significantly larger area, which we take to partially account for the difference in result particularly since the distortion is so large for this system.*

3.3 Conclusions

We have proposed a repeatable, nondestructive PSF measurement technique and validated our results by applying deconvolution to natural images captured by the same device which we measured. We compared our proposed measurement method to a contemporary random noise PSF estimation method [Delbracio2012] and the results show improvement to both methods [Liu2008], [Delbracio2012]. Some unique aspects in each method's approach account for the differences in the results. Delbracio et al's approach [Delbracio2012] requires measurement of a 100x100-pixel area while our method is only restricted by the size of the device's PSF. Also, our method is a direct measurement rather than an indirect estimation using a calibrated target, so we can avoid the ambiguity discussed in the literature related to the assumptions and claims about the best type of target to use to get the most accurate representation of the PSF.

Not only does our increased spatial sampling allow for smoother transitions in the boundaries between regions, but applying our method to a variety of systems could allow for better understanding of the PSF in general regarding differences in lens families or arrangements of real systems that may not have been discovered previously by mathematical analysis of models.

3.3.1 Competitiveness of Method

Our method was developed in response to potential side effects that the authors saw in Liu and Chen's work [Liu2008], that measurement of the PSF using their method involved destruction of the optical system. We address this facet, both by eliminating the cost and time required for disassembly as well as by aligning with Shih et al's observations

[Shih2012], which state that due to manufacturing tolerances, different lenses of the same prescription will have different PSFs. Also of note are the alignment effects of the lens with its paired sensor, both in the effects of the translational alignment with the pixel surface [Liu2008] as well as handling focal-plane alignment and any tilt alignment issues. By leaving the lens paired with the sensor it was intended to work with, we eliminate all of these undesirable side effects and ambiguities. The demonstration in Figure 3.25 shows the sensitivity of deconvolution to an accurate knowledge of the PSF of the lens.

Further, our method does not utilize a calibrated target, used in popular methods present in the literature for measurement of the PSF. Joshi et al [Joshi2008] use a sharp edge pattern. Other methods that use a random noise target [Levy1999], [Park1984], [Delbracio2012], [Brauers2010] also do not directly measure the PSF, and many do not consider the spatially variant nature of the PSF across the FOV of the optical system. One benefit of using these types of methods is the enhanced SNR of each measurement image captured, which was a concern handled as part of our work.

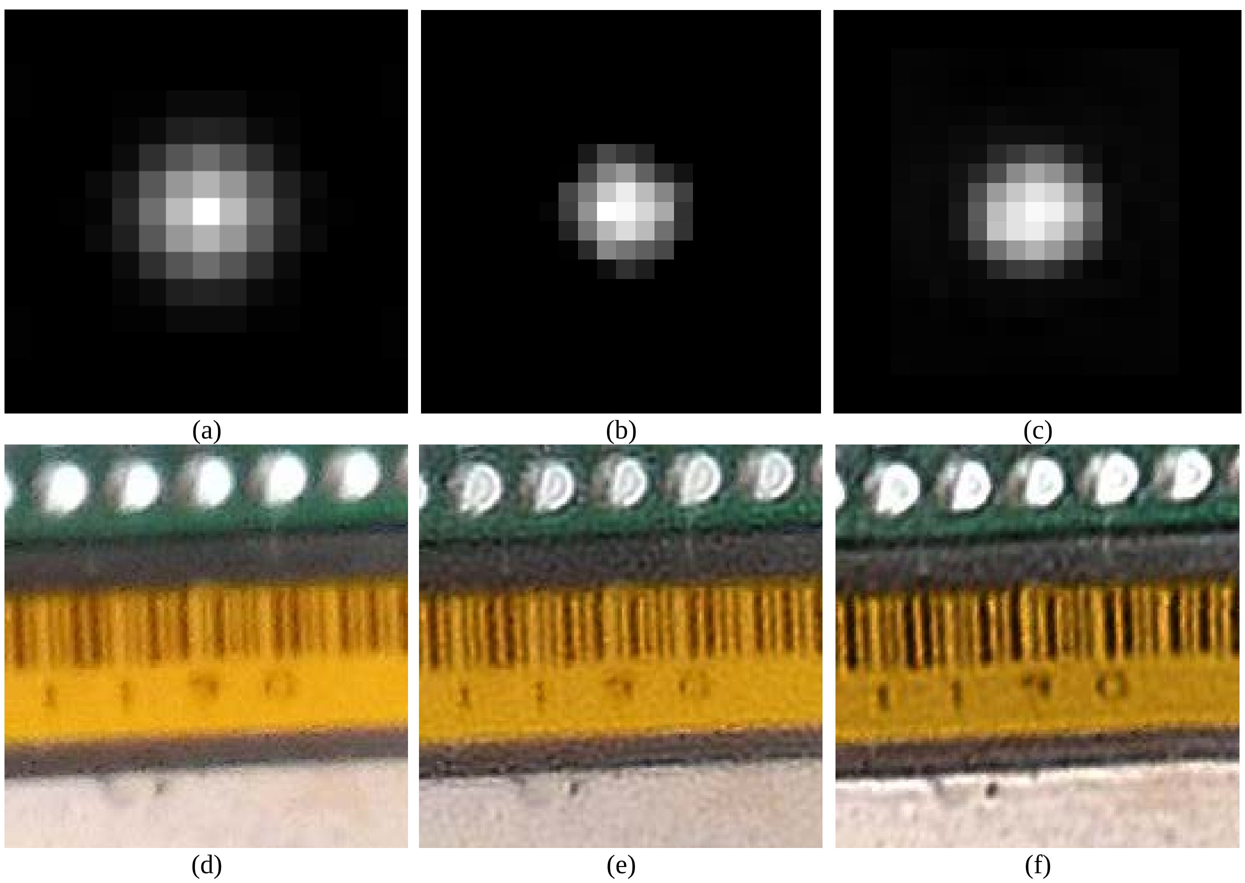


Figure 3.25: *Accurate knowledge of the PSF is vital to successful removal of the blur effects of a lens. Due to manufacturing tolerances the lens prescription does not give sufficient information for blur removal. Shown above are an ideal representation of the PSF (a), the PSF measured by the method of Delbracio et al. [Delbracio2012] (b), and the measurement obtained from our proposed PSF measurement method (c). For a captured image (d) we see how small differences in the accuracy of the PSF measurement can greatly influence the clarity of the restored image. Using Delbracio et al.'s measurement (e) is an improvement over the original but the result does not give as significant an increase in clarity as does using the measurement using the proposed method (f).*

3.3.2 Future Work

Additional work in applying our PSF measurement method includes analyzing the use of our measurements in other related applications: non-stationary image deconvolution, validation of lens quality compared to a prescription or in detection of counterfeit lenses, etc. As we obtain more information about the performance of our method in a variety of applications, updates and corrections in the process will further enhance the general applicability.

Development of a fully automated system would allow users of smartphones to be able to calibrate the internal image correction for lens effects by enabling rapid calculation of the PSF of their imaging system. Since consumer use of mobile phone cameras for everyday capture of personal memories is ever increasing, applying our method would widely benefit modern image processing methods by enabling the best quality recording of images and video to enhance the inter-communication of millions. Further, extending this automation would allow researchers to analyze the effects of changes in imaging systems subject to extreme environmental conditions. Generally these are expensive, purpose-built systems for which accuracy and longevity are key. Analyzing the change in PSF over the course of the device's lifetime would ensure accurate correction of the data obtained, and validate the device's continued use, either to serve as a lesson learned in design change or to renew its term of service.

4 Image Simulation

Although the intent of development of any image processing algorithm is to provide accurate results in practical application, often the first step is to evaluate a proposed methodology on image of known construction. For instance, a good image may be degraded using known parameters in order to obtain knowledge of where the algorithm's areas of robustness and brittleness lie. In this section we describe the process used and the scope of parameters explored to test the limits of our synthetic image generation workflow. Results will be shown to demonstrate the utility of first operating on synthetic image sets.

4.1 Target Path Information

For our application, the source information from which we base our synthetic image generation does not derive from simplistic 'cartoon'-style images to which noise and blur will be added, but rather from three-dimensional path information of the target's position. The original information is composed of regular time-steps with XYZ position and XYZ velocity vectors at each time-step. Time-steps are irregular, ranging from between 5 seconds before the next measurement to 60 seconds. This information is then recomposed and interpolated into a regular series of XYZ positions at ¼ second time-steps between each position vector. There are several target paths which are processed in this way where we generate dense position information to be used in our experimentation.

These positions and a model of the target are used in generation of our simulated images. We explore the situation in which the target is viewed from above by a sensor which is mounted on a gimbal and equipped to be capable of tracking the target when within its trackable area. The sensor has a given resolution and noise characteristics, and is instrumented with a lens which introduces its own degradations.

4.2 Image Acquisition Model

For the situation described in Section 1.1 we must derive an image acquisition model. We desire to have our simulated images as representative as possible of those images we would expect to acquire from a real system and therefore our acquisition model will be as comprehensive as possible. This is a departure from many synthetic image generation schemes which often simply assume additive Gaussian white noise and Gaussian blur [Sroubek2008]. What follows is a graphical depiction of the construction of our described situation for which we simulate the image acquisition, a mathematical formulation of the situation depicted, and finally a discussion of the scope of error level magnitude considered for each error source.

4.2.1 Depiction of Image Acquisition

For our considered image acquisition situation, we have a target we desire to track passing through the atmosphere of Earth. The target is always above the surface and the sensor platforms are downward-facing, always above the target as depicted in Figure 4.1. Given that the target is warmer than the atmosphere surrounding it or the surface of the Earth behind it, the infrared (IR) portion of the electromagnetic spectrum is selected for simulating the images. Table 4.1 gives a summary of the variable names which will be used.

Table 4.1: *Description of the variable names used in description of the tracking situation and in mathematical formulation of the image acquisition model. Variable names are in bold for the parameters which appear in the mathematical formulation in the 'Name' column, the data type is also listed as well as a description of the contents of the variable/parameter.*

Name	Type	Description
Target position (**TP**)	3x1 vector	x,y,z value of target
Target size	2x1 vector	Height and width of target in meters
Sensor position (**SP**)	3x1 vector	x,y,z value of sensor
Sensor direction (**SD**)	3x1 vector	Direction indicating the pan/tilt of the sensor
Sensor **FOV**	Scalar	The field of view (in degrees) of the sensor
Sensor resolution	2x1 vector	Height (**SH**) and width (**SW**) of sensor in pixels

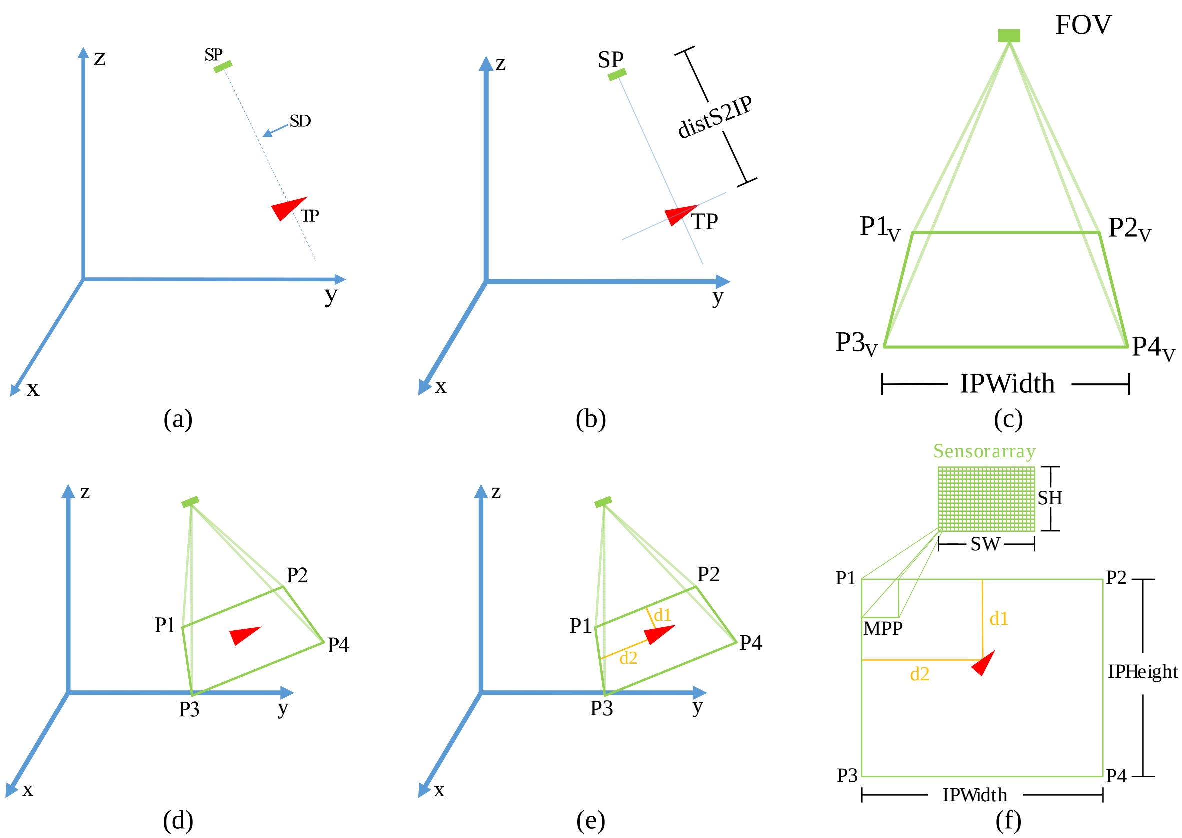


Figure 4.1: *Depiction of the geometric formulation of the imaging situation. (a) shows the coordinate axes, sensor position (SP), target position (TP), and sensor direction (SD). (b) shows the formulation of dist2IP, the distance of the target to the image plane, which is the difference of TP and SP. (c) shows the extents of the FOV of the sensor at dist2IP, which establishes the corners of the image plane. (d) shows the FOV superimposed on the coordinate space surrounding the target. (e) shows the physical meaning of d1 and d2 which are the coordinates of the target in image space represented in physical space. (f) shows the same FOV grid in (e) from an image-centric view, demonstrating the analog of image space at the plane define in physical space.*

4.2.2 Mathematical Formulation of Geometric Model

The coordinate space of our system model uses the origin at the center of the Earth. The z-axis goes up through the North Pole and the x- and y-axes lie along the equator. Figure 4.1(a) shows the sensor position (SP) and target position (TP) with the orientation of the sensor or sensor direction (SD) represented as the alignment of the sensor with the target. Whenever the target is within range of the sensor's trackable area (defined as the limit or extents of the sensor's pan-tilt capabilities) the sensor is oriented to face the target directly. TP is known at each time-step and SP is defined prior to the initiation of a given experiment. SP has an orbital path, but for purposes of formulation we will assume that SP does not change appreciably during the period of time for which TP is within the trackable area of the sensor.

Figure 4.1(b) indicates our first calculated variable, distS2IP, which is the distance from the sensor to the image plane defined at the center of the target and is defined as

$$dist2IP = \|SP - TP\| \tag{4.1}$$

At this distance, we are able to define the imaging plane with respect to the center of the target and the FOV of the sensor. This gives values for the corners of the viewable area around the target and the image plane width and height.

$$IPWidth = 2 * dist2IP * tan\frac{FOV}{2} \tag{4.2}$$

If the horizontal FOV (HFOV) and the vertical FOV (VFOV) differ, HFOV and VFOV may be substituted into (4.2) in order to calculate IPWidth and IPHeight, respectively. Once we know the appropriate extents of the image plane using (4.2), we can now rotate the plane from its coordinate system centric orientation and instead orient it with respect to SP. We do this using the following equations, where each of points P{1-4} are represented in the equations by their XYZ coordinates. The transformation of each point is identical.

$$\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & cos\,\theta_x & -sin\,\theta_x \\ 0 & sin\,\theta_x & cos\,\theta_x \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} \tag{4.3}$$

$$\begin{bmatrix} x'' \\ y'' \\ z'' \end{bmatrix} = \begin{bmatrix} cos\,\theta_y & 0 & -sin\,\theta_y \\ 0 & 1 & 0 \\ sin\,\theta_y & 0 & cos\,\theta_y \end{bmatrix} \begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} \tag{4.4}$$

(4.3) and (4.4) perform three-dimensional rotations of the image plane about the x- and y-axes. They each use the values θ_x and θ_y, which are defined as

$$\theta_x = tan^{-1}\frac{TP_y - SP_y}{TP_z - SP_z} \tag{4.5}$$

$$\theta_y = tan^{-1}\frac{TP_x - SP_x}{TP_z - SP_z} \tag{4.6}$$

The xyz subscripts indicate which of the values in the position vectors should be used in determining the appropriate value for rotation. The properly-oriented resulting image plane is indicated in Figure 4.1(d).

As seen in Figure 4.1(e), the point $P1$ is designated as the image origin. We now wish to find the distance from the center of the target to the top of the image plane. The vector along the top of the image plane is given by

$$v = \begin{bmatrix} P1_x + (P2_x - P1_x)t \\ P1_y + \left(P2_y - P1_y\right)t \\ P1_z + (P2_z - P1_z)t \end{bmatrix} \tag{4.7}$$

The value for t which minimizes the distance between TP and the top of the imaging plane is

$$t = \frac{(P1 - TP)(P2 - P1)}{|P2 - P1|^2} \tag{4.8}$$

The same steps are used to find the distance from TP to the left edge of the image plane. These two vectors called v in (4.7) define $d1$ and $d2$ as shown in Figure 4.1(e) and (f).

With the location of the center of the target now defined in the physical image plane we must now represent the target in image coordinates. To do this, we must know what the correspondence of units is between the two spaces. The number of meters per pixel is given as

$$MPP_H = \frac{IPHeight}{SH} \tag{4.9}$$

$$MPP_w = \frac{IPWidth}{SW} \tag{4.10}$$

for the image width and height, respectively. SH and SW are the sensor height and width in pixels. Then in image coordinates, the center of the target is

$$i = \frac{d1}{MPP_H} \tag{4.11}$$

$$j = \frac{d2}{MPP_W} \tag{4.12}$$

About these coordinates in image space, we must construct a representation of the target. For now we will assume the simplistic case of a binary image for which the value is set high if the pixel contains some portion of the target and the value is set low otherwise. This gives us the number of pixels that the target will map to:

$$NPi = \frac{Target\ height}{MPP_H} \tag{4.13}$$

$$NPj = \frac{Target\ width}{MPP_W} \tag{4.14}$$

These values allow us to generate a binary image of the target as it would be seen from the perspective of the sensor, shown in Figure 4.2.

4.2.3 Image Acquisition Model

Any physically realizable imaging system will introduce errors in the acquired data. These errors largely come from three different sources. The two intrinsic sources of error that are particular to the imaging system and do not change significantly from one acquisition to the next stem from the lens and sensor. Use of a lens introduces blur, both from diffraction blur due to the finite aperture and the various components of optical blur from the lens elements. These blurs are generally described in the form of what is known as the point-spread function (PSF). The pixels of the sensor also contribute to what is called sensor blur. Rather than a true impulse sampling of the incoming light, sensors have limited resolution and finite pixel sizes. This leads to blurring effects, coupled with aliasing effects which limit the effective resolution of the system.

The third source of error is extrinsic and varies significantly between image acquisitions and are related to the motion of the optical system and motion of subjects within the scene. This motion gives rise not only to the necessary between-frame motion vital to the success of SR algorithms, but also to motion blur. Motion blur occurs when either the image acquisition system or subjects within the scene, often both are present, move during the finite acquisition time of the sensor.

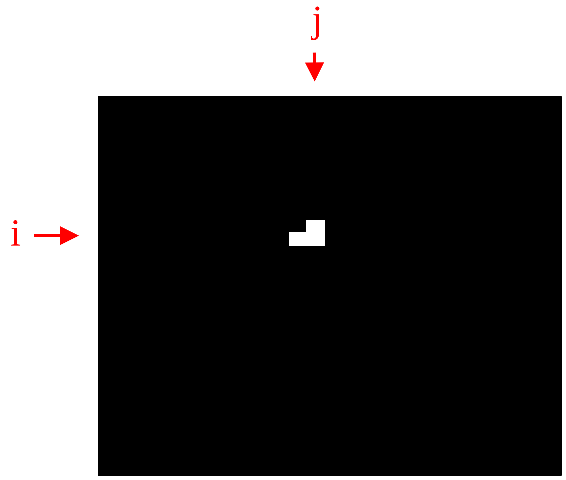


Figure 4.2: *Depiction of the target as viewed from the sensor. The* ***i*** *and* ***j*** *coordinates of the center of the target are indicated and the extents of the target are shown in a binary representation where any pixel with information about the target is set to high. If the target were to be contained entirely within a single pixel, but blurring of the lens system spread this light energy to neighboring pixels, the centroid of the target location can be calculated from these relative values. We also consider the effects of noise contamination and detection of the target based on its spectral signature in a multispectral image acquisition scenario.*

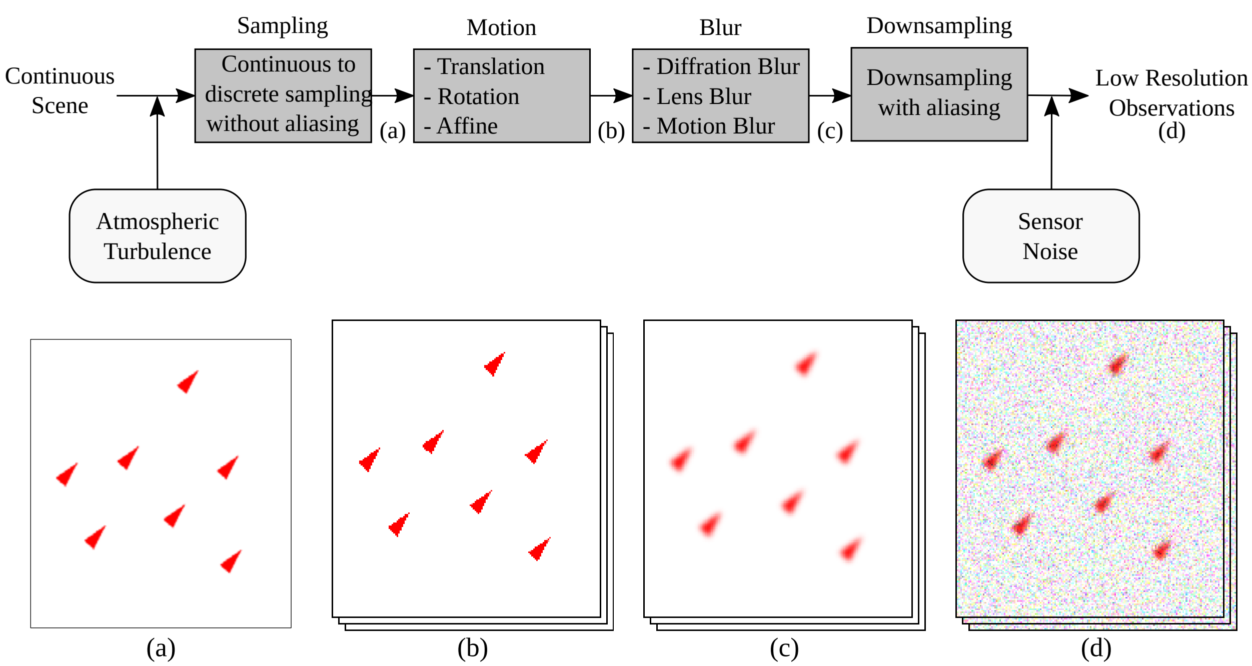


Figure 4.3: *The image acquisition model, also known as the observation model [Milanfar 2010]. The flowchart shown describes the relationship of the original scene to the observed images. The perfect observed image differs only from the continuous scene it represents by the discretizing of the scene into pixels (a). In a multi-acquisition framework, the individual images acquired differ from each other by small amounts and therefore we have images of the same scene with small amounts of motion between them (b). The motion model used depends on the acquisition system used and the motion of objects within the scene. An imaging system which utilizes a lens with a finite aperture and acquisition time will have some level of blur (c). Finally, the images are recorded by a sensor of limited resolution which further introduces some amount of noise in the measurements (d).*

The intrinsic sources of error may be calculated prior to image acquisition, a calibration for each lens-sensor pair can be made and saved for recall in processing the images to remove the negative effects of these sources. The extrinsic source of error must be calculated for each image acquired. Techniques have been developed for both blind [Sroubek2008] and non-blind [Kotera2013] recovery of the intra-frame motion blur, while a class of registration or alignment methods have been developed to calculate the inter-frame motion [Milanfar2010].

As shown in Figure 4.3, noisy, blurry, low resolution images (d) – versions of our desired representation (a) of the continuous scene – are the only data we are able to acquire. We must use our knowledge of how these sources of error were introduced, coupled with methods developed to correct the particular error, and use knowledge of our system and the conditions under which the images were acquired in order to arrive at our best estimate of (a). Two sources of error not yet discussed are typically handled as statistical modeling problems [Parenti1994]: sensor noise and atmospheric turbulence.

4.2.4 Scope of Error Sources in Image Acquisition

An appropriate range must be established for which the synthetic image generation applies the various sources of degradation. We will consider situations ranging from ideal to representative to worst-case for each of the sources of error mentioned in order to show typical and limiting results for our considered situation in order to get an accurate picture of our capabilities with respect to our major output – the position of the target we are tracking. In this document, we will explore the types and ranges of error for each source and show results where conditions are ideal and error is low, where conditions are typical and error is moderate, and we will explore the worst case scenarios, or rather we will discover the breaking point of our approach with respect to each error source and perform a sensitivity analysis of our recovered target position as the error increases.

4.3 Source of Error: Lens

The use of lenses with image acquisition systems is near-ubiquitous, and lenses are an implied component of digital optical systems. There are countless varieties of lenses that may be used and each lens comes with tradeoffs in cost, complexity, size, weight, and quality. In this section, we review the major lens families and discuss correcting for the degrading effects introduced by the lens on acquired images.

4.3.1 Lens Families

“The lens is generally the most expensive and least understood part of any camera”[Kingslake1989], and though a great number of advancements and improvements have been made in the history of photographic lens design only six main families of lenses exist today from which nearly all modern lenses may trace their lineage [Kingslake1989]. We will briefly review these six prototypical base lens designs from which our comparison will stem.

4.3.1.1 Petzval Portrait

This design was created sometime before 1850, and was in popular use for over 50 years. Shown in Figure 4.4, the design consists of two doublet lenses with an aperture stop between. Spherical aberrations are corrected fairly well by the front group at the expense of introducing significant coma. The second group corrects the coma while the buried aperture stop corrects the astigmatism at the expense of magnified field curvature at the image plane. The main purpose of this lens was in taking portraits since a reduced exposure time was made possible compared to other contemporary lenses and its use in this way did not highlight the restricted field of view and vignetting typical of this design [Ray2002].

4.3.1.2 Rapid Rectilinear

The Rapid Rectilinear lens was patented by John H. Dallmeyer in 1866 to fill the need of a lower distortion intermediate lens that was relatively free from distortions at about F/6 within about a 50 degree field of view. This design, shown in Figure 4.5, is also known as a Steinheil Aplanat, developed by a German designer working with Seidel who developed a nearly identical design just a few weeks prior to Dallmeyer's patent.

Similar to the earlier Petzval Portrait design, the Rapid Rectilinear consists of two doublet groups with an aperture stop in between. However, the important difference is the symmetric design and choice of glass type in the groups. The symmetrical design gives near complete elimination of distortion, coma, and lateral chromatic aberration. The paired glass types needed to be as different as possible in refractive index while having relatively similar dispersion. Dallmeyer placed the lower-index positive elements inside near the aperture stop with the higher-index negative elements placed outside. Though difficult to manufacture due to the strong curvature at the cemented internal interface, the design had minor aberrations over a range of object distances with minor distortion. [Kingslake1989]

4.3.1.3 Double Gauss

The Double Gauss derives its name from the 1817 Gauss objective lens. The design shown in Figure 4.6 also takes advantage of the benefits of symmetry in reduction of aberrations. The first known pairing of these lenses and creation of the Double Gauss was done by Alvan Clark in 1888 while working for Bausch and Lomb.

Most modern versions of the Double Gauss contain six elements, which design change was made in 1895 to correct for chromatic aberrations by cementing the inner two elements of each side into doublets. This basic design is one of the most utilized in photography and is still commonly used today in low-cost but high quality fast lenses. The only major aberration that is typical of this lens family is some remaining oblique spherical aberration [Cox1971].

4.3.1.4 Cooke Triplet

The Cooke Triplet has been called the most important lens design ever created [Cicala2011]. This is partly due to its near-ubiquitous use: more derivative designs have been based upon the Cooke Triplet than any other. More so is the fact that the simple and elegant design is found in the vast majority of modern lenses. The Cooke Triplet, shown in Figure 4.7, is simple and inexpensive to produce while giving reasonably good correction of all aberrations.

The Cooke Triplet was designed in 1893 by Dennis Taylor while working at T. Cooke & Sons. The central element is flint glass with the outer elements composed of a crown glass. The design allows for zero Petzval field curvature (the sum of the refractive indices of the elements multiplied by their curvatures, giving a flat field of focus). It was the first design to eliminate most of the optical distortion at the outer edge of the lens. This is partly due to the fact that the negative central element can be as strong as the combination of the outer two elements but still have an overall convergence effect since the light strikes the central element near the optical axis.

The Cooke Triplet is still in heavy use in inexpensive lens systems due to its simple three-element design coupled with its overall good quality image reproduction. Another benefit is the ease with which the focal length of the lens may be adjusted, by moving the central element closer to the front or rear element the field of view of the system changes. [Kingslake1989]

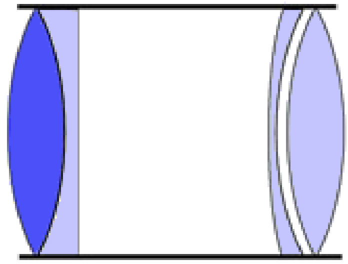


Figure 4.4: *Petzval Portrait lens, image adapted from [Cicala2011]*

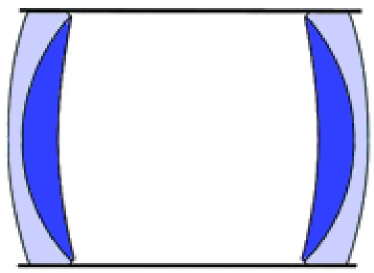


Figure 4.5: *Rapid Rectilinear lens, image adapted from [Cicala2011]*

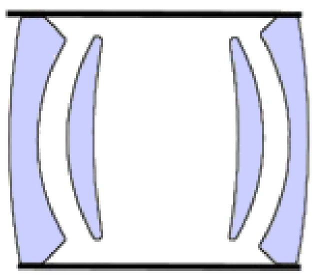


Figure 4.6: *Double Gauss lens, image adapted from [Cicala2011]*

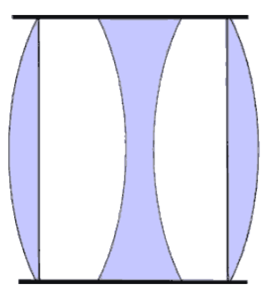


Figure 4.7: *Cooke Triplet lens, image adapted from [Cicala2011]*

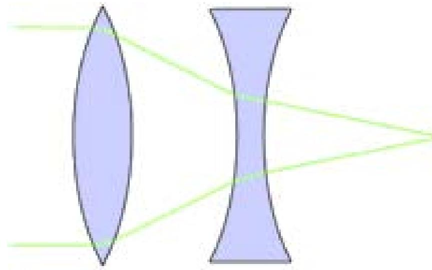


Figure 4.8: *Telephoto lens, image adapted from [Cicala2011]*

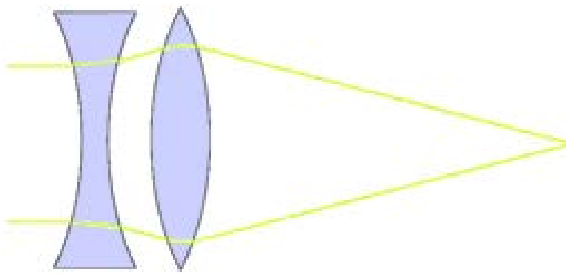


Figure 4.9: *Reverse Telephoto lens, image adapted from [Cicala2011]*

4.3.1.5 Telephoto

The telephoto lens shown in Figure 4.8 is a long-focus lens where the physical length is shorter than the focal length. Several inventors developed what can be called telephoto lenses nearly simultaneously. British optician Thomas R. Dallmeyer, son of John H. Dallmeyer mentioned previously for his work on the Rapid Rectilinear lens; New Zealand photographer Alexander McKay; and German scientist Adolph Miethe all created markedly similar designs within about five years of one another. Their designs are fundamentally a positive front element followed by a strong negative rear element.

Earlier on in the history of lens design, lensmakers found that a single strong element could be split into two or more elements with the advantage of lower expense to produce multiple weak elements as well as the advantage of using different glass types in order to reduce chromatic aberration in the system. Even more important was the fact that different curvatures could be employed within the same shape (convex, concave, meniscus, etc.) in order to reduce certain aberrations due to the added degrees of freedom. Modern telephoto lenses have heavily taken advantage of this principle, so much so that in order to notice the pattern one must look closely as modern designs can easily have more than a dozen individual elements.

4.3.1.6 Reverse Telephoto

The reverse telephoto is, as implied, the reverse of a telephoto lens, with a strong negative front element and a positive rear element. This lens design family is the most recent, being invented in the 1920s to fill the needs of the burgeoning film industry while the other five families were created during the 1800s. The distinctive feature of the design shown in Figure 4.9 is the increased back focal distance of the lens, sometimes more than the focal length. Not to be confused with the focal length of the lens, the back focal distance is the distance between the back surface of the last lens element and the imaging plane. This feature was needed for movie cameras where the beam splitting prism behind the lens took up a significant amount of space, and the reverse telephoto design solved this problem.

When the Single-Lens Reflex (SLR) camera was developed, this design feature also helped with the issue of the position of the lens relative to the reflex mirror. Other advantages include a relatively high aperture compared to other designs as well as minimal vignetting and spherical aberration. However, there are always tradeoffs in design and some of the ones for the reverse telephoto include increased field curvature and severe wide-angle aberrations (including distortion, coma, and lateral chromatic aberration). The aberrations can be corrected with additional lens elements, but this adds to the cost and complexity of the system. The distortion is particularly difficult to correct, so much so that the fisheye lens is considered by some to be an uncorrected wide-angle reverse telephoto. [Kingslake1989]

4.3.2 Lens Degradations

In our experimentations, we investigated some lenses which had only small departures from the basic design of their source family. These simple lenses are superior for our evaluation particularly in regard to the lower weight and cost of the lens system as well as the stability of the lens, meaning the simpler assembly due to the simplicity of the design and decreased likelihood of misalignment of the lens elements due to environmental effects. Further, the simpler lenses often exhibit the types of aberrations that computational systems are particularly suited towards correcting. We demonstrate the simple lenses selected for comparison and demonstrate their effect on the image data.

The selection of a lens has a great deal of influence on the data recorded by the sensor behind it. Accurate knowledge of the point-spread function (PSF) which describes the way a point-like source of light would be seen by the system can improve the quality of data available for further processing. We investigated two methods of PSF estimation: the first being from simulation using the software package ZEMAX, a program developed for the design and simulation of lenses, and the second method was created in order to directly measure the PSF of a real digital optical system. In all of the experiments presented here, we demonstrate our processes on the visual spectrum though the extension to IR wavelengths is similar for the application of our developed method.

First, in our simulated imaging calculation of the PSF we demonstrate on a cemented doublet. More complex lenses often have improved imaging characteristics, including smaller PSFs, and we wish to demonstrate that our work can be successful using smaller, lighter, cheaper, simpler lens systems which are less prone to damage and alignment errors despite their often complicated distortion fields and larger PSFs. This cemented achromatic doublet has seven degrees of freedom: three lens curvatures, two lens thicknesses, and two glass types. The layout is seen in Figure 4.10 and includes a depiction of three field angles. The total area for which the focal plane at the right of the figure is covered is greater than the footprint of the lens elements, making this lens a wide-angle lens though just barely so.

The PSF shown in Figure 4.11 is of the central field (the blue lines in Figure 4.10) which means that there is a 0^{o} angle between the optical axis and the rays considered in calculating the PSF. Figure 4.12 shows the 2nd and 3rd fields (5.6^{o} and 8^{o}, respectively). It is important to consider the isoplanatic and spatially variant nature of the PSF because, as shown in Figure 4.13, the PSF can have a vastly different shape in different regions of the imaging plane. For purposes of our simulated experiments, we use the PSF grid and select samples such that the difference between adjacent samples is small while maintaining enough space between that adjacent samples do not overlap. The representation of the PSF for this lens is shown in Figure 4.13.

In order to view what a certain target will look like through our lens, we can perform ray tracing to simulate the actual imaging process of photons passing through the lens and striking the sensor. This process, though accurate, is quite time-consuming. Another way to view simulated images of scenes is to sample the PSF space via the PSF grid as shown in Figure 4.13, then to perform a region-based convolution of the input scene with the PSF valid for the particular region. Each pixel of a digital optical system does have its own unique PSF which accounts for both the effects of the lens and sensor. However, the PSF usually changes slowly enough that we are able to define isoplanatic regions for which a given PSF sample is valid within a given region. The result of this sampling process for an example image using this lens is shown in Figure 4.13.

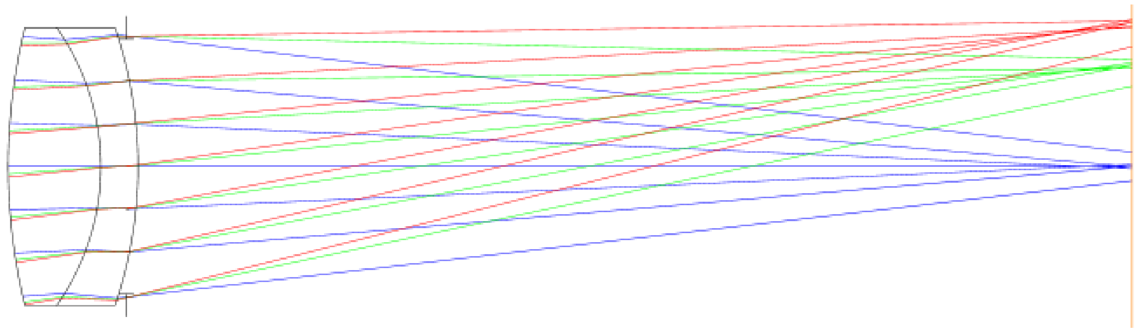


Figure 4.10: *Lens used as example for aberration, cemented doublet. The colored rays indicate three field angles of equal area at the image plane. The lens chosen is just an example to get a first-order but realistic example of errors caused by the imaging system. Other lens arrangements may be considered in order to determine the influence on resulting accuracy of the position estimate.*

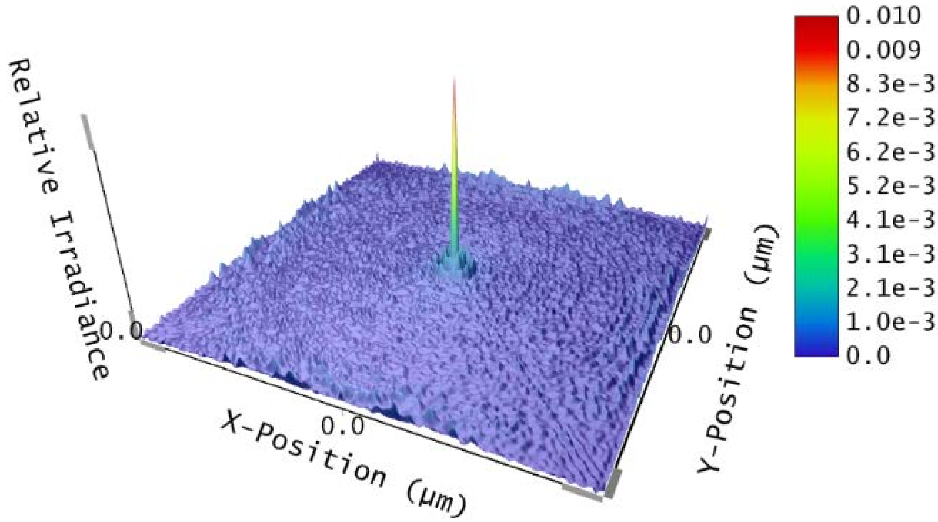


Figure 4.11: *FFT PSF of the lens shown in Figure 4.10 - 0 degree field. The values are normalized and the greatest concentration of energy is in the center as expected. However, the effects of using a low-quality lens are apparent as a significant amount of energy is contained in the surrounding area away from the desired central spike.*

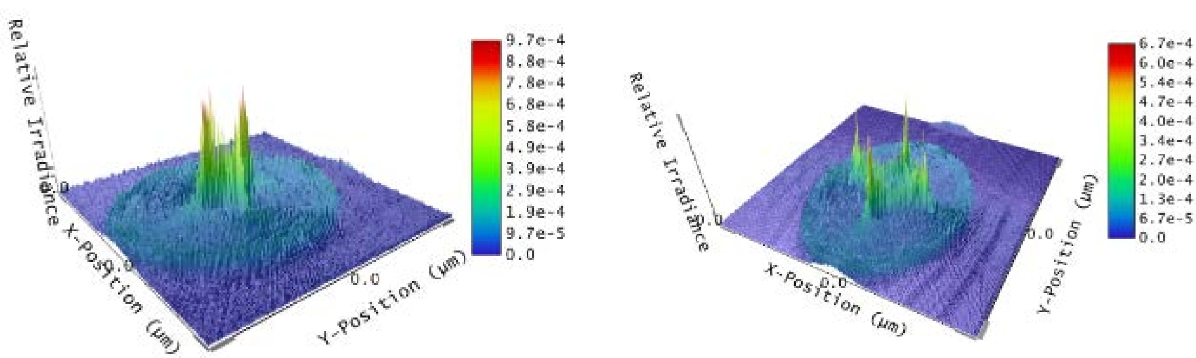


Figure 4.12: *FFT PSF of the lens shown in Figure 4.10 - 5.6 (left) and 8 (right) degree field. The PSF of any real system will change across the FOV, complex systems often use many elements to minimize the effect of the PSF at the cost of added weight and cost and the greater complexity introduces risk for misalignment during the life of the lens. Shown here, we see two examples of the PSF at locations off the optical axis which depict wildly different distribution of the energy compared to the desired single spike in general and to the central PSF of this lens in particular.*

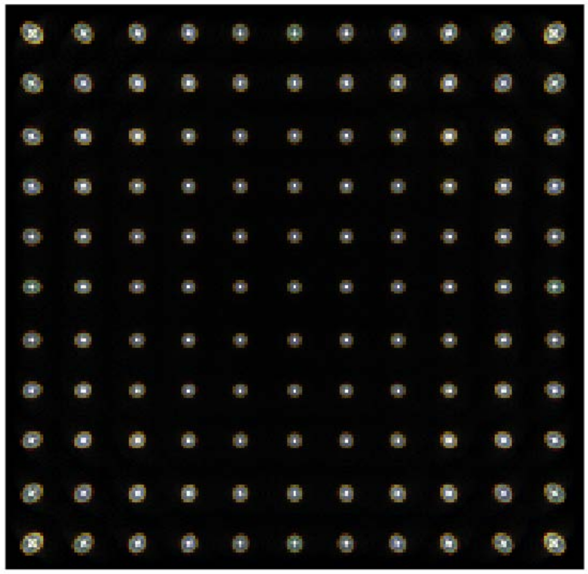


Figure 4.13: *PSF grid for the doublet lens shown in Figure 4.10. The sampling is chosen to provide adequate coverage of the scene relative to the change from one measurement to the adjacent ones while maintaining an appropriate band between measurements. The size of the sample at the center is 11x11 pixels while the size of the sample in the corners comprises an 18x18 pixel region.*

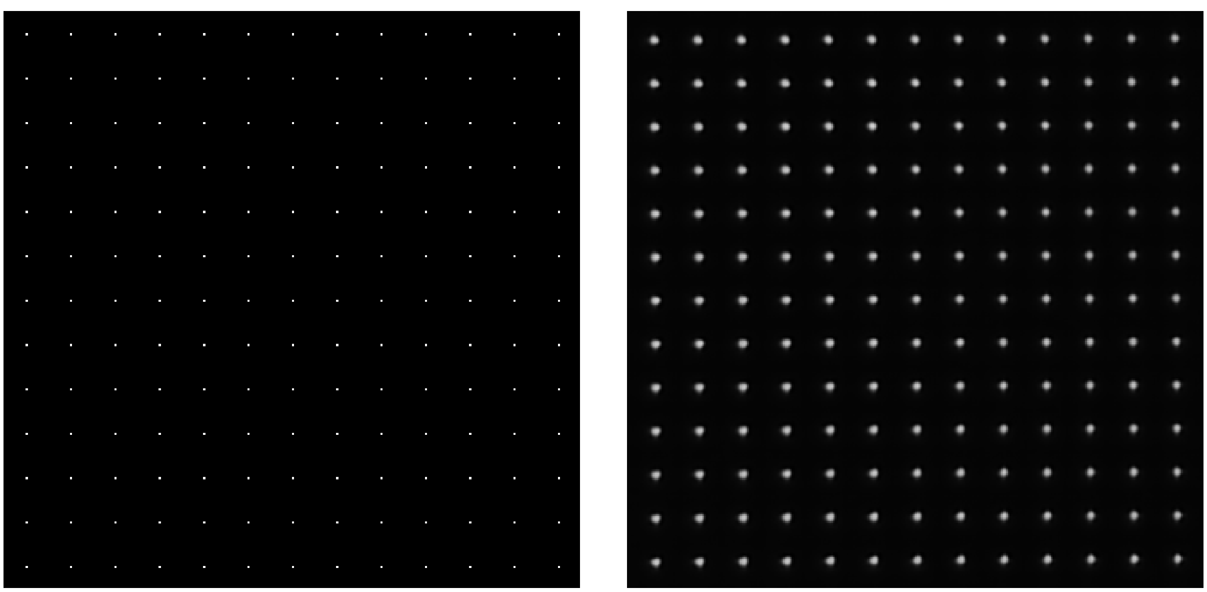


Figure 4.14: *The PSF grid, displayed as point-like sources (left) and after passing through the lens and recorded by the sensor (right). The region shown is about the center of the sensor, resulting in fairly uniform degradation due to the effects of the lens. This comparison demonstrates the effect of a single point of light expanding, or spreading, into the regions shown to the right. The size of the sample at the center is 8x8 pixels while the size of the sample in the corners comprises a 10x9 pixel region.*

Note how although the overall quality of the image is diminished, as it will be to some extent for any real system, the light energy is spread over a larger pixel area which increases the sensitivity of the system to a given target. If target localization is all that is desired, we need not correct for the lens aberrations. However, if we desire to have a more acute sense of the image of the target we may use the PSF measurements in order to de-convolve the acquired image to have a more accurate representation of the scene.

The second method which was developed for measurement of the PSF of a physical lens utilizes a grid of point-like sources in order to measure the isoplanatic, spatially variant PSF of the lens and coupled sensor. This method differs from the previously demonstrated method in that it operates on a realized system. If the lens prescription values are known, a simulated calculation of the PSF may be performed and is recommended if a design under consideration for manufacture is under development. Then the following method may be employed to determine how well the manufactured lens met the design specifications with respect to the PSF. If a lens is being used blind, so to speak i.e. the prescription is unknown, then the following method will work just as well without any prior knowledge of the lens.

As a lens will generally only be used when paired with a sensor, we measure the view of the PSF grid through the lens with the paired sensor. As a single measurement is unlikely to give an accurate measurement useful for all scenes, we shift the point-like grid in 9 subpixel increments. This is to account for the difference in alignment of the optical axis of the lens with the center of a pixel, which in general will not be known about the digital optical system. In order to combine our 9 independent measurements of the PSF grid we apply super-resolution (SR) algorithms in order to achieve a higher-resolution representation of the PSF grid from the view of the sensor. This measurement is then downsampled to the pixel resolution of the sensor in order to allow direct application of the extracted PSFs. A representative result is shown in Figure 4.14, where the paired lens and sensor under test came from a camera on the Asus Transformer Infinity TF700T. Note the similar size, but slightly varying change in shape of the individual PSFs shown. The effect of lens PSF on the final image quality is also demonstrated in Figure 4.15.

Figure 4.15: *Demonstration of the images of a scene as viewed through our simulated lens. The bright points in the input image (left) are all still visible as blurred regions in the output image (right). Points less apparent in the image, near the top, are more visible despite the introduction of blur and the relative contribution of the background in more apparent.*

Once we have an accurate measurement of the isoplanatic, spatially variant PSF of the paired lens and sensor we may perform a correction of the data captured. As the measurement method does not require disconnection of the lens for measurement nor destruction of the system, we may reapply our measurement method regularly in the event that the digital optical system is subjected to extreme conditions such as vibration or changes in pressure or temperature which often effect the alignment of the lens with respect to the individual elements or the sensor or deformation of the lenses themselves depending on the material from which they are made.

A discussion of the methods suitable for correcting the blur effects introduced by the lens is contained in Section 6.3.

4.3.3 Geometric Distortion

Another source of image degradation due to the lens is the geometric distortion. This effect arises from the nonlinear mapping of the 3D light ray incident on the lens surface, across each element of the lens, to the planar surface of the sensor. Wide-angle lenses in particular exhibit this effect, but all lens systems contain some level of distortion, down to the pinhole camera model for which there is a perfect mapping of light to recording medium. Lucchese and Mitra [Lucchese2003] describe the geometric distortion field as representing the internal geometry with regard to the parameters of a mathematical model where the aberration is produced by the lens.

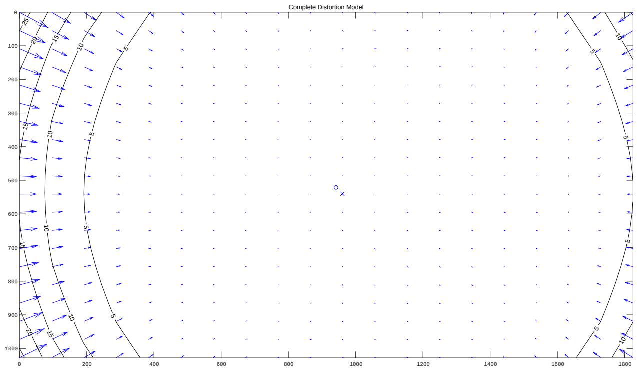
Geometric distortion correction describes the process of removing the effects optical distortion that results from the lens elements. Changes in the objects in images captured by the camera are perceived most easily in observing objects with straight lines being bent by the distortion. This effect is most noticeable in the edges and corners of the images. The geometric distortion field can change over time due to changes in environmental conditions (temperature, pressure) or from physical actions such as vibration or jarring of the lens system causing elements to go out of alignment with respect to one another or the alignment with respect to the sensor changing. Typically the field is considered constant as it does not change significantly from one acquired image to the next, but in practice if significant time has passed or environmental conditions or events have occurred a re-calibration is advised.

In order to correct the geometric distortion field of a lens, the field must first be measured. There are two main approaches to distortion correction:

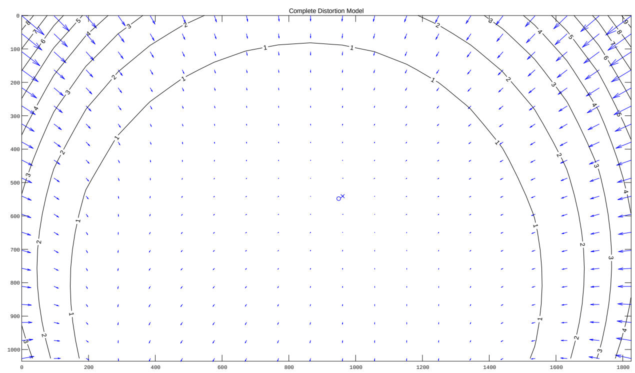
- Calibration prior to data acquisition
- Real-time calculation of distortion parameters as part of the motion estimation step

Altunbasak et al [Altunbasak2003] indicate observing changes in the lens distortion from one sequence to another, in addition to observing a change in the relative distortion seen in images based on distance to targets in the scene. Their proposed approach to include distortion calculations as part of the motion estimation process adds computational complexity and the targets we observe are optically at infinity, so neither of the concerns they address are present in our scenario. We propose performing all possible calibration steps prior to data capture to minimize the processing time, but recognize the need for regular re-calibration operations to account for potential drift in the parameters at a rate not yet determined.

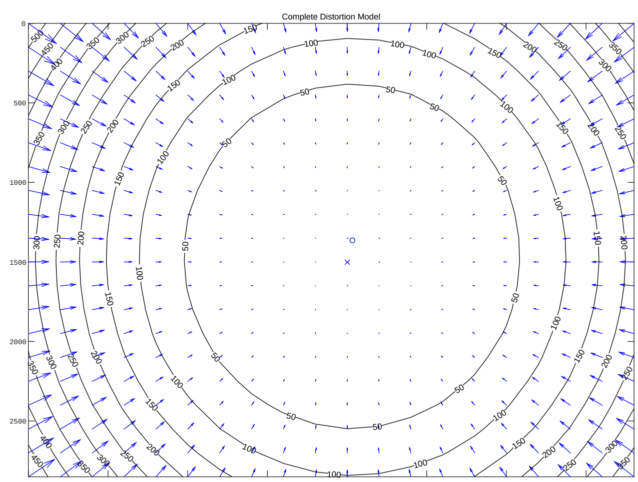
We include geometric correction in order to fix distortion introduced by the lens [Delbracio2012], considering both radial and tangential distortion. The effect of radial distortion is seen in the displacement of image points radially in the image plane [Zhang1999]. Tangential distortion is also referred to as decentering distortion and can be seen in the imperfection of the radial curves seen in Figure 4.16. Tangential distortion occurs when the incident cone of light is projected on the image plane [Zhang1999]. In a perfectly aligned system, there would be no tangential distortion and the radial component alone would dominate the total distortion.



(a)



(b)



(c)

Figure 4.16: *Illustration of the geometric distortion field for three different image acquisition systems. The 'x' near at the center of the figure represents the center of the sensor while the nearby circle represents the optical axis of the lens. (a) Total geometric distortion field of the Galaxy S4. The optical axis of the lens is only slightly translationally misaligned with the center of the sensor, and the rotational misalignment is likewise low, as evidenced by the fairly good balance of the lines of equal distortion seen at the left and right edges of the image. (b) Total geometric distortion of the Asus Transformer Infinity. The translational alignment is particularly good in this lens-sensor combination, but there is significant upward pitch to the rotational alignment as seen by the uneven distortion lines in the upper section of the image. (c) Total geometric distortion of the GoPro Hero 3+ Black. There is slight translational misalignment of the lens and sensor but there appears to be quite excellent alignment of the optical axis perpendicular to the sensor plane.*

Some existing methods such as simple warping can fix the distortion, but may change the scale of the image. The image-warping class of methods will use a model of the distortion with a small set of parameters that often cannot fully describe the complex distortion field of real imaging systems. A method developed for application in the popular Photoshop [PTLens] gives a clear demonstration of the results when using methods that make more or less restrictive assumptions about the complexity of the distortion field.

Camera calibration is introduced in order to calculate the parameters that depend on the lens optics. A popular approach involves the use of a calibration pattern or calibration grid [Dorfmuller2002]. We apply a calibration method that utilizes the data from multiple images of a calibration target to model the distortion and thereby iteratively find the best estimate for parameters of the Brown-Conrady model of lens distortion [Brown1966][Conrady1919] which accounts for higher-dimensionality representations of both the radial and tangential distortions.

This effect further illustrates our concern with the method developed by Liu and Chen [Liu2008] in removing the lens from the system, because the distortion effect will propagate in the PSF measurement and in the images captured after measurement for which we desire to remove the effects of the intrinsic camera blur quantified by the PSF. We considered three methods to correct the distortion: image warping [Gonzalez2008], the Caltech calibration toolbox [Caltech], and MATLAB's built-in toolbox [MATLAB] and selected [Caltech] to calculate the intrinsic parameters based on prior experiments. In order to correct images, three groups of parameters need to be considered: focal length, the focal length of horizontal and vertical pixels; principal points, the points where the principal planes cross the optical axis; and distortion coefficients, the radial and tangential distortion parameters. For color images, we convert to HSI and apply correction to the intensity channel due to the single-channel limitation of the method as implemented.

4.4 Source of Error: Sensor

Digital optical systems suffer from intrinsic sources of error due to their two principal components. This section contains a discussion of the errors inherent from use of a digital sensor with respect to both the effects of sampling and the types of noise which are present in the resulting recorded data.

4.4.1 Errors from Sensor: Sampling

Digital image sensors divide up the light-sensitive area of a device into an array of individual regions called pixels. Although there exist devices, called line-scan cameras which have a single row of pixels, instead of a matrix of them, we focus on devices which have a two-dimensional array of pixels. Each pixel is assumed to have the same response to incident light [Reichenbach1991] and the light-sensitive region of the device is limited by the fill factor, or the percentage of light-sensitive regions versus the total coverage area. The fill factor is of necessity less than 100% due to manufacturing constraints.

4.4.2 Errors from Sensor: Noise

System noise in a CMOS digital image sensor is composed of several categories or sources of noise. Each type can be handled or attenuated using different methods. Herein we discuss these sources, their impact on image quality and certain processing algorithms, and the approaches we applied in order to handle the noise present in the data we collected.

Further details about the modeling of noise sources present in a digital image sensor are covered in Chapter 5.

4.5 Source of Error: Trajectory Disturbance

The trajectory of a sensor's orbit is modeled as a conic section. The departure from the ideal orbital path is one component of what we describe herein as trajectory disturbance. The second potion of what we describe as trajectory disturbance refers to the influence of the atmosphere on the path of the light as it travels from the target to the sensor.

The method employed to apply degradation to the image varies the spatial position of the image data across the scene according to a vector field model with 4 parameters (θ, φ, R, and r – see Figure 4.19) each with their own randomness settings, and an additional smoothness parameter. The purpose of applying this degradation is to represent the effects of the environment when imaging from such a great distance. Typically, the effects of the medium through which the light travels to the sensor may be disregarded in all but the most extreme conditions. In close-up photography the impact of the atmosphere can generally be neglected, but in our situation where a great amount of atmosphere exists between the target and sensor we must account for its contribution in the image acquisition process.

For our purposes, we do not apply any methods to remove or mitigate the effects of trajectory disturbance. We do show that our other processing steps are not significantly burdened by representative degradation from this source, and that we are able to establish a reliable estimate of the target's location despite the presence of the introduced error.

4.5.1 Path Disturbance

Ideal orbital motion can be modeled as a conic section, with defined size and eccentricity [Goncalves2013]. If there were no sources of disturbance of this ideal path, then a perfect knowledge of the sensor's position would be known at any given time and no correction for the introduction of path disturbance forces would be necessary. However, forces exist which influence the sensor to depart from its ideal path and we summarize the major contributors.

4.5.1.1 Celestial Gravitational Fields

The gravitational field of the moon exerts a pull on objects in orbit around the Earth, causing a disturbance in their expected trajectory [Goncalves2013]. Other celestial bodies also influence the path of orbiting objects, such as the sun and, to a lesser extent, other planets in the solar system. Objects with a period greater than 225 minutes (higher than LEO) are more likely to be affected by the gravitation of the sun, moon and planets [Sat1998].

4.5.1.2 Mass Asymmetry

Earth is not a perfect sphere, and the variation in the distribution of mass due to this fact causes variations in the local gravitational field experienced by orbiting sensors. This effect is amplified particularly for objects in low Earth orbit (LEO). The term for this effect on orbiting objects is nodal regression, and means that not only does the object go around the Earth on its orbit, but the orbit itself rotates around the Earth [Sat1998]. One of the effects of the mass asymmetry of the earth is seen in the east-west drift of the orbit, the rate of which depends on the object's altitude and inclination [Weeden2010].

4.5.1.3 Atmospheric Drag

For objects in LEO the effects of the atmosphere are more severe due to its increasing density [Sat1998]. The orbital period limit for defining LEO is generally accepted to be under 225 minutes. Not only does the introduction of drag due to the atmosphere cause for increased use of fuel in order to maintain orbit, but both varying densities of the atmosphere and the introduction of acceleration when thrusters are fired cause unexpected orbital position variation [Weeden2010].

4.5.2 Atmospheric Turbulence

We use the Non-rigid Registration work of [Thirion1996], also known as Demons' Method, a reference to Maxwell's demons [Maxwell1871]. The original demons' algorithm used gradient information from a static reference image to determine the demons' force required to deform the moving image. This may not be efficient, especially when the gradient on the reference image is low [Wang2005]. Conceptually, the diffusing model assumes that local demons at every voxel location are applying invisible forces that push the voxels of the moving image into matching up with the reference (static) image [Cox2007]. In some other studies, the optical flow formula was used to estimate the demons forces either at image feature points or at voxel positions on a grayscale image. In [Wang2005] the passive force is

$$\vec{u} = \frac{(m-s)\vec{\nabla}s}{\left|\vec{\nabla}s\right|^2 + (m-s)^2} \tag{4.15}$$

as in previous formulations, whereas the active force is proposed as

$$\vec{f_m} = \frac{-(s-m)\vec{\nabla}m}{\left|\vec{\nabla}m\right|^2 + (s-m)^2} \tag{4.16}$$

The combined force proposed is then calculated as

$$\vec{f} = \vec{f_m} + \vec{f_s} = \vec{f_m} + \vec{u} \tag{4.17}$$

which is expressed as

$$\vec{f} = (m-s) \times \left(\frac{\vec{\nabla}s}{\left|\vec{\nabla}m\right|^2 + (s-m)^2} + \frac{\vec{\nabla}m}{\left|\vec{\nabla}m\right|^2 + (s-m)^2} \right) \tag{4.18}$$

$\vec{u}$ is the estimated displacement to match a point in the static scene with a point in the moving scene, m and s are moving and static scenes respectively. $\vec{\nabla}\cdot$ is the gradient operator, $(s-m)$ or $(m-s)$ is the differential force between scenes, $\vec{f_m}$ is the active force, $\vec{f_s}$, originally labeled $\vec{u}$, is the passive force, $\vec{f}$ is the total force. Because of this new demons force, the authors claim that the algorithm converges more quickly and requires fewer iterations [Wang2005]. An illustration of the deforming nature as modeled is shown in Figure 4.17.

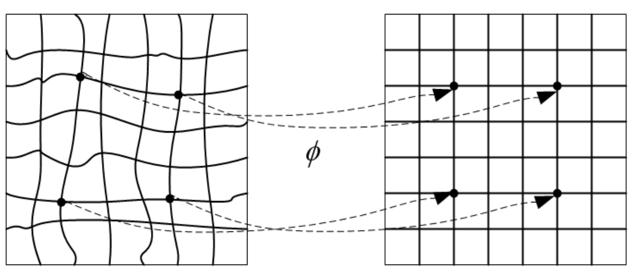


Figure 4.17: *Illustration of the application of [Wang2005] to correct the effect of turbulence on image acquisition. The effects of a large volume of atmosphere between the target and the imaging system allows the opportunity for turbulence in the atmosphere to nonuniformly distort the location of information within the FOV. The illustration indicates calculation of the demon's force to correct for the turbulence effects whereas we use the same method in order to representatively introduce turbulence effects in our simulated imagery in order to ensure we are able to accurately recover the position of our desired target despite the effects of atmospheric turbulence.*

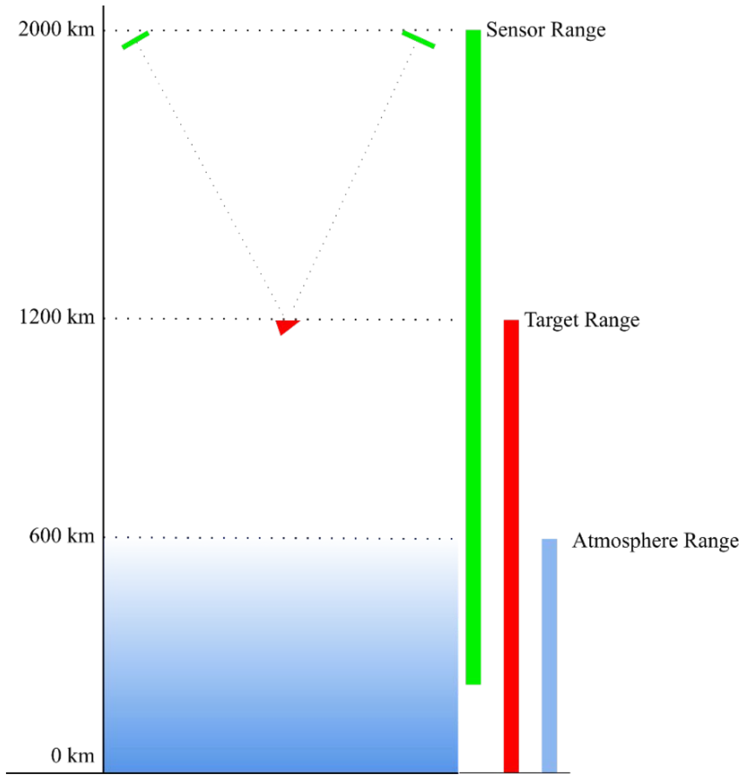


Figure 4.18: *Demonstration of altitude range in considering the effects in our simulated imagery. The target altitude can range from a complete obscuration by the atmosphere to an area where large distances without atmosphere to obscure the imaging process. Though the potential altitude of the sensor could place it below the target altitude, we only consider cases where the target is below the sensors.*

In order to accurately determine the severity of turbulence we're likely to encounter, we must consider the changing imaging conditions with respect to the altitude range of the targets and sensors relative to the density of the atmosphere. This is depicted in Figure 4.18. The target altitude ranges from sea level to 1200 km, the sensors can have a height of 250-2000 km, and the generally accepted range of the atmosphere gives a limit of 600 km. For the atmosphere, the density is not uniform within that range but the impact of the turbulence effects diminishes as the volume and density of the atmosphere drops as the target altitude increases. When the target crests the layer of atmosphere, there still exists a large distance between the target and sensor, but there will be no contribution of atmospheric turbulence. We therefore evaluate the influence of atmospheric turbulence on the range of no turbulence for the period of time when there is no influence to a high degree of turbulence when the greatest amount of atmosphere exists between the target and sensor.

4.5.3 Implementation

Our random geometric deformation software was developed based on demons' method [Maxwell1871]. To achieve synthetic geometric deformation a translation vector is applied to each pixel of the reference image. This maps the original location of the data of the scene into a turbulent representation as though the light from the target had passed through the atmosphere with its random perturbations at each location of the image. The output image is then smoothed with a Gaussian Filter. The creation of the individual random vectors requires generation of a number of parameters. The organization of these parameters is demonstrated in Figure 4.19. The point p_{ij} is remapped to the location p'_{ij} as shown.

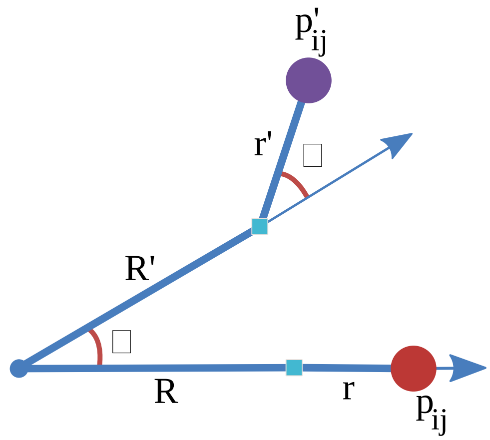


Figure 4.19: *Illustration of the random vector translation creation. The corresponding* $\boldsymbol{R}$ *and* $\boldsymbol{R}'$*, on the range* $[\mathbf{0}, \mathbf{30}]$ *pixels, and* $\boldsymbol{r}$ *and* $\boldsymbol{r}'$ *vectors, on the range* $[\mathbf{0}, \mathbf{1}]$ *pixel, are of the same random length, respectively.* $\boldsymbol{\theta}$ *and* $\boldsymbol{\varphi}$ *are independently randomly distributed on the range* $[\mathbf{0}, \mathbf{180}]$ *degrees. There is an additional randomness value, selectable for each of the* $\boldsymbol{R}$, $\boldsymbol{r}$, $\boldsymbol{\theta}$, $\boldsymbol{\varphi}$ *parameters which controls the randomness of the parameter. In this way, we are able to simulate greater or less intensity in the atmospheric turbulence effects. As we are viewing a target from outside the atmosphere looking in, the effects will be diminished as compared to a terrestrial station looking upward, but the effect is still there.*

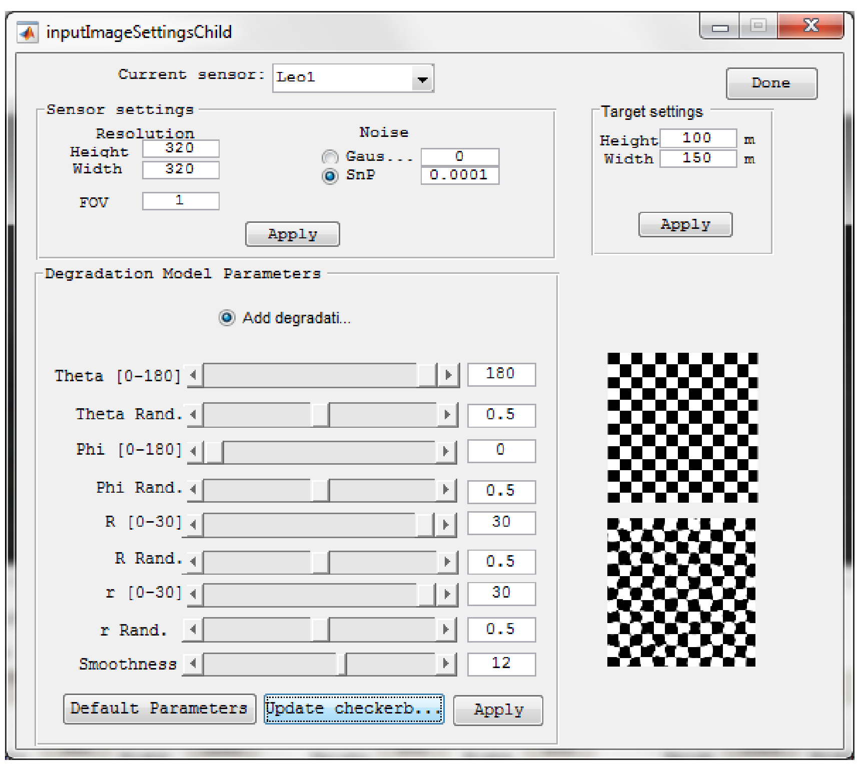


Figure 4.20: *Settings for the target, sensor, and trajectory disturbance levels are adjusted here. Values may be individually manipulated for each sensor and the main parameter groups are resolution, noise, FOV, and trajectory disturbance. Noise models include Gaussian and Salt and Pepper (SnP) and the size of the target centered about the given position may be controlled. Trajectory disturbance (labeled Degradation Model Parameters) effects may be seen in the distorted checkerboard seen in the bottom right.*

The GUI developed for application of synthetic turbulence illustrated in Figure 4.19 is shown in Figure 4.20. This GUI not only contains settings for controlling the severity of induced atmospheric turbulence effects, seen in the region marked ‘Degradation Model Parameters’, but also includes other settings relevant for our simulation. For instance, under ‘Sensor settings’ we are able to control the resolution of the sensor, the FOV of the lens, and the noise type and level of the sensor. Under ‘Target settings’, we are able to control the size of the target centered along the interpolated path. The distance between the target and sensor is controlled elsewhere in the description of the sensor’s orbital path.

Development of this software to simulate the effects of atmospheric turbulence allows us to create imagery as representative as possible of the type of data we would see from the sensors used. Though we do have the demons’ method software implemented, we do not present the use of this method to remove the effects of turbulence as part of our work but rather show that for representative situations and representative data collected we are able to generate a relatively accurate estimate of the target position and an even more accurate fusion of these estimates to improve our estimate despite the sources of error in our image-based approach. In this way we need not burden the hardware to apply the necessary processing that these methods require as our desired output is the location of the target rather than pleasant images of it. We are able to accurately recover this information in spite of the error sources present.

5 Noise Mitigation

As the image restoration methods we will use are ill-posed inverse operations, noise is a significant concern with regards to the successful application of the selected approaches. This comes also with an application where the image acquired of the scene is contaminated with high levels of noise. In this situation we must select methods which handle high amounts of noise well or apply pre- or post-acquisition processes to reduce the effective noise level in the images before further processing can take place.

In this chapter we investigate three potential approaches for achieving this goal: altering the image acquisition hardware or process to achieve higher SNR, applying noise reduction methods which will increase the effective SNR of a given image without significantly degrading our ability to locate vital information about the target, and determining which methods (SR, deconvolution, tracking, etc.) are most robust to increasing levels of noise. A comparison of the results in these three areas will drive a recommended best approach for achieving a high level of success without excessive cost in equipment changes or in processing time.

5.1 Noise Modeling

Image noise is a random signal which manifests as an undesirable artifact due primarily to various parts of the sensor hardware, but can be intensified by various environmental conditions. In this chapter, we identify the major sources or noise and the factors which can increase their intensity, discuss the types of noise distributions, and demonstrate method for removing or diminishing their effects in the image, thereby increasing the effective signal-to-noise ratio (SNR).

5.1.1 Sources of Noise

Noise in an image is primarily introduced by the sensor, but from multiple effects and increases based on certain conditions. We explain each in more detail below.

5.1.1.1 Fixed Pattern Noise

Fixed pattern noise (FPN) is most often caused by manufacturing errors or physical damage to the pixel, but is also influenced by the temperature of the sensor. Though this type of noise can change over the life of the sensor, it will typically remain constant from one image acquisition to the next. Because of this property, we are able to perform a calibration measurement of the fixed noise pattern and remove this type of noise from our images. This noise removal process is called flat-field correction (FFC) which addresses the non-uniform response of the sensor to complete darkness (called dark signal non-uniformity, DSNU) and the response to even illumination across the field of the sensor (photo response non-uniformity, PSNU).

5.1.1.2 Banding Noise

The source of this type of noise is most commonly the readout circuitry of the sensor, but is influenced by the brightness of the scene. Therefore, it is highly sensor dependent and will manifest in different ways depending on the sensor used and the scene conditions during acquisition. Some common causes are when there is damage or a manufacturing error in the amplification circuitry. When the pixel values are read out from the sensor, each row (or

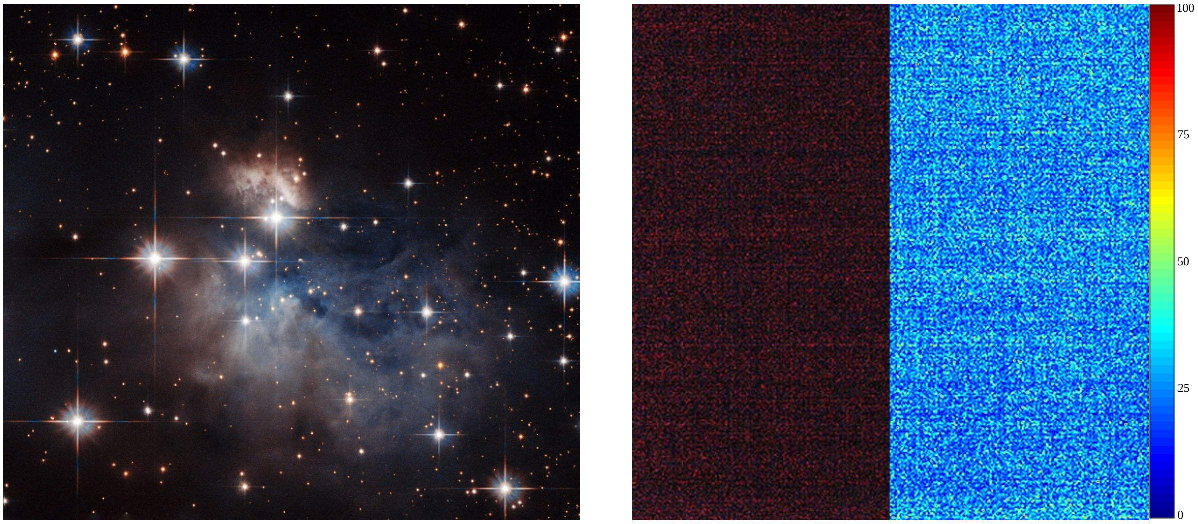


Photo credits: NASA [Hille2015] — (a)

Photo credits: Jon Rista [RistaDec2013] — (b)

Figure 5.1: *Examples of horizontal and vertical banding noise (HVBN). In (a) we see horizontal and vertical bands surrounding the brighter stars. In (b) we see an example of both horizontal and vertical banding effects as a part of FPN in an image taken with a very dark scene. The ISO sensitivity or gain of the sensor is increased, making the effect of the noise significantly amplified. The right half of the image has been pseudocolored for better visibility.*

column, depending on design) is read one at a time, and the whole array is shifted so that the next row (or column) may be read. If the amplifier for a given row or column is damaged, the gain may be higher compared to the rest of the image, creating a horizontal or vertical band. This is why for higher ISO settings or higher gain settings, or in low-light situations for auto exposure – auto gain (AEAG), will amplify the effect. This is depicted in Figure 5.1.

Some sensors have built-in mechanisms to combat HVBN. One way is to eliminate the bias offset, which will halve the banding effect at the expense of detail in the image. Another way to reduce HVBN is to perform the ADC operation within the same sensor die, since digital signals are less sensitive to the introduction of this type of error compared to analog signals.

5.1.1.3 Random Noise

Many sources of noise corrupt the data obtained from a CMOS sensor. These include shot noise, generation (or recombination) noise, and popcorn noise but thermal and flicker noise are the most dominant. [Lundberg2002]. Noise in images from a CMOS sensor may be broadly classified as white noise and colored noise and comes from two physical locations on the device: the image pixel collecting light from the scene and the analog circuit transferring the information for readout [Hoshino2007]. Hoshino and Nishimura [Hoshino2007] suggest computing the time-dependent pixel autocorrelation function (ACF) in order to detect and remove white noise from images captured from CMOS sensors.

Table 5.1: *Relationship of the physical location where noise types manifest and grouping according to whether the noise changes randomly over time for each image acquisition (white) or is stable to be measured during a calibration step and removed for subsequent captures.*

	Image Pixel	Analog Circuit
White	Reset 1/f Dark shot Photon shot Amplifier	Amplifier 1/f
Colored	Offset Dark Condensing Gm	Offset Condensing Gm

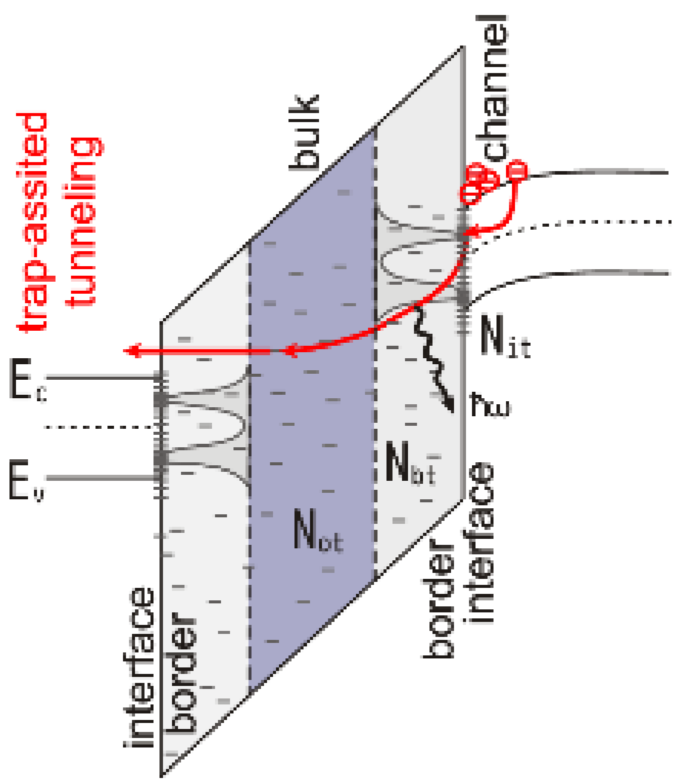


Figure 5.2: *Interface, border and oxide bulk traps. Trap-assisted tunneling is represented by red arrows. Image credits [Starkov2013]*

Table 5.2: *Noise source comparison, each with their mathematical formulation and a description of the source and its effects. The sources with the greatest contribution are thermal and flicker noise.*

Shot Noise	$I_{ns}^2 = 2qI$	Caused by current not smoothly flowing across a junction, individual electrons arrive at random times. Broadband white noise that increases with average current. 1mA current has ~2pA/root(Hz) shot noise. I_{ns}^2 is spectral density of shot noise. I is junction current. [Lundberg2002]
Generation/Recombination Noise	$R_n^2 = \frac{\sigma^2}{N^2} \cdot \frac{4\tau R^2}{1+\omega^2\tau^2}$ Lorentzian spectrum	Trapping centers (Figure 5.2) in the bulk of the device can cause G/R noise. This is in addition to flicker noise as G/R noise results from oxide traps within the bulk rather than across the gate, as shown [Lundberg2002]. The resulting fluctuation of carriers causes a fluctuation in the resistance of the device and the spectral density of the resistance fluctuation is modeled as R_n^2. [Hooge1994]
Popcorn Noise		Also called burst or random telegraph signal (RTS) noise. Caused by capture and emission of a channel carrier resulting in discrete modulation of the channel current [Lundberg2002].
Thermal Noise	$V_{nt}^2 = 4kTR$ Gaussian [Hoshino2007]	The voltage fluctuation due to Brownian motion of electrons in a resistive medium. Broadband white noise, increases with resistance and temperature. V_{nt}^2 is the thermal noise across a resistor. A 50ohm resistor has ~1nV/root(Hz) [Lundberg2002]
kT/C Noise	$v_{no} = \sqrt{V_{nt}^2 \Delta f} = \sqrt{\frac{kT}{C}}$ $\Delta f = \frac{1}{2\pi}\int_0^\infty \frac{d\omega}{1+(\omega RC)^2} = \frac{1}{2\pi}\frac{\pi}{2RC} = \frac{1}{4RC}$ $V_{nt}^2 = 4kTR$	Not a fundamental noise source, describes thermal noise in the presence of a filter capacitor [Lundberg2002]. Discussion is included as the reduction of noise by applying a low pass filter of this design is common practice and knowing how it effects the presence of noise in the data is important. This effect is seen not only when a low pass filter is explicitly designed into the device but also due to the effect of the channel resistance and drain capacitance of a MOSFET. In consideration of the noise present in a system, either the thermal noise or the kT/C noise are considered since they stem from the same fundamental source [Lundberg2002].
Flicker Noise (1/f Noise)	$\frac{1}{f^\alpha}$	Dominates the noise spectrum at low frequency. $\alpha = 1 \pm 0.2$ [Lundberg2002] Oxide traps (Figure 5.2) result from imperfections in the construction of the transistor material, flicker noise resulting from charge tunneling across the gate [Jayaraman1989]. Two main theories strive to explain the source, [McWhorter1957] and [Hooge1969]

Colored noise is FPN and thus can be measured and removed due to its time-invariant properties by subtracting the calibrated "dark" image from the data captured of the desired scene [Hoshino2007]. White noise, however, is time-dependent and differs with each acquired image. The noise sources classified in Table 5.1 are detailed in Table 5.2.

5.1.1 Noise Distributions

Methods for mitigating or removing noise in images can depend on the distribution of the noise spectrum. In this section we review the types of noise distributions commonly seen during the image acquisition process.

5.1.1.1 Gaussian (Normal)

The Gaussian distribution shown in Figure 5.3 is the most commonly encountered type in digital imaging due in part to its presence due to thermal effects. The SNR will degrade as the amplitude of the noise increases. This can come about due to increased on-board resistance from the construction of the specific device and heat from ambient environmental conditions. The mathematical form of the Gaussian distribution is shown below [Scherzer2009].

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} exp\left(-\frac{|x-\bar{x}|^2}{2\sigma^2}\right), \sigma > 0 \qquad (5.1)$$

The Gaussian distribution is a continuous probability distribution that expresses the expected value of a random process with a known mean value $\bar{x}$ and standard deviation σ. Knowledge of the standard deviation allows for worst-case predictions of the influence of the noise as for a Gaussian distribution the values bounded within two standard deviations of the mean will contain 95% of the range of values while accounting for three standard deviations of the mean will contain 99.7% of the possible values. Thus a measurement of a device with respect to the noise from the circuitry at a variety of temperatures and lighting conditions will give bounds on the contribution of noise to the SNR of the image. Nasrollahi and Moeslund [Nasrollahi2014] note that the Laplacian distribution has been shown to be more accurate in representing image noise than using a Gaussian as is typical in the literature. However, as the majority of the literature focuses on the Gaussian both in modeling and mitigation, we will not address the Laplacian distribution in our experiments.

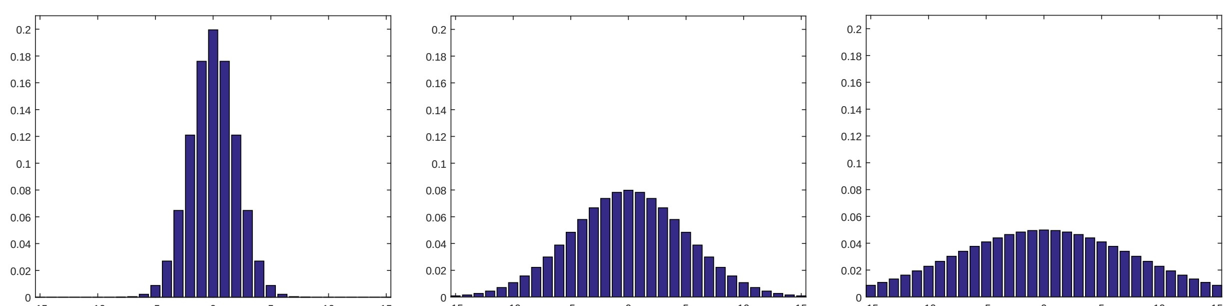


Figure 5.3: *Graphical depictions of the Gaussian distribution function. The images show values of* $\bar{\boldsymbol{x}} = \mathbf{0}$ *for all distributions and* $\boldsymbol{\sigma} = \mathbf{2}$ *(left),* $\boldsymbol{\sigma} = \mathbf{5}$ *(middle), and* $\boldsymbol{\sigma} = \mathbf{8}$ *(right). The specific values of* $\bar{\boldsymbol{x}}$ *and* $\boldsymbol{\sigma}$ *are unique to a particular device and will fluctuate with environmental conditions at time of capture.*

5.1.1.2 Poisson

The primary source of noise in an image that follows a Poisson distribution is photon shot noise. This type of noise, shown in Figure 5.4, dominates the darker portions of the image and is due to the nature of entering photons from a source. In higher intensity regions the Poisson distribution looks very much like a Gaussian distribution. The mathematical form of the Poisson distribution is shown below [Scherzer2009].

$$P(k) = \frac{\lambda^k}{k!} exp(-\lambda), \qquad \lambda \geq 0, k \in \mathbb{N} \cup \{0\} \tag{5.2}$$

The Poisson distribution is a discrete probability distribution that expresses the probability of a given number of independent event occurring within a fixed interval of time, given knowledge of a known average. Thus as the collection of photons on a given pixel is inherently discrete, the Poisson distribution describes this part of the image formation process.

5.1.1.3 Salt and Pepper

The Salt and Pepper distribution demonstrated in Figure 5.5 has also been called a 'heavy-tailed' distribution, in that the range of values seen with this type of noise will occur at the extreme ends of the pixel range as suggested by its name. This type of noise generally is a result of data transmission and coding errors by the hardware or manufacturing defects in the imaging surface. Exposure to ambient radiation such as cosmic rays can increase the effect. Typically this type of noise is quantified by the percentage of corrupted pixels and therefore has no mathematical formula. An example image is shown below.

5.1.1.4 Comparison of Noise

In order to visually see the way that different types of noise can corrupt an image, a demonstration of each noise type is shown in Figure 5.6 below.

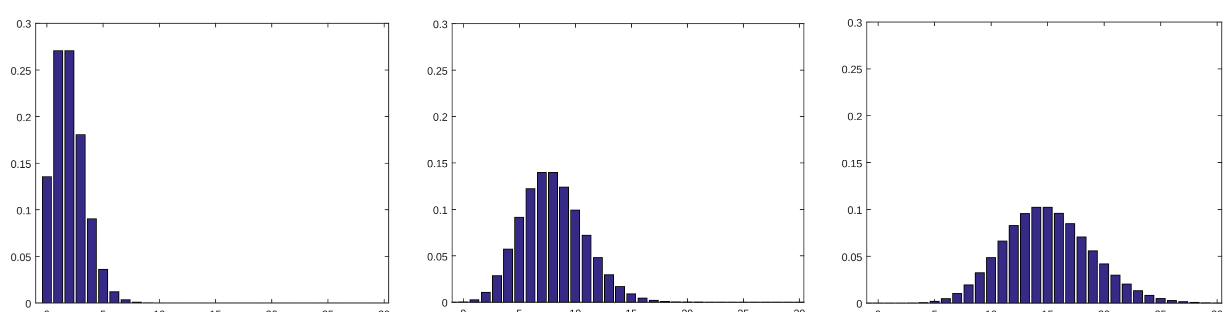


Figure 5.4: *Graphical depictions of the Poisson distribution function. The images show values of* $\boldsymbol{\lambda = 2}$ *(left),* $\boldsymbol{\lambda = 8}$ *(middle), and* $\boldsymbol{\lambda = 15}$ *(right). Note how as* $\boldsymbol{\lambda}$ *increases the distribution more closely resembles a Gaussian function.*

Figure 5.5: *Example of salt-and-pepper noise. The noise manifests as pixel values being set to the minimum (black) or maximum (white) values. Can occur randomly due to environmental conditions or more predictably from manufacturing defects in the sensor. For this image, 5% of the pixels were selected to have noise.*

Figure 5.6: *Comparison of images corrupted by differing noise types. Gaussian (left), Poisson (middle), and Salt and Pepper (right) are all forms of noise commonly encountered in digital imaging. The approaches for removing noise differ and depend on the type of noise corrupting the image.*

5.2 Methods of Mitigation

The challenge for any noise removal method is one of eliminating the undesirable noise signal while simultaneously preserving the fine details in an image. This is still an open problem in the field and while no perfect solution exists, there are a number of effective approaches for improving the SNR in images corrupted by noise.

5.2.1 Gaussian and Poisson: Spatial Filtering

Filtering noise from an image is an inherently low-pass operation, but many methods exist in order to improve the desired edges of the image which are also a high-frequency component. Each pixel in an image has what is known as a neighborhood, or the surrounding pixels in close proximity to the pixel of interest. In order to filter out noise a basic approach is to calculate the combination of pixel values in the neighborhood in order to determine the value of the output image. Shown below are the input image corrupted by noise and the filtered output from the application of two different methods.

The linear average filter employed operates in the spatial domain and utilizes a Gaussian profile in the kernel. The structure of the kernel is shown in Figure 5.7. The disadvantage of this simpler filter type is the selection of filter coefficients is constant across the entire image. This often can result in the kind of oversmoothing as seen in Figure 5.8(b). The adaptive filter developed by Luisier et al [Luisier2011] is an improvement to this approach which utilizes local pixel statistics that are used as part of the filtering process. More details of the scene can be preserved as seen in Figure 5.8(c).

The equations for calculating the local mean and variance of the pixel's neighborhood are:

$$\mu = \frac{1}{NM} \sum_{n_1, n_2 \in \eta} I(n_1, n_2) \tag{5.3}$$

and

$$\sigma^2 = \frac{1}{NM} \sum_{n_1, n_2 \in \eta} I^2(n_1, n_2) - \mu^2 \tag{5.4}$$

where η is the N-by-M local neighborhood of each pixel in the image I. Using these estimates, the adaptive filter employed creates a local filter:

$$b(n_1, n_2) = \mu + \frac{\sigma^2 - v^2}{\sigma^2} (I(n_1, n_2) - \mu) \tag{5.5}$$

where v^2 is the noise variance. If not known, we can use the average of all the estimated local variances.

$$\begin{bmatrix} f & e & d & e & f \\ e & c & b & c & e \\ d & b & a & b & d \\ e & c & b & c & e \\ f & e & d & e & f \end{bmatrix}$$

Figure 5.7: *Form of a 5x5 Gaussian linear spatial average filter. All the elements sum to 1 and are distributed according to a Gaussian profile. The values operate as a mask over the neighboring pixels and the output value of the central input pixel is the weighted sum of the input pixels scaled according to the coefficients in the filter kernel.*

(a)

(b)

(c)

Figure 5.8: *Comparison of images corrupted by differing noise types. Input image corrupted by Gaussian noise (a), output filtered using a Gaussian average filter with 15x15 neighborhood (b), and output filtered with a locally adaptive filter (c). The average filter uses the same parameters for the entire image while the adaptive approach [Luisier2011] utilizes statistics of the local pixel neighborhood in calculating the resulting pixel values. The uniform regions of the image retain similar amounts of noise, but notice the improved edge content around the face and tripod legs.*

5.2.2 Gaussian and Poisson: Synthetic Extended Exposure

An alternative approach to removing noise in images involves a modification of the capture process. Since the white noise in the images is different from one capture to the next, by capturing a sufficient number of images under otherwise identical capture conditions we are able to recover a result with diminished noise by in effect extending the exposure time to let the noise signal average out while retaining the desired signal of the scene.

When the images obtained contain a large amount of noise, we can use multiple motionless acquisitions of the scene in order to enhance the SNR. In Figure 5.9 we show a comparison of the resultant image when an increasing number of images are used in the image stack. The noise statistics remain the same but each image is unique. Because it may be difficult to see the quality change from image to image as the number of samples increases, we include a chart comparing these images enhanced using the described method by evaluating them compared to the original source image.

As demonstrated in Figure 5.10, the greatest increase in improvement of image quality is seen early on according to both metrics, it becomes a tradeoff based on the particular application. If accurate information is of paramount importance, then a capturing a large sequence of images may be done at the expense of a longer acquisition time. If a tradeoff in quality vs. acquisition time is needed, or if objects in the scene move relatively quickly, then a short burst of even as few as 10 images can significantly improve the quality of the result. Another possible approach that we have investigated is the use of multiple cameras arranged in a grid close together with all cameras operating simultaneously so that even if the scene contains objects of interest with a high relative speed there is a time correlation in the date from each camera and unique samples of the scene are still available for processing. There is a different class of methods called super-resolution which handle the offset from using multiple cameras.

5.2.3 Salt and Pepper

Not all types of noise are best served with the same mitigation approach. Figure 5.11 demonstrates this for salt and pepper noise. An understanding of the underlying sources of noise is important for designing appropriate mitigation strategies. The most common approach to addressing salt and pepper noise is an application of the median filter. The neighborhood size for constructing a candidate value for the output pixel is set, then the values are sorted and the median is assigned to the pixel in the output image. If the area is largely uniform, then little or no change will occur, if however the source pixel is corrupted by noise, then a more appropriate output pixel value will be assigned as selected from the neighborhood by the filter. This results in an effective removal of noise while preserving more edges and detail in the image as compared to a linear averaging operation.

5.3 Conclusions

Noise corruption of real images is an inevitable challenge that must be addressed in order to make accurate measurements from image data. Particularly in high-noise environments, the contribution of noise relative to the signal level can be such that the useful information in a single image can be lost beneath the noise floor. Methods exist to address the types of noise we tackle, but at a cost in processing time and in image quality. The best situation is less noise in recorded images, where reduction can take place by improving the hardware or shielding of the acquisition device. Modification of the image capture process rather than hardware alteration is also promising, but at a cost of increased acquisition time, or if a multi-aperture approach is employed, at a cost of marginal modification to the optics side of the setup, coupled with adapted processing to take advantage of the richer data. If no modification is made to the acquisition process or hardware, computational methods do offer marginal improvement to the SNR but this is the least desirable situation to be in with respect to resulting image quality vs. processing time.

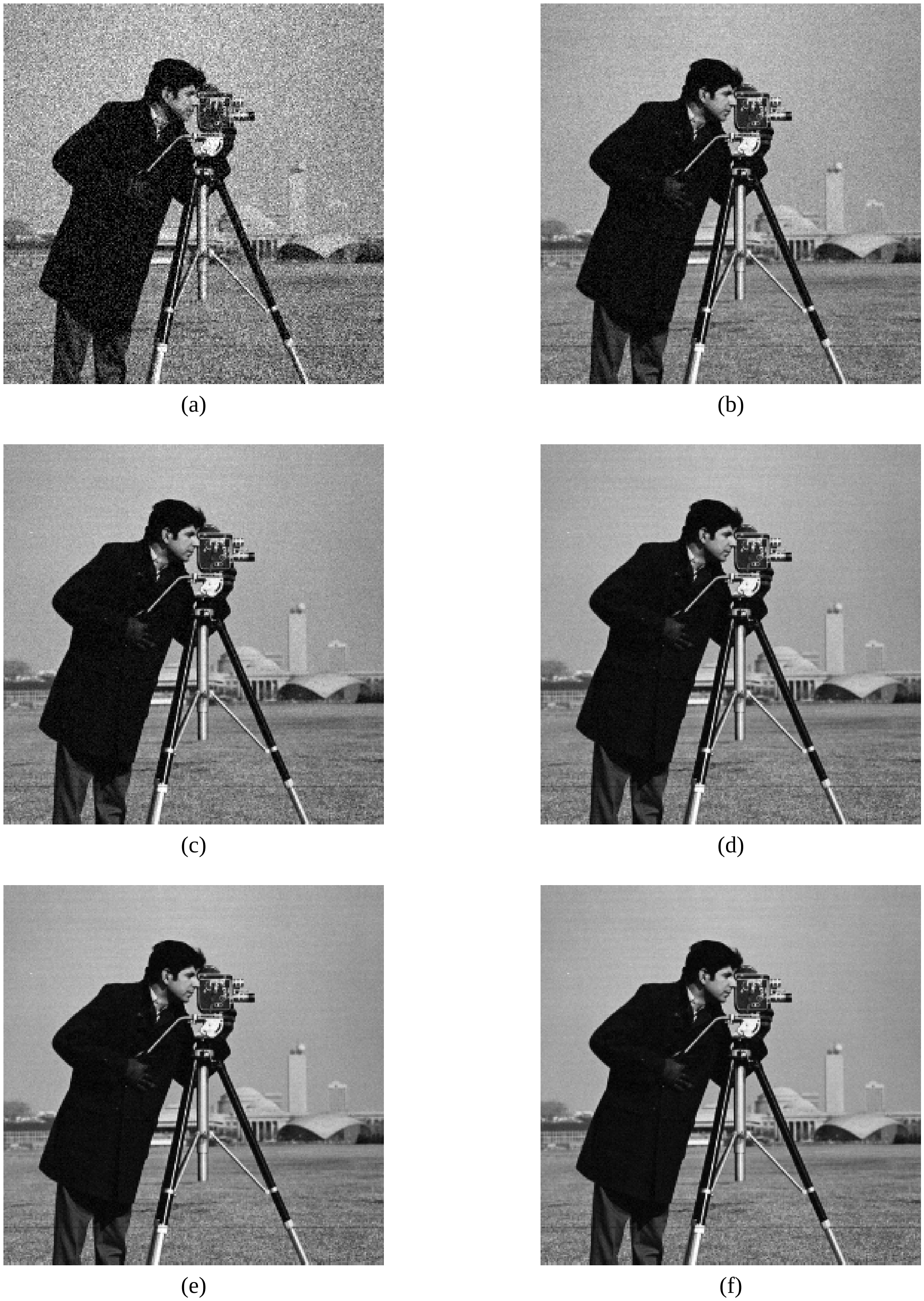
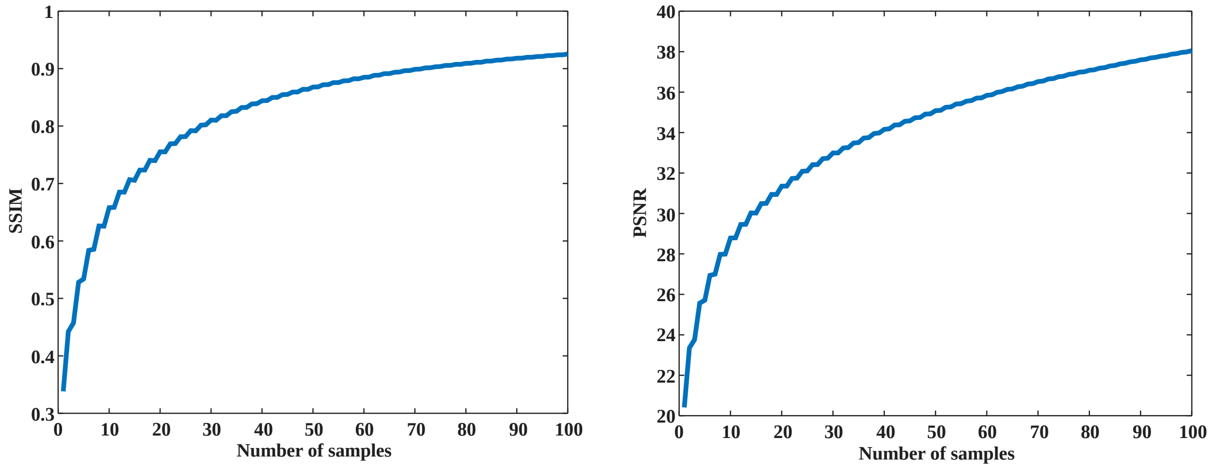


Figure 5.9: *Comparison of the impact of increasing the number of images combined on the quality of the result. Despite a very high noise level in the original images, a near-perfect recovery of the scene is possible due to the independent nature of the individual samples of the scene. The number of images averaged per result shown is: one (a), ten (b), 25 (c), 50 (d), 75 (e), and 100 (f).*



(a) (b)

Figure 5.10: *Comparison of the image quality increase that results from increasing the number of images combined. For both the structural similarity image metric (SSIM) shown in (a) and the peak SNR (PSNR) shown in (b) there is a clear correlation in increasing the number of images and an increase in the quality of the image obtained.*

(a) (b) (c)

Figure 5.11: *Comparison of differences in noise removal approach for different noise types. The image in (a) is corrupted by salt and pepper noise with 5% of the pixels affected. Application of a Gaussian linear averaging filter (b) smooths the effect of the noise but still preserves artifacts of the corrupted information. Utilizing a median filter with a 3x3 neighborhood removes nearly all of the detrimental effects.*

The recommendation of the authors with respect to design of an acquisition system involves utilizing a multi-aperture approach with sufficient shielding of the electronics and selection of reasonably low-noise components in order to balance the cost in procurement, acquisition time, and computation time with respect to the resulting image quality available to use in tracking the position of a target of interest. A complete discussion on the influence of varying noise levels on system performance is contained in Chapter 8.

6 Super Resolution

The aim of super resolution (SR) algorithms is to reconstruct a high-resolution (HR) representation of a scene from multiple low-resolution (LR) observations. This class of methods has a number of applications, and it is important to note that our conclusions are based on the results which most benefit our application. A basic overview of the concept of SR is shown in Figure 6.1.

6.1 Introduction

A perfect digital image acquisition system would capture all information of a continuous HR scene and allow perfect knowledge of the scene within the domain of the sensor. Realizable digital image acquisition systems have fundamental limitations in their ability to capture all information about the scene of interest. These limitations include a limited frequency spectrum, sampling accuracy determined by the number of sensor pixels, finite frame acquisition time, and other limitations arising from the physical implementation of the device. SR algorithms mitigate the effect of limited sampling accuracy by application of mathematical methods to incorporate multiple measurements of the scene to reconstruct an enhanced representation of the scene at higher resolution than otherwise would be possible. It is possible to image beyond the diffraction limit using SR when we cannot achieve this with conventional imaging systems.

Applications for SR image enhancement include satellite and aerial imaging, medical image processing, facial image enhancement, automated road sign and license plate reading, iris recognition, fingerprint image enhancement, high-dynamic range (HDR) imaging, and many other areas [Nasrollahi2014]. Each of these areas will have specialized needs based on the type of data available. Considerations include the available source resolution, desired SR factor, noise level and distribution, blur type and level, presence and degree of motion blur, scale change of the target of interest in the source images, type and level of inter-frame motion, and motion/occlusion of targets with respect to both camera motion and with respect to the scene background. Data acquisition conditions for which many of these considerations may be controlled will simplify the approach needed and improve the likelihood of a quality result, whereas general acquisition conditions in which no guarantees can be made will increase complexity, processing time, and decrease likelihood of a successful result.

SR has a place in any application which operates on images used to gather data about a scene. These applications include microscopy, astronomic and terrestrial telescope, and photographic areas. Conventional light microscopy is limited in resolution by diffraction, as first noted by Abbe [Abbe1873] to reach a maximum of approximately 250nm. SR astronomy often encounters the barriers of noise due to low illumination and blur when long exposure is used to overcome low light levels [Bauer2011]. Both ground-based astronomical and terrestrial telescope domains encounter turbulence from the atmosphere. Typically the effect (known as astronomical seeing) is mitigated for the astronomical case by capturing data when the seeing effect is low. Terrestrial telescope applications must address this effect due to the large volume of atmosphere between the scene and imaging system. The ‘lucky imaging’ approach was developed to address turbulence in this application, which operates by taking many images of the scene of interest and selecting regions of the image in sharp contrast for which ‘atmospheric lensing’ enables a nearly diffraction limited fusion result [Mackay2013]. Due to the nature of the solution, only static scenes can be imaged successfully using this approach, even then only diffraction limited images are possible. The other applications of SR fall under the photographic designation. These include medical imaging, surveillance, face recognition, and a host of other applications.

Consideration of each particular application is important, as the relative impact of the factors influencing the success of an SR algorithm vary and selection of an appropriate algorithm to match these conditions is vital. Our application most closely matches the telescopy variety and as such, the consideration of a low-light scene, high noise levels, and

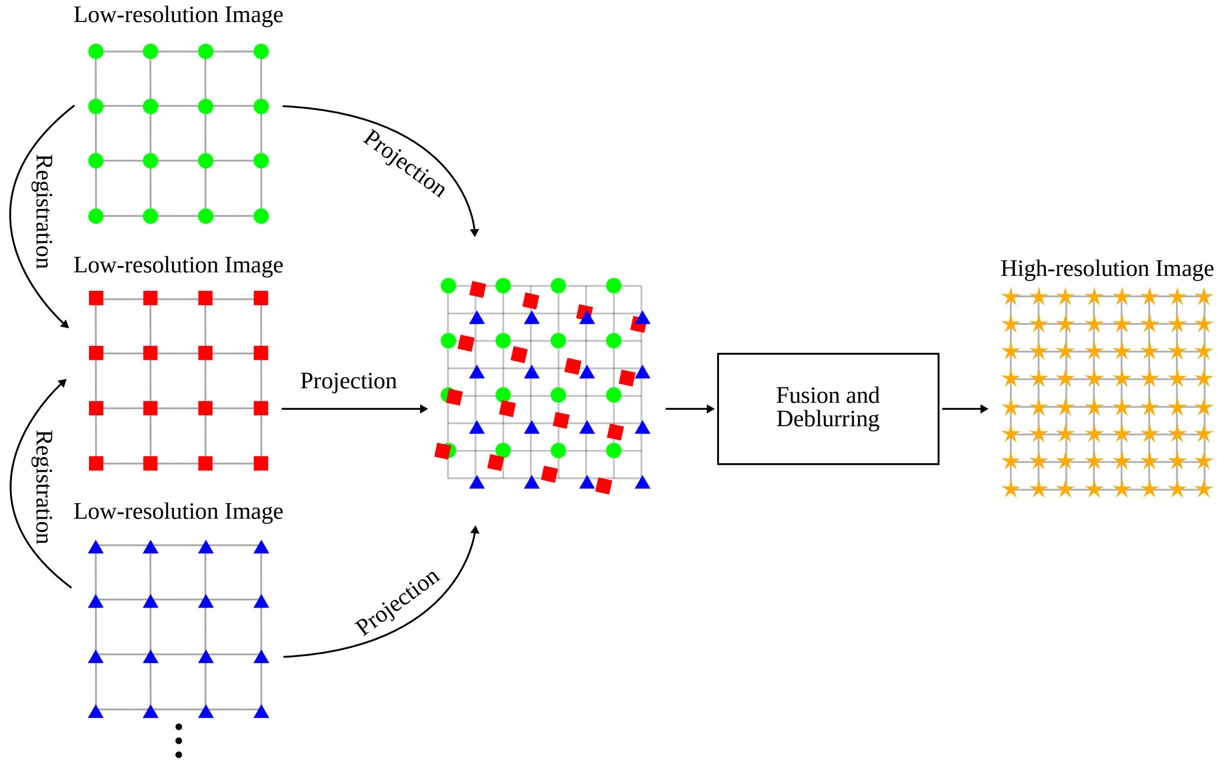


Figure 6.1: *Conceptual illustration of SR. LR images of a scene are registered (the motion between frames is calculated to represent the relationships of the samples of the HR scene) and aligned to the selected HR grid. SR algorithms then fuse these samples, some include deblurring, into a HR representation of the scene. Reproduced from [Tian2011].*

atmospheric turbulence are of importance, as well as the necessity of relatively low processing time in order to increase total system stability. Also of importance is consideration of the effect of the point-spread function (PSF) in limiting effective resolution. Many authors have proposed SR methods which also address the effect of blur on image quality [Borman2004][Tanaka2005][Sroubek2008][Bai2010]. We also include consideration of the system PSF as part of our approach and further include consideration of atmospheric turbulence [Droege2012]. Although Lin and Shum [Lin2004] find that the limits on the performance of SR algorithms do not depend heavily on the system PSF, as our application seeks to accurately locate the coordinates of a target in the scene we must account for all factors which improve the metrological accuracy of the image data.

We have also considered the impact of modified hardware and other novel image acquisition systems. Hardie et al [Hardie2015] determined that a reduced fill factor and selection of a circular, rather than rectangular active pixel area is one possible modification of the hardware that improves the success of SR algorithms. One concern with SR, particularly for images from quickly changing scenes, is that significant changes from one frame to the next can cause a reduction of accuracy in the motion estimation and SR steps of the process. Finite image integration time is closely tied to the available light from the scene and sensitivity of the sensor, and too long of an exposure can introduce motion blur into the captured images. Too much time between frames can limit the available number of frames, and significant perspective changes can limit the usable data available for processing. One approach we developed involves development of a multi-aperture system such that the inter-frame time is negligible and the geometric relationship between the frames is known and only scales with respect to object depth. In our application, objects in the scene are at optical infinity so the need for accurate registration drops as the process becomes a calibration rather than a necessity at each time step.

6.1.1 Conceptual Overview

Multiple LR observations of a scene may be combined to create a HR representation of the scene, even beyond the limitation established by the diffraction limit [Born1999]. SR is a subset of the image fusion class of methods, where algorithms are specifically applied to enhance the resolution of 2D images. Some methods use a dual approach where information about the PSF of the system used to acquire the images is incorporated as part of the enhancement process. The basic idea behind SR is to combine the non-redundant information contained in multiple LR frames to generate a HR image [Milanfar2010].

SR should not be confused with the related methods of interpolation or upscaling. These methods are applied with only a single source image. Interpolation uses simpler mathematical models to interpolate unknown pixel values from known samples in a HR grid. Upscaling, also commonly called single image super resolution (SISR), apply more complex approaches to improving the resolution of a single input image, often employing information from a database of low/high resolution image patches. Nasrollahi and Moeslund [Nasrollahi2014] also note the potential for confusion with image restoration and image rendering techniques. SR algorithms can improve both the quality and resolution of an image with respect to its source data while image restoration techniques such as deconvolution or sharpening operations can improve the quality of an image while the result retains the resolution of the source data. Some SR algorithms also include a deconvolution component to further enhance the result. Image rendering is a computer graphics approach where a scene is modeled with a given set of known scene geometry and imaging parameters, essentially a generation rather than a recovery like we achieve with SR.

The successful application of SR typically requires (1) subpixel inter-frame motion (2) the presence of aliasing in the LR frames and (3) lack of motion blur in the frames used [Milanfar2010][Hardeep2013][Gunturk2012]. Some authors have proposed motionless SR based on defocus or varying the blur present in the LR frames [Rajan2001][Rajagopalan2003]. Aliasing is a visually unappealing artifact in images which is why many, if not most image acquisition devices are designed with antialiasing filters; however, some level of aliasing in the LR images is required [Nasrollahi2014]. Ben-Ezra et al [BenEzra2005] note that the presence of motion blur causes significant degradation of SR, even when the motion blur is known.

6.1.2 Image Formation Model

An understanding of the model for image acquisition is an important first step when approaching any image processing or computer vision problem. Most conventional cameras will form the image similarly, i.e. light enters the lens which focuses the light on the film or sensor. Some specialty image acquisition systems (light field cameras, coded aperture systems, multi/hyper-spectral imaging systems, etc.) have an image acquisition model which departs significantly from the type of system we present here and which are in most common use. The basic important components of a classical digital optical imaging system are: aperture, lens system, and sensor.

Using this physical model of the image acquisition system, we will now look at the mathematical models described in the literature. The most basic model of image formation is the pinhole camera model, comprising a perspective projection of a 3D scene onto a 2D sensor. The pinhole camera is similar to the system shown in Figure 6.2, but without lenses and using a very small aperture. Though this type of camera has no aberration, the problem with this construction is the excessive amount of time needed to capture an image. With the introduction of lenses to the system, a wider aperture can be used, introducing more light to the sensor at the expense of obtaining a blurred and distorted image. Delbracio et al [Delbracio 2012] describe the image formation process as

$$\boldsymbol{v} = \boldsymbol{S}_1\big(g(F(H(u) * h_{ex}) * h)\big) + \boldsymbol{n} \tag{6.1}$$

Where $\boldsymbol{v}$ is the measured image,

$\boldsymbol{S}_1$ is the bi-dimensional ideal sampling operator,

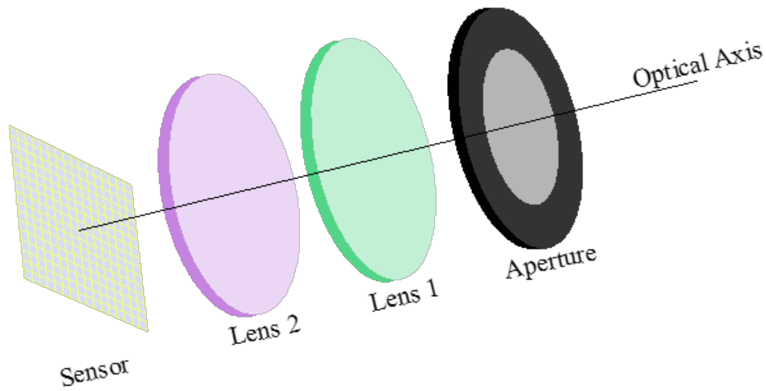


Figure 6.2: *Exploded view of a simplified representation of a digital optical image acquisition system. Light enters from the right, and is either passed, blocked, or diffracted by the aperture. The path of the light is then altered by the lenses. These components serve to focus the light from the scene, but also introduce undesirable aberration and distortion. Finally, the light is recorded by the sensor and binned into pixels with the relative intensity levels discretized into a value representing the amount of light incident on a given pixel.*

g is a monotonically increasing function that describes the nonlinear sensor response or camera response function (CRF),

F is the geometric distortion operator,

H is the perspective projection, reduced to a planar homography,

u is the ideal planar representation of the scene,

$*$ is the local convolution operator,

h_{ex} is the kernel comprising all extrinsic blur effects (e.g. motion blur, atmospheric turbulence),

h is the kernel comprising all intrinsic blur effects (e.g. diffraction, lens blur, antialiasing filters), also called the PSF, and

$\boldsymbol{n}$ models the sensor noise.

Šroubek et al [Sroubek2008] model the image formation process as

$$g_k(i,j) = D\big([h_k * W_k(o)](x,y)\big) + n_k(i,j) \tag{6.2}$$

Where $\mathrm{g_k}$ is low-resolution, digitized, noisy image,

$(\mathrm{i,j})$ are the image coordinates of $\mathrm{g_k}$,

D is the decimation operator which models the function of a CCD/CMOS sensor array,

$\mathrm{h_k}$ is the spatially variant PSF,

$\mathrm{W_k}$ is the geometric deformation (warping),

o is the intensity function of the original, continuous, high-resolution scene,

$\mathrm{n_k}$ is additive noise, and

k is the acquisition index.

Šroubek et al [Sroubek2008] combine modeling of both the intrinsic and extrinsic blurs into a single parameter. They mention turbulence, lens blur, and relative camera-scene motion. They further note the volatile nature of the blur sources, despite not separating the volatile extrinsic blur sources from the more stable intrinsic blur sources.

Tian and Ma [Tian2011] describe the observation model as

$$\boldsymbol{y}^{(k)} = \boldsymbol{D}^{(k)}\boldsymbol{P}^{(k)}\boldsymbol{W}^{(k)}\boldsymbol{X} + \boldsymbol{V}^{(k)} \tag{6.3}$$

Where **y** is the low-resolution image,

D is the decimation matrix,

P is the blurring matrix,

W is the warping matrix,

X is the original high-resolution image,

V represents white Gaussian noise encountered during the image acquisition process , and

k is the acquisition index.

Tian and Ma [TianMa2011] state that the operations **D**, **P**, and **W** may be combined into a single transformation matrix. They also note that **V** is assumed to be independent of **X** and that **P** and **W** are commutable if the imaging blur is spatio-temporally invariant and there exists only global translational motion between the frames.

The image acquisition models presented [Sroubek2008][Tian2011][Delbracio2012] differ most in their selection of notation and how they handle lens blur and sensor sampling. Delbracio et al [Delbracio2012] include the CRF and separate the intrinsic and extrinsic blurring effects. In our formulation, we consider a flat-field correction as sufficient to address the effects of the CRF and omit this contribution to the acquisition model. Also, for our imaging scenario the effects of motion blur are negligible so the extrinsic source of blur is due to atmospheric turbulence. Intrinsic sources of blur are due to diffraction from the aperture and blur from the lens. We adapt the notation of (6.2) and utilize the structure of (6.1) to develop our model.

$$g_k(i,j) = D\big([W_k(o * h_{ex}) * h_{in}](x,y)\big) + n_k(i,j) \tag{6.4}$$

Where $\mathrm{g_k}$ is low-resolution, digitized, noisy image,

$(\mathrm{i,j})$ are the image coordinates of $\mathrm{g_k}$,

D is the decimation operator which models the function of the sensor array,

$\mathrm{W_k}$ is the geometric deformation (radial and tangential) due to the lens system,

o is the intensity function of the original, continuous, high-resolution scene,

$\mathrm{h_{ex}}$ represents the blur from extrinsic sources (simply atmospheric turbulence in our scenario),

$\mathrm{h_{in}}$ is the spatially variant PSF representing all intrinsic blur effects of the acquisition device,

$(\mathrm{x,y})$ are the image coordinates of the sampled version of the recovered high-resolution image,

$\mathrm{n_k}$ is additive noise, and

k is the acquisition index.

The model in (6.4) represents a formulation which implies a multiframe approach to enhancement of the observed images. A simplified formulation representing the model used to generate the synthetic images used as part of the pipeline is shown in (6.5).

$$g = D([W(o * h_{ex}) * h_{in}]) + n \tag{6.5}$$

Where the parameters are as described above.

In Figure 1.2, the element of the pipeline labelled 'Lens Parameters' corresponds to the intrinsic blur h_{in} and the geometric distortion W, while the element labelled 'Sensor Parameters' corresponds to the decimation D and the noise n. The original scene is formed based on the target position and the sensor position and orientation. The extrinsic blur (atmospheric distortion) parameter h_{ex} is considered independent of these known or estimated parameters. Droege et al. [Droege2012] also include the effects of atmospheric distortion, but do so via the MTF rather than utilizing a time- and space-variant representation.

Some researchers [Farsiu2004][Wang2004B][Wang2005] have discussed the order of applying the convolutional blur and the geometric warping operations. Wang and Qi [Wang2004B][Wang2005] indicate that the typical approach (performing the warping operation followed by the convolutional blur operation) is consistent with imaging physics where the blur degradation dominates. They do warn, however, that this order of operations is only useful if the inter-frame motion is known a priori. The authors reach this conclusion based on an analysis of the results of methods for which the accuracy of the motion estimation is diminished due to the degradations present in the LR images. Farsiu et al. [Farsiu2004] indicate that the operations are commutative if one assumes that the PSF is space invariant, an incorrect assumption for real captured images as demonstrated in Chapter 3.

6.1.3 Image Registration

Szeliski [Szeliski2006] discusses image registration in great detail, including 2D and 3D parametric motion models. For our purposes, the target motion and the egomotion of the camera in a short duration can be reliably modeled as a 2D planar transformation. Once a model is chosen, there are a selection of methods for solving for the parameters of the motion model. The exact method of solving for these parameters is not important for our application, but rather the accuracy of the estimates. In general the computational time for the geometric registration is nontrivial but as we formulate our image acquisition system in the multi-aperture case, we assume that this operation is a calibration step that can be performed a priori. We include a comparison of the results of several methods with respect to their accuracy compared to a known ground truth motion vector in order to recommend a method for use in calibration for images obtained using this kind of system.

Aliasing must be present in the LR input images for success of the SR algorithms. Meaning that at least one parameter in the imaging model must be unique about each of the LR input images to contribute to enhancement of the resolution of the result. We focused on inter-frame motion as the source of uniqueness but also considered the application of multi-aperture imaging which would contribute unique views as well as slight differences in the blur kernel due to using distinct lens elements at each aperture location. The difference from one frame to the next can also be a zoom or scale change or utilizing the unique spectral bands from a multichannel image. The most common considered are the inter-frame motion and blur kernel estimation [Nasrollahi2014]. In this section we discuss the geometric image registration while covering blur kernel estimation in Section 6.3.

Geometric registration compensates for the misalignment between a set of images. These misalignments, or inter-frame motion, typically result from global or local motion. Global motion arising from the egomotion of the camera used to acquire the images, or from motion of the object of interest if the camera is stationary and the object comprises a large percentage of the FOV. Local motion estimation is necessary when there are many objects of interest within the FOV or when the object of interest is deformable, such as with a human face. Atmospheric distortion can also be modeled as local motion [Nasrollahi2014].

In addition to pixel-based registration approaches like optical flow, there are also methods which utilize edges (typically via gradient operations or another edge detection operation), corners (typically via the Harris or another corner detector), edges and corners using the sampling theory of finite rate of innovation (FRI), normalized cross correlation (NXC), Fourier-based matching, mesh-based warping, depth information, and PCA for hyperspectral image sets [Nasrollahi2014]. Selecting a motion estimation approach depends greatly on the goals of the application and the type of motion in the LR input images. The type of motion seen in the images we encounter contain a single target of interest when the resolution is low enough that all potential targets are in a single cluster such that the individual items contained are indistinguishable. If the resolution is higher (whether in experiments considering higher resolution source data or from the target cluster approaching the altitude of the sensors) then the individual items in the cluster may be distinguished and the global motion assumption becomes less valid. Global motion estimation methods can be successful if the targets in a cluster are first isolated using segmentation, and in a sense the global motion is applied as a local motion estimation. Most of the scenarios deemed realistic for our application were determined to be sufficiently reasonable to apply global motion estimation, but a discussion of local methods is included for comparison, the knowledge of the reader, and to illustrate potential for future research directions.

6.1.3.1 Local

Modeling of local motion is accomplished using a non-rigid motion model, which uses a set of control points calculated for each image. These control points are weighted utilizing position information from the image pair (current image and reference image). Local motion estimation is desirable when there are multiple rigid objects in the scene with independent motion, a single (or multiple) deformable object such that different parts have their own motion, such as a human walking or a human face.

Optical flow [Bouguet2000] is a popular local motion estimation method, which operates by dividing the image into blocks which are then processed at multiple resolutions in order to construct a flow map which represents the motion vector of the local motion of pixels in the image compared to a reference [Nasrollahi2014]. This approach works particularly well when the objects in the scene are non-rigid, non-planar, non-Lambertian, and subject to self-occlusion. Such objects include human faces, human bodies, or other living beings. Typically candidate feature points are obtained using SIFT [Lowe1999] (or another process) which produces better results than dense flows which calculate a motion map for each pixel [Nasrollahi2014].

6.1.3.2 Global

Typical global motion models include global translational, affine, and projective [Nasrollahi2014]. We focus on global translation estimation since our imaging scenario is particularly well-suited to this model, but also include algorithms that perform rotation estimation about the optical axis to account for the egomotion of the imaging platforms used.

The equation in (6.6) demonstrates the relationship between the LR and HR images, allowing for translation and rotation.

$$\begin{aligned} x &= x_k^t + q_x m \cos\theta_k - q_y n \sin\theta_k \\ y &= y_k^t + q_x m \sin\theta_k + q_y n \cos\theta_k \end{aligned} \tag{6.6}$$

Where x_k^t, y_k^t is the translation of the kth image,

θ_k is the rotation, and

q_x, q_y are the sampling rates along the x and y direction, respectively.

Some authors employ a Taylor series expansion to solve for the parameters but this approach is only valid for small translations and rotations [Nasrollahi2014]. It is possible to apply multiple approaches, however. Meaning that a rough “pre-registration” can be performed prior to application of an approach which is only successful at calculating the registration parameters if they are small. This type of approach is embedded in the idea of resolution pyramids, where downsampled versions of the input are aligned at progressively finer scales. This is accomplished by first finding the parameters in the smallest resolutions of the pyramid and then using those values as the starting point of the next scale up. Block-based application of this resolution pyramid idea is known as optical flow, a popular type of local motion estimation.

There are two approaches to global motion estimation: differential (progressive), and cumulative (anchored). Differential global motion estimation calculates the motion between each successive frame whereas cumulative calculates the motion of an image set relative to a single reference (anchor) image [Nasrollahi2014]. The differential approach is particularly well-suited to processing a stream of frames coming from an input video, though the cumulative approach may be adapted for this situation by applying a windowed operation where a sliding window of input frames of a given width are selected for calculating an SR output at each time step, such that the result is only delayed by the number of frames that need to be accumulated before processing begins, assuming negligible processing time. An important consideration when applying the cumulative approach is appropriate selection of the reference. If the image selected is of particularly bad quality due to image degradation or too much important information in the scene is occluded then motion estimation will be unreliable. Nasrollahi and Moeslund [Nasrollahi2014] have suggested methods for selecting an appropriate reference frame utilizing image quality metrics.

6.1.4 Taxonomy of SR Methods

A variety of approaches have been developed since Tsai and Huang's [Tsai1984] seminal method. The taxonomy of methods in Figure 6.4 is adapted from that proposed by Nasrollahi and Moeslund [Nasrollahi2014]. As we require results that are more rigorous by metrology standards rather than driven by aesthetic concerns we focus on only those methods which develop their results from input observations of the scene and exclude the hallucination or machine-learning approaches which use data obtained from other images to build their dictionaries or databases. This effect is shown in Figure 6.3 where the results of a competitive single image SR approach, though a drastic improvement in the quality of the input supplied, in many cases lacks sufficient similarity to the original HR source to be considered appropriate for our application.

Classification of SR algorithms may be conducted in a handful of ways, we will focus on methods which utilize multiple LR images. The approach shown in Figure 6.4 which we will use in this chapter classifies SR approaches based on the domain in which the method operates, whether in the spatial or frequency domain. Tsai and Huang [Tsai1984] utilized a frequency-domain approach, but as may be seen, spatial domain has received much more attention, primarily due to the ability to utilize more flexible image registration methods as discussed in Section 6.1.3.

The imaging acquisition model presented in Equation (6.5) is almost identical to the approach presented by Nasrollahi and Moeslund [Nasrollahi2014] as being the structure used by a majority of reconstruction-based SR methods. We account for the intrinsic and extrinsic blur functions, as described previously due to our incorporation of both the atmospheric distortion effects and the potential for motion blur during image acquisition. If these factors are deemed negligible then our model agrees with the majority of SR researchers.

6.2 Description of Methods

This section contains a description of the SR methods described in the taxonomy of methods shown in Figure 6.4.

6.2.1 Frequency Domain

The first popular SR method, developed by Tsai and Huang [Tsai1984], was a frequency domain approach, where the authors developed an approach for enhancement of noiseless LR images. A frequency domain approach is intuitive as the fine details in an image are contained in the high-frequency components of the Fourier representation. Aliasing in the LR images is particularly vital for the success of frequency domain approaches as the relationship between the continuous and discrete Fourier transforms is exploited in this class of algorithms [Hardeep2013].

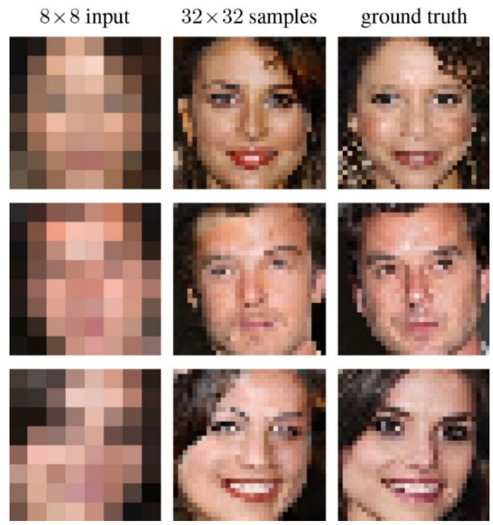


Figure 6.3: *The success of single image SR methods depends heavily on the training data used. This can drastically alter the result. In comparing the result in the center column with the ground truth in the right column, significant details are changed and in the case of the first result shown in the top row, a completely different person would be identified. For our application, it is of great important to have accurate results which represent the actual scene being inspected. For this reason we consider only multi image SR methods that utilize data from the scene in calculating the result. Reproduced from [Dahl2017].*

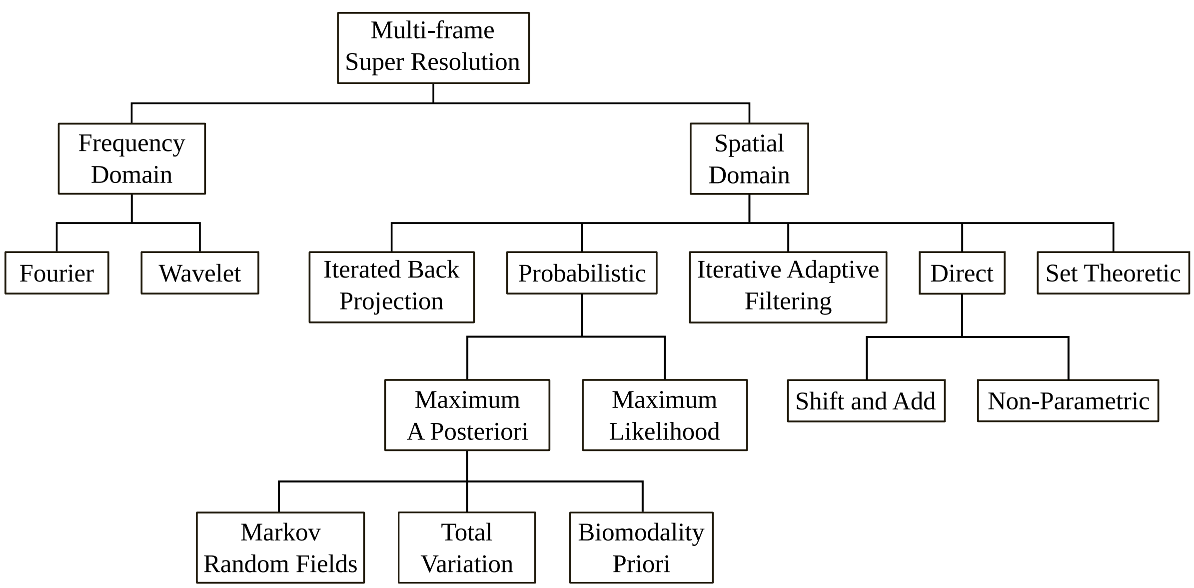


Figure 6.4: *Taxonomy of SR methods. Interpolation and upscaling methods also increase the pixel resolution of an image but the approaches do not contribute additional information from the scene in calculating the final HR image and therefore are unreliable for metrology-type accuracy needed for our application. The multi-frame SR structure presented by Nasrollahi and Moeslund [Nasrollahi2014] is used in presenting an analysis of multi-frame SR methods. The two broad categories into which multi-frame SR algorithms fall are the frequency domain approaches and the spatial domain approaches. Significantly more attention has been shown in the spatial domain, in part due to the flexibility of choice in motion estimation compared to the limitation of frequency domain methods to only global shift estimation.*

The shared approach for which SR methods are designated as frequency domain is that their algorithms primarily operate after transformation to the frequency domain and then perform an inverse transformation of the calculated HR image [Nasrollahi2014]. One drawback of this class of methods is that they can only operate on images for which the inter-frame motion is global translational. This is not of particular concern in our application however as the image acquisition conditions we investigate can be reasonably assumed to be global translational. In general for many applications it is vital to retain a remembrance of this limitation for this class of methods.

6.2.1.1 Fourier

The first SR methods [Gerchberg1974][Santis1975][Tsai1984] were iterative frequency domain approaches which could extend the spectrum of the image signal beyond its diffraction limit and therefore increase its resolution [Nasrollahi2014]. Tsai and Huang's [Tsai1984] method was demonstrated on satellite images which were related by global translation due to the large distance between the continuous scene and acquisition device. Their method took advantage of the shifting property of the Fourier transform to correlate the unique information contained in the available images. Extensions of their method expanded on the assumption of noiseless, blur-free images and included least squares estimation of the inter-frame motion, along with assigning weight to the LR images based on their SNR.

The effect of aliasing in the frequency domain is shown in Figure 6.5, in the spatial domain. When the bandwidth of the signal is limited during acquisition, the higher-frequency components will alias, or shift to a lower-frequency region and corrupt the signal. When we increase the bandwidth of the signal beyond this limit we remove this effect and the finer details in the signal are recovered.

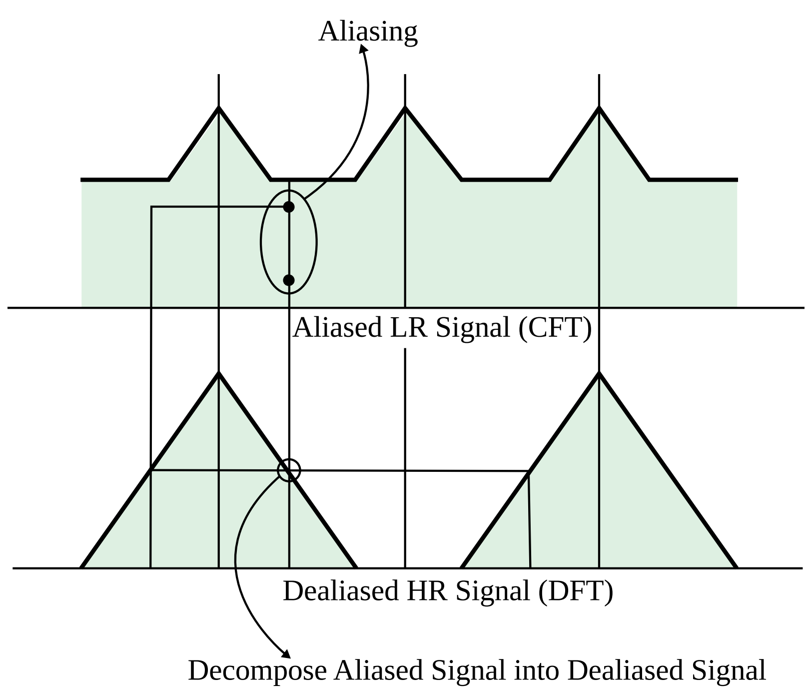


Figure 6.5: *Demonstration of the aliasing effect with respect to improved resolution possible with application of SR. Reproduced from [Hardeep2013]. Frequency domain approaches to SR rely on the shifting property of the Fourier transform, the aliasing relationship between the continuous and discrete form of the Fourier transform, and an assumption that the desired HR image is band-limited.*

6.2.1.2 Wavelet

Wavelet-based SR algorithms are a frequency based adaptation of the Fourier transform approach which is used to decompose the input images into structurally correlated sub-images. These sub-images are a multiscale representation of the input image from which it was calculated, as shown in Figure 6.6. Tian and Ma [Tian2011] describe this class of SR algorithms as considering the LR input images as low-pass filtered subbands of the unknown wavelet-transformed HR image. With this in mind, the approaches seek to estimate these finer scale subband coefficients followed by an inverse transform to arrive at the HR reconstruction.

6.2.2 Spatial Domain

A majority of SR approaches operate in the spatial domain due to the flexibility offered in selection of an inter-frame motion model and for the ability to model the various degradations present in the LR images [Nasrollahi2014]. A majority of multi-frame SR algorithms are reconstruction-based, meaning that they utilize the image formation model and consider aliasing artifacts in LR images as results from the decimation operation such that their SR approach can be considered as recovering from an under-sampling of the continuous scene. This section contains a description of the methods as depicted in Figure 6.4.

6.2.2.1 Iterative Back Projection

One of the first spatial domain approaches, the iterative back projection (IBP) class of SR methods operates by calculating a HR output and minimizing the error by iteratively refining the HR estimate. An initial estimate is usually composed of an average of the registered LR input images. This estimate is updated by utilizing the image formation

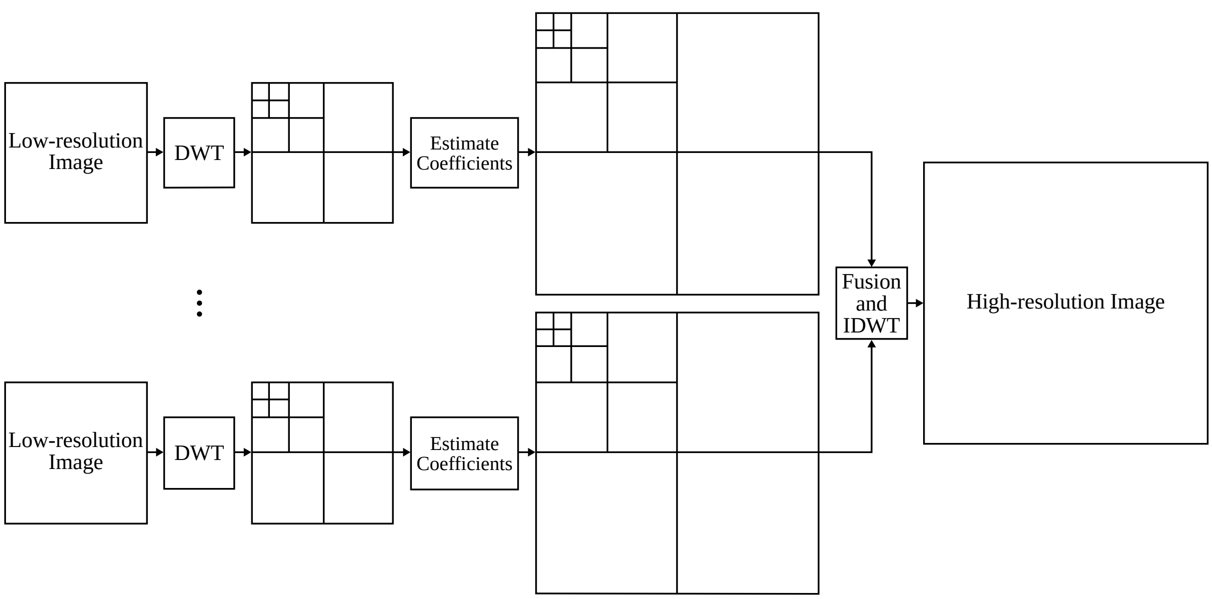


Figure 6.6: *A conceptual framework of wavelet domain SR approaches. Reproduced from [Tian2011]. The LR input images are converted to the wavelet domain by way of the discrete wavelet transform (DWT). Wavelet-based SR algorithms estimate the coefficients for the wavelet domain version of the HR image from each LR sample. Finally, these coefficient estimates are fused by the SR algorithm and the inverse discrete wavelet transform (IDWT) is applied, resulting in the spatial domain representation of the HR image.*

model to re-project the HR estimate and compare the LR version to the input LR images to determine if the current HR output estimate is more likely to have generated the LR inputs. This cycle iterates until no improvement is seen or until a specified number of iterations has occurred [Nasrollahi2014]. In order to speed up the algorithm, if a given pixel value has not changed for a certain number of iterations the pixel will not be included in subsequent iterations [Irani1993]. The back-projected error is calculated as the mean of the error contributed by each LR input image as shown in Equation (6.7).

$$f^{(t+1)}(x,y) = f^{(t)}(x,y) + \frac{1}{K}\sum_{k=1}^{K} w_k^{-1}\left(\left(\left(g_k - g_k^{(t)}\right)\dot{d}\right) * \dot{h}\right) \tag{6.7}$$

Where w_k^{-1} is the warping kernel,

$\dot{d}$ is an upsampling operator,

$\dot{h}$ is a deblurring kernel,

K is the number of LR input images g, and

f is the value of the error at iteration t.

Zomet et al [Zomet2001] proposed an improvement to this approach by using the median rather than the mean in order to improve robustness of the approach to noise in the LR input images and to improve the convergence time. Nasrollahi and Moeslund [Nasrollahi2014] observe the problem of the algorithm in there being multiple solutions which can be reached depending on the limit set on the number of iterations or the potential for the algorithm to oscillate between potential solutions if this parameter is set high enough. They also note that a priori constraints on the solution have been proposed by several authors and different regularization terms have been incorporated to aid in convergence. Further, they note the suggestion to replace the l_2 norm with the l_1 norm which improves robustness to outlier solutions due to noise or errors in motion estimation and also has the effect of improving the speed of convergence.

6.2.2.2 Probabilistic

Maximum Likelihood (ML)

This approach to SR maximizes the log-likelihood of (6.10) for all LR input images. The imaging model used is

$$\mathrm{g} = Af + \eta. \tag{6.8}$$

Where g is the LR observation,

A models the system degradation,

f is the original HR scene, and

η is Gaussian noise with zero mean and variance σ.

Given an estimate $\hat{f}$ of the SR output, the total probability of an observed input image g_k is

$$\mathrm{p}\left(g_k \middle| \hat{f}\right) = \prod_{\forall m,n} \frac{1}{2\sqrt{\pi}} exp\left(-\frac{(\hat{g}_k - g_k)^2}{2\sigma^2}\right). \tag{6.9}$$

The log-likelihood function of this estimate is

$$L(g_k) = -\sum_{\forall m,n} (\hat{g}_k - g_k)^2. \tag{6.10}$$

Thus the ML solution seek a SR result $\hat{f}_{ML}$ which maximizes (6.10) for all LR input images. Thus

$$\hat{f}_{ML} = (A^T A)^{-1} A^T g. \tag{6.11}$$

The ML solution is an ill-conditioned problem, which when applied to SR means that it is sensitive to increasing noise and errors in imaging parameter estimates. The problem becomes ill-posed if there are too few LR input images, following the same requirement as outlined by Šorel and Šroubek in [Gunturk2012] who specify a minimum number of LR input image as the square of the SR factor. In this situation, a priori information can constrain the solution space and select from a number of non-unique solutions. If a sufficient number of LR input images are available, the ML and MAP (introduced in the next section) solutions are identical and the ML solution is more computationally efficient. Depending on the information used, the ML problem may become a MAP problem and a MAP solution is preferred when there are not enough LR input images available.

Maximum a Posteriori (MAP)

MAP methods are related to ML methods via the application of Bayes' rule to handle the LR input images as incorporating a priori knowledge of the SR result, as shown in (6.12).

$$\mathrm{p}\left(\hat{f}\middle|g_1, g_2, \cdots, g_k\right) = \frac{\mathrm{p}(g_1, g_2, \cdots, g_k|f)\mathrm{p}(f)}{p(g_1, g_2, \cdots, g_k)} \tag{6.12}$$

Where $\hat{f}$ is the SR result, and

g_k are the k LR input images.

The LR input images are assumed to be independent samples of the original HR scene, so when the same assumption of white Gaussian noise with zero mean, and taking logarithms of each side of the equation and simplifying, then the SR result obtained using the MAP approach in (6.12) becomes (6.13) [Nasrollahi2014].

$$\hat{f}_{MAP} = \arg\min_f \left(\sum_{k=1}^{K} \|\hat{g}_k - g_k\|_2^2 + \lambda \Gamma(f) \right) \tag{6.13}$$

Where λ is a regularization parameter, and

$\Gamma(f)$ is the a priori energy function.

Different approaches have been developed for determining the best value of λ to solve the SR problem with a MAP method [Nasrollahi2014]. Regularization in solving SR problems as there may be many possible solutions the algorithm could find and reducing the solution space to a more desirable result or guiding the algorithm to reach the result quicker may be accomplished by this tactic. Where regularization is employed to utilize a priori information to guide the algorithm to a subset of possible results, many different approaches have been developed and the reader is encouraged to consult the work of Nasrollahi and Moeslund for additional detail [Nasrollahi2014]. Regularization terms employed in SR methods such as smoothness and edge-preserving terms are commonly found in use, but others are also found such as a natural image prior, corner preservation, multichannel smoothness (particularly for video SR), among others [Nasrollahi2014].

Markov Random Fields (MRF)

SR approaches which impose similarity structure a priori terms typically employ MRF models to represent the values of the local neighborhood in a given image. These models are expressed in the form

$$\Gamma(f) = \sum_{r \in R} V_r(f) = \sum \sum \sum \rho_r(f), \tag{6.14}$$

where V_r is a function of a set of local points r called cliques,

R is the set of all such cliques, and

ρ_r are the potential function defined over the pixels in each clique r.

The potential functions ρ_r are homogenous, isotropic, and quadratic in the pixel values for the most common case, which results in the Tikhonov regularization term

$$p(f) = \frac{1}{C} exp(-f^T Qf), \tag{6.15}$$

where Q is a symmetric, positive definite matrix.

One problem with using the Tikhonov regularization term is that it does not allow for discontinuities in the solution which impacts the preservation of edges in the result. Some authors have introduced modified approaches to preserve sharp edges. These include Huber MRFs which are a convex, non-quadratic prior resulting in nonlinear cost functions [Nasrollahi2014].

When Q is non-diagonal, the off-diagonal elements model spatial correlation between neighboring pixels. This gives a multivariate Gaussian distribution over f and is known as a Gaussian MRF. This approach helps with noise reduction at the expense of smoother edges. These effects can be balanced with appropriate choice of weight on the term.

Total Variation (TV)

When the weighting parameter of the Huber MRF tends to zero, the a priori term converts to a TV norm that penalizes both over smoothing and edge discontinuities. This results in edge preservation while minimizing ringing effects in the output. An approximation called bilateral TV forms the a priori term by generating multiple scales of derivatives, which can result in saturated data when applied to data collected from UAVs unless it is paired with the Hubert function shown in (6.16).

$$\rho(|x|) = \begin{cases} \frac{|\nabla x|^2}{2} & if\ A < \alpha \\ \frac{\partial A}{\partial x} & otherwise \end{cases} \tag{6.16}$$

Where A is the bilateral TV regularization term, and

α is calculated as $\alpha = \text{median}[A - median|A|]$.

This approach keeps the uniform areas smooth while preserves edges in the image [Nasrollahi2014].

Biomodality Priori (BP)

The BP approach has been shown to work particularly well for images such as text, where the data in the LR input images can typically be divided into two classes [Nasrollahi2014. BP is modeled as the exponentiated 4^{th}-order polynomial seen in (6.17).

$$\rho(f) = \frac{1}{C} exp\left(\frac{\left(f_{i,j} - \mu_0\right)^2 \left(f_{i,j} - \mu_1\right)^2}{\bar{\omega}^4}\right) \tag{6.17}$$

Where C is a normalization constant,

μ_0, μ_1 are the centers of the peaks of the polynomial (estimated using EM), and

$\bar{\omega}$ is the width of the foreground and background distributions.

Other appropriate problems include separating portions of the image into foreground and background regions.

6.2.2.3 Iterative Adaptive Filtering

This class of SR algorithm develops a solution as a state estimation problem and is typically employed for video SR, where a LR video is used as input to generate a HR output video. The Kalman filter [Elad1999] is typically employed, using the image formation model in (6.18) and the state equation of the Kalman filter in (6.19).

$$g(m,n) = \frac{1}{q^2} \sum_{x=qm}^{(q+1)m-1} \sum_{y=qn}^{(q+1)n-1} f(x,y) \tag{6.18}$$

Where g is the LR observation,

f is the original HR scene,

q is the decimation factor, assumed to be equal in the x, y directions.

This models the image formation process as an averaging of the HR image over a $q \times q$ area to form the LR image. The state equation for the Kalman filter is

$$f_k = B_k f_{k-1} + \zeta_k, \tag{6.19}$$

where B_k is the model of the relationship between the current estimate of the HR image and the previous estimate,

ζ is the error of estimating B_k, and

f is the HR image.

Some methods have also employed additional regularization terms in order to incorporate a prior knowledge of the relationship between LR and HR images, modeling the motion estimation in order to speed convergence of the algorithm [Nasrollahi2014].

6.2.2.4 Direct

Direct SR methods share similarity in the approach that they take to calculating the result, as follows:

1. One of the LR input images is selected as a reference
2. The remaining LR input images are registered against the reference image
3. The LR input image selected as a reference is scaled by the chosen SR factor
4. The registration information for the remaining LR input images is used to transform them to the scaled reference image
5. The SR image is calculated by fusing information from all of the available input data, including an optional deblurring step

Direct SR methods differ from one another in their approach to fusion of the scaled inputs; such approaches include mean and median filters, adaptive normalized averaging, Adaboost classifier, and SVD-based filters [Nasrollahi2014]. The direct SR methods are faster than IBP, and median fusion of the scaled inputs is equivalent to the ML approach. This approach is robust if the motion is purely translational (often referred to as a *shift-and-add* approach) and the blur kernel does not change significantly across the image.

Nonparametric

The distinguishing characteristic of this subset of direct SR approaches is in the combination of the motion estimation and fusion components of the process described above. They are particularly well-suited to video SR and utilize fuzzy motion estimation which is more robust to issues of occlusion and local motion such as found in human faces or other complex scenes [Nasrollahi2014]. This motion estimation approach involves patch-based segmentation of the LR input images, followed by a comparison of the similarity of patches, weighted by the calculated importance to producing a HR representation of that portion of the scene.

6.2.2.5 Set Theoretic

There are quite a few SR methods which employ the set theoretic projection onto convex sets (POCS) approach. The POCS approach assumes the LR input images impose a priori knowledge of the HR result, and that this a priori knowledge is a closed convex set S_k as shown in Equation (6.20).

$$S_k = \{f|\delta_l \leq |dh_k w_k f - g_k| \leq \delta_u\} \tag{6.20}$$

Where g_k is the kth LR image,

f is the SR result, and

δ_l and δ_u are the lower and upper bounds of the model's uncertainty.

With this collection of K convex sets, the iteration function in (6.21) is used to estimate the HR result. Occlusion in particular increases the error of the solution, as well as inaccurate motion estimation for which errors increase with increasing blur and noise in the available LR images. Some approaches detect inaccurate information from LR inputs and remove them from the estimation [Nasrollahi2014].

$$f^{(L+1)}(x,y) = \wp_m \wp_{m-1} \cdots \wp_2 \wp_1 f^{(L)}(x,y) \tag{6.21}$$

Where f is the HR estimate for the Lth iteration, and

$\wp$ is the projection of the HR estimate onto the convex set of the ith LR image.

Since ringing can be introduced from projections at the edges of sets, adding blur at the edges can reduce this effect.

6.2.3 Implementation

The majority of tools developed in gathering, generating, processing, and analyzing data for this dissertation were coded in MATLAB, primarily due to rapid production time in implementation. Also, many of the methods developed by other authors were available in MATLAB and incorporation of their implementations was chosen to ensure appropriate comparison of the results obtained using their methods. Some adaptation of existing methods were performed in order to increase interoperability with other tools developed as well as to parallelize the method to speed up processing given the large amount of data chosen to give a reasonable average time in comparing results. Shown in Figure 6.7 is our implementation of a GUI used in preliminary experimentation to determine the general performance of the methods as well as appropriate parameter ranges to use.

6.2.3.1 Geometric Registration Methods

Background information about geometric registration methods and an overview of the considerations involved in selecting methods for comparison are given in Section 2.2.3 and Section 6.1.3. Four methods each were chosen for comparison in the global translational (shift) and Euclidean (shift and rotate) motion models as shown above. The global translation methods chosen are Lucas-Kanade [Lucas1981], Keren [Keren1988], Marcel [Marcel1997], and Vandewalle [Vandewalle2006]. The Euclidean registration methods chosen for comparison are Keren [Keren1988], Marcel [Marcel1997], Lucchese [Lucchese2000], and Vandewalle [Vandewalle2006]. The authors who developed methods suitable for each model either developed two approaches or the rotation estimates are set to null prior to calculating the SR result.

6.2.3.2 SR Methods

The multi-image SR methods chosen are IBP [Irani1991], Papoulis-Gerchberg [Gerchberg1974], Robust SR (ZRSR) [Zomet2001], POCS [Vandewalle2006], Normalized Convolution [Pham2006], Robust SR (RSR) [Farsiu2004], Fast Robust SR (FRSR) [Farsiu2004], as well as bicubic interpolation for a baseline comparison. The approach of our implementation of IBP is described in Section 6.2.2.1. The Papoulis-Gerchberg method implemented is an iterative frequency domain approach as described in Section 6.2.1.1. Zomet et al's approach (ZRSR) is a modification of the

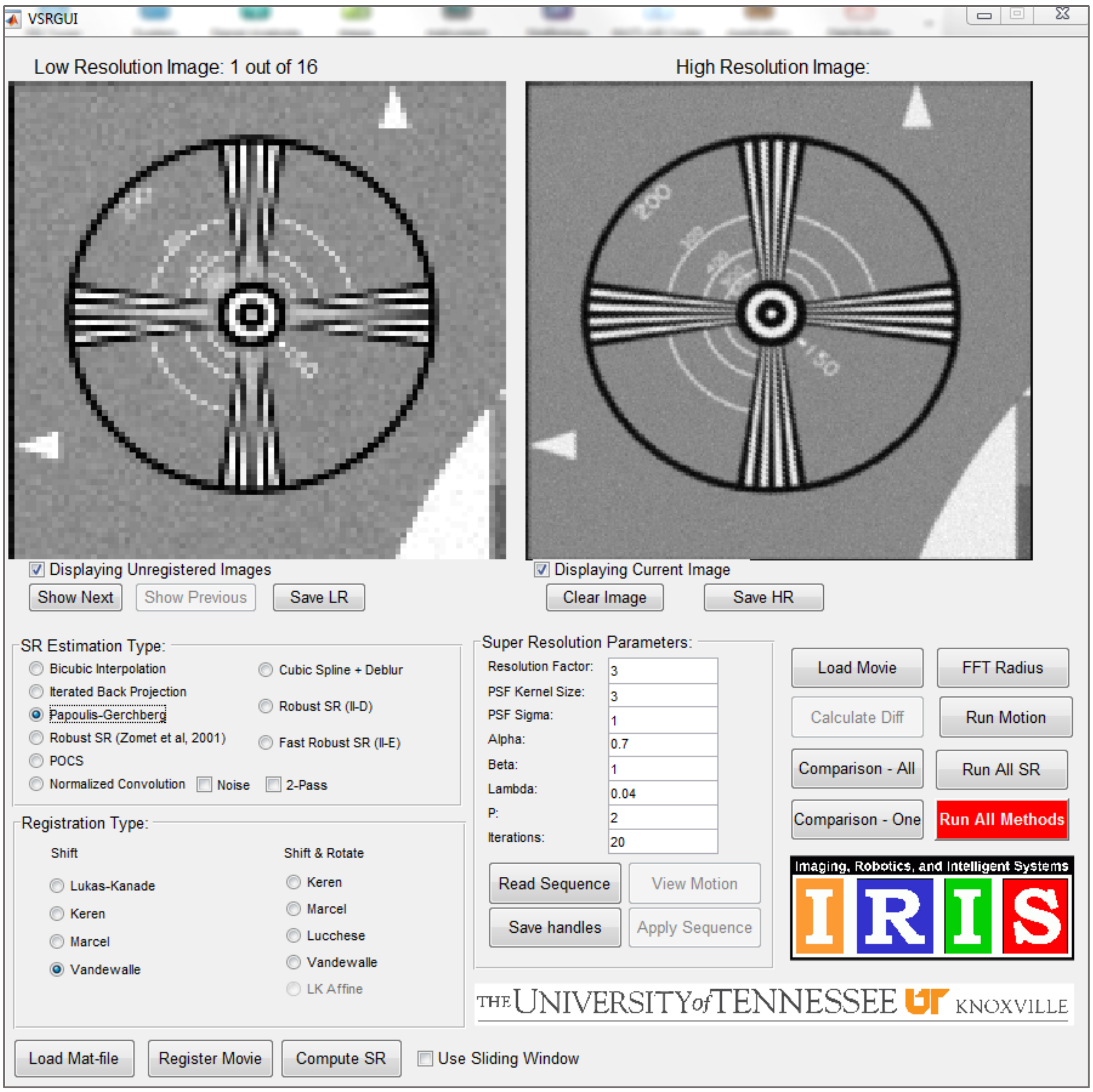


Figure 6.7: *The GUI used to perform preliminary SR experiments. The tool allows for selecting a set of LR input images, selection of a registration method, whether global translational (shift only) or Euclidean (shift and rotate) as shown in Figure 2.1. Selection of SR method is also performed along with setting the parameters to be used during processing. For longer image sequences and smaller SR factors, a sliding window operation was implemented which allows for a sequence of HR outputs rather than a single HR output as is the typical use. Further, analysis, comparison, and automation tools were implemented to ease the comparison of large sets of data.*

typical IBP method described in Section 6.2.2.1 where the mean operation is replaced with a median in order to increase noise robustness. The implementation of POCS is a set theoretic method as described in Section 6.2.2.5. Normalized Convolution is direct spatial SR approach as described in Section 6.2.2.4. The implementation of Farsiu et al's RSR and FRSR methods are a modified form of the IBP approach described in Section 6.2.2.1 where different regularization terms and norms were utilized compared to the original algorithm.

6.2.3.3 Parameters

Each of the SR methods uses the SR factor (labelled "Resolution Factor" in the figure) to establish the size of the HR grid with respect to the LR input image size. The other parameters listed (PSF Kernel Size, PSF Sigma, Alpha, Beta, Lambda, P, and Iterations) are used for the CSD, RSR, and FRSR methods which were developed by Farsiu et al [Farsiu2004] and did not result in competitive results, we therefore do not detail the purpose of these parameters but direct the interested reader to the author's publication on the operation of their methods.

6.2.3.4 Comparison and Analysis

The comparison and analysis tools developed allow for quick inspection of the results to determine the improvement in the quality of an SR result compared to the results of other SR methods and to the LR input images. When computing the result of a single method operating on a set of LR input images, an immediate result is displayed for comparison, along with the option of viewing the most recent previous result, which is helpful for determining if changes in parameters or SR factor contribute to improvements in the result. Also, the user can cycle through each of the LR input images to compare the output in determining a subjective comparison in the improvement of the output. The "Comparison – One" button launches a separate window that provides the same information described, just in a larger view with zoom and pan tools in order to facilitate inspection of find details as seen in Figure 6.8(a).

The "Comparison – All" button launches a separate window that allows inspection of the first SR_{factor}^2 images in the LR input image sequence, along with a single larger version (using nearest neighbor interpolation) of the image selected as reference for calculating the motion estimation, the SR result most recently calculated, and a difference image of the SR result and the interpolated LR reference image for comparison of the areas of the image where fine details were enhanced, as shown in Figure 6.8(b). For color images, we handle processing using the approach of Irani and Peleg [Irani1991], where the color image is converted to HSI space and SR is applied to the intensity channel while simple interpolation is applied to the hue and saturation channels before converting the image back to RGB color space.

6.2.3.5 Automation

In order to facilitate comparison of the available geometric registration and SR methods, automation tools were developed in order to calculate motion estimation for the LR input image sequence using each of the methods. This allows for analysis of the accuracy and a comparison of the performance of the methods. Where the motion vectors are known a priori, a more accurate analysis of the accuracy of a given method may be performed. Similarly for the SR methods, automation of the calculation of each method utilizing each motion estimation allows for inspection of which methods are more sensitive to errors in the registration step as well as which SR methods produce the best results for a given SR factor, image set, etc.

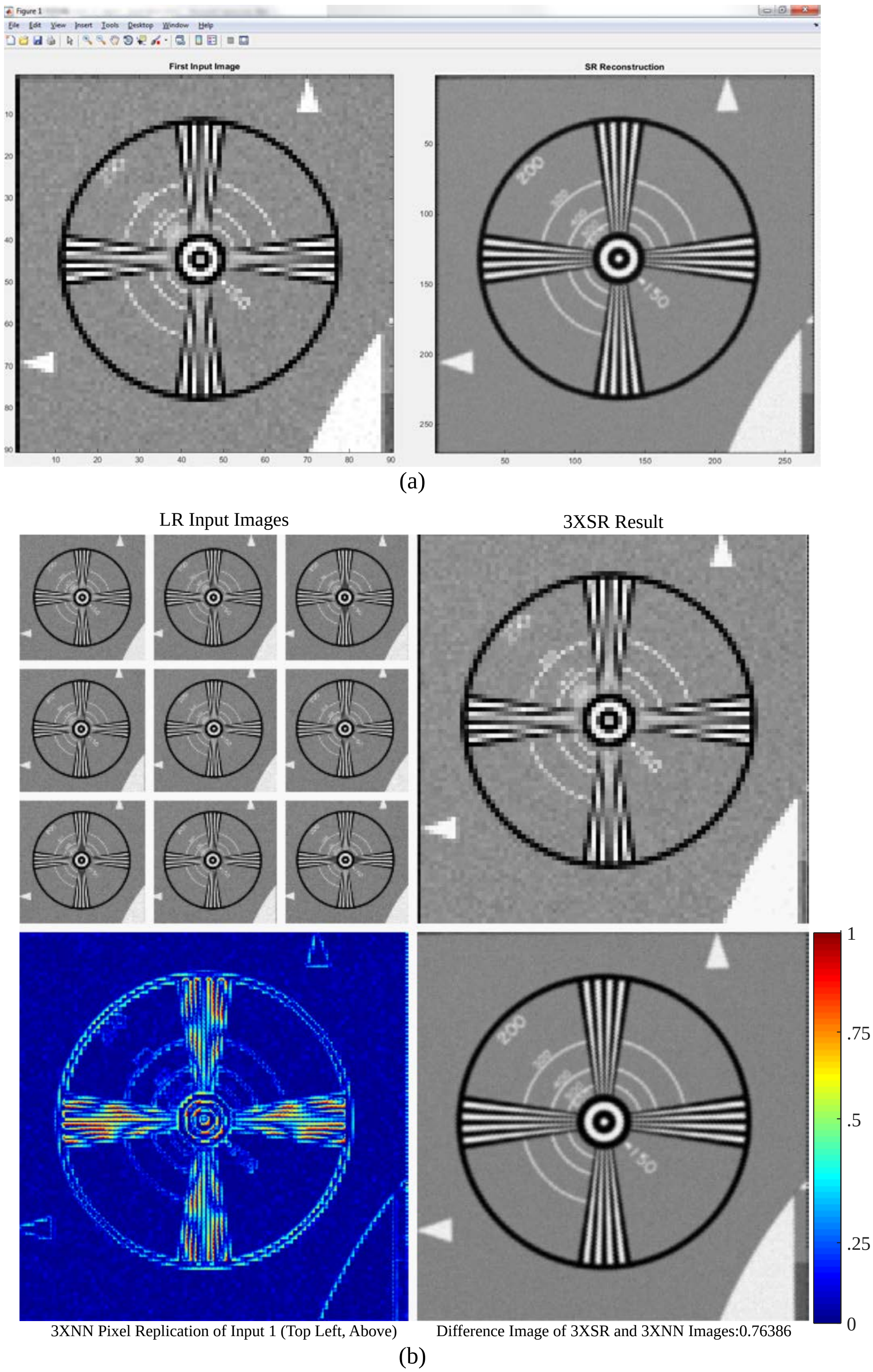


Figure 6.8: *The comparison windows generated for comparison, analysis, and inspection of the SR results. (a) Shown are the nearest-neighbor interpolated LR image used as reference for calculating the motion estimation (left) and the SR result most recently calculated (right). This view allows for a larger portion of the screen to be used in comparing the quality of the experiments being run compared to the parent GUI. (b) A more detailed comparison and analysis tool that also allows for detailed inspection of the results.*

6.3 Deconvolution

Though deconvolution/deblurring are generally separate operations which address independent image degradations, some multi image SR algorithms include deblurring as part of their process, whether by supplying a known PSF or by estimating the system PSF directly from the LR input images. Therefore, a discussion of deblurring in general as well as a description of those SR algorithms which perform both steps simultaneously is included here. Of important note is the observation by Lin and Shum that the effectiveness of SR algorithms is not heavily influenced by the system PSF [Lin2004]. Thus the operations are orthogonal, which is evidenced by the structure of the methods which perform both operations. As the operations can be strictly separated, they should be thought of separately even though they both contribute to an improvement in the clarity of degraded images.

6.3.1 Overview of Deconvolution

In order to successfully apply a deconvolution operation, two pieces of information are typically required: the image degraded by convolutional blur, and the kernel used to induce the blur into that image. While this is the usual arrangement, there are methods which estimate the kernel from the image itself. Therefore in some sense only the source image is needed although it is more helpful to think of these approaches as first estimating the kernel from the available image(s) and then applying a standard deconvolution step using this estimate and the observed image. We first discuss standard deconvolution as indicated and then consider the SR methods which incorporate deconvolution as part of their approach.

6.3.1.1 Deconvolution as a Post-Processing Step

Deconvolution approaches typically assume that the blur kernel is known, and that noise in the process is negligible or nearly so. Recovering the true image from a blurred observation is a straightforward process. The simplest model used to represent this operation is the inverse filter shown in Equation (6.22) in the frequency domain.

$$\hat{\mathrm{F}}(u,v) = \frac{G(u,v)}{H(u,v)} \tag{6.22}$$

Where $\hat{\mathrm{F}}$ is the estimate of the true image,

G is the degraded observation, and

H is the blur kernel.

The main difficulty comes when considering the contribution of noise in an observed image. This artifact adds uncertainty in the ill-posed inverse process which makes restoration of the true image a nontrivial problem. Often the contribution of noise in an image can be significant. Several solutions have been proposed. The model used by Wiener et al [Gonzalez2008] is also known as MMSE filtering and was developed to address the problem of noise in real images. The model used is described in Equation (6.23) in the frequency domain. It is assumed that the noise term and observed image are uncorrelated, that one or the other have zero mean, and that the pixels are linearly related.

$$\hat{\mathrm{F}}(u,v) = \left[\frac{1}{H(u,v)} \frac{|H(u,v)|^2}{|H(u,v)|^2 + \frac{S_\eta(u,v)}{S_f(u,v)}}\right] G(u,v) \tag{6.23}$$

Where S_η is the power spectrum of the noise, and

S_f is the power spectrum of the undegraded image.

Since the power spectrum of the undegraded image is not likely to be available, it is common practice to replace the $\frac{S_\eta(u,v)}{S_f(u,v)}$ term with a constant K [Gonzalez2008].

Because even small levels of noise causes dramatic instability in the output of the method, we must utilize methods for which handling the noise is an integral component. There are three classes of deconvolution algorithms we consider which do this [Mercimek2009]:

1. Statistical
2. Iterative
3. Blind

Statistical Deconvolution

Statistical deconvolution methods introduce a priori information about the statistics of the noise present in the image formation process. Linear deconvolution methods operate by minimizing the norm of an error term. For instance, for the Euclidean norm this is

$$\|G - F * H\|^2. \tag{6.24}$$

Various approaches have been proposed to find a solution to this formulation using e.g. Tikhonov, Total Variation (TV), and other regularization methods [Mercimek2009].

The form of the solution utilizing Tikhonov regularization is

$$T_\alpha = \|G - F * H\|^2 + \alpha\|f\|^2 \tag{6.25}$$

and the form utilizing Total Variation is

$$T_{TV} = \|G - F * H\|^2 + \alpha \cdot TV(f) \tag{6.26}$$

where $TV(f) = \sum_i \left(\Delta_i^h f\right)^2 + (\Delta_i^v f)^2$,

α is the regularization parameter,

Δ_i^h and Δ_i^v are first order difference operators at pixel i in the horizontal and vertical directions, respectively.

Linear deconvolution approaches are appealing because of their straightforward construction and computational efficiency, however they suffer to greater or lesser extent to overshoots in the solution which is seen in the result as what is known as “ringing” artifacts. Careful selection of regularization parameters can mitigate the severity of the effect, but it will still be present to some degree.

Iterative Deconvolution

A priori information can be utilized to constrain the solution space of deconvolution, such as non-negativity of the elements of the blur kernel [Delbracio2012], finite kernel size or support constraints [Mercimek2009]. Lucy-Richardson [Richardson1972][Lucy1974] is an iterative expectation maximization deconvolution approach based on a Bayesian framework [Temerinac2010]:

$$p(X|Y) = p(Y|X) \cdot \frac{p(X)}{p(Y)} \tag{6.27}$$

where $p(Y|X)$ is the likelihood probability,

$p(X|Y)$ is the posteriori probability, and

$p(X)$ is the prior model of the image.

The full derivation is detailed by Temerinac-Ott [Temerinac2010] but the result of the derivation is

$$\nabla J\big(X(v)\big) = H(-v) * \left[1 - \frac{Y(v)}{H * X(v)}\right] \tag{6.28}$$

which, after setting the derivative $\nabla J(X) = 0$ and accounting for the energy of the PSF being one,

$$H(-v) * \frac{Y(v)}{H * X}(v) = 1 \tag{6.29}$$

and assuming the convergence ratio $\frac{x^{p+1}}{x^p}$ is 1, the iteration steps are defined as

$$\hat{X}^{p+1}(v) = \hat{X}^p(v) \cdot C^p(v) \tag{6.30}$$

where $\hat{X}^p$ is the current state estimate of the original image X at iteration p, and

$$C^p(v) = H(-v) * \frac{Y(v)}{S^p(v)} \tag{6.31}$$

where S is the simulated image using PSF H.

Dey et al [Dey2004] apply TV regularization to Lucy-Richardson deconvolution. Their derivation results in the following iteration:

$$\hat{X}^{p+1}(v) = \frac{\hat{X}^p(v)}{1 - \lambda div\left(\frac{\nabla \hat{X}^p(v)}{\left|\nabla \hat{X}^p(v)\right|}\right)} \cdot C^p(v). \tag{6.32}$$

This formulation scales the image $\hat{X}^p(v)$ in (6.30) prior to applying the correction factor.

Blind Deconvolution

Blind deconvolution approaches are typically classified as either direct or indirect [Mercimek2009]. Direct methods perform the restoration process in a single operation while indirect methods converge to a solution iteratively. Direct methods employ a separate blur identification step prior to applying deconvolution. A priori blur identification involves making assumptions about the characteristics of the blur kernel, such as symmetry (a good assumption close to the optical axis, but increasingly less valid as the distance to the optical axis increases at the edges of the image), and parametric form (typically a bad assumption as described in Chapter 3).

Indirect methods utilize an iterative process to simultaneously estimate the blur kernel and the recovered image as part of the same process. This is very similar to the approaches which perform deconvolution as part of their SR approach, which is covered in the next section. It is common for algorithms of this type to require an initial guess of the recovered image and blur kernel. However, in the absence of an available estimate being supplied an estimate can be generated from the input data [Mercimek2009]. Ayers et al [Ayers1988], for instance, alternately iterating in the spatial and Fourier domains, applying a priori constraints in each to converge to a solution.

6.3.1.2 Deconvolution and Super Resolution

Lin and Shum [Lin2004] indicate that the limits of SR algorithms do not depend heavily on the intrinsic PSF of the acquisition system. For multi-image SR approaches, this is especially convenient because there are multiple samples of the effect of the spatially variant PSF in the available image observations from which to estimate the PSF for subsequent or simultaneous deconvolution of the restored image as part of the SR process. There are several authors which include deconvolution and SR as part of their approach to recovery of an estimated HR image of the scene of interest from the available LR images.

Baker and Kanade [Baker2002] consider the system PSF in two components, separating the blur due to the system optics and the blur introduced by the sensor. The optical blur component is further divided into a defocus element that they represent with the pillbox function, and model the remaining optical blur by the square of the first-order Bessel function of the first kind. Their separation of the different sources that contribute to the intrinsic blur is correct, but their fitting of the PSF to a parametric function is not, as discussed in Chapter 3. They do show however that although the effect of intrinsic blur does not limit the SR reconstruction process, that using higher SR factors does result in a blurring effect in the result, suggesting an upper bound on the SR factor. Their results show the limit exists somewhere between 4 and 8.

Tanaka and Okutomi [Tanaka2005] analyze the effect of assuming either a box or Gaussian PSF, developing an *SR condition number* evaluation of the limiting influence of these assumptions on the quality of the SR result. Lin and Shum [Lin2004] also use the pillbox model (they also consider the Gaussian function), but apply it to represent the system PSF. Their analysis shows that the theoretical limit of SR is due to the fine-resolving of the system, such that even though an image may be represented at a higher pixel count, with respect of the goal of SR to increase the effective resolution of the image by restoring detail at a higher resolution, that there exists a fundamental limit of 5.7 for the SR factor under ideal conditions. The authors further indicate a relationship between the SR factor and the necessary number of LR images where if the SR factor is an integer M, then the minimum number of LR images is M^2. If, however the SR factor is halfway between integers (1.5, 2.5, and so on), then the minimum number of LR input images is $4M^2$. Since there is no particular need in our application for half-value SR factors, in light of the penalty we utilize integer SR factors only.

Šorel and Šroubek [Gunturk2012] discuss the relationship between the number of LR input images, the size of the blur kernel support, and the SR factor. We give more detail in Section 6.4.3, but they consider the known intrinsic blur of the system as well as an unknown extrinsic blur of the scene due to camera motion during the integration period of the sensor. They solve this problem by developing a patch-based space-variant SR approach by dividing the image into patches, estimating the PSF in each patch, for each frame of available LR images, refine the estimated blur kernels by comparing the available data, and applying space-variant deconvolution and SR. Kotera et al [Kotera2013] utilize an alternating minimization approach based on the same structure where the error in the blur domain and error in the image domain are alternately minimized until the algorithm converges for a combined blind deconvolution/SR solution.

Ben-Ezra et al [BenEzra2005] develop a novel image acquisition system called “jittercam” which is a combined hardware-software solution. The sensor used is translated so that the image acquisition system does not need to move with respect to the scene, but rather the sensor moves with respect to the light coming in from the lens. However, despite the findings of other authors that the system PSF has little influence on the success of SR, motion blur of any significant amount (anything over ~4 pixels) causes SR to completely fail. Their main difficulty then is to control the acquisition time so that the integration time of the sensor does not coincide with motion of the sensor.

6.3.2 Blur Kernel Estimation

Before application of an appropriate deconvolution method, the system PSF must be known. This is the case for non-blind approaches, blind deconvolution approaches estimate the PSF as part of the deconvolution process. Several methods exist for determining the system PSF, an approach developed by the authors of this work is described in Chapter 3. Most authors who consider deblurring as part of the SR process either utilize a blind estimation method or utilize a fixed Gaussian blur kernel [Tanaka2005][Nasrollahi2014]. In this section, we will briefly describe the process and compare SR methods which include blur correction as a critical component of their approach.

There are three approaches to determining the blur kernel of an acquisition system [Mercimek2009]:

- Computation utilizing a mathematical model
- Experimental measurement utilizing a pointlike source
- Simultaneous estimation of the kernel and the recovered image utilizing the observed image

6.3.2.1 Parametric Blur Models

We have previously discussed the issues with assuming a parametric form for the blur kernel, but it is a popular approach in the literature [Couzinie2013][Xue2015][Campisi2016]. A number of authors, however, also recognize the issue of parametric PSF models and develop their approaches with nonparametric kernels [Delbracio2012][Schelten2014][Sun2015][Campisi2016]. A parametric model for extrinsic motion blur is one situation or source of blur where parameterized motion estimation is appropriate though the representation of the kernel that is moved through space during the sensor integration period should be represented with a nonparametric approach [Michaeli2013][Sun2015]. Often parametric models will be utilized in direct PSF measurement or blind estimation approaches where the fitting of the measurement or estimation to an assumed model is included as one part of the approach.

6.3.2.2 PSF Measurement

A pointlike source is used to measure the PSF directly. In microscopy applications this is typically a luminescent bead of some small size, in photography a source must be constructed according to the conditions described in Chapter 3, for telescopic applications a star of sufficient brightness can be used under suitable astronomic seeing conditions. The authors present an approach for direct measurement of the PSF in Chapter 3 of this dissertation, along with more detailed description of other contemporary methods.

6.3.2.3 Blind Kernel Estimation

Intro words…

Šorel and Šroubek [Gunturk2012] discuss their patch-based approach where a MAP based approach is used to estimate the HR patch and the blur kernel using two observation models. The first is

$$M_1(u, h_k) = \tfrac{\mu_1}{2}\sum_{k=1}^{K}\|D(u * h_k) - g_k\|^2 \tag{6.33}$$

and the second is

$$M_2(h_1, \cdots, h_k) = \tfrac{\mu_2}{2}\|\mathrm{N}\boldsymbol{h}\|^2. \tag{6.34}$$

Both μ_1 in the first model and μ_2 in the second model are inversely proportional to the noise variance, meaning that the MAP approach is equivalent to the optimization

$$\min_{u,\{h_k\}} M_1(u, \{h_k\}) + M_2(\{h_k\}) + R_u(u) + R_h(\{h_k\}) \tag{6.35}$$

Where R_u and R_h are image and PSF regularization terms.

Michaeli and Irani [Michaeli2013] note that the performance of combined deconvolution/SR methods degrades significantly the worse the estimate of the kernel is from the true kernel. They suggest that the camera PSF should not be used for deconvolution in SR applications and propose recovery of an “SR blur kernel” from the LR input image by using an image patch pair approach where a known set of LR/HR image patch pairs are available for substitution. They then use the recurrence of these natural image patch pairs to estimate the optimal blur kernel.

Sun et al [Sun2015] utilize a convolutional neural network (CNN) and a Markov random field (MRF) model to recover the kernel of non-uniform motion blur from a single image. Campisi and Egiazarian [Campisi2016] classify blind deconvolution approaches as either a priori, where the PSF is identified separately from the image reconstruction, or joint, where the reconstructed image and the blur kernel are identified simultaneously. Joint approaches often incorporate a priori information about the image (such as use of natural statistical priors) or blur kernel (such as parametric models) in order to constrain the solution space.

6.3.3 Experimental Results

While any physically-realizable imaging system will have the undesired degradations described, methods have been introduced for mitigating their influence so as to improve their usefulness in other tasks which require metrology-quality images.

6.3.3.1 Deconvolution Methods Considered

In our evaluation of popular deconvolution methods to use in validating our method, we considered three methods: Wiener deconvolution [Gonzalez2008], Lucy-Richardson deconvolution [Richardson1972], [Lucy1974], and Regularized deconvolution [Steve2008],[Kotera2013]. Image deconvolution is an ill-posed inverse problem [Bertero2005] and as such no perfect solution exists; however, numerous methods exist by which the quality of the image can be improved. We test the methods noted to determine which gives the best results using data obtained with our imaging systems.

Wiener deconvolution is a frequency-domain method that seeks to minimize noise effects in frequencies with low SNR [Gonzalez2008]. The method is developed to find a solution that provides a minimum mean-square error (MMSE) estimate of the original image f. In the implementation we use, the image to be deblurred, the measured PSF of the system, and the estimated SNR of the image are supplied to the algorithm. An assumption of Wiener deconvolution that may not apply in all situations is that the image is modeled as a random process, and that 2nd-order statistics and noise variance are known or can be estimated [Steve2008].

Lucy-Richardson deconvolution makes use of the expectation-maximization algorithm in order to maximize the likelihood of the restored image [Biggs1997]. Shepp and Vardi [Shepp1982] have shown that if the algorithm converges, it will converge to the maximum likelihood solution.

Regularization-based methods of deconvolution do not make the same assumptions as those made in Wiener deconvolution [Gonzalez2008] or attempt calculation of an EM-based solution. A side-effect of regularization-based methods is an over-smoothed solution, which while diminishing the influence of undesired noise in the result also limits image sharpness and crisp edges [Steve2008]. Regularizing terms are added to the minimization function in order to penalize selected aspects of unacceptable solutions. Methods exist that seek to enhance edges [Yang2009], diminish noise by penalizing high frequencies, etc. In the implementation of regularized deconvolution that we apply, a high-pass filter is employed in order to minimize the effects of the noise present in the captured images used for testing. Kotera et al [Kotera2013] apply their alternating minimization approach in the nonblind deconvolution method presented and adapt their approach to better correct natural images by using sparse priors ($L^p, p < 1$) that better fit the gradient distribution of natural images.

6.3.3.2 Application of Deconvolution

Image degradation comes from such sources as aperture diffraction, sensor noise, and lens blur. Proper application of deconvolution can reduce or eliminate some or all of these effects. Lens blur is a non-random spreading of the light that occurs before recording the scene at the sensor plane. Lens blur is intrinsic and thus can be modeled a priori. Wiener [Gonzalez2008], Lucy-Richardson [Richardson1972], [Lucy1974], Regularized Deconvolution [Steve2008], and Kotera et al [Kotera2013] methods are applied to sample image captures to demonstrate the improved quality.

Though our method was initially developed in response to issues the authors saw with Liu and Chen's approach [Liu2008], we lack the equipment to directly compare to their method. The approach of Delbracio et al [Delbracio2011], though a calibrated target-based approach, presents a viable comparison as they also consider subpixel measurement of the PSF.

We calculated the centroid of each local measurement and generated a map of the matching area using kNN (with k=1) to generate the map. The value of each pixel corresponds to the index number in the data structure used to store the individual extracted PSF measurements. Though not encountered in this measurement, some systems we tested had some “bad” pixel regions that reported a reading higher than the surrounding areas, which were filtered out during our extraction process. Our spatially variant deconvolution function iterates in the order indicated by the map, extracting the region from the image acquired, and applies the desired deconvolution method on the region, adding the masked result to the output image.

6.3.3.1 Wiener Deconvolution [Gonzalez2008]

Wiener deconvolution takes just under three minutes to complete, and is shown in Figure 6.10(b). The results do not significantly improve on the clarity of features compared to the input. All experiments shown were performed on a system with an Intel Xeon E5507 Processor (Quad-core 2.26GHz) and 12GB DDR3 RAM.

6.3.3.2 Lucy-Richardson Deconvolution [Richardson1972], [Lucy1974]

Lucy-Richardson deconvolution completed in just under three minutes as well. Shown in Figure 6.10(c), the result is much clearer compared to that of Wiener. In our experimentation, we found that restricting the number of iterations of the algorithm to four yielded results that did not significantly diminish in contrast.

6.3.3.3 Regularized Deconvolution [Steve2008]

Regularized deconvolution took just under eleven minutes to complete, and seems to introduce blur rather than remove it. This effect increases in proportion to the distance from the center of the image. For this reason, the result is not included in our comparison shown in Figure 6.10.

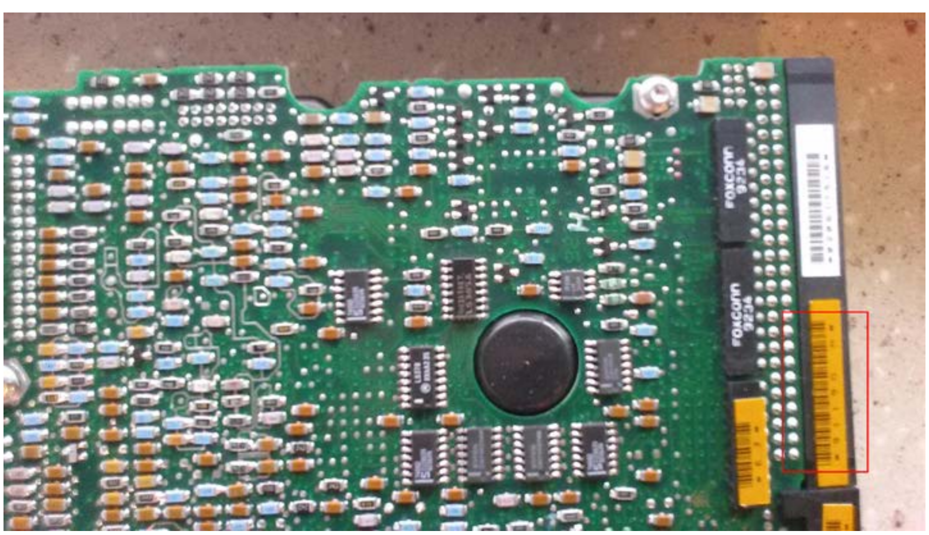


Figure 6.9: *Input image from the Galaxy S4. A text pattern was selected to demonstrate the increased clarity in the results after applying deconvolution. Highlighted regions are shown in Figure 6.10 to enhance visibility. The red cast to the scene is due to the ambient lighting.*

(a) (b) (c) (d)

Figure 6.10: *Comparison of the results of deconvolution applied to the image shown in Figure 6.9. (a) Highlighted section of the image shown in Figure 6.9. (b) Result of the application of Wiener spatially variant deconvolution. The Weiner filter is sensitive to noise and parameters were carefully chosen to avoid this artifact. The result is improved though the parameter selection was time-consuming and the result is not superior to the other methods. (c) Result of the application of Lucy-Richardson spatially variant deconvolution. At a comparable processing time to Weiner, the results are the best of the methods tested. Note in particular the improved definition of the characters. (d) Result of application of Šroubek et al.'s spatially variant deconvolution. Comparable processing time to both Weiner and Lucy-Richardson, and with clarity similar to Lucy-Richardson.*

6.3.3.4 Kotera et al Nonblind Deconvolution [Kotera2013]

Shown in Figure 6.10(d), the method of Kotera et al [Kotera2013] took just under four minutes to complete and though the results are an improvement on the original, the Lucy-Richardson result has slightly greater clarity.

6.3.3.5 Evaluation of Deconvolution

A tool developed by InLite Research [InLite] was used to analyze the output images from each of the methods described above. The results are shown in Figure 6.11 where only Lucy-Richardson’s result was readable. Based on both visual inspection of the results and the confirmation of the tool provided by [InLite], the authors believe that we have sufficiently demonstrated the effectiveness of our approach and justified our choice in supporting algorithms for the successful removal of blur effects.

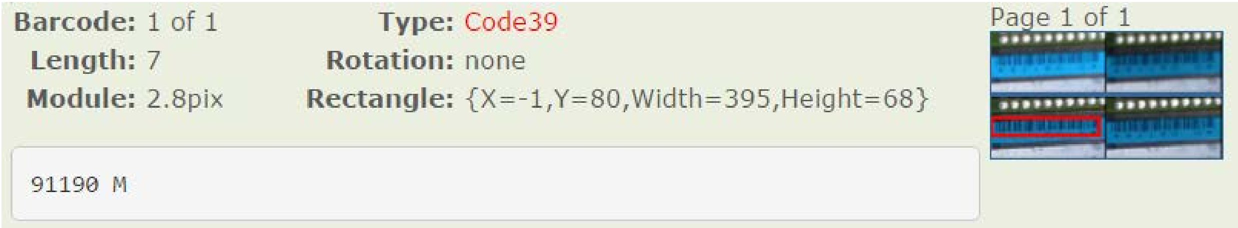


Figure 6.11: *Comparison of the results of applying the barcode reader tool from [InLite] to the input and deblurred images shown in Figure 6.10. Only the result of Lucy-Richardson's algorithm shown in Figure 6.10(c) was recognized as containing a barcode, and the information was successfully decoded.*

6.4 Evaluation of SR Methods

A description of the methods used and considerations taken in comparing the capabilities of the selected SR approaches in contained in this section. Nasrollahi and Moeslund [Nasrollahi2014] note that there is no ideal SR algorithm that has been developed, that selection of an appropriate method is highly application-specific.

6.4.1 Image Metrics

Nasrollahi and Moeslund [Nasrollahi2014] discuss the methods used to evaluate the quality of SR algorithms. While there have been important contributions made by authors who employed subjective evaluation, we will focus on objective metrics for assessing the quality of the results of a given SR method. Further, we will utilize non-blind quality estimation methods which compare the result to a known ground truth image. The methods utilized in our evaluation are peak signal to noise ratio (PSNR) and structural similarity (SSIM) metric, but we also describe another popular method.

6.4.1.1 Mean Square Error

MSE is defined as

$$MSE = \frac{\sum_{k=0}^{qN_1-1} \sum_{l=0}^{qN_2-1} \left({\hat{f}_{k,l}}_{k,l} - f_{k,l} \right)^2}{\sum_{k=0}^{qN_1-1} \sum_{l=0}^{qN_2-1} \left(f_{k,l} \right)^2} \tag{6.36}$$

Where $\hat{f}$ is the SR image being evaluated, and

f is the ground truth image.

With this metric, the smaller values indicate the image under test is closer to the ground truth. Wang and Bovik [Wang2009] give some insight into the issues with MSE which led the authors not to consider its use in evaluating the SR results.

6.4.1.2 Peak Signal to Noise Ratio

PSNR is defined as

$$PSNR = 20\log_{10}\frac{255}{\sqrt{MSE}} \tag{6.37}$$

Criticisms of the MSE and PSNR include a lack of correlation with representing the way the human visual system perceives information. Wang and Bovik [Wang2004A] sought to address this issue with SSIM and earlier approaches.

6.4.1.3 Structural Similarity

SSIM [Wang2004A] is defined as

$$SSIM = \frac{\left(2\mu_f\mu_{\hat{f}} + C_1\right)\left(2\sigma_{f\hat{f}} + C_2\right)}{\left(\mu_f^2 + \mu_{\hat{f}}^2 + C_1\right)\left(\sigma_f^2 + \sigma_{\hat{f}}^2 + C_2\right)} \tag{6.38}$$

Where C_1 and C_2 are constants,

σ_f and $\sigma_{\hat{f}}$ are the standard deviation of the ground truth and image under test, respectively, and

$\sigma_{f\hat{f}}$ is defined as

$$\sigma_{f\hat{f}} = \frac{1}{qN_1 \times qN_2 - 1}\sum_{k=0}^{qN_1-1}\sum_{l=0}^{qN_2-1}\left(f_{k,l} - \mu_f\right)\left(\hat{f}_{k,l} - \mu_{\hat{f}}\right) \tag{6.39}$$

Where μ_f and $\mu_{\hat{f}}$ are the mean of f and $\hat{f}$, respectively.

The SSIM score is between 0 and 1, with a value closer to 1 indicating a better match of the image under test to the ground truth reference. Many other objective image metrics have been used to evaluate the quality of SR output [Nasrollahi2014], but the authors of this dissertation have selected PSNR and SSIM as their metrics of choice. These metrics will be seen in the evaluation of the selected SR algorithms.

6.4.2 Robustness

An important consideration in evaluating the effectiveness of an algorithm is not just the quality of the results under ideal conditions, but also how capable the algorithm is in the presence of noise or errors in the estimation of the blur kernel or motion parameters. This is known as a measure of system robustness. Ideal conditions only exist for synthetically-generated data, and accurate estimations of real-world performance must include reliable estimates of the kind of degradations present in images captured using actual acquisition systems.

Errors in geometric registration can occur when there are areas of the image that do not agree with the assumptions of the motion estimation algorithm selected. Increasing noise and blur, particularly motion blur, also contributes to errors in motion estimation. The way that an algorithm’s robustness is evaluated is by determining its breakdown point [Nasrollahi2014]. The breakdown point represents the smallest percentage of contamination in the input that puts the result of the algorithm beyond an acceptable level of error. Since the final output for which we are interested in evaluating the accuracy against sources of error in our system is the position estimate of the target location, such sources of error can include contamination of the image data from noise and blur and inaccurate information about the position and orientation of the sensor platform.

A full analysis of the impact of these sources of error on the accuracy of our recovery of the target position is contained in Chapter 8.

6.4.3 SR Factor

The SR factor refers to the increase in pixel resolution in each of the spatial dimensions of a single-layer image. This means that for an SR factor of 2 for a $640 \times 480 \times k$ image (k being the number of channels in the image), the SR result would have $1280 \times 960 \times k$ pixels. Nasrollahi and Moeslund [Nasrollahi2014] note that some estimates of the theoretical capability of SR are low when using a deterministic linear model derived from the condition number of the system matrices and number of LR input images. Typically these estimates give a limit of 1.6X improvement in the SR factor.

A stochastic estimate of the limits of SR factor for reconstruction-based SR methods are more complex but more accurate. Lin and Shum [Lin2004] for instance, indicate a theoretical limit of 5.7 and a practical limit of 1.6 for the SR factor. They further relate the SR factor to the minimum number of LR input images. The equation in (6.40) holds for integer SR factors.

$$LR_{min} = SR_{factor}^2 \tag{6.40}$$

These estimates assume that each of the LR input images contains unique information that contributes to the method producing the result. Further, the LR input images much contain subpixel differences when the global translational model is used. Given the limits of SR factor and from results of preliminary experiments performed, integer SR factors of 2, 3, and 4 will be performed. This will allow for analysis for situations in which there are a minimum of 4, 9, or 16 LR input images, which is a small but reasonable number for the application scenario presented.

Šorel and Šroubek [Gunturk2012] give an equation relating the minimum patch size of an image to the number of input images, SR factor, and PSF size since their method simultaneously calculates the SR result and PSF. The equation they present is

$$G \geq \left\lceil \frac{\sqrt{K}(H-2) + H + \varepsilon - 1}{\sqrt{K} - \varepsilon} \right\rceil \tag{6.41}$$

Where G is the patch size of an image,

K is the number of available LR input images,

H is the PSF size, and

ε is the SR factor.

Rearranging to solve for K, we have

$$K \geq \left\lceil \frac{H + (G+1)\varepsilon - 1}{G - H + 2} \right\rceil^2 \tag{6.42}$$

Some similarities to the construction of [Lin2004] are that the minimum number of LR input images required increases by the square of the SR factor. Additional information we learn from the construction provided by Šorel and Šroubek [Gunturk2012] is that an increasing blur kernel has no effect on the limit of the SR factor with respect to the number of LR input images, which is the conclusion reached by [Lin2004] that the blur does not play an important limiting role.

6.4.4 Computational Time

One unifying factor that SR algorithms share compared to interpolation is a relatively high computational processing time. The computational time can be high for several reasons, including complexity of the algorithm, poor scaling to higher resolution input images or number of input images, slow convergence, etc. [Nasrollahi2014]. The computational processing time can also increase if there is particularly complex motion in the scene or if the data does not conform to the assumptions made by the registration algorithm chosen. In general, the simpler the motion (global translational being ideal), the smaller the input images, and number of images will contribute to a lower processing time. While these parameters can be controlled to some extent by adequate design of the image capture system, there will still exist differences between SR algorithms that will need to be established for proper selection of an appropriate method.

The reason these considerations are important for our application is due to its live, active, real-time nature. Thus in our evaluation of the selected SR algorithms we also include a comparison of the processing time required to achieve a given result. This is done in order to consider the tradeoffs in selecting a more accurate method with the increased time causing instability in the tracking step. Where possible, the SR methods have been optimized and parallelized to represent the fastest they can run on an identical system in order to represent an accurate comparison between methods.

7 Tracking

Tracking as we use the term refers to the localization of a target of interest in an image. Establishing a reliable track across a sequence of frames requires domain knowledge of the target's unique features that can be recorded by the sensor used and must be of sufficiently high framerate to allow for stable reorientation of sensors in an active hardware system such as we investigate. The tracking stage is separated into three stages: segmentation, where the raw image data is segmented into regions based on pixel similarity; target separation, where these candidate regions are analyzed to determine the likelihood of containing a target of interest; and 3D pose recovery, where the real-world position of the target is recovered from the image coordinates of the available sensors.

7.1 Segmentation

The data which will come from the sensors representing the target(s) has been indicated to lie on a noisy background with brighter regions representing the target(s). We utilize image segmentation in order to separate the target from the background. In our current implementation, we seek to find a binary image where black pixels indicate background, or the absence of a target within that pixel, and white pixels to indicate foreground, or that a target was found to be contained at least partially within the pixel. Our segmentation methods have parameters which may be selected to indicate candidate pixel regions of the image which may contain a target while the target separation operation described in the next section takes the indicated candidate regions and compares to templates representing the characteristics of the target and makes a final decision. Essentially, the segmentation step is an intelligent preprocessing step while the target separation will employ more accurate methods to reach a final determination about the target's position within the image.

In Figure 7.1 the input image selected for inspection is on the left side of the screen and contains some level of spectral information. Currently we have support for 3-channel images but support for an arbitrary number is planned in a future version. To the right of the input image, near the center of the GUI, is the output after segmentation has been performed on the input image. Note the bright pixel regions and compare to the input image. Application of segmentation allows for identification of regions of interest which are not readily apparent to the naked eye. On the right side of the GUI there is an area which indicates the spectral information of a pixel. The user may place the cursor over a pixel in the input image and the information contained at that location is represented in the indicated region as shown.

Below the area where the input and output images are shown is an information area. First is the image magnification. The user may use their mouse wheel to zoom in and out of either the input or output images and the magnification relative to native resolution is shown. Both images are currently shown at native resolution and thus the magnification for each is 1X. Below the input images magnification indicator are further information regions indicating the cursor's position relative to the bottom-left corner of the GUI, just above the area marked 'Image', and the cursor's position relative to the top-left corner of the image in pixels, marked 'Image'. Below this information region, labeled 'Visible rect', indicates the portion of the input image visible within the display region previously mentioned.

The area in the right-middle of the GUI marked 'Regions' indicates the number of regions present in the image as determined by the segmentation algorithm. Finally, the region of the GUI marked 'Input' contains the controls used for applying varying segmentation algorithms and parameters for extracting the regions of interest in the input image containing a target. The methods available are first, thresholding, whose output is shown in the appropriate region of Figure 7.1, using a parameter of 56 for the threshold value. We also have the k-means clustering algorithm available as well as the statistical region merging (SRM) algorithm. The user has the option of creating a video to save the visualization and choosing whether the output is represented in greyscale or a pseudocolored version for easier viewing.

We have also developed three segmentation methods which is included as part of our main image generation GUI and includes a simple thresholding algorithm (shown in Figure 7.2), SRM, and mean shift. Thresholding refers to a binarization process in which a pixel level, or threshold, is selected as the decision point. All values above this level are set to a logical 1 while all values below this level are set to a logical 0. The image demonstrated in Figure 7.2 is a color image, and the threshold is applied to each of the three channels which results in the multi-colored effect seen.

A hard threshold level may not always be known a priori, and SRM may be applied to determine similar pixels without this prior information. SRM uses a selected set of statistical parameters to determine which neighboring pixels are likely candidates for grouping based on their pixel values. All pixels initially begin as part of their own individual region. If a neighboring pixel falls within the statistical range, then it is merged into the current pixel. Thus, an added definition of neighboring pixel expands from spatial proximity and is enhanced to include information about the content of the pixel as well.

Mean shift is an application independent mathematical method which has found application in image segmentation. A kernel which establishes weights of the neighboring pixels is first selected, for instance a flat disk with equal weight over the coverage of the kernel or a Gaussian for which the weight assigned to neighboring pixels decays as the distance increases. Mean shift is known as a hill climbing algorithm, meaning that the kernel used is shifted to a

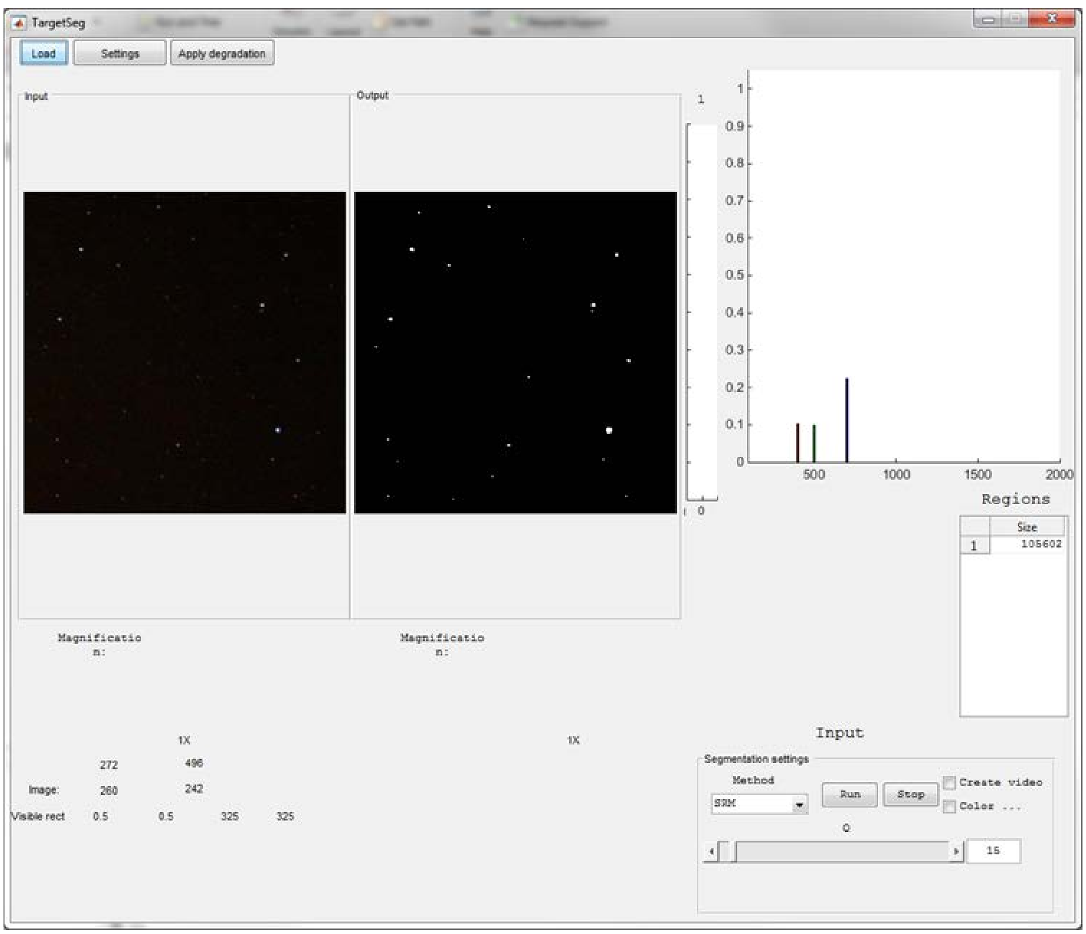


Figure 7.1: *Target segmentation GUI. Image region on the left indicates input image to be processed. Image region to the right indicates result of application of selected segmentation method. The bottom right area contains controls for selecting the method and parameters. The upper right region contains spectral information about the selected pixel.*

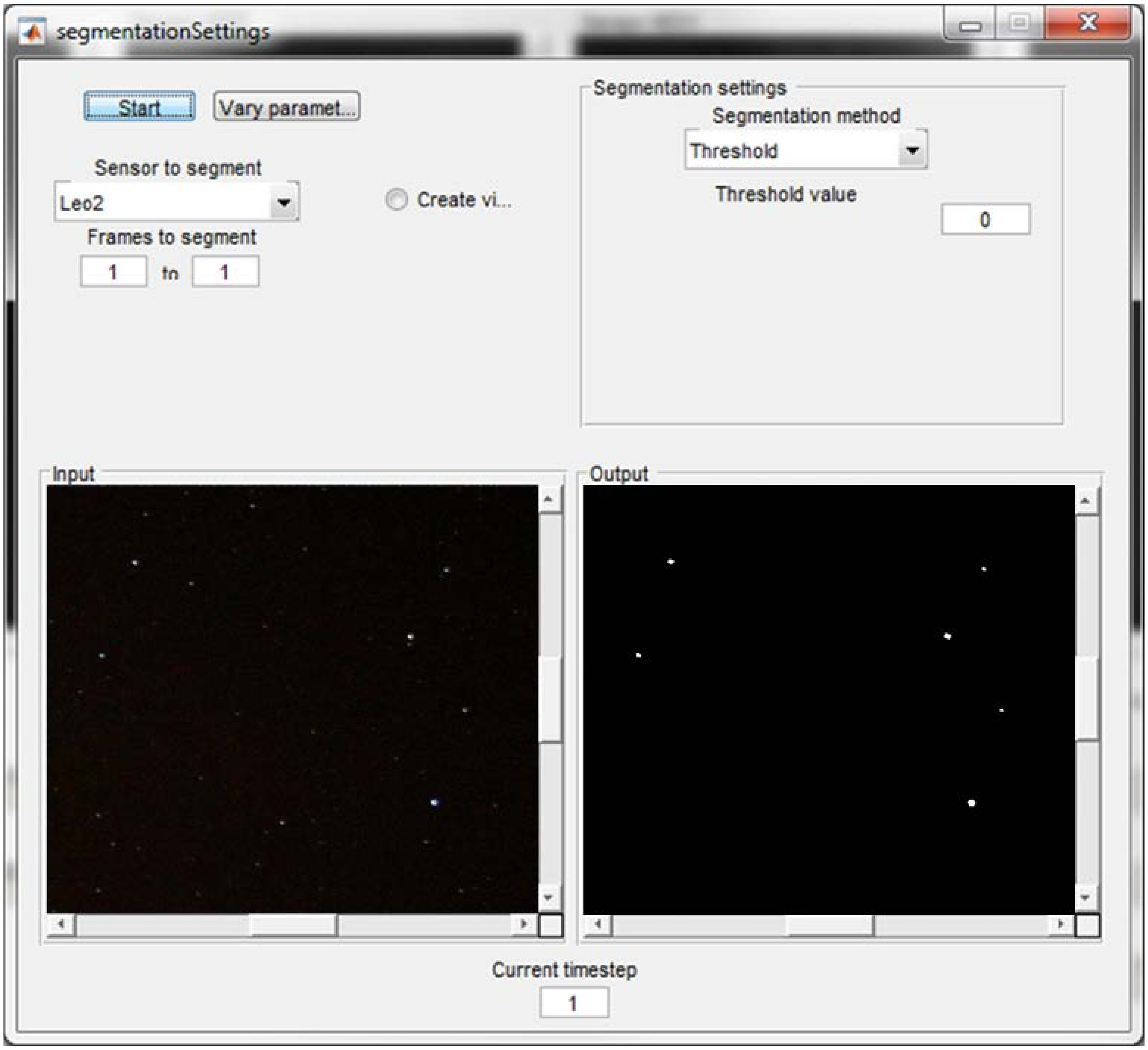


Figure 7.2: *Segmentation settings window. Segmentation occurs on a per-sensor basis, with parameters set based on the user's preference. Results may be saved for documentation and later reference. Segmentation in our workflow is a rapid preprocessing step to indicate candidate regions which may contain a target. In this sense, the step can also be viewed as a background separation step with the greatest importance placed on having a reliable true negative rate with focus placed on certainty that a pixel does not contain a target.*

higher-density region until the algorithm converges. The shift between iterations is defined by a mean shift vector, which always points in the direction of the maximum increase in density. Thus in areas for which the pixel values are indicative of the features which we wish to detect, we are able to generate an output image which has segmented the image to indicate the parts of the image most likely to contain information of interest.

7.2 Target Separation

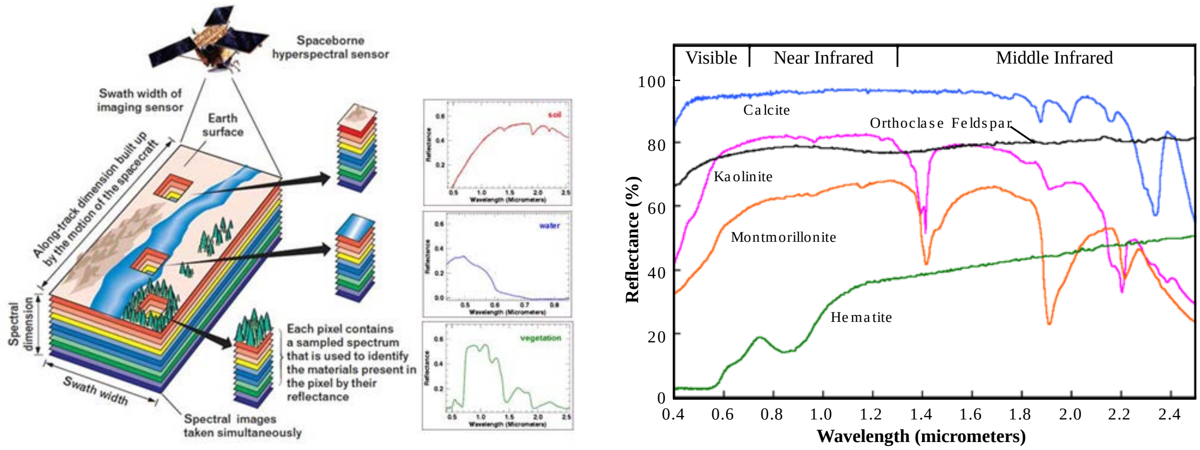
Once we have an image segmented by one of the above methods, we must determine which of the indicated candidate regions truly contains a target or targets which we are searching for. This portion of the process is very application specific and further depends on the characteristics of the target in the image space for which we have obtained data. In essence, we must determine a template which is able to separate the target data from the candidate points given by the naïve segmentation. This can take the form of a multi-spectral signature, possibly combined with information about the expected size of the target, or in a more complex implementation may use multiple frames to factor in the speed of the target over time.

7.2.1 Feature Descriptors

Extracting information about the location of a target of interest in the available images is highly application-dependent, based on the size and shape of the target, distance to the acquisition device, characteristics of the acquisition device (lens and sensor in our formulation), motion model of the target and acquisition device, framerate, illumination or spectral radiation of the target and the system response of the acquisition device. Understanding the target of interest and designing the acquisition system characteristics and placement will allow for design the best approach for gathering data that allows for discrimination of the target of interest from its background and potential false positive imposter targets which may exist nearby that can confuse the developed tracking system into following a decoy rather than the actual target of interest. Traditional feature detectors such as SIFT [Lowe1999], SURF [Knopp2010], etc. rely on there being enough texture in the image in order to apply these methods [Zitova2003]. In our scenario there is high noise and blur and the target of interest often dips below the noise floor, which would lead to repeated loss of the target while tracking. There are three major salient features which we focus on in describing the characteristics of the target in image space:

- Spectral Signature
    - Objects each have their own unique reflectance or emittance at given wavelengths which allows for discrimination of objects from one another depending on where the differences in the target of interest are unique compared to its surroundings. This feature can be exploited for elements, compounds, minerals [Smith2012], metals, ceramics, vegetation [Plaza2006], etc. Figure 7.3 shows an example of this phenomenon for geologic and vegetative applications.
- Target Shape
    - Over the course of the period of time during which the target is being tracked, the apparent shape as viewed by a single sensor may change due to the geometry of the target. However, the maximum extent of the target will not change unless multiple targets are closer than can be resolved due to the sensor resolution. In this case, if the change in the outer boundary is not overly extreme, the centroid may instead be tracked until reliable separation of the cluster is achieved.
- Target Motion Model
    - Knowledge of the expected motion and maneuvering capabilities of the target of interest allow for establishing a prediction of its possible motion and, therefore, an expected volume in which to search for it in subsequent time steps. False positives which pass discrimination by the other two features of interest can be disregarded if they do not exhibit the kind of motion expected of our target of interest.

In focusing on these unique features we are able to establish a reliable approach to extraction of the target of interest in spite of the imaging concerns described previously. By utilizing these three orthogonal characteristics of the target of interest we develop a stable approach to establishing the target location in image space at each time step which allows for correlation of information to maintain a reliable measure of the target's present and prediction of the target's future position.



(a) (b)

Figure 7.3: *Demonstration of utilization of spectral signature of materials in order to isolate targets of interest from potentially similar backgrounds. (a) Illustration of an orbital imaging spectrometer. The acquisition system can gather images in dozens to hundreds of segments of the electromagnetic spectrum such that at each pixel location the spectral distribution of the materials can be measured within each pixel region. Image courtesy Harris Geospatial Systems. (b) Reflectance spectra of representative minerals, naturally occurring chemical compounds that are major components of rocks and soils. Image courtesy MicroImages, Inc. [Smith2012].*

7.2.2 Target Separation

Computer-based understanding of image contents remains an open problem. As addressed in this section, we explore a subset of the problem: identifying which pixels of the image contain information about our chosen target and which pixels contain no information, or belong to the background. While other methods exist which handle the case of multiple targets within the FOV, herein we discuss the case of a single target.

The actual problem of target separation may be separated into two steps. Segmentation of the image into candidate regions (foreground) and regions unlikely to contain the target (background) is a quick process for which the component most needed to be accurate is identification of background pixels. Therefore the segmentation process must have high specificity, or rather, when a pixel in the image is marked as background not containing information likely to be part of a target we seek to find we want to apply a method and settings for which we may be certain that no target will be falsely ignored.

The second part of target separation involves inspecting the regions identified as potentially containing information which would identify a feature as a target. This step is highly application specific and requires processing which requires high sensitivity. In the spectrum in which we are operating, identifying features could involve a given spectral signature unique to the target which would discriminate it against other candidate regions. Another way of separating the true location of the target in the image could involve analysis of the motion of regions between frames over time to determine if the path of the target through space follows a trajectory it would likely exhibit.

7.2.2.1 Thresholding

Thresholding refers to a binarization process in which a pixel level, or threshold, is selected as the decision point. All values above this level are set to a logical 1 while all values below this level are set to a logical 0. Algorithms for determining the cutoff value may be simple, as in the case where the pixel intensity values are used, or more complex such as using image texture, hue, or other more complex features.

7.2.2.2 Mean Shift

Mean shift is an application independent mathematical method which has found application in image segmentation. A kernel which establishes weights of the neighboring pixels is first selected, for instance a flat disk with equal weight over the coverage of the kernel or a Gaussian for which the weight assigned to neighboring pixels decays as the distance increases. Mean shift is known as a hill climbing algorithm, meaning that the kernel used is shifted to a higher-density region until the algorithm converges. The shift between iterations is defined by a mean shift vector, which always points in the direction of the maximum increase in density. Thus in areas for which the pixel values are indicative of the features which we wish to detect, we are able to generate an output image which has segmented the image to indicate the parts of the image most likely to contain information of interest.

7.2.2.3 Statistical Region Merging (SRM)

A hard threshold level may not always be known a priori, and SRM may be applied to determine similar pixels without this prior information. SRM uses a selected set of statistical parameters to determine which neighboring pixels are likely candidates for grouping based on their pixel values. All pixels initially begin as part of their own individual region. If a neighboring pixel falls within the statistical range, then it is merged into the current pixel. Thus, an added definition of neighboring pixel expands from spatial proximity and is enhanced to include information about the content of the pixel as well.

7.2.3 Analysis of Candidate Regions

The data which will come from the sensors representing the target(s) has been indicated to lie on a noisy background with brighter regions representing the target(s). We utilize image segmentation in order to separate the target from the background. In our current implementation, we seek to find a binary image where black pixels indicate background, or the absence of a target within that pixel, and white pixels to indicate foreground, or that a target was found to be contained at least partially within the pixel. Our segmentation methods have parameters which may be selected to indicate candidate pixel regions of the image which may contain a target while the target separation operation described in the next section takes the indicated candidate regions and compares to templates representing the characteristics of the target and makes a final decision. Essentially, the segmentation step is an intelligent preprocessing step while the target separation will employ more accurate methods to reach a final determination about the target's position within the image.

Once we have an image segmented by one of the above methods, we must determine which of the indicated candidate regions truly contains a target or targets which we are searching for. This portion of the process is very application specific and further depends on the characteristics of the target in the image space for which we have obtained data. In essence, we must determine a template which is able to separate the target data from the candidate points given by the naïve segmentation. This can take the form of a multi-spectral signature, possibly combined with information about the expected size of the target, or in a more complex implementation may use multiple frames to factor in the speed of the target over time.

7.3 Three-Dimensional Pose Recovery

When the target is only within the FOV of a single sensor, the accuracy of the estimated location of the target and path in the X and Y coordinates is diminished and an estimate of the Z coordinate is impossible without additional prior information (such as target size) and inaccurate even in such a case. If the target is within the FOV of two sensors, we may perform a 3D pose recovery of the estimated location of the target, using the combined information from the two sensors in order to calculate an estimate of the Z coordinate of the target while improving the accuracy of the estimated X and Y coordinates. This process is known as fusion and our work is to develop a number of methods which will process this fusion data from the available sensors.

We have developed two approaches of calculating the target position from the simulated images, focusing on the second case mentioned above where the target is visible to more than one sensor. The first approach is shown in Figure 7.6 where we see the results of the method, where an update of the visualization of the calculations may be seen via an estimate of the target's location along with an error boundary.

The second method is based on geometric triangulation of the calculated data. The method is based on the known parameters for sensor, lens, baseline (distance between the two sensors), and pixel location of the target. Further discussion on the analysis of error in our estimate continues in the following section.

Each sensor is able to gather independent information about the scene, and a combination of data from more than one sensor can result in a more accurate position estimate than from a single sensor alone. We have already implemented one fusion method which has shown to greatly improve the accuracy of the 3D pose recovery of the target and is discussed in further detail below. Further development is needed to determine which of a number of possible features and processing steps we may employ will result in the greatest accuracy without overly lengthening the computation time.

7.3.1 Three-Dimensional Pose Recovery

Once we have isolated the target within the image, we further desire to georeference the target. In other words, we desire to change a knowledge of the target's position from image coordinates to world coordinates. Where isolation of the target within the image did not necessarily require a priori knowledge, 3D pose recovery is heavily dependent on a priori knowledge of several aspects of the sensor, its orientation in space, and the conditions under which the image was captured.

When the target is only within view of a single sensor, the accuracy of the position estimates of the components which are orthogonal to the optical axis may retain a high degree of accuracy while the component which lies along that axis is inaccurate even with a priori information about the target, such as its size, and even more inaccurate without this knowledge. Some reduction of accuracy is possible over a period of time of observing the target as the orientation of the sensor changes. This single-sensor method we do not discuss further in our work but we focus on the multi-sensor case.

If the target is within the FOV of two sensors, we may perform a 3D pose recovery of the estimated location of the target, using the combined information from the two sensors in order to calculate an estimate of the Z coordinate of the target while improving the accuracy of the estimated X and Y coordinates. This process is known as fusion and our work is to develop a number of methods which will process this fusion data from the available sensors. We have developed two approaches of calculating the target position from the simulated images, focusing on the second case mentioned above where the target is visible to two sensors.

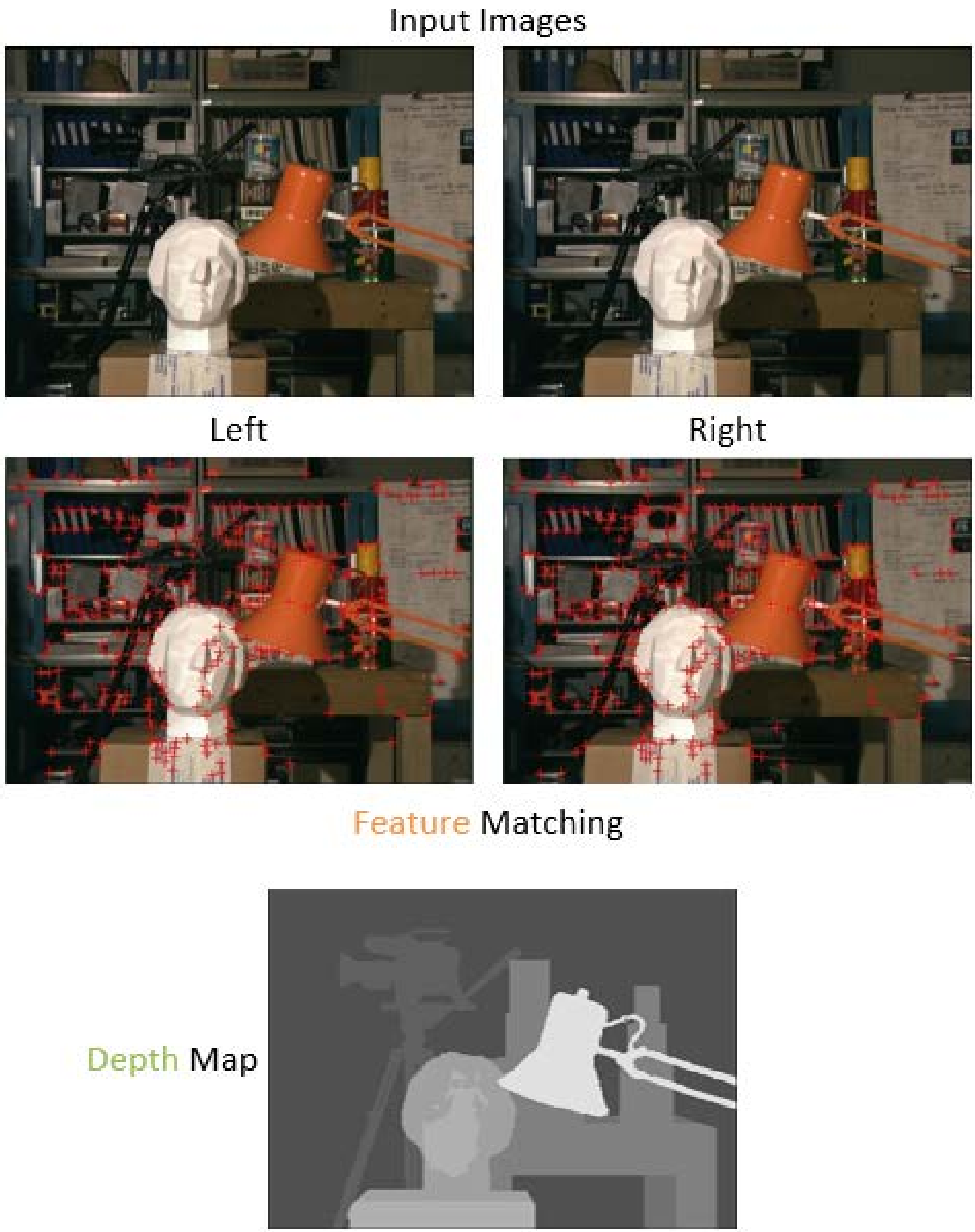


Figure 7.4: *Image demo of 3DP approach. Our approach is based on traditional 3D stereo which calculates a depth map of the scene which is then used to determine the system pose and coordinates of the target. The top row shows the left and right stereo input images of the scene. The middle row demonstrates the use of feature extraction which extracts stable features from each image and matches them to one another, then from this information a depth map can be constructed, shown in the bottom image. The drawback of this approach in our scenario is the high computational time and low scene texture from which to extract features.*

7.3.2 Triangulation

The second method is based on geometric triangulation of the calculated data. The method is based on the known parameters for sensor, lens, baseline (distance between the two sensors), and pixel location of the target. Each sensor is able to gather independent information about the scene, and a combination of data from more than one sensor can result in a more accurate position estimate than from a single sensor alone. We have already implemented one fusion method which has shown to greatly improve the accuracy of the 3D pose recovery of the target and is discussed in further detail below. Further development is needed to determine which of a number of possible features and processing steps we may employ will result in the greatest accuracy without overly lengthening the computation time.

Since we employ simulated imagery based on high-accuracy measurement of target trajectories in our experiments, we have access to the true location of the target and are thus able to determine the effect of varying the parameters of the experiment on the accuracy of the position estimate. Thus far in our experiments we have used representative parameters for what have able to determine are typical for sensor resolution, baseline distance, FOV, etc. and some results are shown below. It is also possible to use the same format we have developed for simulation to analyze worst-case scenarios as well as the enhancement in accuracy that proposed improvements to the sensor hardware, number of sensors, etc. would provide.

In Figure 7.6 we see four samples at the start of the experiment. The green points indicate the target's true position while the red points indicate our raw estimate at each time step. The purple box indicates the confidence bounds in each direction. As may be seen, the confidence in our calculated estimate in the Z position is much larger than our confidence in either the X or Y positions. In order to reduce the error in our estimate, we apply a fusion of the estimated positions, beginning with a simple windowed averaging method.

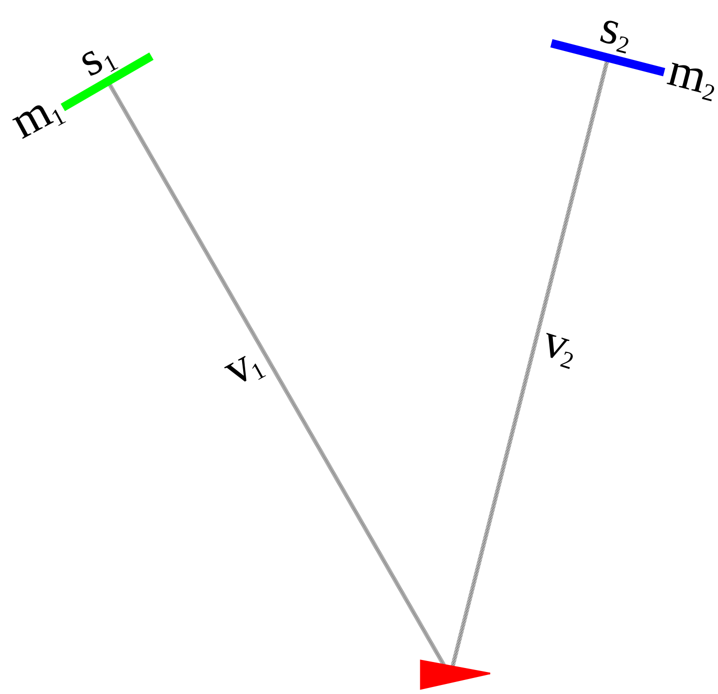


Figure 7.5: *Image demo of projected lines triangulation approach. A line perpendicular to the image plane (along the optical axis) is projected towards the target and a minimum distance to intersection is computed. This approach is computationally much simpler and allows for easy scaling to multiple sensors. Potential sources of error arise if there is significant error in the knowledge of the position and orientation of the sensor.*

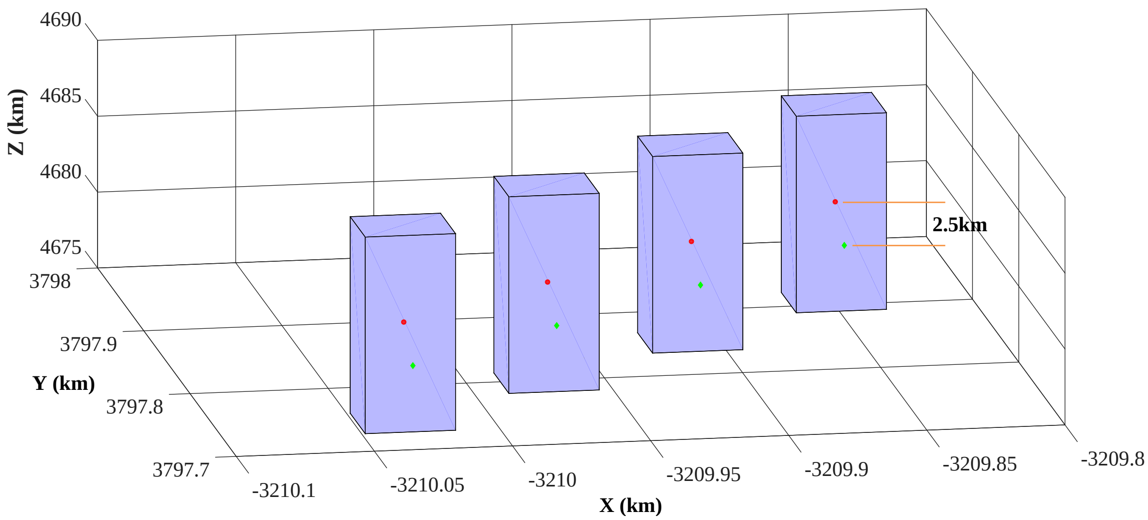


Figure 7.6: *Triangulation based 3D pose estimate of target position - side view. Four raw samples from the beginning of an experiment are shown for clarity. The green points are the true location of the target while the red points are the center of our estimate with the purple box indicating our confidence interval of the estimate. Our estimated recovery of the Z position has a much wider confidence bound and difference from the true location of the target, shown above as approximately 2.5 km away.*

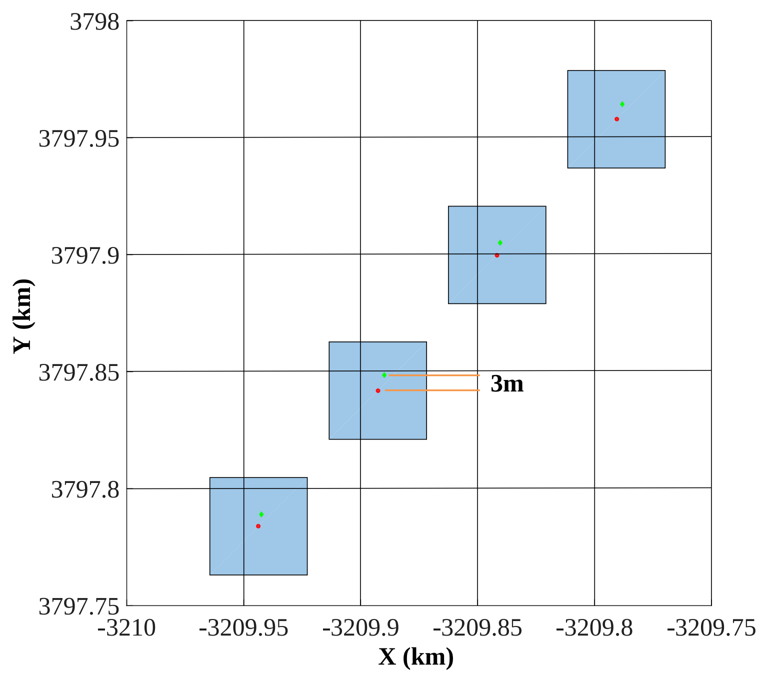


Figure 7.7: *Triangulation based 3D pose estimate of target position - top view. The green points are the true location of the target while the red points are the center of our estimate with the purple box indicating our confidence interval of the estimate. The difference between the estimate of the true X and Y position relative to our estimate of the position is in the range of 3 m, a difference of three orders of magnitude in accuracy. This is a typical problem encountered in triangulation pose recovery due to parallax, resulting in a far less accurate estimate along the axis parallel to the optical axis. The X and Y axes are perpendicular to the optical axis and the accuracy of the Z estimate is effected by the size of the baseline.*

Shown in Figure 7.8 we see the raw measurement data from the triangulation-based 3D pose recovery in red, indicating the individual estimated Z value of the target across the time of the simulation. The raw estimates are noisy and span a range of approximately ±3 km, still a rather decent range for a system with low resolution, noisy sensors located in the range of hundreds to thousands of kilometers from the target. The green line indicates zero error, which we are hoping to achieve and would indicate a completely successful recovery of the target's path from the IR imagery. The blue line indicates the error of our fusion estimate with respect to our known target path. The error starts higher than the remainder of the line at about 1.5 km away from the true position and then quickly drops to a range within approximately ±½ km of the true target position.

Figure 7.7 shows the same data as Figure 7.6 from the top to highlight the accuracy of the X and Y coordinates relative to the Z estimate, which is to be expected due to the parallax problem inherent in 3D pose recovery. As seen in Figure 7.7, the X and Y estimates are much more accurate, coming within the range of 3 m. The higher accuracy is effected more heavily by the parameters of the sensor, with increased resolution corresponding to increased positional accuracy in the X and Y components of the position. The accuracy of the Z position component is effected more heavily by the size of the baseline relative to the X and Y positions.

Figure 7.8 shows results similar in layout to Figure 7.6 but for the X position of the target. As before, the accuracy of our fusion estimate is not only significantly higher, but also quickly settles from the initial improved position estimate to a small range of ±3 m.

7.3.3 Fusion of Position Estimates

We have discussed how to recover an estimate of a target's 3D location in space, demonstrated using multiple processing methods. Once we have these estimates at each time step, we can apply various fusion methods to minimize

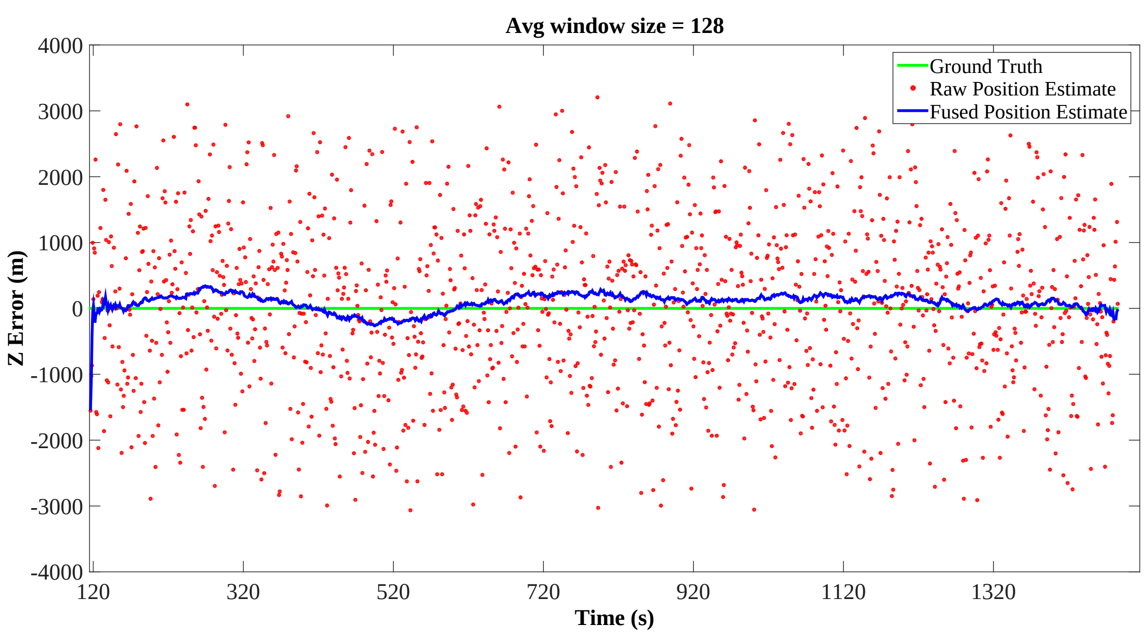


Figure 7.8: *The data shown here comprise an entire experimental run. The red samples are the error of the raw estimates of the target position relative to the green line, which indicates zero error in the estimate of the target's Z position. The blue line is the result of our application of a windowed averaging fusion with a window size of 128 samples. Note how quickly the error drops off and how much closer to the true position the estimate is from the range of the original red samples. The range of the red samples is ±3 km while the blue line ranges between ±½ km.*

the error in our estimate of the target's location with respect to the true location of the target. In this section we demonstrate the effectiveness of several methods on this type of data.

7.3.3.1 Windowed Average Fusion

The simplest approach to fusion of the position estimates is to utilize an average of the individual samples in order to smooth out the effects of noise in the estimates due to errors accumulated during calculation. This approach is very effective in smoothing out the over- and under-shoot errors due to the parallax problem of estimating the target depth which was commonly observed in experimental results. In order to maintain a stable estimate of the position, a window of size n is specified for which the fusion result is comprised of the current raw position estimate and the $n-1$ previous position estimates, up to the maximum number available (a limit is encountered during the early acquisition phase).

The results of this approach are represented in Figure 7.8. This approach is particularly effective when samples of the target position are available with relative temporal frequency. If there is too much delay in the spacing between samples, meaning that the available information being included in calculation of the fusion result is too old, the usefulness of the result degrades.

7.3.3.2 Kalman Filter-based Fusion

The Kalman filter [Kalman1960] is a recursive estimator that was introduced as a method of filtering noisy, error-prone data to produce an estimate of an unknown variable that is more precise than a single given measurement. The filtering process, also known as linear quadratic estimation (LQE) operates by using Bayesian inference and estimates a joint probability distribution of the sample data at each time step [Roweis1999]. The model assumes the true state of the variable at time k evolved from the state of at time $k-1$ using

$$x_{k+1} = F_{k+1}x_k + w_k \tag{7.1}$$

where x_k is the state vector at time k,

F is the state transition model,

w_k is the process noise, typically assumed to be zero mean, multivariate normal with covariance Q_k.

Equation (7.1) is known as the process equation [Cuevas2005]. At time k an observation z_k of the true state x_k is modeled as

$$z_k = H_k x_k + v_k \tag{7.2}$$

where H_k is the observation model that maps the true state space to the observed space, and

v_k is the observation noise, typically assumed to be zero mean, Gaussian white noise with covariance R_k.

Equation (7.2) is known as either the observation or measurement equation [Cuevas2005]. The flow graph in Figure 7.9 demonstrates the relationship between these two equations. The Kalman filtering problem is the process of jointly solving the process and observation equations where for the case of k available samples, at time step i the problem is called filtering when $i = k$, prediction if $i > k$, and smoothing if $1 \leq i < k$.

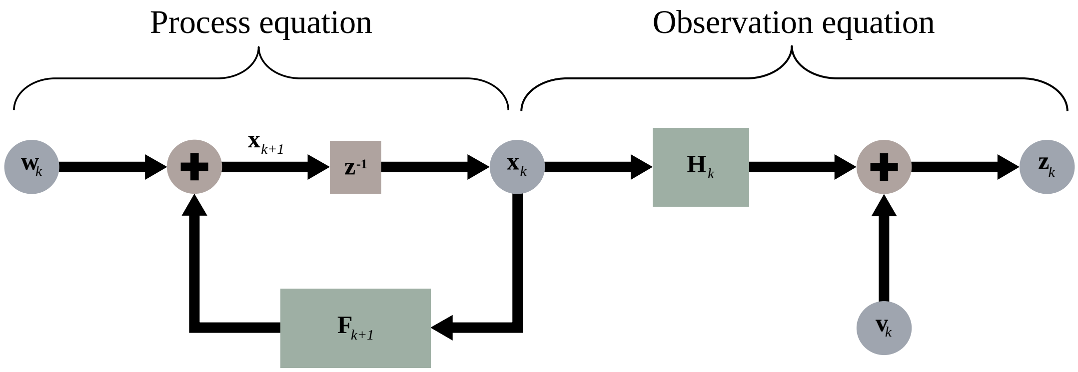


Figure 7.9: *A signal-flow graph which represents the interaction of the linear dynamic system in discrete time as described in (7.1) and (7.2). The state vector is defined as the minimum set of data which is sufficient to represent the unforced dynamic behavior of the system, meaning that the minimum amount of data of the past behavior of the system is needed to predict the future behavior of the system. This information is collected as time* $\boldsymbol{k}$ *progresses and a model of the system is updated using this accumulated information to improve the accuracy of future predicted system states. Reproduced from [Cuevas2005].*

Use of the Kalman filter in visual tracking requires calculating the object position and speed for each time step. The input images are processed to locate the target in each frame, and information about the image acquisition conditions is processed into a representation of the state of the target. Thus the system state x_k represents the target coordinates x, y, and z at time k. This approach likewise lends itself well to establishing a search window given the predicted coordinates of the target [Cuevas2005]. This allows for handing situations where the target can become occluded by an increasing noise floor or lost from the sides or base of the imaging volume.

Kalman Model 1: Velocity

For short durations, a linear estimate is sufficient to model the target trajectory. If the target is lost from the FOV of the system, the predictive capabilities of the system allow for recovery, but the accuracy will diminish if the target cannot be reacquired within a reasonable amount of time. The Kalman filter may be extended to model complex target trajectories, and models exist which account for a forcing control function which represents the ability of the target to alter the path it takes. As seen in Figure 7.10, the application of a Kalman filter utilizing a velocity model is sufficient to reduce the envelope of error compared to the raw observation samples. Where this approach has trouble may be seen in the slow correction to changes in the data, which results in persistent offset from the ground truth.

Kalman Model 2: Acceleration

The same Kalman-based approach is used, but with an adaptation to the state vector to include not only the position and velocity of the target, but also the acceleration as well. The effect of this change is readily seen in Figure 7.11 below. The approach is much more effective at responding to changes in the data, but at a cost of a noisier output. Given our application of locating the target in world coordinates in order to reduce the search volume, this effect is acceptable so long as the average error in the position estimate is reduced compared to the observations and other state estimation methods. The effect of including the estimated acceleration of the target can be seen clearly when comparing Figure 7.10 and Figure 7.11, noting how rapidly the method moves in response to new observations.

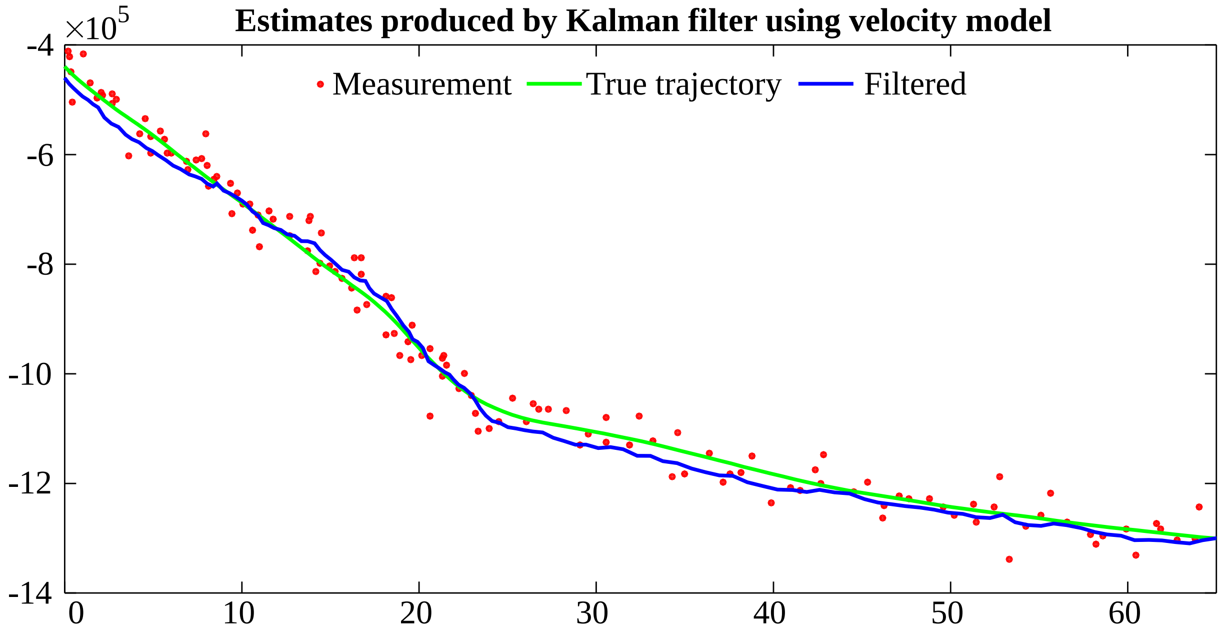


Figure 7.10: *A demonstration of the Kalman state estimation utilizing the velocity model. The green line represents the ground truth, the red dots are the individual raw state observations. The blue line indicates the filtered state estimate using the Kalman filter developed to track the system state vector represents position and velocity. Note the slow correction of the blue Kalman state estimate compared to the green ground truth line.*

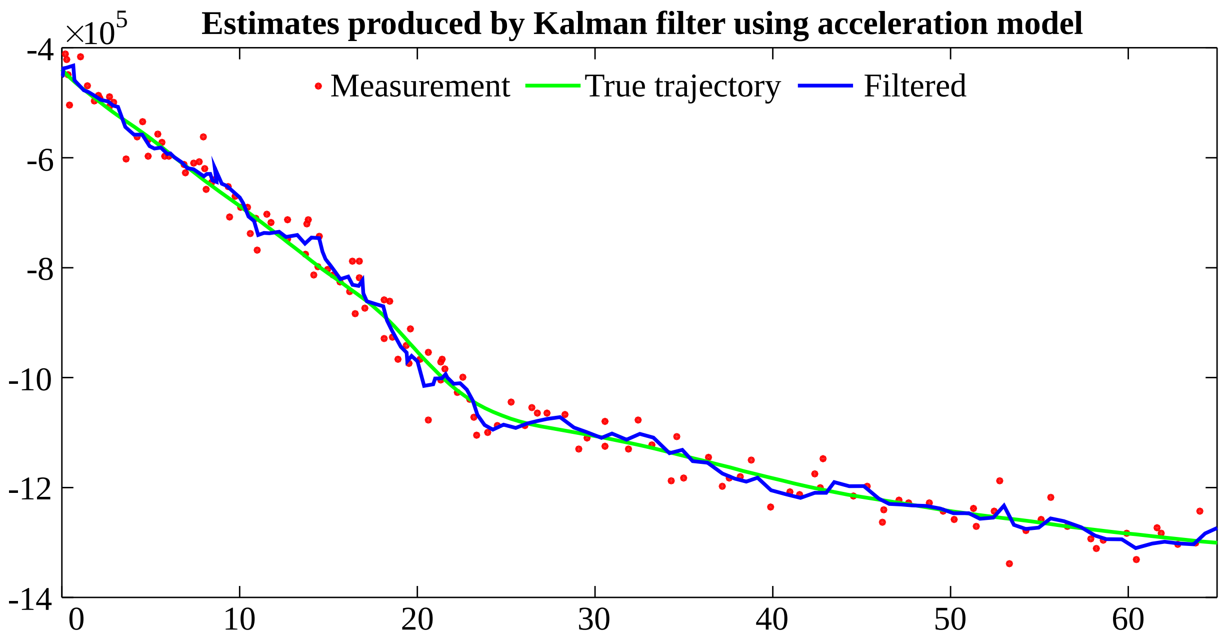


Figure 7.11: *A demonstration of the Kalman state estimation utilizing the acceleration model. The green line represents the ground truth, the red dots are the individual raw state observations. The blue line indicates the filtered state estimate using the Kalman filter developed to track the system state vector represents position, velocity, and acceleration. Note the quicker correction of the blue Kalman state estimate compared to the green ground truth line.*

7.3.3.3 Extended Kalman Filter (EKF)

The EKF is the nonlinear adaptation of the original Kalman state estimation filter. The EKF is the most widely used estimation algorithm for nonlinear systems [Julier2004] and operates under the assumption that all systems are quasi-linear. This assumption drives the approach to linearize nonlinear transformation, opting instead to utilize Jacobian matrices to represent the operation of the system. There are some issues with this assumption, however.

- The linearized transformation form results in a reliable model of the system only when the propagation of error can be represented by a linear function. There are situations where this condition does not hold:
  - Estimation of the parameters of ballistic target paths
  - Computer vision
- The process of linearization is only successful if the Jacobian matrix exists. Situations where this is the case include:
  - Systems with discontinuities
  - System parameters can change abruptly
  - A high level of quantization in sensor measurements
  - The system state is inherently discrete
- Calculation of Jacobian matrices is a difficult and error-prone process.

Julier and Uhlmann [Julier2004] propose an update to the EKF aimed at addressing these issues. Their approach is based on the idea of approximating a probability distribution rather than an arbitrary nonlinear function. Given the difficulties encountered in applying the EKF to our scenario and Julier and Uhlmann's [Julier2004] indication that the EKF is particularly ill-suited to solving the type of problem we investigate, we have selected to omit this result from comparison.

7.3.3.4 Unscented Kalman Filter (UKF)

The Unscented Kalman Filter (UKF) was developed in order to address problems seen with the EKF. The UKF utilizes the unscented transform, which is used to estimate application of a nonlinear transformation when only a finite set of samples are available for calculating the probability distribution [Julier2002]. A set of weighted sigma points are selected such that the selected points match the probability distribution from previous steps. This approach avoids the problems seen in the EKF, where implementation difficulties in calculating the Jacobian and violation of local linearity when approximating a linear representation of a nonlinear process.

The UKF was designed to address the problems with the popular EKF, and as can be seen in Figure 7.12 the method responds well to significant error in the observations while still recovering well to give a much more reasonable estimate of the ground truth position. While the method performs quite well, it comes at a significant penalty in computational time. Potential exists for accelerating the method using suitable hardware or adjusting the implementation to improve the processing time. As a baseline comparison, however, this approach takes on the order of 20X longer to calculate the state estimate compared to the windowed averaging baseline approach.

7.3.3.5 Interacting Multiple Model (IMM)

The IMM [Mazor1998][Bar2001] state estimator is a suboptimal hybrid filter that estimates the state of a dynamic system with several behavior modes and operates by switching between them. The IMM estimator can be implemented as a self-adjusting variable-bandwidth filter, making it ideal for tracking targets capable of self-directed motion. The filter balances complexity and performance. Computational requirements are linear and performance has quadratic complexity [Mazor1998]. Other configurations include cascading or parallel Kalman filter state estimators which use independent models of target motion to reduce the overall error in tracking where the target motion is complex. IMM will calculate weights for each of the independent contributing models and as such is able to make quick adjustments to track target motion.

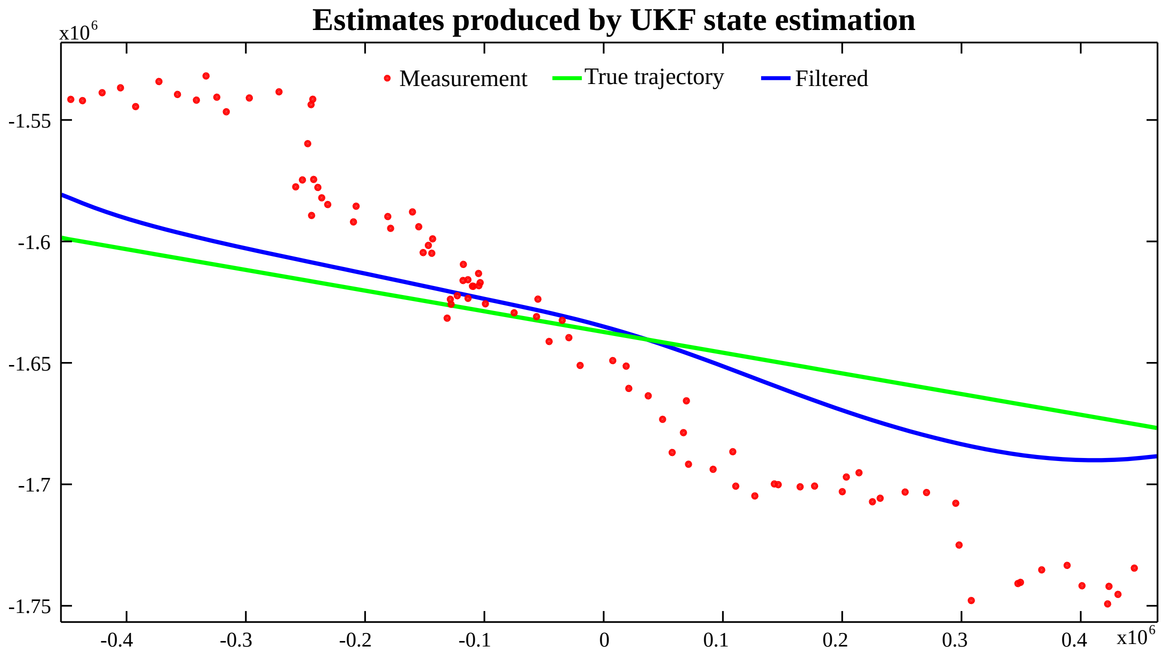


Figure 7.12: *A demonstration of state estimation utilizing the Unscented Kalman Filter (UKF) approach. The green line represents the ground truth, the red dots are the individual raw state observations. The blue line indicates the filtered state estimate using the UKF developed to track the system state vector. Note the offset of the filtered state estimate from the ground truth line, which still achieves greater accuracy than the raw samples which drift from an over to an underestimate of the true target position.*

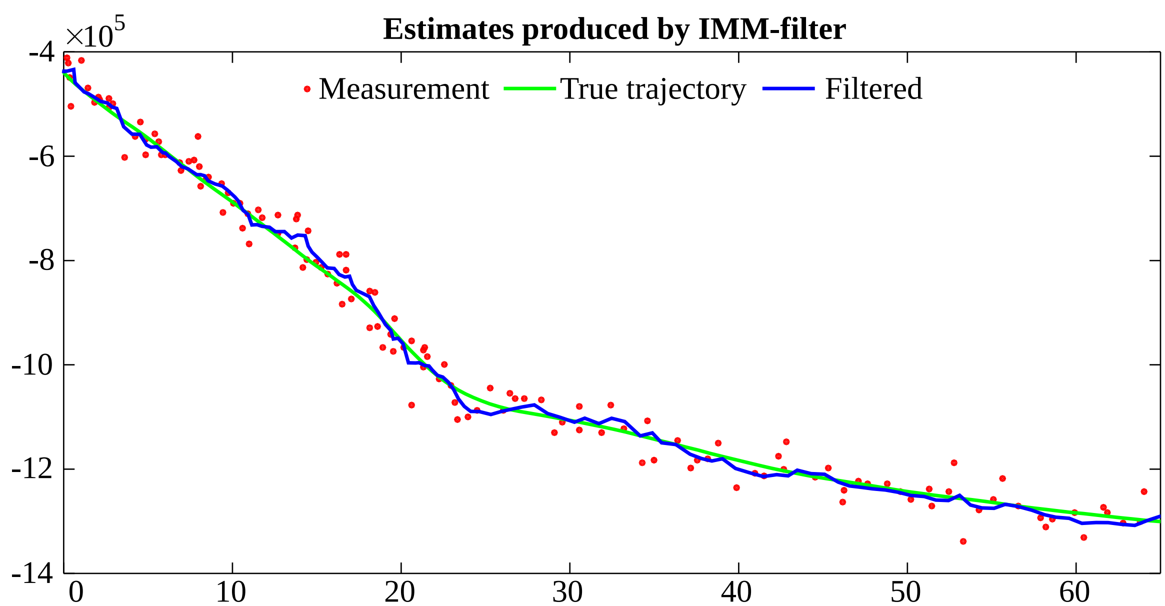


Figure 7.13: *A demonstration of the Kalman state estimation utilizing the Integrated Mixed Model (IMM) approach. The green line represents the ground truth, the red dots are the individual raw state observations. The blue line indicates the filtered state estimate using the IMM filter developed to track the system state vector using two parallel Kalman filter implementations tuned to represent combinations of unique target motion modes. Note the balance of quicker correction of the blue Kalman state estimate compared to the green ground truth line while avoiding overshoot as significant as is present in the acceleration model.*

The results of the IMM filter are an improvement compared to either of the base Kalman approaches or the UKF approach. The IMM results more resemble those of the Kalman approach using the acceleration model, but without the significant noise/jumpiness of the state estimate as seen in Figure 7.11. In comparing the difference from Figure 7.11 to Figure 7.13 above, the improvement in keeping the filtered state estimate closer to the ground truth without responding overly quickly to new observations is handled very well with the two-model IMM implementation utilized here.

7.3.3.6 Results

In evaluating the merits of the several approaches considered, we consider the performance of each in minimizing the error of the target position estimate. The data from several test runs were utilized to get a reasonable average performance of each tracking method. The summary over all experimental runs is contained in Table 7.1, where the results of each method are evaluated via the root-mean-square error (RMSE) metric. As is typical of most enhancement approaches, the simpler approach exhibited worse performance compared to the more computationally expensive methods.

In considering a recommended tracking method for smoothing out the predicted location of the target of interest based on historical observations, we note the results shown in Table 7.1 with respect to the level of accuracy only lead to a suggested application of the IMM approach. However, when considering the small reduction in accuracy with much faster processing time the use of the Kalman filter using the acceleration model provides a decent balance. At less than half the processing time of the IMM approach, there is only a marginal increase in error allowed. In considering our approach versus the radar-based tracking system, even the highly inaccurate windowed averaging fusion approach still delivers an improved result compared to existing solutions, at half the time of the Kalman – Model 2 approach. Specific choice of tracking method will depend on factors of a given real-world implementation. However, we have evaluated several options with quite impressing potential application and indicated metrics useful for weighing the tradeoffs in making this selection.

Table 7.1: *Comparison of tracking methods applied to representative simulation data. Each method is described earlier in the chapter, including figures to represent the raw position observations versus the state estimates utilizing each approach. RMSE is the root-mean-square error of the state estimates compared to the true target position. For windows averaging, three window sizes (128, 64, and 32) were selected for comparison in timing the method. The RMSE score was calculated using a window size of 128.*

Method	**RMSE**	**Computational Time**
Windowed Averaging [128, 64, 32]	8.762×10^3 m	[0.0463, 0.0439, 0.0431] ms
Kalman Filter – Model 1 (Velocity)	2.079×10^3 m	0.0904 ms
Kalman Filter – Model 2 (Acceleration)	0.390×10^3 m	0.0799 ms
Unscented Kalman Filter (UKF)	0.350×10^3 m	0.9252 ms
Integrated Mixed Model (IMM)	0.297×10^3 m	0.1915 ms

8 Error Analysis

Since we are simulating our experiment, we have access to the true location of the target and are thus able to determine the effect of varying the parameters of the experiment on the accuracy of the position estimate. Thus far in our experiments we have used representative parameters for what we have been able to determine are typical for sensor resolution, baseline distance, FOV, etc. and some results are shown below. It is also possible to use the same format we have developed for simulation to analyze worst-case scenarios as well as the enhancement in accuracy that proposed improvements to the sensor hardware, number of sensors, etc. would provide.

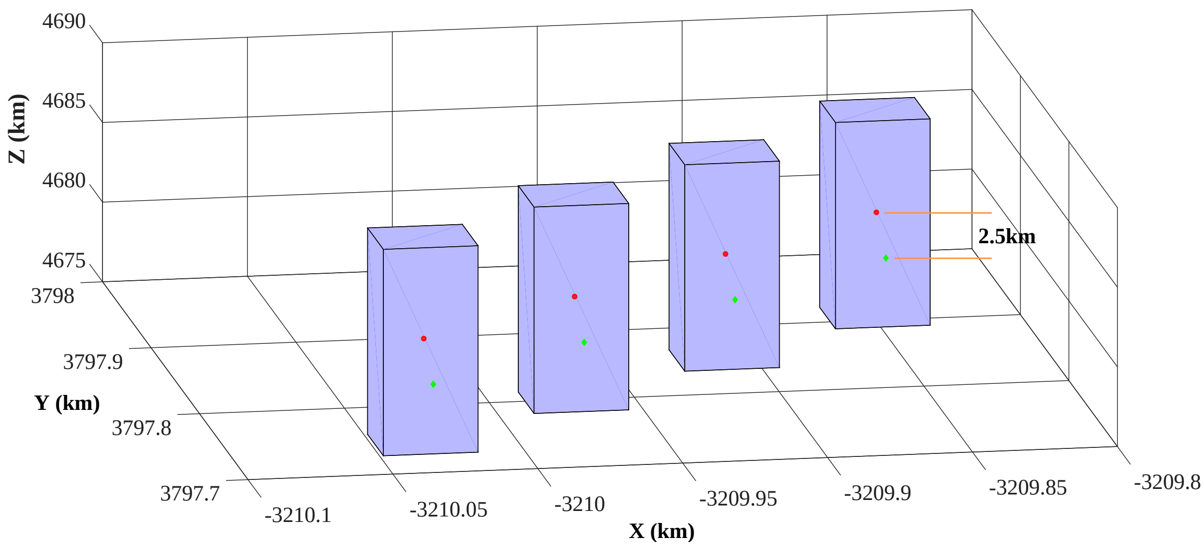
In Figure 8.1 we see four samples at the start of the experiment from both the side (a) and top (b) views. The green points indicate the target's true position while the red points indicate our estimate of the target's position at each time step. The bounding box indicates the confidence bounds in each dimension. The confidence in our calculated estimate in the Z dimension is much larger than our confidence in either the X or Y dimensions. This is expected due to the parallax problem inherent in the 3D pose recovery. In order to reduce the error in our estimate, we apply a fusion of the estimated positions, initially utilizing a simple windowed averaging approach but expanding to use of the Kalman state estimator to improve on our result.

In Figure 8.2 we represent the experimental data in a different way, inspecting the results for an entire scenario showing both the Z dimension (a) and the X dimension (b). As before, the raw measurement data is shown in red, indicating the individual estimated Z value of the target across the time of the simulation. The raw estimates are noisy and span a range of approximately ±3 km, a rather decent range for a system with low resolution, noisy sensors located thousands of kilometers from the target. The green line indicates zero error, which we hope to achieve and would indicate a completely successful recovery of the target's path from the IR imagery. The blue line indicates the error of our fusion estimate with respect to our known target path. The error in (a) starts higher at about 1.5 km away from the true position and then quickly drops to a range within approximately ±½ km of the true target position. In (b) the accuracy of our fusion estimate is not only significantly higher, but also quickly settles from the initial improved position estimate to a small range of ±3 m.

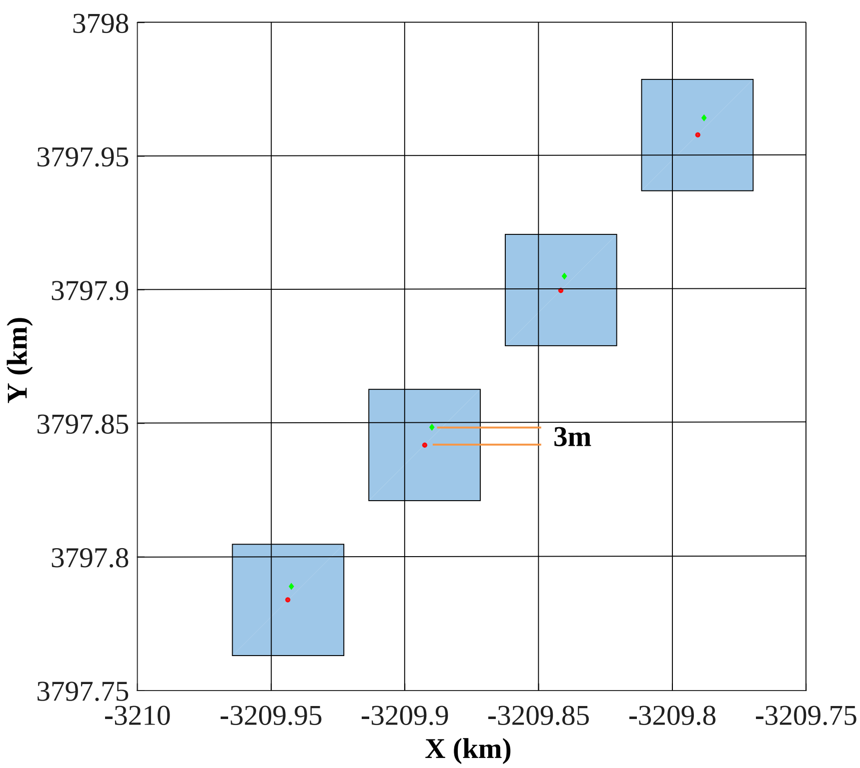
In this chapter, we conduct a performance analysis of a set of system parameters to understand the envelope of performance of an image-based target position recovery system. We do this to understand the tradeoffs of these components and to develop a recommended system design in order to best utilize the type of target position recovery approach we have presented.

8.1 Effect of Baseline

For a given experiment, we assume a fixed sensor baseline. However, we investigate the influence of the chosen baseline on the accuracy of the estimated target position in order to determine a recommended best placement of sensors. This placement will also be influenced by the importance of various areas of interest such that recommended coverage in terms of sensor baseline need not be applied to all regions, but for regions where knowledge of target position is most important, we seek to analyze the influence of the sensor baseline on recommending an arrangement of sensors. Further, our analysis will provide an understanding of how changes in the baseline impact the error so that for areas of less importance that still require coverage, tradeoff for an acceptable level of error can be made in order to establish acceptable coverage without too much cost in each deployed sensor and management overhead in utilizing more sensors than necessary for a given level of coverage and accuracy.

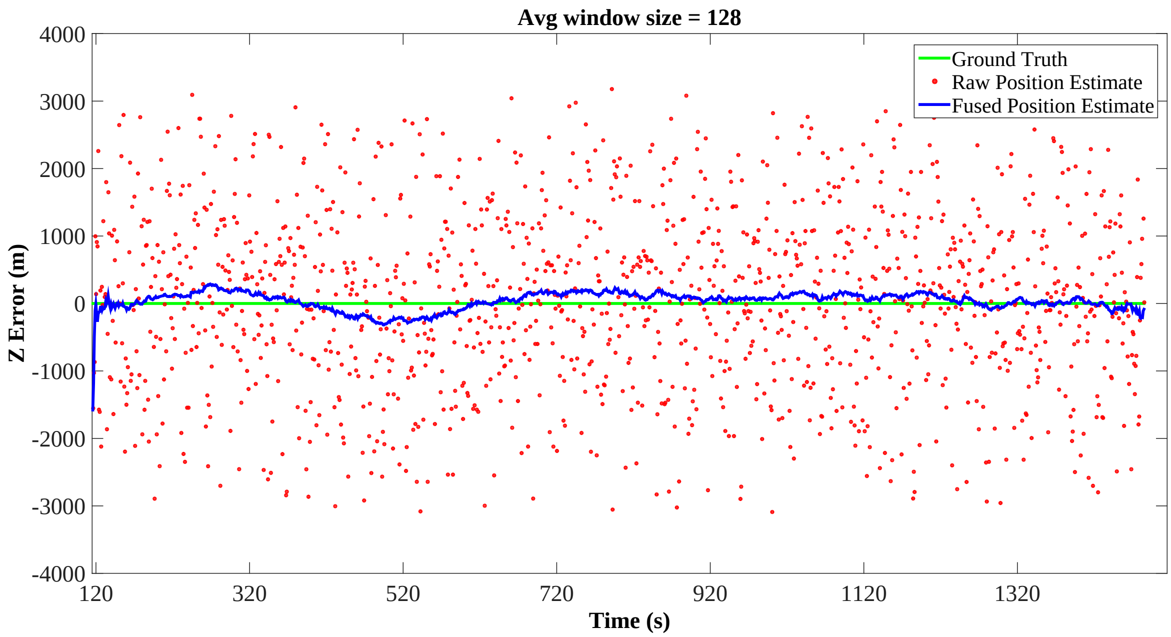


(a)

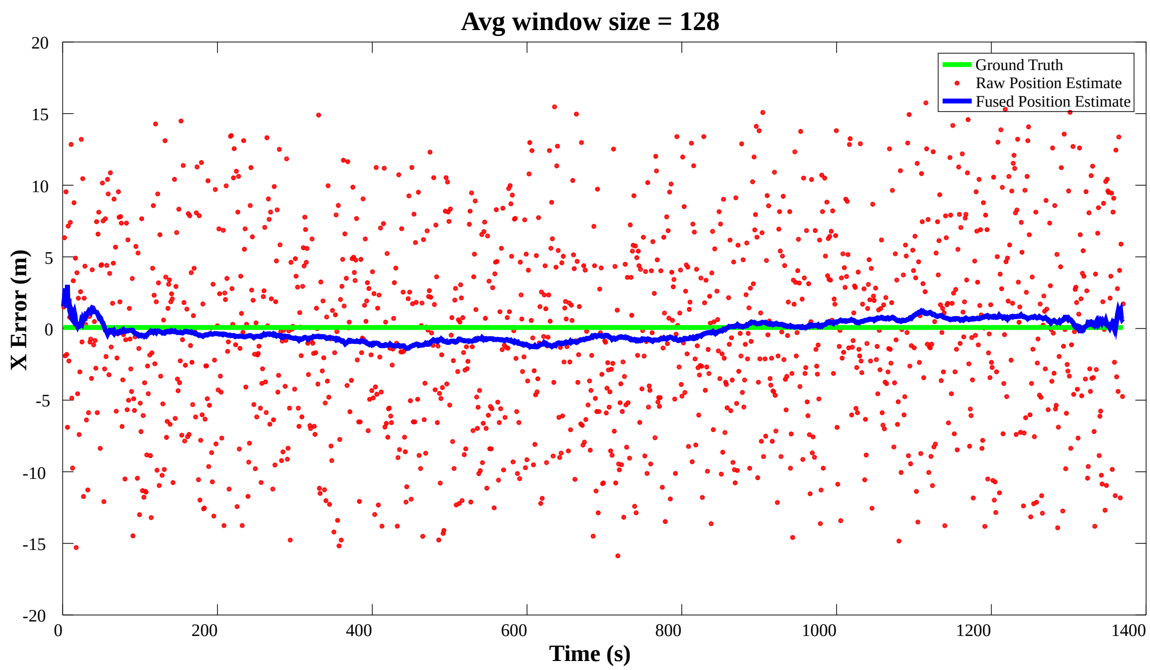


(b)

Figure 8.1: *The plots in this figure were previously shown in Figure 7.6 and Figure 7.7, repeated here for more convenient referencing. Triangulation based 3D pose estimate of target position - side view (a) and top view (b). Four raw samples from the beginning of an experiment are shown. The green points are the true location of the target while the red points are the center of our estimate with the purple box indicating our confidence interval of the estimate. Our estimated recovery of the target position has a much wider confidence bound in the Z-dimension and difference from the true location of the target, approximately 2.5 km away.*



(a)



(b)

Figure 8.2: *The data shown comprises a single experimental scenario. Displayed are the target position at each time step in the Z-dimension (a) and the X-dimension (b). The results in the X and Y-dimension are similar. The red samples are the error of the raw estimates of the target position relative to the green line, which indicates zero error in the estimate of the target's position. The blue line is the result of our application of a windowed averaging fusion. Note how quickly the error drops off and how much closer to the true position the estimate is from the range of the original red samples. The range of the unfiltered samples is ±3 km which is reduced to between ±½ km once filtered in (a) and from a range of ±15 m to ±3 m in (b), around three orders of magnitude less than the scale of error in the Z-dimension.*

We inspect the influence of a changing baseline on the accuracy of the recovered target position for each of the two 3D pose recovery approaches, generalized stereo and 3D projection. In Figure 8.3 we see the comparison of the error represented as a boxplot of error for the entire experiment as opposed to the time-based representation of experimental data shown previously. Each boxplot shows data from a single experiment where the only parameter changed is the baseline of the sensors for a sensor pair. As before, the red samples are the raw position estimates while the blue samples are the fused position estimates. Noting the change in baseline, the error increases drastically as the baseline reduces to where for a seemingly wide margin between sensor vantage points of 50km the median raw error is just under 200km and an expanding baseline to 1000km results in a comparatively more acceptable error level. This arrangement highlights the need for fusion in reducing the error since the accuracy of a single measurement can vary significantly.

For the 3D projection approach, in Figure 8.4 we first note the drastic scale difference in the methods. In the generalized stereo approach the maximum error is 1200km while for 3D projection it is only 50km. Likewise for the 3D projection approach, a narrow baseline has higher error than a wider baseline. Of interesting note, however, is the increased number of outliers for the 3D projection approach as compared to the generalized stereo approach. The maximum outliers are still more accurate than the expected error value of generalized stereo though it does indicate the potential helpfulness of applying one of the predictive models to smooth out these extremes. The results shown are for the simple fusion approach and not the more advanced tracking methods discussed in Chapter 7.

In determining an appropriate recommended spacing of the sensors, we note the diminished reduction of error with each step increasing the baseline. In order to determine the number of sensors required for a given sensor baseline, we approach from the perspective of Tammes problem by determining the number of circles of a given radius according to the baseline can be packed on a sphere [Clare1991]. However, due to the difficulty of finding a solution and the fact that an estimated number of sensors rather than a particular arrangement of sensors is our goal, we approximate the value by comparing the surface area of a sphere at a given altitude above the surface of Earth and the area of a circle of radius determined by the baseline. Our estimate is shown in Figure 8.5, where we consider an altitude of 600km, and inspect a number of baselines.

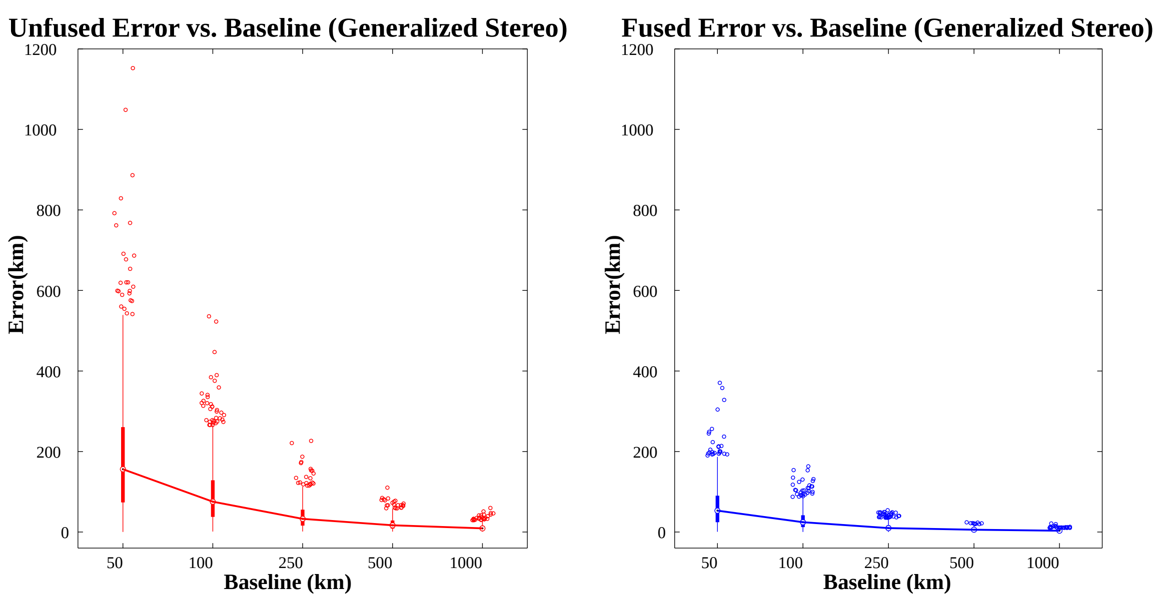


Figure 8.3: *Demonstration of the effect of fusion on reducing the error in target estimates for varying baselines. The charts above are for experimental data using the Generalized Stereo approach, where raw position estimates are shown in red (left) and the fused position estimates are shown in blue (right). An overly-narrow baseline contributes to higher error while a wider baseline has reduced expected error, though application of fusion drastically reduces the error.*

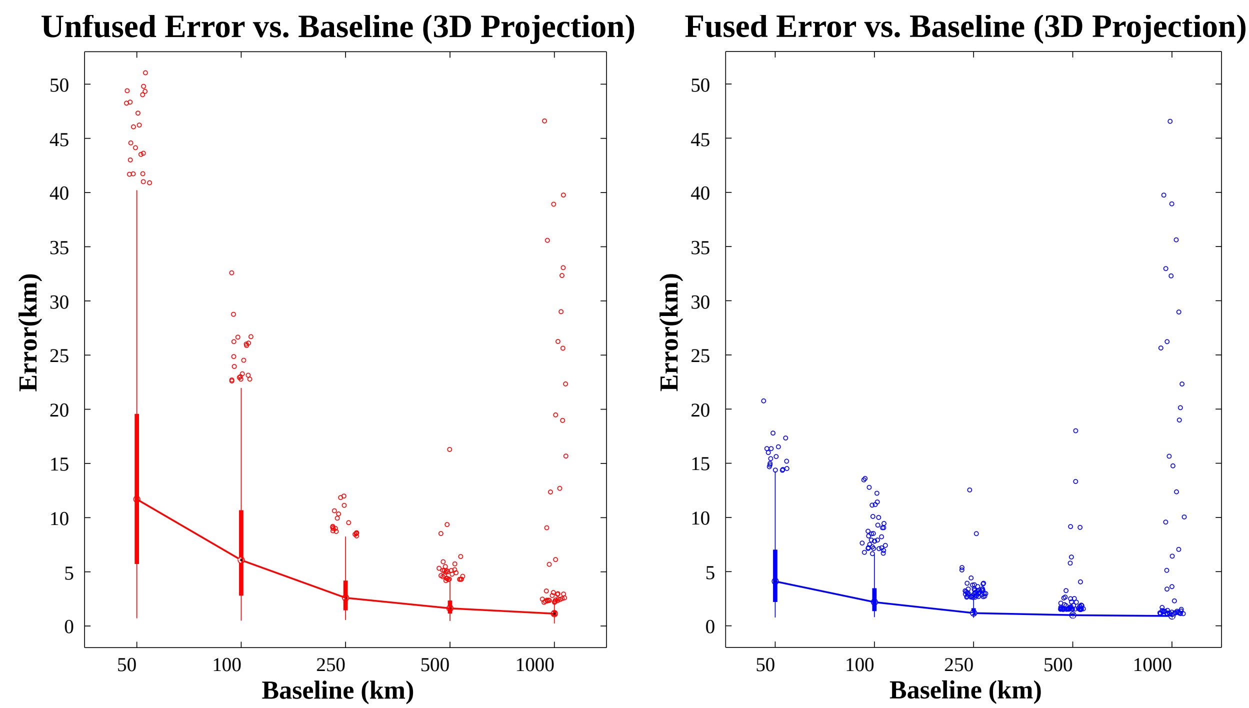


Figure 8.4: *Demonstration of the effect of fusion on reducing the error in target estimates for varying baselines. The charts above are for experimental data using the 3D Projection approach, where raw position estimates are shown in red (left) and the fused position estimates are shown in blue (right). Compared to the results shown in Figure 8.3, the 3D Projection approach is much more accurate. Similar to the Generalized Stereo approach, an overly-narrow baseline contributes to higher error while a wider baseline has reduced expected error.*

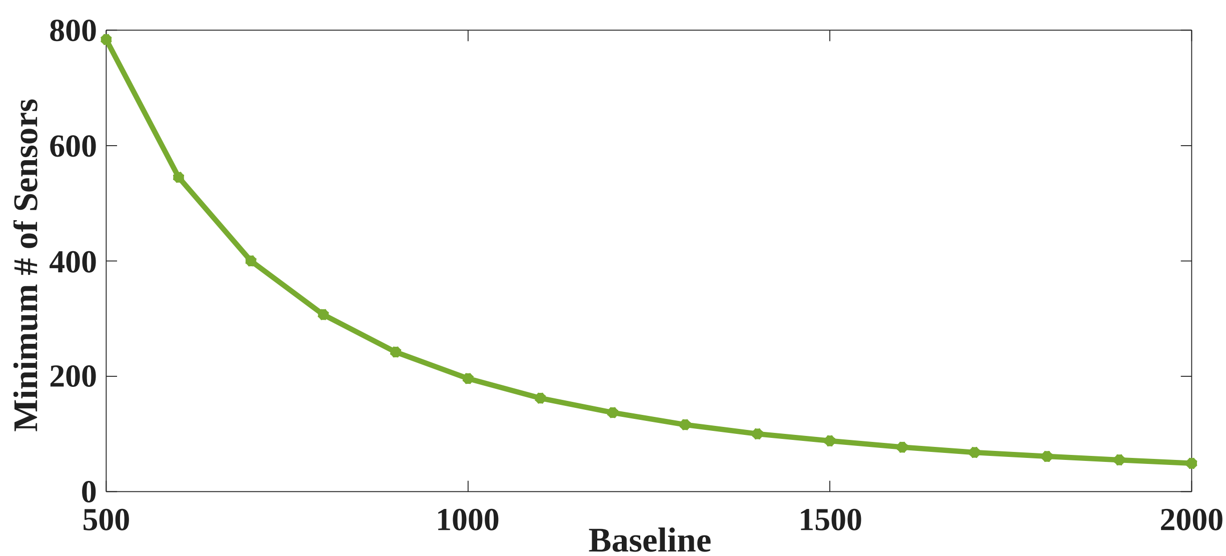


Figure 8.5: *Influence of baseline on sensor coverage area, indicating the minimum number of sensors required for global coverage. We consider a sphere with radius 7000km, based on an altitude of approximately 600km above the surface of Earth. The number of sensors required is based on the ratio of surface area of this sphere, compared with the area of a circle of radius determined by the indicated baseline as estimation of the number of sensors required.*

The minimum number of sensors required for a given altitude and baseline drops rapidly as the baseline increases. Given the configuration of the sensor with respect to resolution, lens equipped, post-processing employed, and individual cost per deployed sensor, a range of configurations can be proposed within acceptable bounds. For instance, consideration of a sensor going offline, handling cases where there are more than two sensors with line-of-sight to increase accuracy, considering configurations with sensors at varying altitudes, and handling the orbital paths of the distributed sensor constellation all weigh in on recommending a final system configuration. Of further note, for how our minimum number of sensors vs. baseline estimation was performed, the information shown in Figure 8.5 provides an upper bound on the number of sensors given that Tammes problem (circle packing on a sphere) indicates the maximum number of circles of a given radius which may be placed on a sphere.

8.2 Effect of Sensor Parameters

Although we typically refer to the entire sensor platform assembly as the sensor, in this section we specifically refer to the digital sensor chip which records the incident light as the sensor and address the considerations which contribute to the performance of the overall work which changes in the parameters of this device cause. Sensor arrays may be constructed using CMOS, CCD, or another technology such as a microbolometer array for thermal imaging. We do not consider the differences in terms of these various technologies but rather analyze the influence of the sensor in terms of two factors, pixel resolution and sensor noise, as being representative of the quality of the sensors for determining the requirements to meet a given level of error.

8.2.1 Resolution

Two uses of the term “resolution” are common when referring to image sensors. The way referred to in this section is the pixel resolution of the image sensor array. The other is the bit depth resolution of a given pixel. This could have an influence on the quality of the work, but we do not consider differing values. The data used to perform all experiments uses 8-bit images with pixel values in the range 0-255.

The resolution of the sensor is important as a higher resolution sensor means that the region each pixel covers within the FOV is smaller, and therefore localization of the target in world coordinates is more accurate. However, higher resolution also means higher computational time at each stage of the post-acquisition processing pipeline. Therefore, striking an appropriate balance is of great importance as overly long processing times allow for too much movement of the target between image acquisitions that will contribute to unstable tracking even if the error at each step is lower.

In target tracking implementations with existing technology, the minimum requirement for knowledge of a target’s position is determined by the angle, where 0.1^{o} is sufficient for tracking [Tracking1998]. Given the configuration constraints previously discussed, in determining an appropriate analog to evaluate our resolution requirements, we consider a maximum target to sensor distance and analyze the uncertainty of a target’s position when it is known to be contained within a single pixel. At a baseline of 2000km, and a sensor altitude of 2000km, the maximum distance from a target to a sensor is 2828km. For purposes of analysis we will round to 3000km to account for sensor drift and other potential issues regarding instantaneous position of sensors.

In Figure 8.6, we see an analysis of the error associated with the uncertainty of the imaging sensor pixel at a given target-sensor distance. This is related to the fact that for a given configuration a single pixel will cover a substantial portion of the FOV. If the target is known to be within a single pixel, the distance from the target to sensor changes the search volume in which we must look for the target given the level of knowledge offered by the image-based system we have proposed. Also of importance is consideration of the FOV of the lens used, where for a given configuration the area covered changes. Shown are the curves for lenses from 15^{o} to 60^{o}. The difference diminishes in higher FOV coverages as shown by the clustering of the curves. Since this coverage is sampled by the sensor, the total FOV is partitioned by the sensor into regions where the coverage of a single region is as shown below.

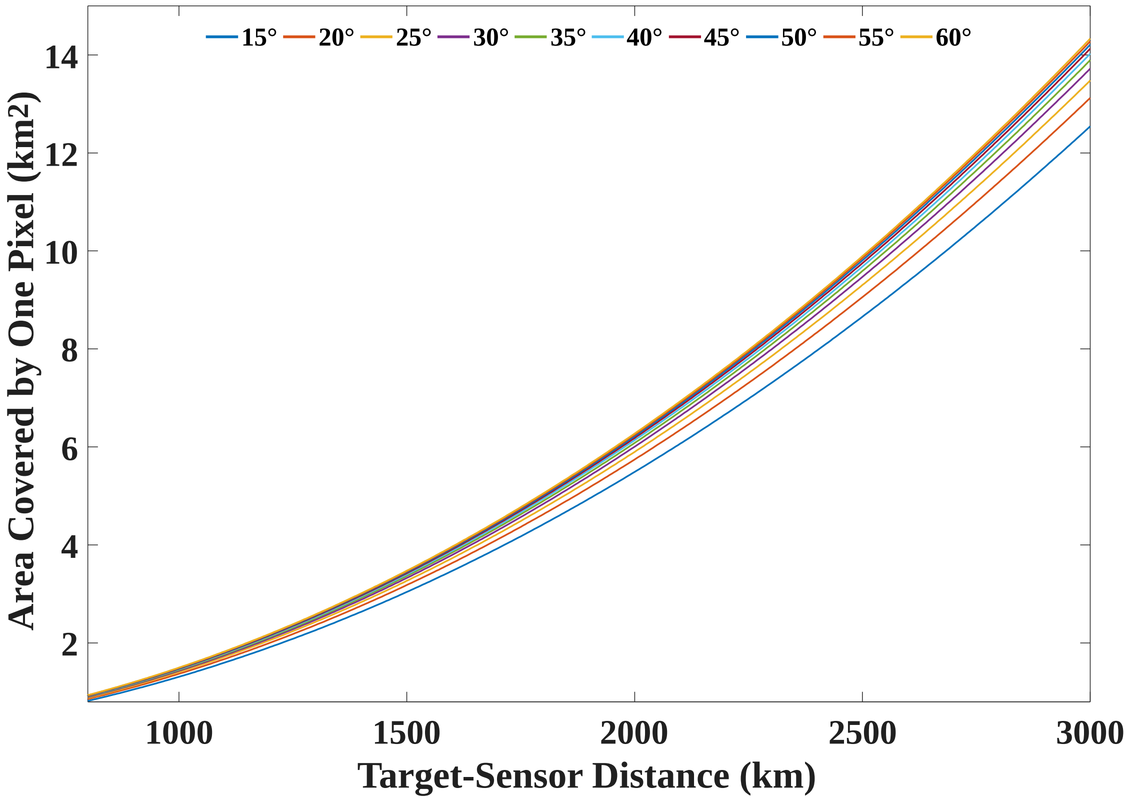


Figure 8.6: *Relationship between the sensor resolution and the distance between the target and sensor. For the size of the targets considered, a majority of the experimental time the target will be contained within a single pixel. Given this relatively low accuracy, we need to understand the error bounds on the position of the target given knowledge of its location in pixel coordinates. This chart demonstrates the relationship in terms of the distance of target to sensor for a set of lens FOVs. As expected, when the target is further away the search volume is larger, but of important note is that the difference in changing the FOV of the lens is low and diminishes as the FOV increases.*

From these results we can see the significant change in search volume that changes in the geometric relationship of the target and sensor has. Further, this suggests important consideration in deployment of a sensor constellation. Since the sensor will be oriented to inspect only a portion of its area in order to retain sufficiently accurate knowledge of the position of the target, two approaches are possible. The first involves the use of an adjustable-zoom lens. This would allow for tuned selection of the sensor's FOV to adjust from a wide-angle 'spotting' mode to a narrower 'tracking' mode. The tradeoff is mechanical complexity and additional weight, cost, and complexity from added control parameters. The second approach involves utilization of multiple layers of sensors in the configured constellation such that higher-altitude sensors could be deployed with wider-angle lenses and once a potential target has been spotted, the lower elevation sensors with narrower FOV could orient to track much more accurately.

Lastly, the relationship between the accuracy of the current technology previously mentioned compared to our proposed approach leads to a restriction on the resolution of the sensor related to the lens FOV. If 0.1^{o} is a sufficient limit for establishing a reliable track, then it stands to reason that the minimum sensor resolution would be related by

$$R_{min} \geq 10 \times \text{FOV}. \tag{8.1}$$

Meaning that for a given FOV, the sensor resolution must be at least ten times higher in pixel count. Or for a given sensor resolution, the maximum FOV of the lens is $1/10^{th}$ of the sensor size in pixels in order to meet the given tracking criterion. The chart in Figure 8.6 was constructed on the assumption of square pixels and a resolution of 480×480. Different sensor geometries will adjust the limits discussed and must be addressed accordingly.

8.2.2 Noise

The problem with any real system is reality, and for image sensors (other sensors too, for that matter) that means noise. Even with shielding, the potential for a high noise floor can inhibit accurate target localization in the image data. We want to understand to what extent the noise in the image acquired influences the accuracy of the system and to determine what approach works best for reducing the degradation in accuracy due to image noise. The images obtained of the target of interest using the existing sensor systems gives a bright point that is just above the noise floor and occasionally gets buried in the image noise. Therefore, there are several important questions we need to address:

- How much noise can we handle?
- How can we reduce the noise to an acceptable level?

An important limiting factor for determining an acceptable noise level is the influence that increased noise in the image acquisition stage has on the performance of subsequent steps. In Figure 8.7 we analyze the effect that an increasing noise level has on the quality of SR results, comparing the quality using both the SSIM and PSNR metrics. In each labeled experiment group, the only consistent parameter is the level of noise in the input images. The groups include data from experiments with different amounts of blur, source images, registration and SR methods, and SR factors. Even with this wide variety of differences, there is a clear trend: more noise correlates to worse performance. Determining a threshold for an acceptable amount of noise depends greatly on the application.

In such applications as face detection for instance, the quality of the image prior to applying this kind of algorithm would necessitate a much higher standard. For our application, the SNR simply needs to be high enough to reliably detect the target within the image. Since the target is expected to be contained within a single pixel, this means only that the target must be visible above the noise floor. Here again we are constrained by the distance of the target to the sensor, as for a uniform amount of light energy emanating from the target, as the target-sensor distance drops, the amount of light that reaches the sensor will increase. As this factor is so application dependent in determining an exact threshold in terms of these factors, we do not explore an analysis of the exact level of noise that can be handled in these terms. However, we do analyze the effect that a given noise level has on the success of subsequent steps as previously shown and also discuss the potential for configuration of the acquisition device in order to reduce the noise such that for a given acceptable noise level as determined by application-specific experimentation, steps may be taken to meet this standard.

There are three methods of noise reduction that we explore:

- Computational approaches to noise reduction
- Multi-exposure (image stacking, or synthetic extended exposure)
- Multi-aperture

The computational noise reduction methods and the multi-exposure approach have been previously described in Section 1.2.2.1, Section 2.3, and Section 3.1. An interesting novel image acquisition arrangement is the multi-aperture imaging device. Traditional image acquisition devices consist of a lens as the focusing element with finite aperture and a recording element of either film or a digital sensor as in our case. A multi-aperture system can either be composed of multiple lens-sensor pairs in some useful configuration, or utilize a single sensor with the imaging area partitioned into subsets with an individual lens and aperture projecting light onto that region. Examples of these kind of systems are shown in Figure 8.8. The advantage of a multi-aperture system is that multiple independent measurements of the scene may be acquired simultaneously from the same vantage point. This is in contrast to the capability previously described where multiple views of the target from unique vantage points allows for recovery of the 3D position of the target.

By obtaining multiple unique samples of the scene simultaneously from the same vantage point, there are several capabilities that can be developed. First, rough depth information can be obtained by utilizing lenses of different focal lengths. Also, by recording multiple samples the image noise can be reduced without encountering issues with fast-moving objects in the scene. Finally, the application of SR will be more accurate since the inter-frame motion doesn't

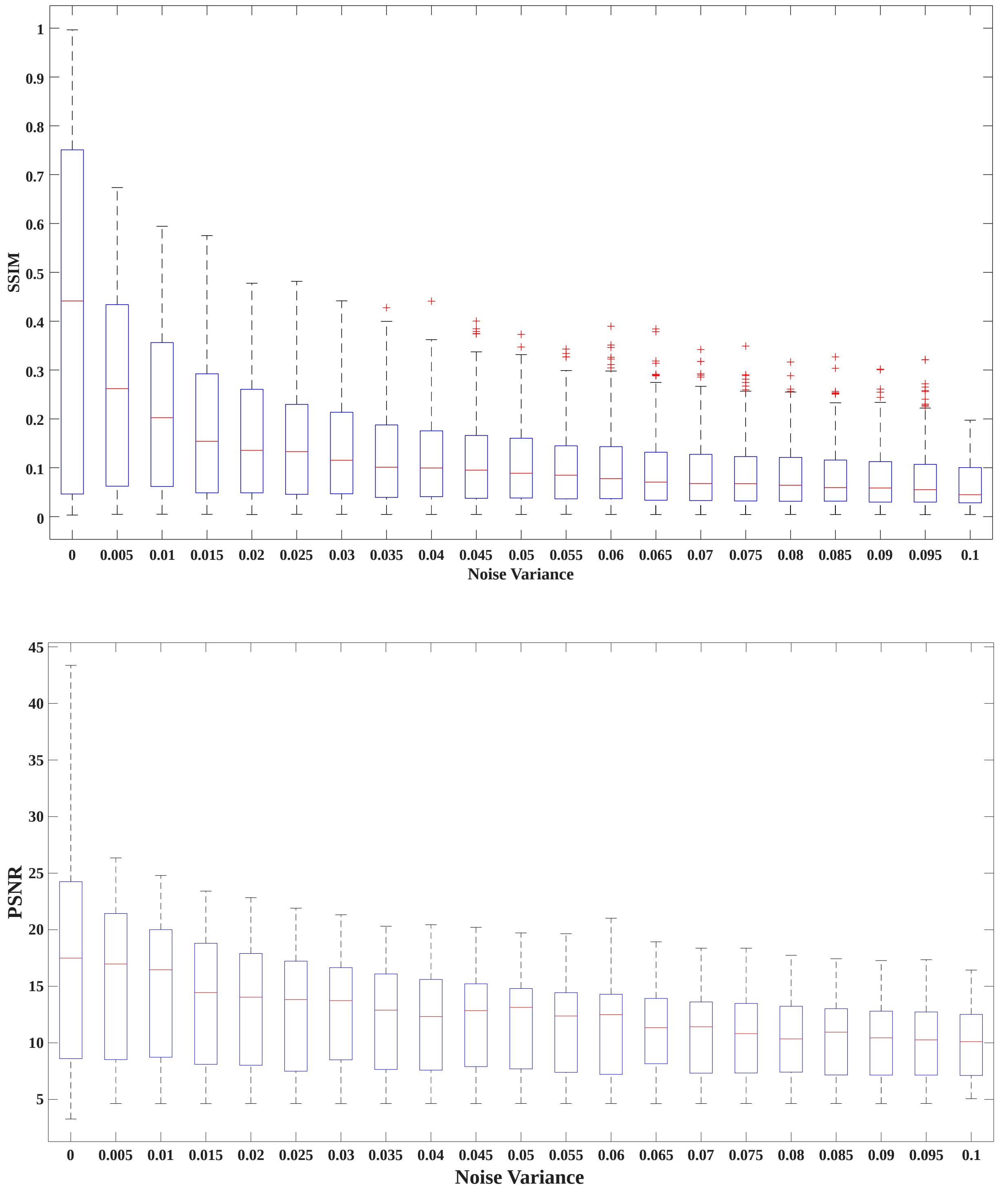


Figure 8.7: *Demonstration of the effect of sensor noise on the success of SR. Both SSIM (top) and PSNR (bottom) are used to analyze the quality of the images with varying noise levels in the input. SR methods are very sensitive to increasing noise levels. The data shown in a single boxplot include varying levels of blur, different configurations of image registration, SR methods, and SR factors. Including all of these features, there is a clear diminishing trend with increasing image noise.*

Figure 8.8: *Example of multi-aperture image acquisition devices. The recently developed Light camera (top) has sixteen individual 13MP sensor modules that give it improved light performance, multiple focal volumes, and an effective 52MP recording ability. Image credits Light.co. The 3x3 Lomography camera (bottom) uses traditional film and can be useful for e.g. action shots, if the timing between shots is staggered, passport photos, where multiple copies are needed, and multispectral imaging, if unique light filters are used at each aperture location. Image credits Light.co, Lomography.*

change, meaning it can be pre-calibrated, and the objects within the FOV are in the same position in each observation. This allows for faster processing of the SR step since the registration is known a priori. Lastly, in situations where unique spectral information of the target is helpful for identification, tuned filters may be introduced into the lens assembly so that each observation also provides a unique sampling of the light spectrum.

8.3 Effect of Blur

The use of lenses allows for a larger aperture and focused image acquisition. However, lenses also introduce blur distortions that can reduce the quality of the image. We want to identify what the influence of blur is on the accuracy of the system. Lens assemblies on existing systems add excessive weight and impact the maneuverability of the system. If simpler, lighter, cheaper optics systems be employed and comparable imaging performance achieved with corrective deconvolution, what is the limit of our capabilities? Or, how does the blur degradation impact accurate location of the target and how much computational time will application of this type of arrangement introduce? Further, can multi-aperture imaging systems allow for reduction in weight and allow for improved accuracy?

Given that the light from the target incident at the sensor is already dim enough to warrant concern about the noise floor, it is especially beneficial that all available light that gives information about the target position is correctly mapped to the appropriate location on the sensor. Well-corrected lenses, particularly of the type presently employed use large apertures and special glass to correct distortion and achieve low blur – at the expense of added weight which increases cost and decreases system maneuverability. Use of cheaper optics will introduce blur, but perhaps careful selection of a lens will enable an overall improved or competitive performance at reduced size, weight, and cost. In particular, our investigation into multi-aperture systems allows the collection of light using multiple independent simple lens systems and sensors. The images from these kind of systems can be combined to reduce the noise and any blur present will have less degrading effect since multiple simultaneous images of the scene taken from the same perspective can be combined to increase the overall SNR, while employing deconvolution of these simple lens systems to reduce the blur in the images through each of the lenses.

In order to address the concern of errors introduced by the lens system, we consider the following questions:

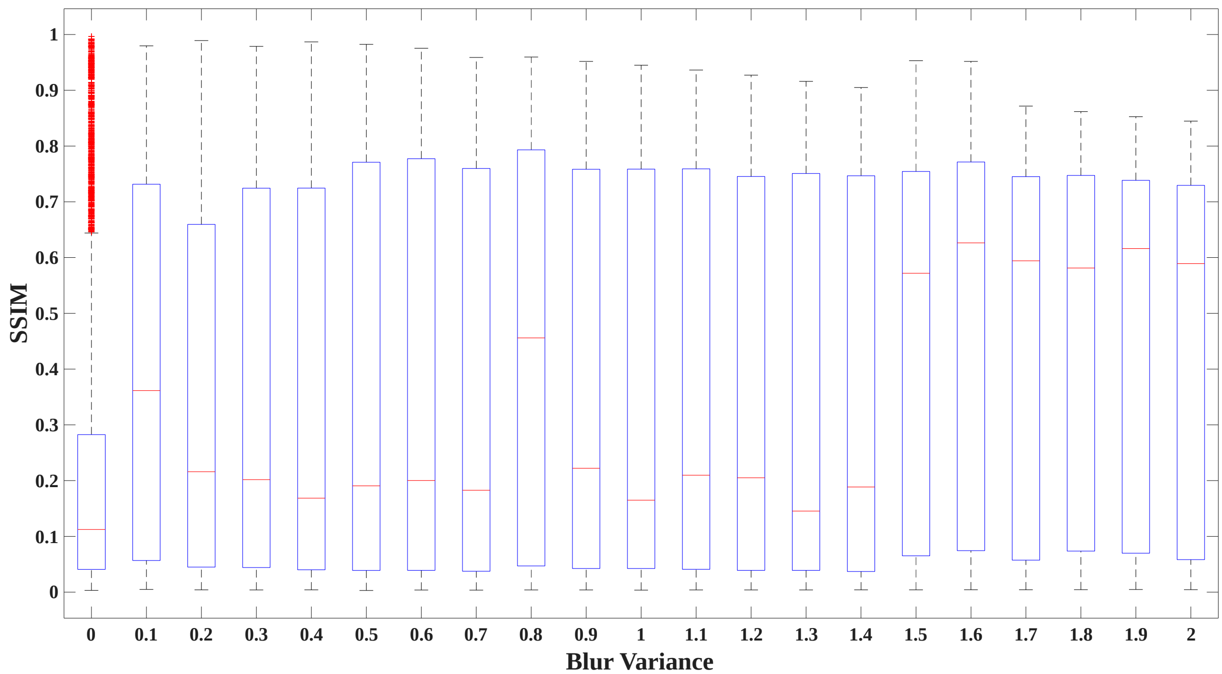
- How does blur impact the quality of the recovered position?
- Does deconvolution help? At what cost?

8.3.1 PSF Size

To address the impact that blur has on the quality of the recovered target position, we consider the bulk of the influence of the blur to be due to the extent of the kernel. For a Gaussian blur model this is controlled via the variance parameter. As discussed in Section 4.3.1, the exact form of the PSF is different based on the lens configuration and also the location in the image. As a complete discussion of every form of PSF would be very nearly prohibitive, we utilize the Gaussian model and adjust the extents via the variance. In this way, our aim is to capture the effect of blur in our application by controlling how much of the local region influences a given pixel. One type of blur that will differ from the results shown is motion blur. Motion blur, rather than being a static local degradation, is a complex measurement-dependent degradation that cannot be known a priori. For all the sources of potential error in our application, we luckily expect the probability of encountering motion blur to be relatively low.

Similar to the approach taken in considering the impact of noise on the system, we likewise use image quality metrics on the output images of SR in order to determine the impact that increasing amounts of blur will have on successful recovery of the target coordinates. Shown in Figure 8.9, we see an analysis of the experiments performed. The individual boxplots include varying noise levels, input images, registration and SR methods, and SR factors. The only constant in each is the level of blur present in the input images. Contrary to the clear impact of noise on the quality of the reconstruction quality seen previously, there appears to be little impact of blur. Most surprising is how much lower the bar is for images with no blur. There is somewhat of a trend in the upper bound seen particularly in the chart utilizing SSIM, where but for a slight increase at 1.5 and 1.6 there is a decline. The median value follows something of a pattern in the 0.1-0.7 and 0.8-1.4 ranges and then levels off at higher values.

This effect is actually quite welcome. Given the expense and difficulty of maintaining large lens systems allowing for some level of blur and being able to maintain a consistent level of quality is very helpful. Or rather it suggests that we can employ some of the multi-aperture systems which could introduce more blur in the images at the benefit of additional light and a lower noise floor, coupled with the capability to perform SR utilizing the multiple observations of the scene. Another feature that can be incorporated is the use of lenses with multiple focal lengths. Not only will we continue to be able to combine the images to increase the SNR (after appropriate scaling), but we will also be able to derive rough depth information that can enhance cueing of the tracking step as well as provide spotting information in the event that only a single sensor platform has the target within its FOV. Also, with application of multiple focal length lenses from the same vantage point, we gain the capability that having a variable focal length lens offers without the mechanical issues or delays that adjustment require as we have access to all information simultaneously.



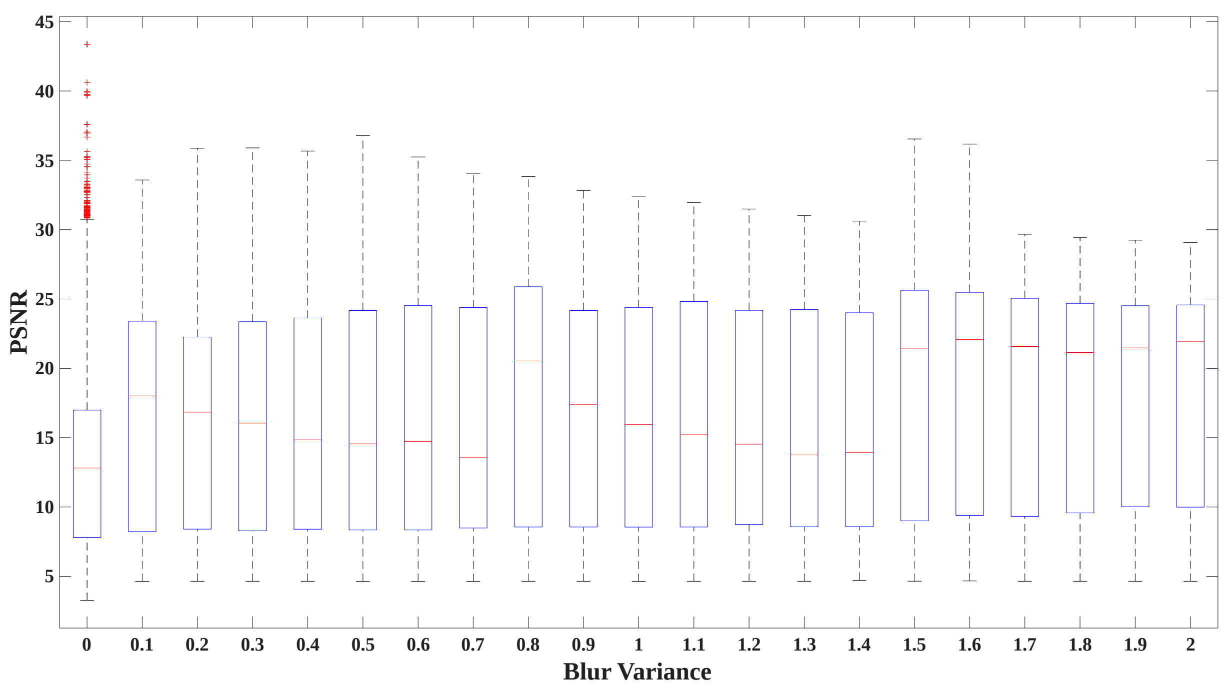


Figure 8.9: *Demonstration of the effect of blur on the success of SR. Both the SSIM (top) and PSNR (bottom) metrics are used to evaluate the image results. Curiously, it appears that if a trend exists, having no blur reduces the quality of the result. There does appear to be some pattern, notice for instance the descending trend on the range 0.1-0.7 and 0.8-1.4. Also of note, the maximum quality indicated does decline overall except for a jump at 1.5 and 1.6. There is not nearly as consistent a trend in the quality of the result as was apparent in the comparison of noise.*

These results are in keeping with the observations of Lin and Shum [Lin2004] who found that the limits of SR do not heavily depend on the lens PSF. It is important to note that although we have shown that the system blur does not heavily influence the successful recovery of the target position, there still exists the potential for the blur effect spreading useful signal information about the target below the noise floor. Therefore, it is important to note that if the tradeoff of some blur is allowed, it is even more important to ensure the noise is reduced as much as possible. This gives further incentive to employ the multi-aperture approach as it strikes a very nice balance on these two sources of degradation while keeping the weight much lower.

8.3.2 Application of Deconvolution

Despite the seemingly low impact of blur on the quality of our imaging scenario, it is still of importance to quantify the effect of deconvolution as the potential for utilizing lenses with larger and more complex PSF kernels could be explored and understanding the computational impact of deconvolution so that those images may be compared to our prior results is still of benefit in understanding the contribution that employing this class of methods will have. Further, should we encounter the situation where the signal of the target dips below the noise floor, application of deconvolution for a method well-adapted for noisy observations could be employed to enable successful recovery of the target despite excessive degradation that would otherwise have prevented calculation of an accurate position estimate.

In Figure 8.10 we show a comparison of the processing time required for deconvolution. Four deconvolution approaches were tested, along with a range of image sizes, PSF kernel sizes, and PSF variances. The processing time was most highly correlated to the image size, though there was some influence based on the support size of the kernel.

Given the active nature of the system in tracking the target, both accuracy in the location's coordinates and timely knowledge of when the target was at that location are essential to the success of our application. Therefore, we are sensitive to not only the improved accuracy contributed by deconvolution in our post image acquisition processing, but also the computational load in achieving a given reduction of error. In minimizing the overall error is appears that for all but the highest levels of blur with respect to the noise floor, that the system blur is not the most significant contributor to increased error in recovering the target coordinates.

8.4 Application of SR

SR algorithms provide the capability of increased resolution when LR images are the only observations available. We can utilize this feature in creating a multiaperture acquisition system utilizing existing sensor technology to obtain HR images of the scene without the expense that a similar resolution monolithic sensor would require. This in addition to the ability to use multiple smaller, lighter, cheaper lenses instead of a single large, heavy, expensive lens. An additional advantage to this approach is that an image of a given size in pixels obtained using a multiaperture system can have a significantly lower noise level even without utilizing the denoising steps we have discussed previously. There is some denoising that occurs naturally with application of SR, though image noise can disrupt the registration step. The benefit of a multiaperture system is that the registration parameters can be calculated a priori, but decreasing the processing time and eliminating the problems that arise with increased noise in non-ideal conditions.

Enhancing the image resolution prior to application of the tracking step reduces the search volume for locating the target. This effect is related to the SR factor in each dimension, or the cube of the SR factor. Though there will be some deviation from this direct relationship, as the error at a given timestep will change and there is some added denoising effect as a tradeoff with some processing lag, we consider the general trend of the reduced search volume to coincide with the cube of the SR factor. E.g. for an SR factor of two, the search volume will decrease by a factor of eight, and for an SR factor of three there will be a reduction in the search volume by a factor of twenty-seven.

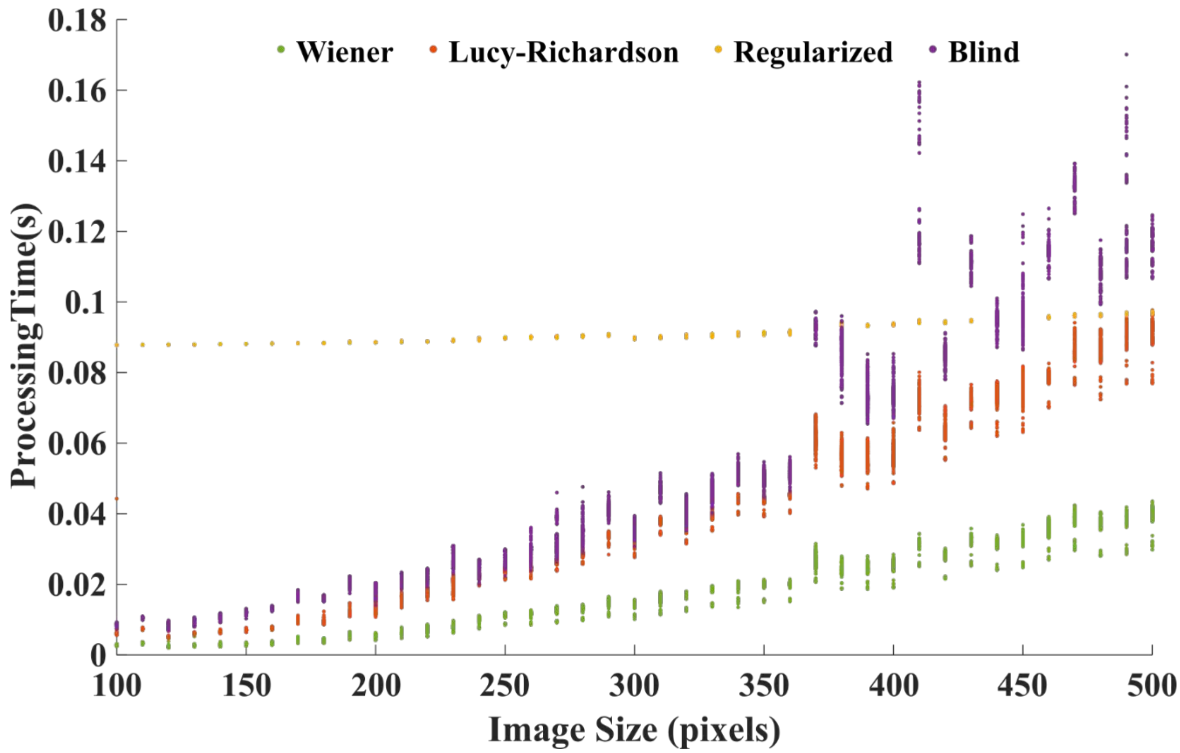


Figure 8.10: *Comparison of the processing time of four deconvolution algorithms. The size of the image being deconvolved is indicated on the chart. Within each stack are the results using varying blur kernel sizes to indicate the range and sensitivity of the algorithm to changes in these parameters.*

To get a sense of the drop in accuracy compared to the ideal, we inspect the difference in quality of an image rendered at a given resolution compared to the quality of the image at that same resolution that was constructed from LR images. As there are different registration methods that supply information about the relationship of the input images with respect to one another, we also inspect the difference that selection of one of these algorithms has on the quality of the output. Shown in Table 8.1 we compare the average quality using the SSIM and PSNR metrics. The information in this table is grouped according to registration method. The averages include differences in SR method, SR factor, noise and blur level, and source image. The data in Table 8.2 represents this same data, but grouped by SR method.

The data shown in these tables is calculated as follows:

1. The chosen source image is convolved with a specified blur kernel
2. The blurred image is downsampled with unique subpixel offset for each image at a sampling rate determined by the desired SR factor
3. The blurred, downsampled images have noise of a specific level added to corrupt the images
4. The blurred, downsampled, noisy images are registered using one of the selected methods
5. The degraded LR images and registration values are supplied to a given SR method
6. The output of the SR method is compared to the source image using the SSIM and PSNR metrics

Then, for each categorical grouping, the results are averaged to obtain the values shown. This gives a rough idea of the capabilities of each method across varying conditions. In Table 8.1 we note the somewhat surprising observation that the shift and rotate registration methods have comparable results to the shift only (global translation) methods. Further inspection reveals that the rotation vector is small relative to the shift estimates.

Table 8.1: *Comparison of image quality with respect to the image registration method used. Data includes variations in SR method, SR factor, source image, and blur and noise levels. The methods fall into one of two classes: global translation (shift only) and shift and rotate (global translation and global rotation considered simultaneously). Despite the LR input images differing by only translation, the methods which considered both global shift and rotation (notably Keren and Vandewalle's approaches) had the best results.*

	Registration Method	Mean SSIM	Mean PSNR
Global Translation	Lucas-Kanade	0.299	14.92
	Keren	0.258	14.36
	Marcel	0.324	15.44
	Vandewalle	0.366	16.36
Shift and Rotate	Keren	0.370	16.42
	Marcel	0.283	14.60
	Lucchese	0.283	14.60
	Vandewalle	0.364	16.30

Table 8.2: *Comparison of image quality with respect to the SR method employed. The data includes variations in registration method, SR factor, source image, and blur and noise levels. The values across all these variables confirm the observations of the authors in selecting SR methods in previous work. One surprising result, however, is the high value of cubic spline and deblur. This approach is an interpolation and deconvolution approach which does not take the inter-frame motion into account. One explanation for this result is the application of deconvolution in boosting the quality of the images compared to other methods.*

SR Method	Mean SSIM	Mean PSNR
Bicubic Interpolation (Bi3)	0.452	19.11
Iterated Back Projection (IBP)	0.208	10.37
Papoulis-Gerchberg (PG)	0.098	10.02
Zomet Robust SR (ZRSR)	0.266	13.07
Projection onto Convex Sets (POCS)	0.462	20.24
Normalized Convolution (NC)	0.493	21.32
Cubic Spline and Deblur (CSD)	0.538	21.62
Fast Robust SR (FRSR)	0.031	7.24

Of the registration methods, it appears from this data that we would expect the approaches developed by Vandewalle et al and Keren et all to give the best performance. In Table 8.2 we see the same results grouped according to the SR method employed. There are some methods which simply give poor results, notably Farsiu's Fast Robust SR approach. One surprising result is the Cubic Spline and Deblur method having such a high value relative to the methods considered. The surprise of the authors is due to the fact that this method does not consider the registration information in evaluating its result. It is possible that the deblurring performed accounts for some of the boost in average performance.

In order to better understand the interaction of these methods, we inspect the methods paired with one another. This comparison is shown in Table 8.3 using the SSIM score and Table 8.4 using the PSNR score. In these tables we highlight the sixty-four individual pairings of registration and SR method and compare the average results across differing source image, SR factor, and noise and blur levels using the indicated quality metrics. We highlight the pairings with the highest rankings, specifically those in the upper quartile. Our previous observations are correct with respect to the methods which tend to do better across varying conditions. Of interesting note is that while bicubic interpolation ranks in the upper quartile when utilizing the SSIM score, it does not when using PSNR. It is interesting that a simpler interpolation method ranks so well with more sophisticated SR methods.

Even though the pairing of methods suggests an expected quality level, we need to understand the computational time required for these methods. We consider the registration process as a calibration step and evaluate only the SR methods indicated as most likely to improve the quality of the images: bicubic interpolation, projection onto convex sets, and normalized convolution. The results of this timing analysis are shown in Figure 8.11, and the experiments were performed on a machine with a quad-core Intel Xeon E5507 2.27Ghz processor with 12Gb of RAM. The Windows 7 OS was used with MATLAB implementations of the algorithms.

Table 8.3: *Comparison of the average SSIM image quality score with the indicated pairing of registration and SR method. This table gives more detail in understanding the best methods to use when combining LR images into a HR representation for purposes of target localization. Values in the upper quartile are in bold.*

		Registration Method							
		Global Translation				**Shift and Rotate**			
		Lucas-Kanade	Keren	Marcel	Vandewalle	Keren	Marcel	Lucchese	Vandewalle
SR Method	Bi3	0.431	0.339	0.409	**0.571**	**0.577**	0.360	0.360	**0.545**
	IBP	0.169	0.152	0.286	0.219	0.217	0.204	0.204	0.216
	PG	0.109	0.091	0.045	0.141	0.157	0.050	0.051	0.141
	ZRSR	0.221	0.191	0.367	0.264	0.266	0.276	0.276	0.267
	POCS	0.428	0.345	0.466	**0.559**	**0.570**	0.385	0.385	**0.559**
	NC	0.472	0.381	0.457	**0.611**	**0.610**	0.405	0.405	**0.603**
	CSD	0.538	0.538	0.538	0.538	0.538	0.538	0.538	0.538
	FRSR	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028

SSIM

Table 8.4: *Comparison of the average PSNR image quality score with the indicated pairing of registration and SR method. This table gives more detail in understanding the best methods to use when combining LR images into a HR representation for purposes of target localization. Values in the upper quartile are in bold.*

		Registration Method							
		Global Translation				**Shift and Rotate**			
		Lucas-Kanade	Keren	Marcel	Vandewalle	Keren	Marcel	Lucchese	Vandewalle
SR Method	Bi3	18.73	17.53	18.28	21.33	21.35	17.25	17.25	21.16
	IBP	9.42	9.35	12.45	10.25	10.14	10.59	10.59	10.16
	PG	10.93	10.85	7.80	11.11	11.54	8.38	8.38	11.14
	ZRSR	11.57	11.14	16.36	12.48	12.53	13.98	13.98	12.58
	POCS	19.32	18.08	19.88	**22.55**	**22.71**	18.42	18.42	**22.53**
	NC	20.63	19.19	19.96	**24.40**	**24.32**	18.98	18.98	**24.10**
	CSD	21.62	21.62	21.62	21.62	21.62	21.62	21.62	21.62
	FRSR	7.14	7.14	7.16	7.14	7.14	7.56	7.56	7.14

PSNR

Similar to our approach shown in Figure 8.10, we analyze the processing time of the selected SR algorithms for a set of image sizes from 100 to 500 pixels in 10-pixel increments. Shown in Figure 8.11 are a comparison of the three selected SR algorithms (bicubic interpolation, POCS, and NC) using SR factors of 2, 3, and 4. As expected, the processing time of each SR method increased with increasing SR factor, seen via the stacked effect for each of the method groupings. Also, increasing input image dimensions increased the processing time, and this effect is enhanced with increasing SR factor. While the processing time for both POCS and NC increase significantly, the bicubic interpolation approach has a fairly low and stable processing time. When comparing these findings with respect to the processing time and the previous results shown in Table 8.3, we reach the conclusion that the drastically increased processing time of NC does not justify the marginal improvement in quality compared to the results obtained using bicubic interpolation.

Of additional interest is the influence that the SR factor plays in the output image quality. In Figure 8.12 we see a comparison of the SR experiments performed with the data grouped by SR factor and using the SSIM and PSNR metrics. The data used to generate each individual boxplot includes samplings of noise and blur level, differences in registration and SR methods, and different source images. The results illustrate a clear trend: better available information gives better results. The experiments performed were designed to give an output of the same size in pixels for all experiments in order to more easily compare results. It is likely that holding the input image size fixed and arriving at different output image sizes would give a different result than that shown. One interesting item seen in the SSIM results is that although the median image quality goes down with increasing SR factor (for a fixed output image size) the maximum quality does still range quite high. It is important to remember that there are many factors which influence the quality of the results obtained. Though we have included many figure to compare the results and have sought to thoroughly analyze the experimental data, a specific application will drive the precise parameters and algorithm selection.

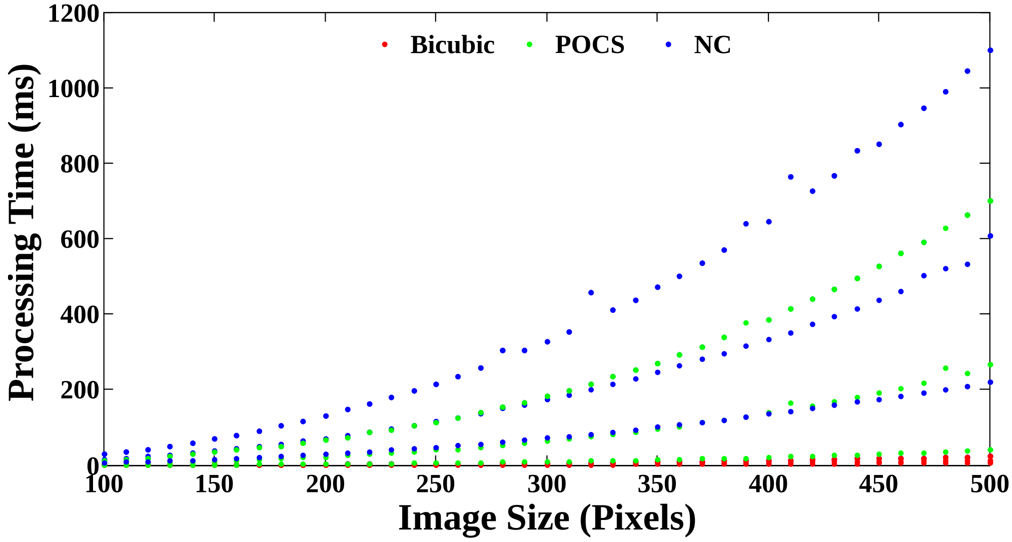


Figure 8.11: *Comparison of processing time of three SR algorithms. Bicubic interpolation (red), projection onto convex sets (POCS) (green), and normalized convolution (NC) (blue) were sampled at the indicated image sizes for 2x, 3x, and 4x experiments. LR input images used were square, thus the total number of pixels increases by the square of the indicated value. Each of the SR factors tested for each method increased in processing time as expected, where a higher SR factor and larger input image size both contributed to increased processing time. Bicubic interpolation was the least sensitive to changes in these parameters and maintained a very consistent processing time. When comparing the results in this figure to the quality measures in Table 8.3, particularly for larger images the drastic increase in processing time seems to not justify the slight difference in image quality.*

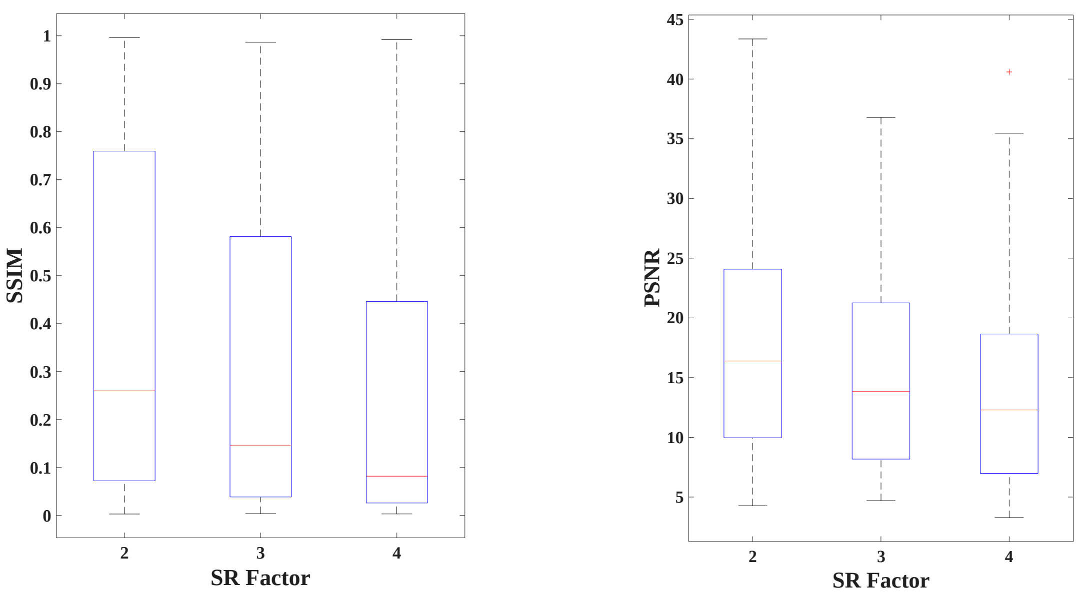


Figure 8.12: *Comparison of the difference in SR factor using the SSIM and PSNR metrics. The summary data shown here includes experiments with varying levels of noise and blur, varying registration and SR methods, and varying source images. Of important note for understanding these results is that the final output image size is fixed, such that for each set of experiments the size of the LR input images is different between boxplots. From these results we learn that the better quality information that is available the better the result will be.*

The image of the target at our sensors is near to being buried by the noise floor and often images to less than one pixel. This gives a wide cone of possible locations for the target given the distance from target to sensor. In a multi-sensor system we can utilize the independent views of the target to significantly narrow down the potential volume in which the target lies. Utilizing SR-enhanced measurements at each sensor platform, we will be able to narrow down the volume of possible solutions and increase our detection accuracy by increasing the SNR of the images obtained from a given observation point. If we make use of behavioral models of the target to predict the next position or set of positions, we can calculate an improved reorientation of the sensors to keep the target in view while reducing the error by correcting our prediction using the measurements of subsequent timesteps.

8.4.1 Comparison of SR Formulation

Given the wide variety of SR methods that have been developed, and the results of our analysis of a selection of methods, we wish to delve into the mathematical backing to explore the underlying reasons for the difference in quality of the results of application of SR. This is done for two reasons: to inform future research directions for those who develop SR algorithms and to develop greater understanding of our own results to further refine selection of an SR algorithm which can produce superior results. Of the methods tested, normalized convolution (NC), projection onto convex sets (POCS), and bicubic interpolation (Bi3) gave the best results.

8.4.1.1 Normalized Convolution

Normalized convolution is a spatial direct SR approach described in detail in Section 6.2.2.4. As mentioned, the implementation uses first-order modeling of local image structures rather than the higher-order approaches which can increase both the processing time and risk of overfitting to the noise component of the image.

8.4.1.2 Projection onto Convex Sets

Projection onto convex sets is a set theoretic method as detailed previously in Section 6.2.2.5. In the comparison of quality metrics by pairing registration and SR methods shown in Table 8.3, Although normalized convolution was strictly significantly higher in quality for the highlighted method pairs compared to bicubic interpolation, the POCS results were not, being lower than bicubic interpolation for two of the three highest pairings. This is in keeping with the observation made in the formulation that the method is particularly sensitive to inaccuracies in the motion estimation and increasing noise and blur.

8.4.1.3 Bicubic Interpolation

Bicubic interpolation is typically used in resampling images for display. The traditional formulation utilizes the values of 16 pixels in a 4x4 neighborhood as compared to bilinear interpolation which only utilizes 4 pixels in a 2x2 neighborhood. The tradeoff for increased data requirements and computational complexity compared to other interpolation approached is the increased image smoothness with reduced presence of undesirable artifacts. For an image f and the derivatives f_x, f_y, and f_{xy}, where f_x and f_y are the single derivatives in the x and y plane, respectively, and f_{xy} is the cross derivative, the interpolated surface representation of f to be resampled is

$$p(x, y) = \sum_{i=0}^{3} \sum_{j=0}^{3} a_{ij} x^i y^j \tag{8.2}$$

The problem, then, is how to determine the 16 coefficients a_{ij}. We need 16 equations for determining these 16 unknown coefficients. Four each from matching $p(x, y)$ with the function values of f and from the cross derivative, and eight from the derivatives in the x and y direction. The matrix formulation of this system of equations is

$$p(x,y) = [1 \quad x \quad x^2 \quad x^3] \begin{bmatrix} a_{00} & a_{01} & a_{02} & a_{03} \\ a_{10} & a_{11} & a_{12} & a_{13} \\ a_{20} & a_{21} & a_{22} & a_{23} \\ a_{30} & a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} 1 \\ y \\ y^2 \\ y^3 \end{bmatrix} \tag{8.3}$$

Though bicubic interpolation is not an SR approach but an interpolation approach, our use of a multiframe implementation and the impressive results of this approach warrant inclusion in our analysis. The closest categorization of bicubic interpolation in the SR taxonomy would be the shift-and-add class of direct SR.

8.4.1.4 Analysis of SR Formulation

In considering the classification of the methods used with respect to the taxonomy shown in Figure 6.4, only two types of methods had reasonable success and the two with the best results were both shift-and-add methods. It is reasonable that the Fourier domain methods did not give the best results as these methods are notorious for their sensitivity to noise. Of the spatial domain methods, iterated back projection is also known to have difficulty with high noise content, though Zomet et al [Zomet2001] modified the approach to introduce more noise robustness. This is reflected in the improvement of ZRSR compared to IBP of between 20-35% seen in Table 8.3.

8.5 Summary of Findings

In this section we review the findings of this chapter with respect to the system components considered and develop a recommended system implementation using our image-based approach. We then compare the expected performance of our approach to the existing approach to solving the target localization problem for the presented scenario.

8.5.1 Review of System Components

We review the recommended blocks discussed in detail previously with regard to the range of influence within each block in consideration of highest accuracy and lowest computational cost. As the exact implementation is not discussed as part of our work, we do not include considerations of component size, weight, financial cost, or power requirements, but caution those who may use our work as a guide to consider these features for any realized system. As we only review the analysis that contributes to the accuracy of the system, we will not review the PSF measurement method, our orientation update approach, or our error analysis block. Of the blocks in the pipeline shown in Figure 8.13 we only review the image simulation, noise modeling, super resolution, deblurring, and tracking steps, as follows.

8.5.1.1 Image Simulation

The image of the target at our sensors is near to being buried by the noise floor and often images to less than one pixel. This gives a wide cone of possible locations for the target given the distance from target to sensor. In a multi-sensor system we can utilize the independent views of the target to significantly narrow down the potential volume in which the target lies. We adapted models popular in the literature to represent the scenario under consideration and in testing subsequent steps in the pipeline we investigated representative ranges of noise and blur to examine the effects these changes would have on the position estimation accuracy.

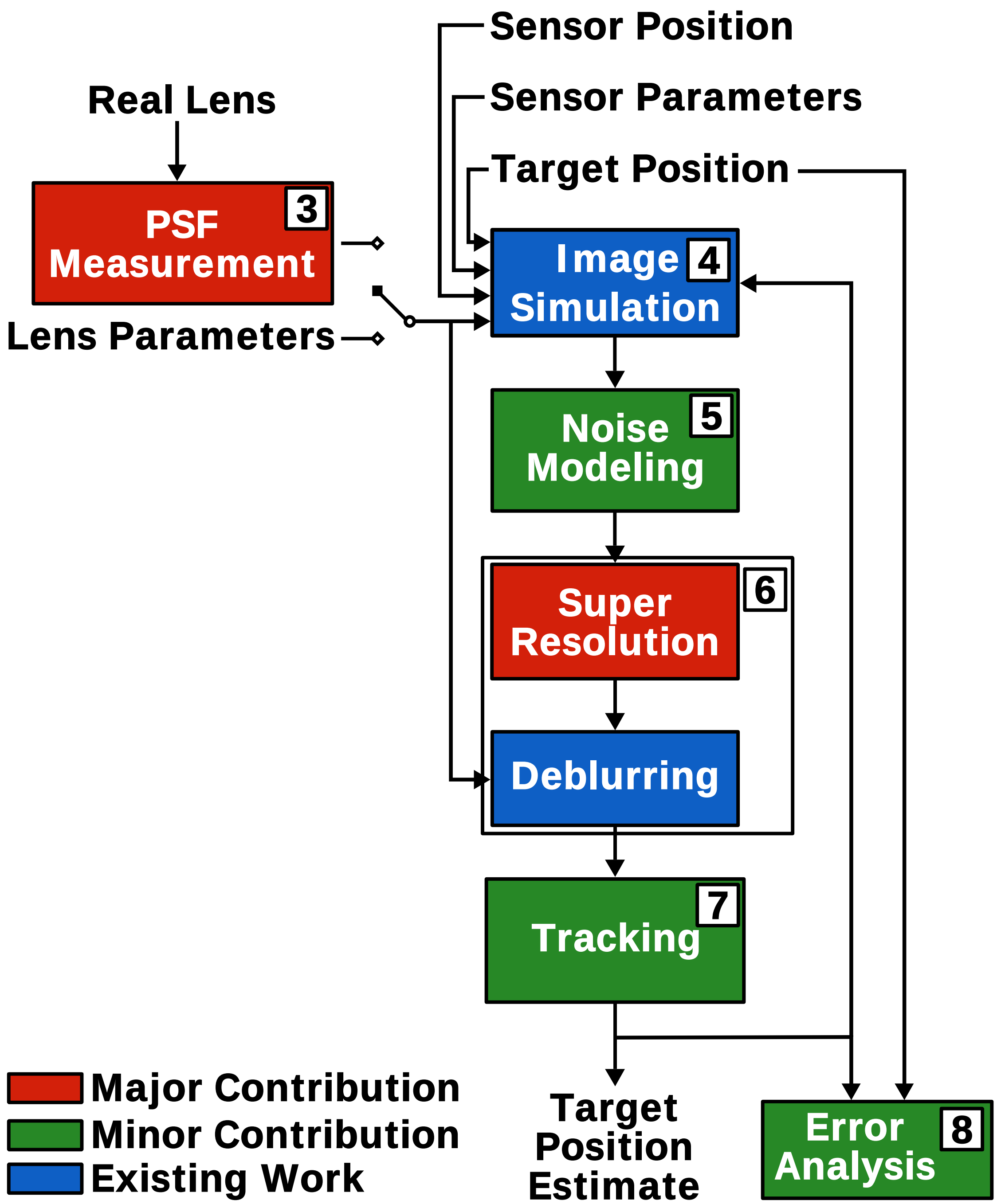


Figure 8.13: *Review of pipeline of contributions. The ideal target image is generated from known information of the sensor's position, resolution, and field-of-view (FOV) combined with the position of the target from the simulated path. We simulate the effects which cause a disturbance to the trajectory of the target, sensor, and infrared (IR) light as it travels between the two locations, and the lens aberration step factors in the effect a real lens has on image acquisition. We then apply our image segmentation methods to separate candidate signal information from the background noise. Next we separate the target from these candidate regions based on appropriate IR signatures. Finally, we estimate the 3D pose of the target in the frame and analyze the error with respect to the known position of the target.*

8.5.1.2 Noise Modeling

We have applied current noise level detection methods to analyze images prior to subsequent processing steps. Knowledge of the noise level is useful in selecting an appropriate SR, deblurring, or tracking approach suited to the level of noise present in the images acquired. We have also demonstrated three noise mitigation approaches that reduce the level of noise present in the images. Further, per our recommendation of a multi-aperture image acquisition approach, not only can the mitigation approaches demonstrated reduce the error in the images presented to the tracking step be better suited, but the SR methods considered also have a noise reduction effect in addition to their resolution enhancement. The reduction of noise in the source images acquired, coupled with further noise reduction by the SR step, contributes to significantly enhanced images from relatively low-quality, noisy, blurred images obtained using low-cost sensors.

8.5.1.3 Super Resolution

The cost of sensors suitable for imaging in the spectral range of interest means that we expect to have little information available due to the use of low-resolution sensors. Economies of scale bring down the unit price, but a multi-aperture acquisition system utilizing SR-enhanced measurements at each sensor platform means we are able to further narrow down the volume of possible solutions and increase our detection accuracy. By increasing the SNR of the images obtained from a given observation point and reducing the effective coverage area of a single pixel we can enhance the amount of data obtained at a given sensor platform and potentially reduce the system size and weight. We have considered eight methods and identified three with promising results including one with a low enough computational load to be included in the system pipeline. The other methods with consistently high quality output but longer computational time could still be utilized, however. Utilization of an FPGA implementation of one of these other methods could allow for more rapid processing while offering higher quality results.

8.5.1.4 Deblurring

We have demonstrated the effects of applying deblurring on a variety of source data and blur levels. Though the deblurring process does improve the quality of the image, there is significant computational cost. One benefit seen is the relatively decoupled nature of the deblurring and SR operations. As the majority of the time we expect our target of interest to occupy a single pixel, and given the high computational load, utilization of this operation will not enhance the result to a degree that warrants inclusion as part of the pipeline. However, should the application of deblurring be determined to be necessary, an FPGA implementation will likewise be useful for speeding up the processing by an expected factor of 20x [Clukey2016].

8.5.1.5 Tracking

Using behavioral models of target motion to predict the next position or set of positions, we can calculate an improved reorientation of the sensors to keep the target in view while reducing the error by correcting our prediction using the measurements of subsequent timesteps. We are also able to minimize the error of a given position estimate by utilizing predictive modeling of the target position. Given the active nature of the system in tracking the target, both accuracy in the location's coordinates and timely knowledge of when the target was at that location are essential to the success of our application. Therefore, the computational time contributed by the image acquisition and post processing steps is of important consideration in addition to the reduction of error. We have demonstrated own simple tracking formulation and four formulations based on the Kalman filter. Our analysis suggests utilization of one of these approaches in balancing the computational time and accuracy.

8.5.2 Recommended System Configuration

The contribution of each system component should be carefully considered with regard to the cost: in dollars for a hardware solution and for added computational hardware, in time to not introduce delay for a more accurate solution if an approximate solution would be of greater value sooner, and the contribution of the hardware to system size, weight and power requirements that each proposed solution would induce. An understanding of the relative merit of each component with regard to these considerations can drive a timeline of real-world implementation in stages, starting with the component that will give the greatest benefit for the least cost, and so on. Further, if financial cost is considered to be of least concern relative to the potential system capabilities that are available, the consideration then becomes the outer envelope of possibility offered by the proposed system component or overall suggested design.

Given the analysis shown, we have developed a recommended system configuration.

- Image Acquisition
    - Monolithic Image Sensor and Single Lens
        - Many existing sensor platforms are equipped with monolithic sensors paired with large, heavy, single-aperture lenses. While we recommend utilizing an alternative construction and have detailed the benefits, the fact that these units are already deployed does provide some incentive to make the best use of them during their operational lifespan. To this end, we recognize the capabilities of an implementation of our system pipeline using these platforms does outperform the existing radar-based approach. We are, however, able to improve on this result and can simultaneously offer added capabilities not possible with existing hardware.
    - Multi-Aperture Lens and Sensor System
        - Adapting the image acquisition system to incorporate a multi-aperture device with several lens and sensor pairs is the most promising approach we have explored. Since the blur component of the possible image degradation sources was shown to have the least influence, we are not rigidly tied to use of large, heavy, well-corrected optics and can instead use multiple smaller apertures to capture more light from the scene in a way that offers enhanced resolution, increased SNR, improves multispectral capabilities, and lower system size, weight, and cost. Based on our analysis of post-processing steps, we recommend use of a 2x2 array, but do recognize that for some targets of interest
- Noise Modeling and Mitigation
    - The proposed noise modeling and mitigation approaches are valuable for selection of appropriate methods in subsequent processing steps. We have shown the considerable difference that changes in noise level can have on output quality and computational time. Therefore, it is important to keep the noise contamination as low as possible to ensure the best expectation of success. This comes in proper design of all system components, including shielding of the electronics from external and self-interference, and utilization of the recommended best acquisition procedures and post-processing to minimize the effects of contamination that do remain.
- SR
    - Given the drastic improvement in accuracy that increasing the resolution of the images available to the tracking step offers, this may be the most vital step. In balancing the improved quality of the images with processing time, and considering the hardware setup or modified acquisition process implications, we recommend utilizing a 2x SR with the bicubic interpolation method. For a fixed arrangement of the acquisition system at each platform, and with the target to sensor distance being at optical infinity, the registration can be calibrated a priori to speed up the processing.
    - Should a more advanced multispectral analysis of an application's target of interest indicate that more than four spectral bands are necessary for discrimination of the target from its surroundings,

the bicubic interpolation approach is less sensitive to timing issues associated with moves to 3x or 4x SR processing so the adaptation in this situation will be the most straightforward.
  - Should increased enhancement of the images be desired, an implementation of the normalized convolution algorithm on FPGA hardware bears promise in keeping the processing time low while offering up to 10% improvement in the image quality over bicubic interpolation.
- Tracking
  - There are several steps to this functional block, and the first two, segmentation and target separation, were not explored in detail due to the fact that the exact optimal approach is highly application-specific, and that the primary work performed in these areas was conducted by others with whom the authors collaborated. The final two steps are within the scope of the work performed by the authors and was considered in much greater detail.
  - The first of these is the 3D Pose Recovery process. After analyzing the results of the two approaches considered, we recommend the triangulation approach due to its computational simplicity, improved accuracy, and capabilities to gracefully scale to multiple sensors. The final step in the tracking process involves target path modeling. We explored five different approaches, one simple windowed average-based fusion approach and four Kalman filtering approaches. Based on the results shown in Table 2.1, we recommend the Kalman Filter - Model 2 (Acceleration) model due to the balance of low average error and low processing time.

As previously mentioned, we omit the deblurring step as it does not significantly influence the results in our scenario. In Figure 8.14 we show a demonstration of the entire acquisition and processing pipeline utilizing our recommended system configuration.

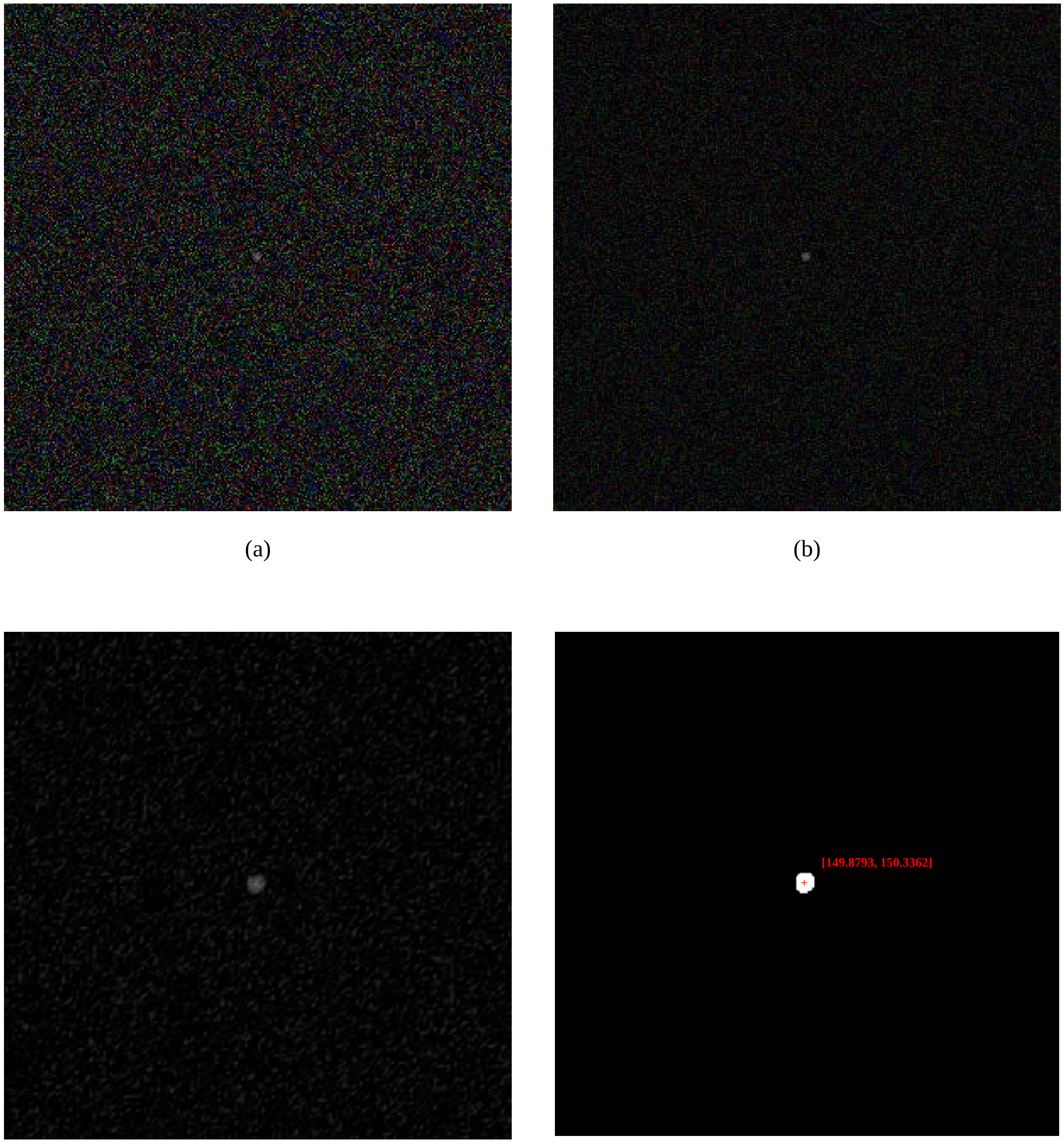
8.5.3 Comparison to Existing Approach

The existing approach to solving the problem of locating a target's coordinates in the scenario under consideration involves use of distributed ground-based radar systems. In Figure 8.15 we see a depiction of a modern radar system along with a representation of the smallest resolvable volume this radar system is capable of discerning. The radar's antenna rotates at 6 rpm, which translates to an effective framerate analog of 0.1 fps. The system can cover an area of 250 NM (463 km), and its minimum resolvable area is defined by 150m accuracy range, 11.3 km azimuth (1.4^o at 250 NM), 16.2 km elevation (2^o at 250 NM) resulting in a 28.7 billion m^3 search volume and a maximum error vector of 19.8 km.

Based on the design of our recommended image-based positional recovery system we are able to achieve RMSE of 390 m utilizing the recommended system configuration. This gives a maximum error vector of 780 m and a search volume of 0.24 billion m^3. The comparison is summarized in Table 8.5 below.

Table 8.5: *Summary of comparison of our approach and the conventional approach presented to the recovery of target coordinates. We use the maximum error vector and the total search volume as representative metrics of the capabilities of each approach. As extensive experimentation has shown the RMSE error of our position estimate, we superimpose a sphere with the indicated radius at the location of the estimate.*

	Conventional Approach	**Our Approach**	**Improvement**
Maximum Error Vector (km)	19.8	0.78	25.4x
Search Volume (m^3)	28.7 billion	0.24 billion	119.6x



(a) (b)

(c) (d)

Figure 8.14: *Illustration of the approach of the method utilizing the recommended system configuration. We start with a target of a single pixel with representative blur and noise (a) after which we apply our synthetic extended exposure for only five frames, shown in (b). Afterward we apply 2XSR using the bicubic interpolation approach (c) and apply the steps pertaining to our tracking approach with regard to locating the target in image coordinates (d).*

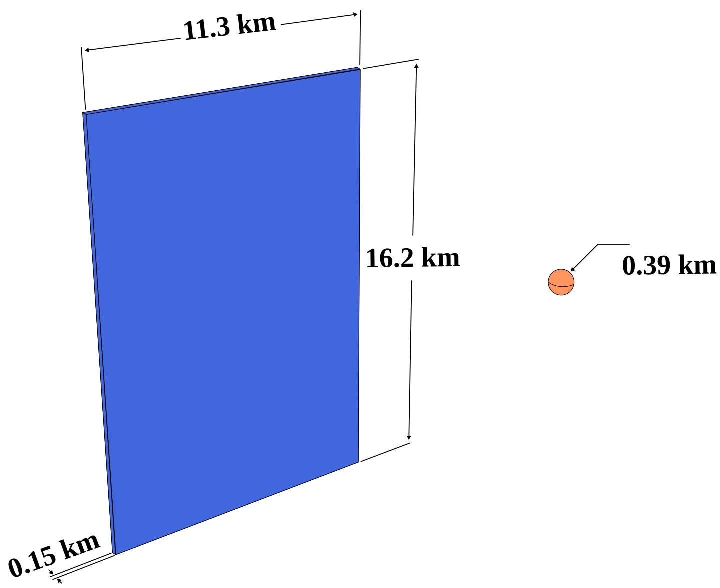


Figure 8.15: *Lanza radar (top) and resolution limit (bottom, blue). The Lanza system is a recently developed 3D radar system which costs \$30 million plus ongoing maintenance and operation expenses. The antenna rotates at 6 rpm, giving an effective framerate analog of 0.1 fps. The system can cover an area of 250 NM (463 km), and has a resolution of 150m range, 11.3 km azimuth (1.4^o at 250 NM), 16.2 km elevation (2^o at 250 NM) resulting in a 28.7 billion m^3 search volume. This is shown in comparison to the expected error performance of our system, shown in orange.*

9 Conclusion and Future Work

This dissertation work was motivated by the need to investigate use of orbital sensor platforms to recover the world coordinates of a target of interest while maintaining reasonable update rate, improved precision and coverage, considering the effects of system size, weight, and power (SWaP) requirements, and overall financial and computational cost. We studied the application of state-of-the-art methods in image simulation, denoising, SR, deconvolution, segmentation, 3D pose recovery, and 3D tracking and documented the results of these methods to our application and have recommended a system implementation that will best solve the presented problem. We have also developed a new method for measuring the PSF of a lens and contributed to the field in our thorough analysis and comparison of the impact of the methods considered. We addressed challenges arising from noisy, blurry, missing data and developed approaches to counter these sources of error. This chapter summarizes the contributions made through our research efforts and addresses potential directions for future research.

9.1 Dissertation Key Points

The main contributions of this dissertation are summarized in this section.

- **Super Resolution**

  We explored eight resolution enhancement methods by analyzing their processing time, quality of their results, and sensitivity to noise and blur degradation. Enhancing the resolution of the data obtained from a sensor platform enables significant reduction in the uncertainty of our target position estimate, and by utilizing SR instead of large sensor monoliths we can obtain a given accuracy at lower financial cost, system weight, reduced noise, and enhanced spectral range.

- **Point Spread Function Measurement**

  We have developed a new method for measurement of the PSF of a prime lens, though our approach will also work for variable-focus lens systems. We have demonstrated the effectiveness of our approach on several lens systems and validated our approach for each, including comparing to the results of competitive methods. Our approach is reliable, repeatable, and cost-effective.

- **Noise Modeling and Mitigation**

  We have explored methods for modeling the noise present in an image for selection of appropriate methods later in the processing pipeline and investigated three approaches for mitigating the noise level in acquired images. By reducing the level of noise contaminating the data available to later stages we enhance the accuracy of our approach and can reduce the processing time as well.

- **Tracking**

  In order to reduce the contribution of error that a single position estimate could contribute, we applied a simple averaging fusion and also investigated several applications of Kalman-based filtering of our raw position estimates. Our analysis of the expected error and processing time of each approach led to a conclusion for application of enhanced predictive tracking to reduce our overall error, improve recovery from temporary occlusion, and to provide a mechanism by which prediction of the future position of the target may be accomplished.

- **Error Analysis**

  We have analyzed the performance of each component of our contribution and the contribution each component has on the success of the overall work. In developing such a complex solution to the scenario under consideration there exists a danger in “not seeing the forest for the trees”. We have seen the trees and the forest and considered the best solution for each part and the best structure for accomplishing our overall goal. Further, we have given cautions and suggestions for adapting our work where changes in the goal may result in different conclusions from those to which we have arrived.

Our contributions, along with the contributions made by collaborators and other contemporary researchers in the field were combined to develop a structured approach to solving the problem of locating a target of interest using orbital sensors. Our work was developed in line with the interests of the authors and adapted to suit the needs of the scenario presented. While the effectiveness of our work was demonstrated in this context, there is greater general applicability to many other problems and researchers interested in related fields will benefit by the work performed and the analysis shown.

9.2 Future Directions for Research

Ideas developed by the authors for potential areas of continued investigation are summarized in this section. Some of these items have received some attention but the degree of completion and structure of the document did not warrant inclusion within the scope of the dissertation, while some items are at this stage ideas only.

- **Multi-Aperture Image Acquisition**

  We have presented some minor commentary on the usefulness and benefits that application of a multi-aperture image acquisition system would provide based on our experiments performed as part of this work. However, we have not provided a full description of the matters concerning design and development of such an approach, including selection of appropriate design for the lens systems, optimal arrangement of the aperture array, choice of lens focal length and filter system characteristics, size, orientation, pixel and bit depth resolution of the sensors, and so forth. Rather in order to keep the scope of this dissertation to a manageable size we have instead made note of the benefit that this type of system would offer in the context of mapping our proposed system onto such a platform rather than exploring in detail the full capabilities of such a platform with the analysis that would require. We do recommend any researchers endeavoring to adapt the system structure we have described to give proper consideration to utilizing a multi-aperture approach with all the benefits such a system offers.

- **Super Resolution**

  Though we investigated a variety of SR methods, ours was a small subset of the many methods developed. A broader selection of existing methods would provide an improved picture of the capabilities of the methods in the field with respect to the kind of data each operates best on. Future work could include adapting the different parts of our implementation. The LR image generator could be adapted to allow for more complex motion and a wider variety of noise and blur types, depending on the target application. The registration step could also be enhanced to allow for more generic motion models, and the SR evaluation step updated to include comparison of methods. Finally, the evaluation step could be updated to include other metrics should a given field use others as a standard.

- **Multi/Hyperspectral Imaging**

  We briefly mentioned the consideration of multispectral imaging in the context of identifying the target of interest by its spectral signature, but did not go into any significant detail concerning this topic. Investigating selection of optimal spectral range, optimal number of bands, and optimal band selection for reliable target identification would be another dissertation in itself. It is a fascinating topic, however and further investigation is warranted. Particularly in light of the natural adaptation of this approach to the multi-aperture implementation suggested, and the existence of SR algorithms which operate on multispectral data, there is great promise in enhancing the capabilities of the system presented.

- **Upward-Looking Aerial Sensors**

The evaluation presented in this dissertation only considered the use of downward-looking orbital sensor platforms. Given the increasing prevalence of platforms that can be placed much closer to the ground and observe both the area below and above the platform, investigation into the use of these platforms as part of the overall system could prove useful for expanding the capabilities shown.

- **Hardware Acceleration**

In evaluating the computational time for a given method, we are able to obtain a rough comparison of the difference between competing approaches to a given problem and their sensitivity to a range of potential parameters. Switching to a machine with different processing capabilities can significantly alter the exact processing time required and some types of processing technology are better suited to certain operations. In light of the scenario under consideration, we have suggested the use of FPGA implementation as an effective way to reduce the size, weight, and power requirements for a given computational operation while accelerating the time required for a given operation. Further exploration into the exact requirements and a full implementation of the pipeline presented should enhance the capabilities possible with the system structure presented in this dissertation.

These are just some ideas for areas of expansion and continuation of this work. There is also much that can be done just in adapting the existing work to alternative targets of interest.

Bibliography

Chapter 1

[Liu2012] X. Liu, M. Tanaka, and M. Okutomi, “Noise level estimation using weak textured patches of a single noisy image,” in *2012 19th IEEE International Conference on Image Processing (ICIP)*, 2012, pp. 665–668.

[Liu2013] X. Liu; M. Tanaka; M. Okutomi, “Single-Image Noise Level Estimation for Blind Denoising,” *ITIP*, vol. 22, no. 12, pp. 5226–5237, Dec. 2013.

[Luisier2011] F. Luisier, T. Blu, and M. Unser, “Image Denoising in Mixed Poisson-Gaussian Noise,” *IEEE Transactions on Image Processing*, vol. 20, no. 3, pp. 696–708, Mar. 2011.

[Zanella2009] R. Zanella, P. Boccacci, L. Zanni, and M. Bertero, “Efficient gradient projection methods for edge-preserving removal of Poisson noise,” *Inverse Problems*, vol. 25, no. 4, p. 45010, Apr. 2009.

Chapter 2

Section 1

[Aguet2008] F. Aguet, D. Van De Ville, and M. Unser, "An accurate PSF model with few parameters for axially shift-variant deconvolution," *Proc. 5th IEEE International Symposium on Biomedical Imaging, ISBI 2008 (May 14-17)*, pp. 157-160, 2008.

[Babu2011] R. Sudheer Babu and K. E. Sreenivasa Murthy, “A Survey on the methods of Super-resolution Image Reconstruction,” *International Journal of Computer Applications*, vol. 15, no. 2, pp. 1–6, Feb. 2011.

[Benjelloun2007] M. Benjelloun, H. J. Tellez Oliva, and R. Prevot, “Edge Closing of Synthetic and Real Images using Polynomial Fitting,” JCIT, vol. 2, pp. 8–19, 2007.

[Capel2004] D. P. Capel, Image Mosaicing and Superresolution. 2004.

[Chen2011] Z. Chen, K. Y. K. Wong, Y. Matsushita, X. Zhu, and M. Liu, “Self-calibrating depth from refraction,” in Computer Vision (ICCV), 2011 IEEE International Conference on, 2011, pp. 635–642.

[Cosel2009] K. von Neumann-Cosel, E. Roth, D. Lehmann, J. Speth, and A. Knoll, “Testing of Image Processing Algorithms on Synthetic Data,” in Fourth International Conference on Software Engineering Advances, 2009. ICSEA ’09, 2009, pp. 169–172.

[Dore2004] D. S. Doré and R. E. Kearney, “Experimental evaluation of computerised tomography point spread function variability within the field of view: Parametric models,” *Med. Biol. Eng. Comput.*, vol. 42, no. 5, pp. 591–597, Sep. 2004.

[Gunturk2012] B. K. Gunturk and X. Li, *Image Restoration: Fundamentals and Advances*. CRC Press, 2012.

[Hendee1997] W. R. Hendee and P. N. T. Wells, *The Perception of Visual Information*. Springer, 1997.

[Joshi2008] N. Joshi, R. Szeliski, and D. J. Kriegman, “PSF estimation using sharp edge prediction,” in *IEEE Conference on Computer Vision and Pattern Recognition, 2008. CVPR 2008*, 2008, pp. 1 –8.

[Kuhl2006] A. Kuhl, C. Wöhler, L. Krüger, and H. Groß, “Monocular 3D Scene Reconstruction at Absolute Scales by Combination of Geometric and Real-aperture Methods,” in *Pattern Recognition. Proc. 28th DAGM Symposium*, 2006, pp. 607–616.

[Kundu2010] A. K. Kundu, B. Bandopadhyay, and S. Sanyal, “A Microwave Imaging and Enhancement Technique from Noisy Synthetic Data,” arXiv:1010.0301 [cs], Oct. 2010.

[Kuthirummal2011] S. Kuthirummal, H. Nagahara, C. Zhou, and S. K. Nayar, "Flexible Depth of Field Photography," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 1, pp. 58–71, 2011.

[Liu2008] C. Liu and X. Chen, "Point spread function (PSF) measurement for cell phone camera with a high resolution PSF of the imaging lens and a sub-pixel digital algorithm," *Proc. SPIE 7072, Optics and Photonics for Information Processing II*, vol. Volume 7072, p. 70720F–70720F, Aug. 2008.

[Lucas1996] P. W. Lucas and P. F. Roche, "Near-infrared observations of L1551-IRS 5 with image sharpening," *MNRAS*, vol. 280, no. 4, pp. 1219–1229, Jun. 1996.

[Marcal2010] A. R. S. Marcal, A. Rodrigues, and M. Cunha, "Evaluation of satellite image segmentation using synthetic images," in Geoscience and Remote Sensing Symposium (IGARSS), 2010 IEEE International, 2010, pp. 2210–2213.

[Piper1901] C. W. Piper, A first book of the lens: an elementary treatise on the action and use of the photographic lens. Hazell, Watson, and Viney, 1901.

[Shih2012] YiChang Shih, Brian Guenter, Neel Joshi, Image Enhancement using Calibrated Lens Simulations. *Proc. 12th European Conference on Computer Vision, ECCV 2012*

[Siegman1986] A. E. Siegman, *Lasers*. University Science Books, 1986.

[Smith1997] S. W. Smith, *The scientist and engineer's guide to digital signal processing*, 1st ed. San Diego, Calif: California Technical Pub, 1997.

[Szeliski2006] R. Szeliski, "Image Alignment and Stitching: A Tutorial," *Found. Trends. Comput. Graph. Vis.*, vol. 2, no. 1, pp. 1–104, Jan. 2006.

[Tang2013] H. Tang and K. N. Kutulakos, "Utilizing Optical Aberrations for Extended-Depth-of-Field Panoramas," in *Computer Vision – ACCV 2012*, K. M. Lee, Y. Matsushita, J. M. Rehg, and Z. Hu, Eds. Springer Berlin Heidelberg, 2013, pp. 365–378.

[Tian2001] Y. Tian and S. G. Narasimhan, "Globally Optimal Estimation of Nonrigid Image Distortion," International Journal of Computer Vision, vol. 98, no. 3, pp. 279–302, 2011.

[Wang2006] F. Wang and M.-Y. Kan, "NPIC: Hierarchical Synthetic Image Classification Using Image Search and Generic Features," in Image and Video Retrieval, H. Sundaram, M. Naphade, J. R. Smith, and Y. Rui, Eds. Springer Berlin Heidelberg, 2006, pp. 473–482.

[Zheng2006] H. Zheng and O. Hellwich, "Introducing dynamic prior knowledge to partially-blurred image restoration," in *Proceedings of the 28th conference on Pattern Recognition*, Berlin, Heidelberg, 2006, pp. 111–121.

[Zitova2003] B. Zitová and J. Flusser, "Image registration methods: a survey," *Image and Vision Computing*, vol. 21, pp. 977–1000, 2003.

[Zuo2006] J. Zuo, N. A. Schmid, and X. Chen, "On performance comparison of real and synthetic iris images." IEEE Int. Conf. Image Process. (ICIP). 1, 305-308 (2006)

Section 2

[He2010] K. He, J. Sun, and X. Tang, "Guided Image Filtering," in *Computer Vision – ECCV 2010*, K. Daniilidis, P. Maragos, and N. Paragios, Eds. Springer Berlin Heidelberg, 2010, pp. 1–14.

[Hoshino2007] K. Hoshino and T. Nishimura, "Noise Reduction in CMOS Image Sensors for High Quality Imaging: The Autocorrelation Function Filter on Burst Image Sequences," *ICGST-GVIP*, vol. 7, no. 3, Nov. 2007.

[Liu2013] Xinhao Liu; Masayuki Tanaka; Masatoshi Okutomi, “Single-Image Noise Level Estimation for Blind Denoising,” *ITIP*, vol. 22, no. 12, pp. 5226–5237, Dec. 2013.

[Luisier2011] F. Luisier, T. Blu, and M. Unser, “Image Denoising in Mixed Poisson–Gaussian Noise,” *IEEE Transactions on Image Processing*, vol. 20, no. 3, pp. 696–708, Mar. 2011.

Section 3

[Farsiu2004] S. Farsiu, M. Robinson, M. Elad and P. Milanfar, "Fast and Robust Multiframe Super Resolution", IEEE Transactions on Image Processing, vol. 13, no. 10, pp. 1327-1344, 2004.

[Irani1991] M. Irani and S. Peleg, "Improving resolution by image registration", CVGIP: Graphical Models and Image Processing, vol. 53, no. 3, pp. 231-239, 1991.

[Pham2006] T. Pham, L. van Vliet and K. Schutte, "Robust Fusion of Irregularly Sampled Data Using Adaptive Normalized Convolution", EURASIP Journal on Advances in Signal Processing, vol. 2006, pp. 1-13, 2006.

[Siu2012] W.-C. Siu and K.-W. Hung, “Review of image interpolation and super-resolution,” in Asia-Pacific Signal Information Processing Association Annual Summit and Conference (APSIPA ASC), 2012, pp. 1–10.

[Sroubek2008] F. Šroubek, G. Cristobal and J. Flusser, "Simultaneous super-resolution and blind deconvolution", J. Phys.: Conf. Ser., vol. 124, p. 012048, 2008.

[Zomet2001] A. Zomet, A. Rav-Acha, S. Peleg, "Robust super-resolution," in Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2001. CVPR 2001, vol.1, pp.I-645-I-650.

Section 4

[Bertero2005] M. Bertero and P. Boccacci, "Image deconvolution", in From cells to proteins: Imaging nature across dimensions, V. Evangelista, L. Barsanti, V. Passarelli and P. Gualtieri, Ed. Berlin Heidelberg: Springer, 2005, pp. 349-370.

[Biggs1997] D. Biggs and M. Andrews, "Acceleration of iterative image restoration algorithms", Appl. Opt., vol. 36, no. 8, p. 1766, 1997.

[Lucy1974] L. Lucy, "An iterative technique for the rectification of observed distributions", The Astronomical Journal, vol. 79, p. 745, 1974.

[Gonzalez2008] R. C. Gonzalez and R. E. Woods, Digital Image Processing, 3rd ed. Prentice Hall, 2008.

[Kotera2013] J. Kotera, F. Šroubek and P. Milanfar, "Blind Deconvolution Using Alternating Maximum a Posteriori Estimation with Heavy-Tailed Priors", in Computer Analysis of Images and Patterns, R. Wilson, E. Hancock, A. Bors and W. Smith, Ed. Berlin Heidelberg: Springer, 2013, pp. 59-66.

[Richardson1972] W. Richardson, "Bayesian-Based Iterative Method of Image Restoration", Journal of the Optical Society of America, vol. 62, no. 1, p. 55, 1972.

[Shepp1982] L. Shepp and Y. Vardi, "Maximum Likelihood Reconstruction for Emission Tomography", IEEE Transactions on Medical Imaging, vol. 1, no. 2, pp. 113-122, 1982.

[Steve2008] "Image deblurring using regularization » Steve on Image Processing", 2008. [Online]. Available: http://blogs.mathworks.com/steve/2008/07/ 21/image-deblurring-using-regularization. [Accessed: 29- Apr- 2015].

[Yang2009] J. Yang, W. Yin, Y. Zhang and Y. Wang, "A Fast Algorithm for Edge-Preserving Variational Multichannel Image Restoration", SIAM J. Imaging Sci., vol. 2, no. 2, pp. 569-592, 2009.

Section 5

[Aguet2008] F. Aguet, D. Van De Ville, M. Unser, "An accurate PSF model with few parameters for axially shift-variant deconvolution," in 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro, 2008. ISBI 2008, pp.157-160, 14-17 May 2008.

[Brauers2010] J. Brauers; C. Seiler; T. Aach; Direct PSF estimation using a random noise target. Proc. SPIE 7537, Digital Photography VI, 75370B, 2010.

[Claxton2007] C. Claxton and R. Staunton, "Measurement of the point-spread function of a noisy imaging system", Journal of the Optical Society of America A, vol. 25, no. 1, p. 159, 2007.

[Delbracio2012] M. Delbracio, P. Musé, A. Almansa and J. Morel, "The Non-parametric Sub-pixel Local Point Spread Function Estimation Is a Well Posed Problem", International Journal of Computer Vision, vol. 96, no. 2, pp. 175-194, 2012.

[Gunturk2013] B. Gunturk and X. Li, Image restoration. Boca Raton, FL: CRC Press, 2013.

[Howell1996] S. Howell, B. Koehn, E. Bowell and M. Hoffman, "Detection and Measurement of Poorly Sampled Point Sources Imaged With 2-D Array", The Astronomical Journal, vol. 112, p. 1302, 1996.

[Hu2012] W. Hu, J. Xue, and N. Zheng, "PSF Estimation via Gradient Domain Correlation", IEEE Transactions on Image Processing, vol. 21, no. 1, pp. 386-392, 2012.

[ISO12233] ISO 12233: Photography -- Electronic still picture imaging -- Resolution and spatial frequency responses. Genève, Switzerland: International Organization for Standardization, 2014. Print.

[Levy1999] E. Levy, D. Peles, M. Opher-Lipson and S. Lipson, "Modulation transfer function of a lens measured with a random target method", Appl. Opt., vol. 38, no. 4, p. 679, 1999.

[Liu2008] C. Liu ; X. Chen; Point spread function (PSF) measurement for cell phone camera with a high resolution PSF of the imaging lens and a sub-pixel digital algorithm. Proc. SPIE 7072, Optics and Photonics for Information Processing II, 70720F (September 03, 2008)

[NavasMoya2013] F. Navas-Moya, J. Nieves, E. Valero and E. Garrote, "Measurement of the optical transfer function using a white-dot pattern presented on a liquid-crystal display", JEOS:RP, vol. 8, 2013.

[Schuler2012] C. Schuler, M. Hirsch, S. Harmeling, and B. Schölkopf, “Blind Correction of Optical Aberrations,” in Proceedings of the 12th European conference on Computer Vision – Volume Part III, Springer-Verlag, Berlin Heidelberg, Proc. ECCV 2012, pp. 187–200.

[Shih2012] Y. Shih, B. Guenter and N. Joshi, "Image Enhancement Using Calibrated Lens Simulations", in 12th European Conference on Computer Vision, 2012, Proceedings, Part IV, 2012, pp. 42-56.

[Sroubek2008] F. Šroubek, G. Cristobal and J. Flusser, "Simultaneous super-resolution and blind deconvolution", J. Phys.: Conf. Ser., vol. 124, p. 012048, 2008.

[SVI] "Recording Beads | Scientific Volume Imaging". [Online]. Available: http://www.svi.nl/RecordingBeads. [Accessed: 29- Apr- 2015].

[Tang2013] H. Tang and K.N. Kutulakos, “What Does an Aberrated Photo Tell Us about the Lens and the Scene?”, In Proc. 5th Int. Conf. on Computational Photography (ICCP), Boston, MA, 2013.

[Zheng2006] H. Zheng and O. Hellwich, “Introducing dynamic prior knowledge to partially-blurred image restoration,” in Proceedings of the 28th conference on Pattern Recognition, Berlin, Heidelberg, 2006, pp. 111–121.

Section 6

[Bouchrika2014] I. Bouchrika, J. N. Carter, and M. S. Nixon, “Towards automated visual surveillance using gait for identity recognition and tracking across multiple non-intersecting cameras,” *Multimed Tools Appl*, vol. 75, no. 2, pp. 1201–1221, Nov. 2014.

[Chen2006] C. H. Chen, C. Cheng, D. Page, A. Koschan, and M. Abidi, "A Moving Object Tracked by a Moving Robot with Real-time Obstacles Avoidance,"*Proc. IEEE International Conference on Pattern Recognition ICPR 2006*, Vol. III, Hong Kong, pp. 1091-1094, August 2006.

[Dalal2005] N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection,” in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05)*, 2005, vol. 1, pp. 886–893 vol. 1.

[Elhayek2012] A. Elhayek, C. Stoll, N. Hasler, K. I. Kim, H. P. Seidel, and C. Theobalt, “Spatio-temporal motion tracking with unsynchronized cameras,” in *2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012, pp. 1870–1877.

[Fuhl2015] W. Fuhl, T. Kübler, K. Sippel, W. Rosenstiel, and E. Kasneci, “ExCuSe: Robust Pupil Detection in Real-World Scenarios,” in *Computer Analysis of Images and Patterns*, G. Azzopardi and N. Petkov, Eds. Springer International Publishing, 2015, pp. 39–51.

[Harris1988] C. Harris and M. Stephens (1988). "A combined corner and edge detector" (PDF). *Proceedings of the 4th Alvey Vision Conference*. pp. 147–151.

[Hedayati2015] M. Hedayati, M. J. Cree, and J. Scott, “Combination of Mean Shift of Colour Signature and Optical Flow for Tracking During Foreground and Background Occlusion,” in *Image and Video Technology*, T. Bräunl, B. McCane, M. Rivera, and X. Yu, Eds. Springer International Publishing, 2015, pp. 87–98.

[Kazik2012] T. Kazik, L. Kneip, J. Nikolic, M. Pollefeys, and R. Siegwart, “Real-time 6D stereo Visual Odometry with non-overlapping fields of view,” in *2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012, pp. 1529–1536.

[Knopp2010] Jan Knopp, Mukta Prasad, Gert Willems, Radu Timofte, and Luc Van Gool, "Hough Transform and 3D SURF for Robust Three Dimensional Classification", *European Conference on Computer Vision* (ECCV), 2010.

[Leal-Taixe2012] L. Leal-Taixé, G. Pons-Moll, and B. Rosenhahn, “Branch-and-price global optimization for multi-view multi-target tracking,” in *2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012, pp. 1987–1994.

[Lim2012] H. Lim, S. N. Sinha, M. Cohen, M. Uyttendaele. “Real-time Image-based 6-DOF Localization in Large-scale Environments,” in *Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Washington, DC, USA, 2012, pp. 1043–1050.

[Lowe1999] Lowe, David G. (1999). "Object recognition from local scale-invariant features". *Proceedings of the International Conference on Computer Vision*. pp. 1150–1157.

[Orwell1999] J. Orwell, P. Remagnino, Jones G (1999) Multi-camera colour tracking. In: Second IEEE Workshop on Visual Surveillance, pp 14–21.

[Pham2007] N. T. Pham, W. Huang, and S. H. Ong, “Probability Hypothesis Density Approach for Multi-camera Multi-object Tracking,” in *Computer Vision – ACCV 2007*, Y. Yagi, S. B. Kang, I. S. Kweon, and H. Zha, Eds. Springer Berlin Heidelberg, 2007, pp. 875–884.

[Soto2009] C. Soto, B. Song, and A. K. Roy-Chowdhury, “Distributed multi-target tracking in a self-configuring camera network,” in *IEEE Conference on Computer Vision and Pattern Recognition, 2009. CVPR 2009*, 2009, pp. 1486–1493.

[Tomasi2004] C. Tomasi and T. Kanade (2004). "Detection and Tracking of Point Features". *Pattern Recognition* **37**: 165–168.

[Vaca-Castano2012] G. Vaca-Castano, A. R. Zamir, and M. Shah, “City scale geo-spatial trajectory estimation of a moving camera,” in *2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012, pp. 1186–1193.

[VanVerth2008] J. M. Van Verth and L. M. Bishop, Essential Mathematics for Games and Interactive Applications: A Programmer’s Guide, Second Edition. CRC Press, 2008.

[Vepakomma2015] P. Vepakomma, D. De, S. K. Das, and S. Bhansali, “A-Wristocracy: Deep learning on wrist-worn sensing for recognition of user complex activities,” in *2015 IEEE 12th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*, 2015, pp. 1–6.

[Yassin2015] M. Yassin, Y. Dirani, M. Ibrahim, S. Lahoud, D. Mezher, and B. Cousin, “A novel dynamic inter-cell interference coordination technique for LTE networks,” in *2015 IEEE 26th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, 2015, pp. 1380–1385.

[Zou2009] D. Zou, Q. Zhao, H. S. Wu, and Y. Q. Chen, “Reconstructing 3D motion trajectories of particle swarms by global correspondence selection,” in *2009 IEEE 12th International Conference on Computer Vision*, 2009, pp. 1578–1585.

Section 7

[Desurmont2006] X. Desurmont, C. Machy, C. Mancas-Thillou, D. Severin, and J. f Delaigle, “Effects of Parameters Variations in Particle Filter Tracking,” in *2006 International Conference on Image Processing*, 2006, pp. 2789–2792.

[Hadigheh2006] A. G. Hadigheh and T. Terlaky, “Sensitivity analysis in linear optimization: Invariant support set intervals,” *European Journal of Operational Research*, vol. 169, no. 3, pp. 1158–1175, Mar. 2006.

[Jalobeanu2000] A. Jalobeanu, L. Blanc-Feraud, and J. Zerubia, “Estimation of adaptive parameters for satellite image deconvolution,” in *15th International Conference on Pattern Recognition, 2000. Proceedings*, 2000, vol. 3, pp. 318–321 vol.3.

[Kwisthout2012] J. Kwisthout, L. C. van der Gaag, “The Computational Complexity of Sensitivity Analysis and Parameter Tuning,” *Proceedings of the 24th Conference on Uncertainty in Artificial Intelligence, UAI 2008*, 2012.

[Molina2006] R. Molina, M. Vega, J. Mateos, and A. K. Katsaggelos, “Parameter Estimation in Bayesian Reconstruction of Multispectral Images using Super Resolution Techniques,” in *2006 International Conference on Image Processing*, 2006, pp. 1749–1752.

[Muhammed2010] H. H. Muhammed and F. Bergholm, “Sensitivity Analysis of Multichannel Images Intended for Instantaneous Imaging Spectrometry Applications,” *SIAM J. Img. Sci.*, vol. 3, no. 1, pp. 79–109, Feb. 2010.

[Nanda1994] P. K. Nanda, U. B. Desai, and P. G. Poonacha, “A homotopy continuation method for parameter estimation in MRF models and image restoration,” in *Proceedings of IEEE International Symposium on Circuits and Systems - ISCAS ’94*, 1994, vol. 3, pp. 273–276 vol.3.

[Pappenberger2006] F. Pappenberger, I. Iorgulescu, and K. J. Beven, “Sensitivity analysis based on regional splits and regression trees (SARS-RT),” *Environmental Modelling & Software*, vol. 21, no. 7, pp. 976–990, Jul. 2006.

[Racu2005] R. Racu, M. Jersak, and R. Ernst, “Applying sensitivity analysis in real-time distributed systems,” in *11th IEEE Real Time and Embedded Technology and Applications Symposium*, 2005, pp. 160–169.

[Racu2006] R. Racu, A. Hamann, and R. Ernst, “A formal approach to multi-dimensional sensitivity analysis of embedded real-time systems,” in *18th Euromicro Conference on Real-Time Systems (ECRTS’06)*, 2006, p. 10 pp.–12.

[Vega2003] M. Vega, J. Mateos, R. Molina, and A. K. Katsaggelos, “Bayesian parameter estimation in image reconstruction from subsampled blurred observations,” in *2003 International Conference on Image Processing, 2003. ICIP 2003. Proceedings*, 2003, vol. 2, pp. II–969–72 vol.3.

[Vytyaz2008] I. Vytyaz, D. C. Lee, U. K. Moon, and K. Mayaram, “Parameter variation analysis for voltage controlled oscillators in phase-locked loops,” in *2008 IEEE International Symposium on Circuits and Systems*, 2008, pp. 716–719.

[Wong2010] K. C. L. Wong, F. Billet, T. Mansi, R. Chabiniok, M. Sermesant, H. Delingette, and N. Ayache, “Cardiac Motion Estimation Using a ProActive Deformable Model: Evaluation and Sensitivity Analysis,” in Statistical Atlases and Computational Models of the Heart, O. Camara, M. Pop, K. Rhode, M. Sermesant, N. Smith, and A. Young, Eds. Springer Berlin Heidelberg, 2010, pp. 154–163.

[Yang2000] S. Yang, Y.-H. Hu, D. L. Tull, and T. Q. Nguyen, “Maximum likelihood parameter estimation for image ringing artifact removal,” in *2000 International Conference on Image Processing, 2000. Proceedings*, 2000, vol. 1, pp. 888–891 vol.1.

[Yao2006] Y. Yao, B. Abidi, N. Doggaz, and M. Abidi, “Evaluation of sharpness measures and search algorithms for the auto-focusing of high magnification images,” in *Proceedings of SPIE - The International Society for Optical Engineering*, May 2006.

[Yu2015] J. J. Q. Yu and V. O. K. Li, “Parameter sensitivity analysis of Social Spider Algorithm,” in *2015 IEEE Congress on Evolutionary Computation (CEC)*, 2015, pp. 3200–3205.

[Zhang2007] Y. Zhang, D. B. Goldgof, S. Sarkar, and L. V. Tsap, “A sensitivity analysis method and its application in physics-based nonrigid motion modeling,” *Image and Vision Computing*, vol. 25, no. 3, pp. 262–273, Mar. 2007.

[Zhang2010] H.-X. Zhang and J. Goutsias, “A comparison of approximation techniques for variance-based sensitivity analysis of biochemical reaction systems,” *BMC Bioinformatics*, vol. 11, no. 1, pp. 1–17, May 2010.

Chapter 3

[Aguet2008] F. Aguet, D. Van De Ville, M. Unser, "An accurate PSF model with few parameters for axially shift-variant deconvolution," in 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro, 2008. ISBI 2008, pp.157-160, 14-17 May 2008.

[BenEzra2005] M. Ben-Ezra, A. Zomet and S. Nayar, "Video super-resolution using controlled subpixel detector shifts", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 27, no. 6, pp. 977-987, 2005.

[Bertero2005] M. Bertero and P. Boccacci, "Image deconvolution", in From cells to proteins: Imaging nature across dimensions, V. Evangelista, L. Barsanti, V. Passarelli and P. Gualtieri, Ed. Berlin Heidelberg: Springer, 2005, pp. 349-370.

[Born1999] M. Born and E. Wolf, Principles of optics. Cambridge: Cambridge University Press, 1999.

[Brauers2010] J. Brauers; C. Seiler; T. Aach; Direct PSF estimation using a random noise target. Proc. SPIE 7537, Digital Photography VI, 75370B, 2010.

[Claxton2007] C. Claxton and R. Staunton, "Measurement of the point-spread function of a noisy imaging system", Journal of the Optical Society of America A, vol. 25, no. 1, p. 159, 2007.

[Delbracio2012] M. Delbracio, P. Musé, A. Almansa and J. Morel, "The Non-parametric Sub-pixel Local Point Spread Function Estimation Is a Well Posed Problem", International Journal of Computer Vision, vol. 96, no. 2, pp. 175-194, 2012.

[Farsiu2004] S. Farsiu, M. Robinson, M. Elad and P. Milanfar, "Fast and Robust Multiframe Super Resolution", IEEE Transactions on Image Processing, vol. 13, no. 10, pp. 1327-1344, 2004.

[Gunturk2013] B. Gunturk and X. Li, Image restoration. Boca Raton, FL: CRC Press, 2013.

[Gonzalez2008] R. C. Gonzalez and R. E. Woods, Digital Image Processing, 3rd ed. Prentice Hall, 2008.

[Hornberg2006] A. Hornberg, Handbook of machine vision. Weinheim: Wiley-VCH, 2006.

[Howell1996] S. Howell, B. Koehn, E. Bowell and M. Hoffman, "Detection and Measurement of Poorly Sampled Point Sources Imaged With 2-D Array", The Astronomical Journal, vol. 112, p. 1302, 1996.

[Hu2012] W. Hu, J. Xue, and N. Zheng, "PSF Estimation via Gradient Domain Correlation", IEEE Transactions on Image Processing, vol. 21, no. 1, pp. 386-392, 2012.

[IPCamera2015] "IP Camera Adapter". [Online]. Available: http://ip-webcam.appspot.com. [Accessed: 29- Apr- 2015].

[IPOL1] "IPOL Journal · Non-parametric Sub-pixel Local Point Spread Function Estimation". [Online]. Available: http://demo.ipol.im/demo/admm_non_
blind_psf_estimation/archive?key=1FEC3383E2E4F1469583C67ECF1CEFF7. [Accessed: 29- Apr- 2015].

[IPOL2] "IPOL Journal · Non-parametric Sub-pixel Local Point Spread Function Estimation". [Online]. Available: http://demo.ipol.im/demo/admm_non_
blind_psf_estimation/archive?key=7C1434CED5525CCD7B676D07129F8ABC. [Accessed: 29- Apr- 2015].

[IPOL3] "IPOL Journal · Non-parametric Sub-pixel Local Point Spread Function Estimation". [Online]. Available: http://demo.ipol.im/demo/admm_non_
blind_psf_estimation/archive?key=1DCC802F3D78DFB091F88838F398613B. [Accessed: 29- Apr- 2015].

[Irani1991] M. Irani and S. Peleg, "Improving resolution by image registration", CVGIP: Graphical Models and Image Processing, vol. 53, no. 3, pp. 231-239, 1991.

[Joshi2008] N. Joshi, R. Szeliski, D. Kriegman, "PSF estimation using sharp edge prediction," in IEEE Conference on Computer Vision and Pattern Recognition, 2008. CVPR 2008., pp.1-8, 23-28 June 2008

[Keren1988] D. Keren, S. Peleg, R. Brada, "Image sequence enhancement using sub-pixel displacements," in Computer Society Conference on Computer Vision and Pattern Recognition, 1988. Proceedings CVPR '88., pp.742-746, 5-9 Jun 1988.

[Kotera2013] J. Kotera, F. Šroubek and P. Milanfar, "Blind Deconvolution Using Alternating Maximum a Posteriori Estimation with Heavy-Tailed Priors", in Computer Analysis of Images and Patterns, R. Wilson, E. Hancock, A. Bors and W. Smith, Ed. Berlin Heidelberg: Springer, 2013, pp. 59-66.

[LCDSpectra] "DisplayMate Light Spectra for the LCD Tablets". [Online]. Available: http://www.displaymate.com/Spectra_4.html. [Accessed: 04- Apr- 2013].

[Levy1999] E. Levy, D. Peles, M. Opher-Lipson and S. Lipson, "Modulation transfer function of a lens measured with a random target method", Appl. Opt., vol. 38, no. 4, p. 679, 1999.

[Lin2004] Z. Lin and H.-Y. Shum, "Fundamental limits of reconstruction-based superresolution algorithms under local translation", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 26, no. 1, pp. 83-97, 2004.

[Liu2008] C. Liu ; X. Chen; Point spread function (PSF) measurement for cell phone camera with a high resolution PSF of the imaging lens and a sub-pixel digital algorithm. Proc. SPIE 7072, Optics and Photonics for Information Processing II, 70720F (September 03, 2008)

[Liu2012] X. Liu, M. Tanaka, M. Okutomi, "Noise level estimation using weak textured patches of a single noisy image," in 19th IEEE International Conference on Image Processing (ICIP), 2012, pp.665-668, 2012.

[Lucy1974] L. Lucy, "An iterative technique for the rectification of observed distributions", The Astronomical Journal, vol. 79, p. 745, 1974.

[Milanfar2011] P. Milanfar, Super-resolution imaging. Boca Raton: CRC Press, 2011.

[Nasrollahi2014] K. Nasrollahi and T. Moeslund, "Super-resolution: a comprehensive survey", Machine Vision and Applications, vol. 25, no. 6, pp. 1423-1468, 2014.

[NavasMoya2013] F. Navas-Moya, J. Nieves, E. Valero and E. Garrote, "Measurement of the optical transfer function using a white-dot pattern presented on a liquid-crystal display", JEOS:RP, vol. 8, 2013.

[Park1984] S. Park, R. Schowengerdt and M. Kaczynski, "Modulation-transfer-function analysis for sampled image systems", Appl. Opt., vol. 23, no. 15, p. 2572, 1984.

[Pham2006] T. Pham, L. van Vliet and K. Schutte, "Robust Fusion of Irregularly Sampled Data Using Adaptive Normalized Convolution", EURASIP Journal on Advances in Signal Processing, vol. 2006, pp. 1-13, 2006.

[Richardson1972] W. Richardson, "Bayesian-Based Iterative Method of Image Restoration", Journal of the Optical Society of America, vol. 62, no. 1, p. 55, 1972.

[Schuler2012] C. Schuler, M. Hirsch, S. Harmeling, and B. Schölkopf, “Blind Correction of Optical Aberrations,” in Proceedings of the 12th European conference on Computer Vision – Volume Part III, Springer-Verlag, Berlin Heidelberg, Proc. ECCV 2012, pp. 187–200.

[Shih2012] Y. Shih, B. Guenter and N. Joshi, "Image Enhancement Using Calibrated Lens Simulations", in 12th European Conference on Computer Vision, 2012, Proceedings, Part IV, 2012, pp. 42-56.

[Sroubek2008] F. Šroubek, G. Cristobal and J. Flusser, "Simultaneous super-resolution and blind deconvolution", J. Phys.: Conf. Ser., vol. 124, p. 012048, 2008.

[Steve2008] "Image deblurring using regularization » Steve on Image Processing", 2008. [Online]. Available: http://blogs.mathworks.com/steve/2008/07/ 21/image-deblurring-using-regularization. [Accessed: 29- Apr- 2015].

[SVI] "Recording Beads | Scientific Volume Imaging". [Online]. Available: http://www.svi.nl/RecordingBeads. [Accessed: 29- Apr- 2015].

[Tang2013] H. Tang and K.N. Kutulakos, “What Does an Aberrated Photo Tell Us about the Lens and the Scene?”, In Proc. 5th Int. Conf. on Computational Photography (ICCP), Boston, MA, 2013.

[Zheng2006] H. Zheng and O. Hellwich, “Introducing dynamic prior knowledge to partially-blurred image restoration,” in Proceedings of the 28th conference on Pattern Recognition, Berlin, Heidelberg, 2006, pp. 111–121.

[Zomet2001] A. Zomet, A. Rav-Acha, S. Peleg, "Robust super-resolution," in Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2001. CVPR 2001, vol.1, pp.I-645-I-650.

Chapter 4

[Altunbasak2003] Y. Altunbasak, R. M. Mersereau, and A. J. Patti, “A fast parametric motion estimation algorithm with illumination and lens distortion correction,” *IEEE Transactions on Image Processing*, vol. 12, no. 4, pp. 395–408, Apr. 2003.

[Bertero2005] M. Bertero and P. Boccacci, "Image deconvolution", in *From cells to proteins: Imaging nature across dimensions*, V. Evangelista, L. Barsanti, V. Passarelli and P. Gualtieri, Ed. Berlin Heidelberg: Springer, 2005, pp. 349-370.

[Biggs1997] D. Biggs and M. Andrews, "Acceleration of iterative image restoration algorithms", Appl. Opt., vol. 36, no. 8, p. 1766, 1997.

[Brown1966] D. C. Brown, "Decentering distortion of lenses". Photogrammetric Engineering. 32 (3): 444–462, May 1966.

[Caltech] "Camera Calibration Toolbox for Matlab". [Online]. Available: http://www.vision.caltech.edu/bouguetj/calib_doc. [Accessed: 29- Apr- 2015].

[Cicala2011] R. Cicala, “LensRentals.com - Lens Genealogy Part 1.” [Online]. Available: http://www.lensrentals.com/blog/ 2011/08/lens-geneology-part-1. [Accessed: 16-Oct-2012].

[Conrady1919] A. Conrady, "Decentred Lens-Systems", *Monthly Notices of the Royal Astronomical Society*, vol. 79, no. 5, pp. 384-390, 1919.

[Cox1971] A. Cox, *Photographic optics, a modern approach to the technique of definition*, Expanded ed. London, New York: Focal Press, 1971.

[Cox2007] J. D. Cox, J. Y. Chang, and R. Komaki, *Image-Guided Radiotherapy of Lung Cancer*. CRC Press, 2007.

[Delbracio2012] M. Delbracio, P. Musé, A. Almansa and J. Morel, "The Non-parametric Sub-pixel Local Point Spread Function Estimation Is a Well Posed Problem", *International Journal of Computer Vision*, vol. 96, no. 2, pp. 175-194, 2012.

[Dorfmuller2002] K. Dorfmuller-Ulhaas, "Optical Tracking: From User Motion to 3D Interaction", Ph.D., Vienna University of Technology, 2002.

[Goncalves2013] L. D. Gonçalves, E. M. Rocco, and R. V. de Moraes, “Orbital Disturbance Analysis due to the Lunar Gravitational Potential and Deviation Minimization through the Trajectory Control in Closed Loop,” *Journal of Physics Conference Series*, vol. 465, no. 1, p. 2013–, 2013.

[Gonzalez2008] R. C. Gonzalez and R. E. Woods, Digital Image Processing, 3rd ed. Prentice Hall, 2008.

[Hooge1969] F. N. Hooge. “1/f noise is no surface effect”. Physics Letters, 29A(3):139-140, April 1969.

[Hooge1994] F. N. Hooge. “1/f noise sources”. IEEE Transactions on Electron Devices, 41(11):1926-1935, November 1994.

[Hoshino2007] K. Hoshino and T. Nishimura, “Noise Reduction in CMOS Image Sensors for High Quality Imaging: The Autocorrelation Function Filter on Burst Image Sequences,” *ICGST-GVIP*, vol. 7, no. 3, Nov. 2007.

[Jayaraman1989] R. Jayaraman and C. G. Sodini, “A 1/f noise technique to extract the oxide trap density near the conduction band edge of silicon,” *IEEE Transactions on Electron Devices*, vol. 36, no. 9, pp. 1773–1782, Sep. 1989.

[Kingslake1989] R. Kingslake, *A History of the Photographic Lens*. Academic Press, 1989.

[Kotera2013] J. Kotera, F. Šroubek and P. Milanfar, "Blind Deconvolution Using Alternating Maximum a Posteriori Estimation with Heavy-Tailed Priors", in Computer Analysis of Images and Patterns, R. Wilson, E. Hancock, A. Bors and W. Smith, Ed. Berlin Heidelberg: Springer, 2013, pp. 59-66.

[Liu2008] C. Liu ; X. Chen; Point spread function (PSF) measurement for cell phone camera with a high resolution PSF of the imaging lens and a sub-pixel digital algorithm. *Proc. SPIE* 7072, Optics and Photonics for Information Processing II, 70720F (September 03, 2008).

[Lucchese2003] L. Lucchese and S. K. Mitra, “Correction of geometric lens distortion through image warping,” in *3rd International Symposium on Image and Signal Processing and Analysis, 2003. ISPA 2003. Proceedings of the*, 2003, vol. 1, p. 516–521 Vol.1.

[Lucy1974] L. Lucy, "An iterative technique for the rectification of observed distributions", The Astronomical Journal, vol. 79, p. 745, 1974.

[Lundberg2002] K. Lundberg. *Noise Sources in Bulk CMOS*, Unpublished, October 2002. [Online]. Available: http://web.mit.edu/klund/www/papers/UNP_noise.pdf. [Accessed: 20-Jul-2015].

[MATLAB] "Stereo Calibration and Scene Reconstruction - MATLAB & Simulink Example". [Online]. Available: http://www.mathworks.com/help/vision/ examples/stereo-calibration-and-scene-reconstruction.html. [Accessed: 29-Apr- 2015].

[Maxwell1871] J. C. Maxwell and P. Pesic, *Theory of heat*, Dover ed. Mineola, NY: Dover Publications, 2001.

[McWhorter1957] A. L. McWhorter. "1/f noise and germanium surface properties". In R. H. Kingston, editor, *Semiconductor Surface Physics*, pages 207-228. University of Pennsylvania Press, Philadephia, 1957.

[Milanfar2010] Peyman Milanfar, *Super-Resolution Imaging*. CRC Press, 2010.

[Nasrollahi2014] K. Nasrollahi and T. B. Moeslund, "Super-resolution: a comprehensive survey," *Machine Vision and Applications*, vol. 25, no. 6, pp. 1423–1468, Aug. 2014.

[Parenti1994] R. R. Parenti, "Application of Fourier-space image-restoration techniques to turbulence compensation and sparse-aperture imaging," Jun. 1994.

[PTLens] "PTLens | Adobe Lens Correction Filter". [Online]. Available: http://www.epaperpress.com/ptlens/ lensCorrect.html. [Accessed: 29- Apr- 2015].

[Ray2002] S. F. Ray, *Applied photographic optics: lenses and optical systems for photography, film, video, electronic and digital imaging*. Oxford: Focal, 2002.

[Reichenbach1991] S. E. Reichenbach, S. K. Park, and R. Narayanswamy, "Characterizing digital image acquisition devices," *Opt. Eng*, vol. 30, no. 2, pp. 170–177, Feb. 1991.

[Richardson1972] W. Richardson, "Bayesian-Based Iterative Method of Image Restoration", Journal of the Optical Society of America, vol. 62, no. 1, p. 55, 1972.

[Sat1998] "What Are The Different Kinds Of Satellite Orbits?," *Visual Satellite Observing FAQ*, Feb-1998. [Online]. Available: http://www.satobs.org/faq/Chapter-04.txt. [Accessed: 24-Jun-2015].

[Shepp1982] L. Shepp and Y. Vardi, "Maximum Likelihood Reconstruction for Emission Tomography", IEEE Transactions on Medical Imaging, vol. 1, no. 2, pp. 113-122, 1982.

[Sony2011] Sony Corporation, "Digital Camera Image Noise: Concept and Types," *Advantage of the CMOS Sensor*, 2011. [Online]. Available: http://www.cambridgeincolour.com/tutorials/image-noise.htm. [Accessed: 30-Jun-2015].

[Sroubek2008] F. Šroubek, G. Cristobal, and J. Flusser, "Simultaneous super-resolution and blind deconvolution," *Journal of Physics: Conference Series*, vol. 124, p. 012048, Jul. 2008.

[Starkov2013] I. Starkov, "Comprehensive Physical Modeling of Hot-Carrier Induced Degradation," Vienna University of Technology, Wien, Austria, 2013.

[Steve2008] "Image deblurring using regularization » Steve on Image Processing", 2008. [Online]. Available: http://blogs.mathworks.com/steve/2008/07/21/image-deblurring-using-regularization. [Accessed: 29- Apr- 2015].

[Thirion1996] J.-P. Thirion, "Non-rigid matching using demons," in Proceedings CVPR '96, 1996 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1996, 1996, pp. 245–251.

[Wang2005] H. Wang, L. Dong, J. O'Daniel, R. Mohan, A. S. Garden, K. K. Ang, D. A. Kuban, M. Bonnen, J. Y. Chang, and R. Cheung, "Validation of an accelerated 'demons' algorithm for deformable image registration in radiation therapy," Phys Med Biol, vol. 50, no. 12, pp. 2887–2905, Jun. 2005.

[Weeden2010] B. Weeden, “The Space Review: Dancing in the dark: The orbital rendezvous of SJ-12 and SJ-06F (page 1),” 30-Aug-2010. [Online]. Available: http://www.thespacereview.com/article/1689/1. [Accessed: 13-May-2015].

[Yang2009] J. Yang, W. Yin, Y. Zhang and Y. Wang, "A Fast Algorithm for Edge-Preserving Variational Multichannel Image Restoration", SIAM J. Imaging Sci., vol. 2, no. 2, pp. 569-592, 2009.

[Zhang1999] Z. Zhang, "Flexible camera calibration by viewing a plane from unknown orientations," in *The Proceedings of the Seventh IEEE International Conference on Computer Vision*, 1999., vol.1, pp.666-673.

Chapter 5

[Hille2015] K. Hille, “Hubble and a Stellar Fingerprint,” *NASA: Hubble Space Telescope Images*, 21-Jan-2015. [Online]. Available: http://www.nasa.gov/mission_pages/hubble/multimedia/index.html. [Accessed: 05-May-2016].

[Hooge1969] F. N. Hooge. “1/f noise is no surface effect”. Physics Letters, 29A(3):139-140, April 1969.

[Hooge1994] F. N. Hooge. “1/f noise sources”. IEEE Transactions on Electron Devices, 41(11):1926-1935, November 1994.

[Hoshino2007] K. Hoshino and T. Nishimura, “Noise Reduction in CMOS Image Sensors for High Quality Imaging: The Autocorrelation Function Filter on Burst Image Sequences,” *ICGST-GVIP*, vol. 7, no. 3, Nov. 2007.

[Jayaraman1989] R. Jayaraman and C. G. Sodini, “A 1/f noise technique to extract the oxide trap density near the conduction band edge of silicon,” *IEEE Transactions on Electron Devices*, vol. 36, no. 9, pp. 1773–1782, Sep. 1989.

[Luisier2011] F. Luisier, T. Blu, and M. Unser, “Image Denoising in Mixed Poisson-Gaussian Noise,” *IEEE Transactions on Image Processing*, vol. 20, no. 3, pp. 696–708, Mar. 2011.

[Lundberg2002] K. Lundberg. *Noise Sources in Bulk CMOS*, Unpublished, October 2002. [Online]. Available: http://web.mit.edu/klund/www/papers/UNP_noise.pdf. [Accessed: 20-Jul-2015].

[McWhorter1957] A. L. McWhorter. “1/f noise and germanium surface properties”. In R. H. Kingston, editor, *Semiconductor Surface Physics*, pages 207-228. University of Pennsylvania Press, Philadephia, 1957.

[Nasrollahi2014] K. Nasrollahi and T. B. Moeslund, “Super-resolution: a comprehensive survey,” *Machine Vision and Applications*, vol. 25, no. 6, pp. 1423–1468, Aug. 2014.

[RistaDec2013] Jon Rista, “Digital Noise: What is it?,” *Nature Photography*, 17-Dec-2013. [Online]. Available: http://jonrista.com/2013/12/17/digital-noise-part-1-what-is-it/. [Accessed: 30-Jun-2015].

[Scherzer2009] O. Scherzer, M. Grasmair, H. Grossauer, M. Haltmeier, and F. Lenzen, “Image and Noise Models,” in *Variational Methods in Imaging*, Springer New York, 2009, pp. 27–49.

[Starkov2013] I. Starkov, “Comprehensive Physical Modeling of Hot-Carrier Induced Degradation,” Vienna University of Technology, Wien, Austria, 2013.

Chapter 6

[Abbe1873] E. Abbe, “Beitrage zur Theorie des Mikroskops und der mikroskopischen Wahrmehmung,” *Archiv für Mikroskopische Anatomie* (in German), vol. 9, pp. 413–420, 1873.

[Ayers1988] Ayers, G. R., Dainty J. C., “Iterative blind deconvolution method and its applications”, *Optics Letters*, vol. 13, no.7, pp. 547 - 549, July 1988.

[Bai2010] Y. Bai, J. Hu, and Y. Luo, “Self-adaptive blind super-resolution image reconstruction,” in *2010 3rd International Congress on Image and Signal Processing (CISP)*, 2010, vol. 3, pp. 1208–1212.

[Baker2002] S. Baker and T. Kanade, “Limits on super-resolution and how to break them,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 9, pp. 1167–1183, Sep. 2002.

[Bauer2011] T. Bauer, “Super-Resolution Imaging: The Use Case of Optical Astronomy,” *Proceedings of the IADIS International Conference on Computer Graphics, Visualization, Computer Vision and Image Processing, CGVCVIP 2011*, p. 49-59, Rome, Italy, 2011, ISBN 978-972-8939-48-9.

[BenEzra2005] M. Ben-Ezra, A. Zomet, and S. K. Nayar, “Video super-resolution using controlled subpixel detector shifts,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 6, pp. 977–987, Jun. 2005.

[Borman2004] S. Borman and R. L. Stevenson, “Linear models for multiframe super-resolution restoration under nonaffine registration and spatially varying PSF,” *Proc. SPIE Computational Imaging II*, vol. 5299, pp. 234–245, 2004.

[Born1999] M. Born and E. Wolf, *Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light*. Cambridge University Press, 1999.

[Bouguet2000] J. Bouguet, “Pyramidal implementation of the Lucas Kanade feature tracker,” *Intel Corporation, Microprocessor Research Labs*, 2000.

[Campisi2016] P. Campisi and K. Egiazarian, *Blind Image Deconvolution: Theory and Applications*. CRC Press, 2016.

[Couzinie2013] F. Couzinié-Devy, J. Sun, K. Alahari, and J. Ponce, “Learning to Estimate and Remove Non-uniform Image Blur,” in *2013 IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 1075–1082.

[Dahl2017] R. Dahl, M. Norouzi, and J. Shlens, “Pixel Recursive Super Resolution,” *arXiv:1702.00783 [cs]*, Feb. 2017.

[Delbracio2012] M. Delbracio, P. Musé, A. Almansa and J. Morel, "The Non-parametric Sub-pixel Local Point Spread Function Estimation Is a Well Posed Problem", *International Journal of Computer Vision*, vol. 96, no. 2, pp. 175-194, 2012.

[Dey2004] N. Dey, L. Blanc-Feraud, C. Zimmer, P. Roux, Z. Kam, J.-C. Olivio-Marin, and J. Zerubia. “3D Microscopy Deconvolution Using Richardson-Lucy Algorithm with Total Variation Regularization,” Technical Report, INRIA, 2004.

[Droege2012] D. R. Droege, R. C. Hardie, B. S. Allen, A. J. Dapore, and J. C. Blevins, “A real-time atmospheric turbulence mitigation and super-resolution solution for infrared imaging systems,” 2012, vol. 8355, p. 83550R–83550R–17.

[Elad1999] M. Elad and A. Feuer, “Superresolution restoration of an image sequence: adaptive filtering approach,” *IEEE Transactions on Image Processing*, vol. 8, no. 3, pp. 387–395, Mar. 1999.

[Farsiu2004] S. Farsiu, M. D. Robinson, M. Elad, and P. Milanfar, “Fast and robust multiframe super resolution,” *IEEE Transactions on Image Processing*, vol. 13, no. 10, pp. 1327–1344, 2004.

[Gerchberg1974] R. W. Gerchberg, “Super-resolution through Error Energy Reduction,” *Optica Acta: International Journal of Optics*, vol. 21, no. 9, pp. 709–720, Sep. 1974.

[Gunturk2012] B. K. Gunturk and X. Li, *Image Restoration: Fundamentals and Advances*. CRC Press, 2012.

[Hardeep2013] P. Hardeep, P. B. Swadas, and M. Joshi, “A Survey on Techniques and Challenges in Image Super Resolution Reconstruction,” *IJCSMC*, vol. 2, no. 4, pp. 317–325, Apr. 2013.

[Hardie2015] R. C. Hardie, D. R. Droege, A. J. Dapore, and M. E. Greiner, "Impact of Detector-element Active-area Shape and Fill Factor on Image Sampling, Restoration, and Super-Resolution" (2015). *Electrical and Computer Engineering Faculty Publications.* Paper 55. http://ecommons.udayton.edu/ece_fac_pub/55

[Irani1991] M. Irani and S. Peleg, “Improving Resolution by Image Registration,” *CVGIP: Graph. Models Image Process.*, vol. 53, no. 3, pp. 231–239, Apr. 1991.

[Irani1993] M. Irani and S. Peleg, “Motion Analysis for Image Enhancement: Resolution, Occlusion, and Transparency,” *Journal of Visual Communication and Image Representation*, vol. 4, no. 4, pp. 324–335, Dec. 1993.

[Keren1988] D. Keren, S. Peleg, R. Brada, "Image sequence enhancement using sub-pixel displacements," in Computer Society Conference on Computer Vision and Pattern Recognition, 1988. Proceedings CVPR '88., pp.742-746, 5-9 Jun 1988

[Lin2004] Z. Lin and H.-Y. Shum, “Fundamental Limits of Reconstruction-Based Superresolution Algorithms under Local Translation,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 1, pp. 83–97, Jan. 2004.

[Lowe1999] Lowe, David G. (1999). "Object recognition from local scale-invariant features". *Proceedings of the International Conference on Computer Vision*. pp. 1150–1157.

[Lucas1981] B. D. Lucas and T. Kanade, “An iterative image registration technique with an application to stereo vision,” in *In IJCAI81*, 1981, pp. 674–679.

[Lucchese2000] L. Lucchese and G. M. Cortelazzo, “A noise-robust frequency domain technique for estimating planar roto-translations,” *IEEE Transactions on Signal Processing*, vol. 48, no. 6, pp. 1769–1786, Jun. 2000.

[Mackey2013] C. Mackay, “High-Efficiency Lucky Imaging,” *Monthly Notices of the Royal Astronomical Society*, vol. 432, no. 1, pp. 702–710, Jun. 2013.

[Marcel1997] B. Marcel, M. Briot, and R. Murrieta, “Calcul de translation et rotation par la transformation de Fourier,” *TS. Traitement du signal*, vol. 14, no. 2, pp. 135–149, 1997.

[Mercimek2009] Muharrem Mercimek, “Image Deconvolution: A Technical Overview,” The University of Tennessee-Knoxville, Imaging, Robotics, and Intelligent Systems Laboratory, Technical Report, Spring 2009.

[Michaeli2013] T. Michaeli and M. Irani, “Nonparametric Blind Super-resolution,” presented at the *Proceedings of the IEEE International Conference on Computer Vision*, 2013, pp. 945–952.

[Milanfar2010] P. Milanfar, *Super-Resolution Imaging*. CRC Press, 2010.

[Nasrollahi2014] K. Nasrollahi and T. B. Moeslund, “Super-resolution: a comprehensive survey,” *Machine Vision and Applications*, vol. 25, no. 6, pp. 1423–1468, Aug. 2014.

[InLite] “Barcode Reader. Free Online Web Application.” [Online]. Available: https://online-barcode-reader.inliteresearch.com/. [Accessed: 13-Feb-2017].

[Pham2006] T. Pham, L. van Vliet and K. Schutte, "Robust Fusion of Irregularly Sampled Data Using Adaptive Normalized Convolution", EURASIP Journal on Advances in Signal Processing, vol. 2006, pp. 1-13, 2006.

[Rajagopalan2003] A.N. Rajagopalan and V.P. Kiran, “Motion-free super resolution and the role of relative blur,” *Journal of the Optical Society of America A*, vol. 20, no. 11, pp. 2022–2032, 2003.

[Rajan2001] D. Rajan and S. Chauhuri, “Simultaneous Estimation of Super-Resolved Intensity and Depth Maps from Low Resolution Defocused Observations of a Scene,” *Proc. Int’l Conf. Computer Vision*, pp. 113-118, July 2001.

[Santis1975] P. D. Santis and F. Gori, “On an Iterative Method for Super-resolution,” *Optica Acta: International Journal of Optics*, vol. 22, no. 8, pp. 691–695, Aug. 1975.

[Schelten2014] K. Schelten and S. Roth, “Localized Image Blur Removal through Non-parametric Kernel Estimation,” in *2014 22nd International Conference on Pattern Recognition*, 2014, pp. 702–707.

[Sroubek2008] F. Šroubek, G. Cristobal, and J. Flusser, “Simultaneous super-resolution and blind deconvolution,” *Journal of Physics: Conference Series*, vol. 124, p. 12048, Jul. 2008.

[Sun2015] J. Sun, W. Cao, Z. Xu, and J. Ponce, “Learning a Convolutional Neural Network for Non-Uniform Motion Blur Removal,” presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 769–777.

[Szeliski2006] R. Szeliski, “Image Alignment and Stitching: A Tutorial,” *Found. Trends. Comput. Graph. Vis.*, vol. 2, no. 1, pp. 1–104, Jan. 2006.

[Tanaka2005] M. Tanaka and M. Okutomi, “Theoretical analysis on reconstruction-based super-resolution for an arbitrary PSF,” in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005. CVPR 2005*, 2005, vol. 2, pp. 947–954 vol. 2.

[Temerinac2010] M. Temerinac-Ott, “Tile-based Lucy-Richardson Deconvolution Modelling a Spatially-Varying PSF for Fast Multiview Fusion of Microscopical Images,” University of Freiburg(260), Technical Report 260, 2010.

[Tian2011] J. Tian and K.-K. Ma, “A survey on super-resolution imaging,” *SIViP*, vol. 5, no. 3, pp. 329–342, Sep. 2011.

[Tsai1984] R. Tsai, T. Huang, “Multi-frame image restoration and registration,” *Advances in Computer Vision and Image Processing*, vol. 1, no. 2, JAI Press Inc., Greenwich, CT, 1984, pp. 317–339.

[Vandewalle2006] P. Vandewalle, S. Süsstrunk, and M. Vetterli, “A Frequency Domain Approach to Registration of Aliased Images with Application to Super-resolution,” *EURASIP Journal on Advances in Signal Processing*, vol. 2006, pp. 1–15, 2006.

[Wang2004A] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image quality assessment: from error visibility to structural similarity,” *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[Wang2004B] Z. Wang and F. Qi, “On ambiguities in super-resolution modeling,” *IEEE Signal Processing Letters*, vol. 11, no. 8, pp. 678–681, Aug. 2004.

[Wang2005] Z. Wang and F. Qi, “Analysis of Multiframe Super-resolution Reconstruction for Image Anti-aliasing and Deblurring,” Image Vision Comput., vol. 23, no. 4, pp. 393–404, Apr. 2005.

[Wang2009] Z. Wang and A. C. Bovik, “Mean squared error: Love it or leave it? A new look at Signal Fidelity Measures,” *IEEE Signal Processing Magazine*, vol. 26, no. 1, pp. 98–117, Jan. 2009.

[Xue2015] F. Xue and T. Blu, “A Novel SURE-Based Criterion for Parametric PSF Estimation,” *IEEE Transactions on Image Processing*, vol. 24, no. 2, pp. 595–607, Feb. 2015.

[Zomet2001] A. Zomet, A. Rav-Acha, and S. Peleg, “Robust super-resolution,” in *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2001. CVPR 2001*, 2001, vol. 1, p. I-645-I-650 vol.1.

Chapter 7

[Bar2001] Y. Bar-Shalom, X.-R. Li, and T. Kirubarajan, *Estimation with applications to tracking and navigation*. New York: Wiley, 2001.

[Cuevas2005] E. V. Cuevas, D. Zaldivar, and R. Rojas, “Kalman filter for vision tracking,” 2005.

[Julier2002] S. J. Julier, “The scaled unscented transformation,” in *Proceedings of the 2002 American Control Conference (IEEE Cat. No.CH37301)*, 2002, vol. 6, pp. 4555–4559 vol.6.

[Julier2004] S. J. Julier and J. K. Uhlmann, “Unscented filtering and nonlinear estimation,” *Proceedings of the IEEE*, vol. 92, no. 3, pp. 401–422, Mar. 2004.

[Kalman1960] R. E. Kalman, “A New Approach to Linear Filtering and Prediction Problems,” *J. Basic Eng*, vol. 82, no. 1, pp. 35–45, Mar. 1960.

[Knopp2010] Jan Knopp, Mukta Prasad, Gert Willems, Radu Timofte, and Luc Van Gool, "Hough Transform and 3D SURF for Robust Three Dimensional Classification", *European Conference on Computer Vision* (ECCV), 2010.

[Lowe1999] Lowe, David G. (1999). "Object recognition from local scale-invariant features". *Proceedings of the International Conference on Computer Vision*. pp. 1150–1157.

[Mazor1998] E. Mazor, A. Averbuch, Y. Bar-Shalom, and J. Dayan, “Interacting multiple model methods in target tracking: a survey,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 34, no. 1, pp. 103–123, Jan. 1998.

[Plaza2006] A. Plaza, D. Valencia, J. Plaza, and P. Martinez, “Commodity cluster-based parallel processing of hyperspectral imagery,” *Journal of Parallel and Distributed Computing*, vol. 66, no. 3, pp. 345–358, Mar. 2006.

[Roweis1999] S. Roweis and Z. Ghahramani, “A Unifying Review of Linear Gaussian Models,” *Neural Comput.*, vol. 11, no. 2, pp. 305–345, Feb. 1999.

[Smith2012] Randall B. Smith, “Introduction to Hyperspectral Imaging,” MicroImages Inc., Jan. 2012.

[Zitova2003] B. Zitová and J. Flusser, “Image registration methods: a survey,” *Image and Vision Computing*, vol. 21, pp. 977–1000, 2003.

Chapter 8

[Clare1991] B. W. Clare and D. L. Kepert, “The optimal packing of circles on a sphere,” *J Math Chem*, vol. 6, no. 1, pp. 325–349, Dec. 1991.

[Clukey2016] S. A. Clukey, "Architecture for Real-Time, Low-SWaP Embedded Vision Using FPGAs. " Master's Thesis, University of Tennessee, 2016.

[Tracking1998] “Tracking Systems,” *Introduction to Naval Weapons Engineering*, 20-Jan-1998. [Online]. Available: https://fas.org/man/dod-101/navy/docs/es310/trksys/trksys.htm. [Accessed: 09-May-2017].

Vita

Michael Vaughan was born to David and Marva Vaughan in Provo, Utah and moved to Knoxville in March 1991. He received his Bachelor of Science degree in Computer Engineering from The University of Tennessee in Knoxville, Tennessee in May 2006. After graduating, Michael spent two years in Singapore and Sri Lanka as a volunteer missionary for The Church of Jesus Christ of Latter-day Saints. He joined the Imaging, Robotics, and Intelligent Systems (IRIS) Laboratory as a graduate research assistant in August 2008 in the direct Ph.D. program with a major in Computer Engineering, specializing in computer vision. He will complete the requirements for the doctorate degree in 2017.

His research interests include methods in computational imaging, novel image acquisition systems, imaging for surveillance and security applications, image and data fusion, autonomous navigation, process automation, face recognition and tracking, statistical and correlative data analysis, and photogrammetry.