Model Agnostic Extreme Sub-pixel Visual Measurement and Optimal Characterization

by

Michael R. Munroe

A Thesis Presented in Partial Fulfillment
of the Requirements for the Degree
Master of Science

Approved June 2012 by the
Graduate Supervisory Committee

Patrick Phelan, Chair
Eric Kostelich
Alex Mahalov

ARIZONA STATE UNIVERSITY

August 2012

ABSTRACT

It is possible in a properly controlled environment, such as industrial metrology, to make significant headway into the non-industrial constraints on image-based position measurement using the techniques of image registration and achieve repeatable feature measurements on the order of 0.3% of a pixel, or about an order of magnitude improvement on conventional real-world performance. These measurements are then used as inputs for a model optimal, model agnostic, smoothing for calibration of a laser scribe and online tracking of velocimeter using video input. Using appropriate smooth interpolation to increase effective sample density can reduce uncertainty and improve estimates. Use of the proper negative offset of the template function has the result of creating a convolution with higher local curvature than either template of target function which allows improved center-finding. Using the Akaike Information Criterion with a smoothing spline function it is possible to perform a model-optimal smooth on scalar measurements without knowing the underlying model and to determine the function describing the uncertainty in that optimal smooth. An example of empiric derivation of the parameters for a rudimentary Kalman Filter from this is then provided, and tested. Using the techniques of Exploratory Data Analysis and the Formulize genetic algorithm tool to convert the spline models into more accessible analytic forms resulted in stable, properly generalized, KF with performance and simplicity that exceeds “textbook” implementations thereof. Validation of the measurement includes that, in analytic case, it led to arbitrary precision in measurement of feature; in reasonable test case using the methods proposed, a reasonable and consistent maximum error of around 0.3% the length of a pixel was achieved and in practice using pixels that were 700nm in size feature position was located to within ± 2 nm. Robust applicability is demonstrated by the measurement of indicator position for a King model 2-32-G-042 rotameter.

DEDICATION

I dedicate this to my friends and family:

to my wife, Micah, your exceptional encouragement, support, and faith, have been a mountain and a refuge – you are the gift of Grace to me and the noblest person I know;

to Frank and Sabina Peyton who said “do the unthinkable, go to college”;

to Stewart and Becky Bruner who made a bridge across the uncrossable;

to Jeff and Gwen Lavell who showed me how to look up on high;

to Michele Milano who opened up a door in the impossible future;

to my grandfather, Jack Clifford Hansen, who was killed long before we ever could have discussed a shared interest in Kalman filtering;

to the Providence who brought me here and carries me on eagles wings.

You have been my ‘sine qua non’ – I could have done nothing if not for you.

ACKNOWLEDGMENTS

I would like to thank Dr. Patrick Phelan for taking the risk of sponsoring a non-traditional student, and for a balanced approach to managing my work. By allowing me to suffer more of the consequences of my own decisions, your restraint directed my independence in healthy and effective directions, and motivated me to more action than anything else could have.

I would like to thank Dr. Ronald Adrian for teaching me how to approach problems in visual measurement in a technically systematic, rigorous, and aggressive manner. Yours are the shoulders of a giant even though I was not looking farther I was able to see well where I looked because of you.

I have been the beneficiary of the excellent proofreading, grammar, and human readability expertise of Elizabeth Nicol. Your ability to speak both the language of technology and the language of human beings is admirable. Thank you for your help.

I would like to thank Charles Singleton and Xavier Brun for bringing me the problems that inspired this work. You taught me that the value of cultivating the reputation as a solver of exotic problems is I will be brought many interesting challenges.

It is hard to make a work-life-education balance in a world that doesn't want to allow even a work-life balance. I would like to thank Mike MacGregor, and Ibrahim Bekar – managers who said "keep going" in actions and not just words.

TABLE OF CONTENTS

Page

Page

LIST OF FIGURES

Figure Page

INTRODUCTION

The following describes two "model agnostic" tools for use with the scientific method: 1) a measurement procedure, described in terms of measuring physical dimension on an object using digital images but extensible to non-image metrology domains described by the diffusion of information from a central point of interest, and 2) a model-free but model-selection-optimal smooth that results in analytic expressions describing the underlying system. The nature of tools would be more immediately recognized if a different label was used, but their nature and function would be falsely represented as well, so although they can go by other labels, the term "model-agnostic" is preferred. "Non-parametric" is the most common substitute term, but it implies a lack of parameters when in fact there are many more parameters than traditional models such as polynomials. Given the number of parameters, the term "hyper-parametric" is slightly more appropriate, but it is still misleading because although there are many more parameters, they act together in what statisticians call a "non-informative" manner. To say it is "non-informative" means that the initial form does less to lead one from unarticulated initial assumptions down a path that is more prone to error, and it does not mean that it provides any less information than other methods – it provides more. The term that fits appropriately is "model-agnostic" because it encapsulates a large number of common, useful, and simpler tools in a framework that can reduce the initial assumptions and resultant errors, and because it indicates that these tools have a different set of baggage than those initially developed during the 17th century, which are still in common use.

The order of presentation of these tools is structured so that it will parallel, as far as it can, the last two steps of the flowchart of the scientific method with the intention of supporting the paradigm presented by (Anand, 2010) in the framework he presents for

"Decision Oriented Design of Experiments". Anand argues that the uniform weighting on variation is equivalent to an improper and negatively informative prior and then provides a framework for approaching Design of Experiments (DOE) with results that optimally inform the decision of the hypothesis – reweighting the variation appropriately. This work parallels the intention of Anand in that the results of application are meant to, through their "non-informative" utility, or "agnosticism", maximally support the next steps in scientific investigation or data-driven improvement of production processes.

Some Background on Sub-pixel Metrology

Visual metrology is the science of making measurements using phenomena that parallel the human visual apparatus: light, optics, images, and image analysis. Metrology, the science of measurement, is also an important part of process design, process health tracking, and process improvement, and it is used in product quality assurance. Not only is it a part of research in science but it is a fundamental part of profitability for industrial production. Effective production process development requires information whose certainty or uncertainty is well characterized, and has been engineered to provide relevant and actionable results. Optimized production requires the operating process to be held as closely as possible to the ideal process parameters established during design phase. Visual metrology is applicable in many of these cases and is desirable as a measurement tool because it utilizes the most data-rich of the human senses: sight. It is for this reason that the model-agnostic measurement is presented and described in terms of visual metrology: the measurement of physical locations on an object using a digital image.

Naïve feature measurement using images and human eyes requires an absolute minimum of 5-8 pixels to comprise a trivial image feature (Shelepin & Bondarko, 2004)

due both to the packing of photoreceptors in the eye and the downstream neurological processors, which are optimized to handle that particular format of information. (Extensive description of this is contained in the numeric experiments section and is further along in this work.) Slightly less simple features can require hundreds to thousands of pixels for humans to detect and process. An example is the number of pixels required to differentiate one human face from another – a fact that drives consumer cameras to mega-pixels and beyond. Typical numbers of pixels required for bare-eye feature detection of production typical features can be on the order of several hundred pixels which comprise a region that is on the order of ten's of pixels to a side. Use of first year calculus and computer programming will often reduce the typical number of pixels required for reliable computed metrology to approximately 10-20 pixels for a relatively simple feature. Advanced math, including frequency domain methods and techniques of "textbook" sub-pixel image registration, will allow the measurement of feature locations with the significant reduction to the order of the size of a single pixel (Reed, 2010).

Improvement in visual metrology now requires that sub-pixel methods be more aggressively explored. The uncertainty of a measurement is dependent on the order of the feature size. This fact directly competes with the definition of quality. Quality, as an engineering term, requires the continuous and systematic reduction of uncertainty in the process and the product. This reduction in uncertainty, Edward Deming asserts, allows the producer to take advantage of unexpected opportunities and maximize the perceived value of the product in the marketplace. In visual metrology, the limitation imposed by pixel size places a floor on the uncertainty of a measurement and therefore puts a ceiling on the quality that the metrology system is capable of supporting. The use of sub-pixel

methods can remove the discretization imposed by the camera (or other sensor) as an artificial limit on the quality of the process.

A brief exploration of measurement using sub-pixel methods reveals that it has been successfully applied to a number of areas including structural elastometry (Kelly, Azeloglu, Kochpura, Sharma, & Gaudette, 2007), thermometry (Laval, 2008), velocimetry (Yamamoto, 2009), anthropometry (Yu, 2008), densitometry (Liebgott, 2008), econometrics (Wang, 2007) and epidemiology (Hughes, 2008). These demonstrate that systems which discretize continuous data over a domain and for which relevant reference allows output profiles to be created can utilize this method both for calibration and for ongoing metrology.

Alternative Approaches

There are a number of formalisms used to approach measurement and modeling. In order to evaluate the merits of each component it is valuable to deconstruct the formalisms into the fields of: "basis", "scope", "metric", "fit method" and "validation". Some bases include "polynomial", "trigonometric", "radial", "affine transformed sigmoid," and others. In some cases, these are assembled in networks, for example, the "affine transformed sigmoid" is used to construct a number of Neural Network topologies (Bishop, 1996). Some scopes for these bases include: "global", in which all samples are used to determine parameters; "minimally local", in which only enough adjacent samples are used to determine local interpolation; "hard-bounded intermediate," or windowed; "soft bounded intermediate," or weighted. A number of ways of computing the error between the analytic expressions used to approximate the data and the data itself are used, including the L-norm family members (L1 or "taxicab", L2 or "Euclidean", L-infinity norm or "Max Error"). Other error relationships include Mahalanobis distance,

Cross-correlation, and a number of Winsorized statistics operated on the preceding error measures. These error values are related to parameters of the basis-functions such that parameter adjustments are found that reduce, as far as possible, the value of the error metric using procedures that include: gradient descent, expectation maximization, information criteria, genetic algorithms, particle swarm optimization, and linear simplex optimization. These lists are by no means exhaustive, but they indicate the very broad scope of candidate methods.

Three of the more popular approaches to problems such as this are neural networks, Kernel Methods, and compressed sensing. Neural networks have the advantage of being able to generalize high dimensional data well, but can be challenging to properly train, and sometimes capture fine (local-scale) features poorly. "Kernel methods" (locally weighted methods) can be computationally cheap, allow analytic expressions for gradients and confidence intervals, and can adequately represent fine-scale functions well and easily, but they do not handle very high dimension, or large data, efficiently in time. Compressed sensing (Candes & Tao, 2006) is particularly interesting because it was the first concrete way to "bypass" the Shannon Limit (Shannon, 1948) in a wide variety of systems by randomly sampling in the domain then performing linear optimization using an L1-norm, a very computationally inexpensive method, to fit the samples to the basis-functions.

In this work a Neural Network basis was not used because though they can easily generalize the data, they have more difficulty in representing the fine details. The model-optimal spline-smoothing (described below) is a Kernel method informed by a model-optimal information criterion (detailed below). The extreme sub-pixel measurement method (also described extensively below) can be considered a close relative to the compressed sensing in that it assumes no basis outside that provided by the data itself,

however because of the highly constrained nature of the domain a random sampling strategy was less adequate than a uniform sampling strategy. Although it was not derived using the formalization of compressed sensing, and though it could be deconstructed and analyzed using that sort of framework, this is not done here because it is substantially beyond the scope of this work.

EXTREME SUB-PIXEL MEASUREMENT

The reason that the extreme sub-pixel metrology described herein is considered model agnostic is that it does not presume a basis outside of that provided by the data. The method described below takes a segment of intensities, either aligned with the part or with the rows, and as long as the information spectrum generating that segment or "template" is a reasonable characterization of the edge; it allows measurement to the fullest limit of the information in the image. Each new measurement makes no assumption about the previous one, and each new measurement uses every bit of information available. The single assumption of this method is that the information spectrum of the object being measured is either unchanging or slowly changing. If the profile used is "reasonably characteristic" – if it has the low frequency components are consistent across the domain, then profile mismatch cannot occur.

The following is derived primarily from empirical experience and not from first principles. It is based on the common sub-pixel technique used in stereo 2-pulse particle image velocimetry (PIV) and is convolution based (Adrian, 1991). The informing technique is used to measure laser-illuminated particle positions in a fluid, allowing computation of an approximation of the velocity field. A first image, used as a reference, is captured and broken down into sub-images. Each sub-image is then convoluted with similarly captured and sectioned sub-images of the target. The location of the maximum

of the convolution indicates the offset between the two images. Iterative variations on this procedure allow the velocity field to be resolved down to a level typically greater than 90% of the individual particles in the field, and allow position measurement that is accurate to around 10% of the size of a pixel.

In the large majority of industrial process, continuous material is measured rather than particles. This enables some variations on the process that significantly improve the quality of the measurement. The most significant variation in application between industrial use and PIV is that in particle velocimetry case, the majority of the image space is empty for the purpose of avoiding occlusion of illuminated particles by other particles in the fluid, while in the case of a solid-metrology, a large majority of the image can be comprised of information – the total information density is much higher. For an XGA image (480x640 pixels) it is reasonable for the information to live in a band that is 100 pixels by 640 pixels in size – allowing ~64,000 measurements to inform the output. In a typical continuous-material measurement a 2-dimensional reference image of a geometric structure (edge, curve, or section) is used as the template. Ideal conditions occur when the template is taken from and used to measure one or more "peak-like" shapes with maxima toward the interior, and minima toward the boundaries and enveloping the maxima. This assumes that the information spectrum is "mostly" constant, but does not assume anything about its nature beyond that. It is also preferred for the method if the orientation of the pixels in the camera is positioned parallel to the axis of the template.

The numeric result obtained from using a convolution provides a clear peak indication. This process is repeated as required over the domain of the measurement. This allows a large number of sequential measurements of the data to identify an estimate of the position. Several "adjustments" to the typical method are shown below, including extreme super-sampling and negative offset of the template so that the convolution peak

has higher local (quadratic) curvature than either the template or the region it is being used to measure.

For the material described below, it is assumed that the parameter being measured is parallel to the x-axis of the digital image and that the measurements occur on the y-axis.

Measurement Procedure:

1. Determine template profile
    a. Adjust raw sampling size to be of appropriate density
    b. Smooth as needed (slightly) to improve generalization of the template
    c. Translate template in a negative direction by an appropriate magnitude
    d. Use interpolation to appropriately over-sample the reference
    e. Appropriately zero-pad sample “ends” to eliminate Gibbs “ringing” phenomena
2. Preprocess target profile
    a. Set values outside window of interest to zero
    b. Smooth appropriately
    c. Use interpolation to appropriately improve sampling density of image
3. Fit in y
    a. Perform discrete convolution between target and reference with output same size as target
    b. Perform least-squares analytic fit to quadratic over peak of convolution values
        i. Find peak using ‘max’
        ii. Use twice as many points as parameters in the fit

c. Compute continuous analytic root of fit

4. Post-processing

    a. Adjust by comparing to calibrated reference as appropriate.

Exploration of Measurement

This experiment is run over two families of test-cases: Analytic/Synthetic, and Actual. Analytic/Synthetic means that a software tool, in this case MatLab, was used to generate a more useful case to explore the phenomenology of the process. Actual indicates that a number of real-world samples, typically digital pictures derived from video footage, are used. The analytic and synthetic are used to show the development process for the heuristics or to demonstrate "benchmarking" of the methods. The results section containing the actual values follows the numeric experiments and heuristic derivation sections.

Using the central limit theorem, it can be asserted that many real world phenomena follow approximately normal distributions in their parameters. The central limit theorem asserts that the mean of arbitrary finite distributions is itself distributed normally. This says that the distribution of the mean of means asymptotically approaches a normal distribution. For example, correctly registering Gaussians can accurately represent actual phenomena that have Gaussian noise factors. It is also possible to approximate many non-Gaussian continuous systems using a sum of many superimposed Gaussians in the form of "radial basis functions". It is for these reasons that the analytic models explore elementary functions that can inform parameter settings and enable extensibility to less pristine problems.

A standard normal distribution is used as the reference, the "yardstick" that performs measuring, for the analytic case. Zero mean and unity variance are used as the

reference normal distribution, while a normal distribution with mean of pi, and standard deviation of the square root of two as the object to measure. The object distribution has parameters that are non-integer, so sampling points will not accidentally be placed on it. However, they are well known and will allow an estimation of the effectiveness of the method.

Four parameters are used to control the analytic case: initial sampling density, negative vertical offset of template intensity, interpolation density, and smoothing of discretized data. These are the "dials" to characterize for effective control of the method.

Initial sampling of the reference and object are set at 17 samples per feature – a typical minimum feature of features for the human mind (Shelepin & Bondarko, 2004). According to Shelepin and Bondarko it takes 5 receptors in the eye to clearly separate two points – to recognize that there are two objects instead of one. In this circumstance the four outer receptors are more highly activated than the center one. The transition of a horizontal surface through an intermediate incline to a separate surface – the simplest visual profile – is minimally represented by a structure of 3 separable 2-feature objects and requires 15+2 pixels where 15 are for the objects and the two are the "glue" that identify them as separate but connected. It is a cognitively useful value because it is a typical useful edge size for human-useful selection of references – it is the most efficient scale for building the visual intuition that informs algorithm evaluation. This is the smallest "template" that the eye can process, and it is therefore the most efficient reference for the downstream (in terms of information flow) neural processing hardware to evaluate.

The vertical offset applied to the reference curve has the advantage of creating higher curvature in the convoluted result, thereby giving a better indicator of maximum location. This is counter-intuitive because the classic framework for understanding these

convolutions is probabilistic and a negative probability has no physical meaning. The convolution of two non-offset Gaussian distributions yields another Gaussian whose variance is the sum of the squares of their standard deviations. If the reference Gaussian is negatively offset, then the curvature of the convolution near its mean is significantly smaller than that of either input. If the negative offset is too high, it inverts the concavity of the convolution, and the estimate of the mean, the maximum in convolution, jumps to the edge of the target. A balanced approach to the negative offset maximizes concavity of the convolution near the mean without inverting its value.

Interpolation density improves estimation of maximum location, but can introduce errors where the underlying function is not precisely cubic. The use of high-oversampling is counter-intuitive because it does not add information to the system. It acts, in effect, like a quadrature assuring that the distance between the mean of each Gaussian and its sample points tends toward zero – and that the number of samples on either side of the mean tend toward having equal weight in the convolution. An odd number of points distributed uniformly and symmetrically across the analytic Gaussians and exactly locating one sample point at the mean yield a convolution peak that is analytically exactly at the point of maximum correlation. When the Gaussian is known a-priori then an exact sampling can be contrived to result in perfect measurement, but when the sum of Gaussians comprising the asymptotic Galerkin approximation of the real-world intensity profile is not known the blind cubic interpolation in the limit of many samples delivers results that approach ideal sampling. The higher sampling density also improves the robustness of the algorithm by not allowing under-sampling to detract from the results, while also constraining the noise-energy captured. The use of few spatially segregated samples acts to reduce the total noise energy represented in the information.

Too much over-sampling adversely impacts compute-times, and a calibration of the measurement is always required.

Smoothing overcomes both noise in the data and the effect of discretization. When the image is converted from “analytic”, a 64 bit ‘double’ representation, as described by IEEE 754 (Moler, 1996), to the unsigned 8, 16, or 24 bit integer values stored in images, there is information loss that adversely impacts method performance over the analytic case. The use of an appropriate smooth allows for locally unique intensity values, smooth slope, and a smaller noise impact to the measuring process. It can effectively repair the impact of the discretization on the information content.

Figure 1 shows example reference and object functions, as well as their resulting convolution and its maximum. Several features can be observed from this figure, including that the reference is zero-padded. This observation indicates that no false-indication was induced by Gibbs phenomena; the sampling begins and ends exactly on the endpoints. When the zero padding is not considered, this particular implementation falsely suggests that the best value for pi is 3.0, but when the padding is used, the exact result is recovered. If the reference Gaussian were not shifted on the y-axis, it could be observed that the convolution of two Gaussians is itself a Gaussian function. The radius of curvature of the convolution result on the figure is less than either of the source distributions. This has significant impact on the analytic centering. The analytic centering is accomplished by first performing LSQ fit of a quadratic to the maximum and several of its nearest neighbors and then computing the analytic root of that quadratic. If the radius of curvature is large, then the impact of noise is amplified and the error is increased. If the radius of curvature is small, then the region tends to have a better fit to a quadratic function over the points, and the impacts of noise on the estimate are reduced.

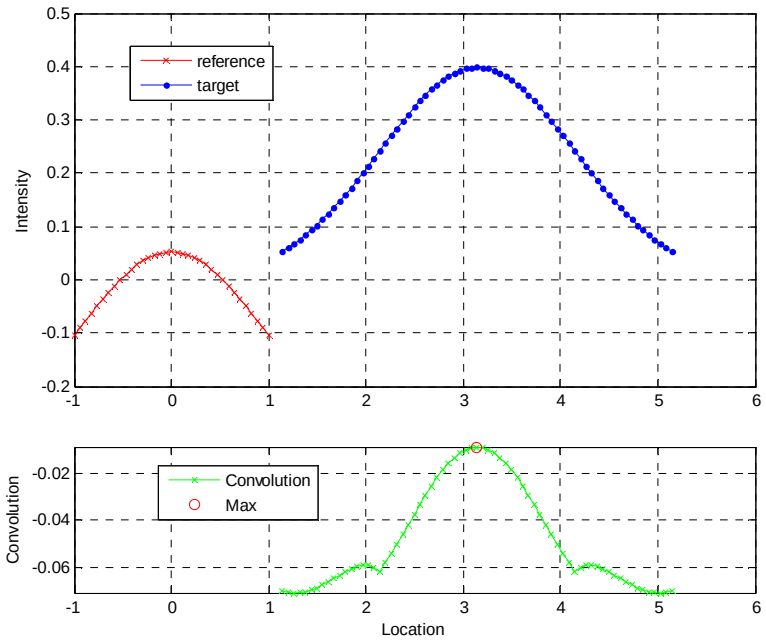

Figure 1 – Sub-pixel Measurement: Convoluting Gaussians

Initial Sampling Density

A very important question raised by this process is which spatial discretization to use since the pixels in an image are equivalent to a uniform spatial sampling. In the analytic form used, the value should be able to be approximated to arbitrary precision. Figure 2 shows the accuracy of the fit function when the sampling is swept from a range of 0.1 to 0.02. Many of the values are substantially below the error ceiling (red "+"), but one of the assumptions in this work, and one very consistent with both the premise for H-infinity filtering, is that it is engineered against a universe where maximum perversity is assumed. Constraint of the maximum error is also consistent with the framework for compressed sensing. The focus is on minimizing the absolute worst-case performance with the assumption that average results will be considerably improved.

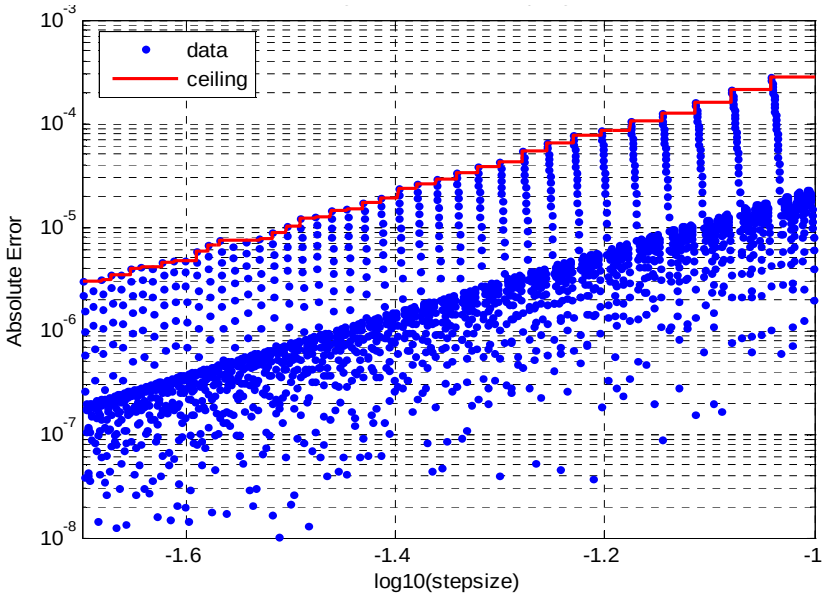

Figure 2–Sub-pixel Measurement: Error vs. Initial Sampling Density

The fit equation for error ceiling was determined, using the Eureqa genetic algorithm tool (Schmidt & Lipson, 2009), to be:

(1) $$y = 0.2738 \cdot \Delta x^3$$

This equation implies that without interpolation the maximum error expected for 17 samples per reference should be on the order of 0.044%. It can be observed that the typical error is about 1.2 decades, or about 16 times smaller than this maximum. When the number of sample measurements is odd and the spacing of those samples brings the left-most one very close in approach to the exact endpoint, then the method accuracy approaches zero error. The method comes close to this once per sample count and this happens repeatedly and more frequently as the initial sampling density is increased. Dividing expression 1 by the spacing yields an inverse quadratic relationship between sampling size and error per pixel.

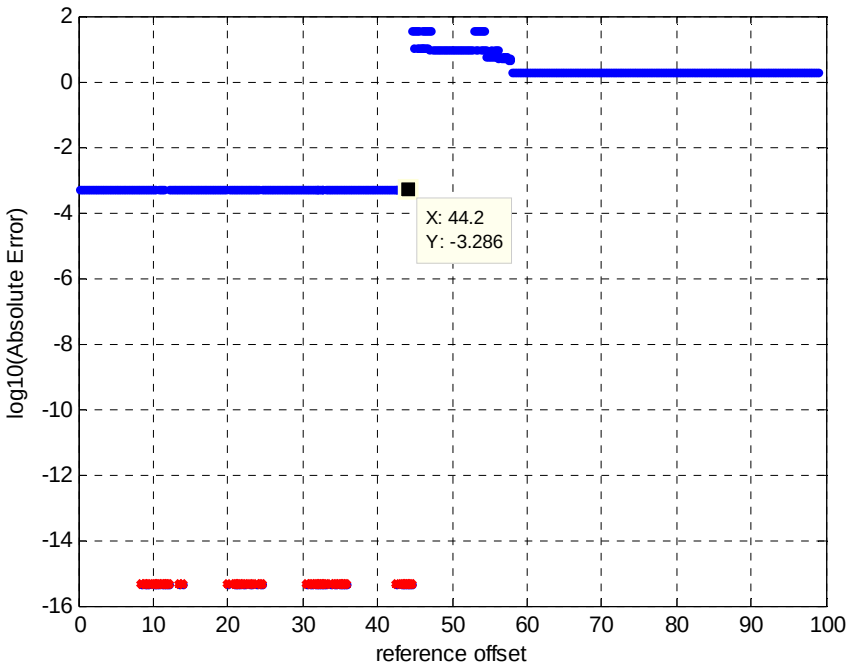

Figure 3 – Sub-pixel Measurement: Error vs. Reference Offset

Negative Offset of Template

Figure 3 shows the estimation error as a function of the offset of the reference. The domain is the percent downward shift, while the range indicates the scale of the error. A perturbed sampling value of 18, instead of 17 points as mentioned above, was used here to show the error at approximately 10^-3.2. Use of 17 samples resulted in the artificially low value of "eps", the smallest positive number the computer can represent, for the method – an artifact of the location of a sample at the exact mean and symmetric location of samples along both sides of the Gaussian. Notice how appropriate downward shifts caused the error to go down to round-off. Using this figure, a downward shift by the 23rd percentile value of the intensity was selected since it is toward the center of the region of good improvement caused by shift. After the 44th percentile, the shift of the template begins to harm the estimation of the mean.

Super-sampling Rate

Finite precision metrology and data representation does not allow for arbitrary sampling spacing. The continuous nature of the real world allows the peak of an observed Gaussian shaped object to exist in real-valued locations, but the camera can represent only a finite subset of them – the fixed midpoints of its pixels. It is for this reason and the reasons previously mentioned that a minimum energy, Hermite, piecewise cubic interpolation between points is used. This makes better use of the information from the image without necessarily requiring a higher pixel density camera. The Hermite interpolation also avoids artificial Gibbs effect variation associated with other cubic interpolation methods.

Figure 4 shows the error for interpolation. Again notice the very low values (red), essentially "eps" that originate from the analytic case. It can also be observed that the error initially increases to a peak and then decays along a trajectory that is approximately hyperbolic. Most of the variation as a function of interpolation factor happens close to zero, suggesting the use of a log-log scaled plot.

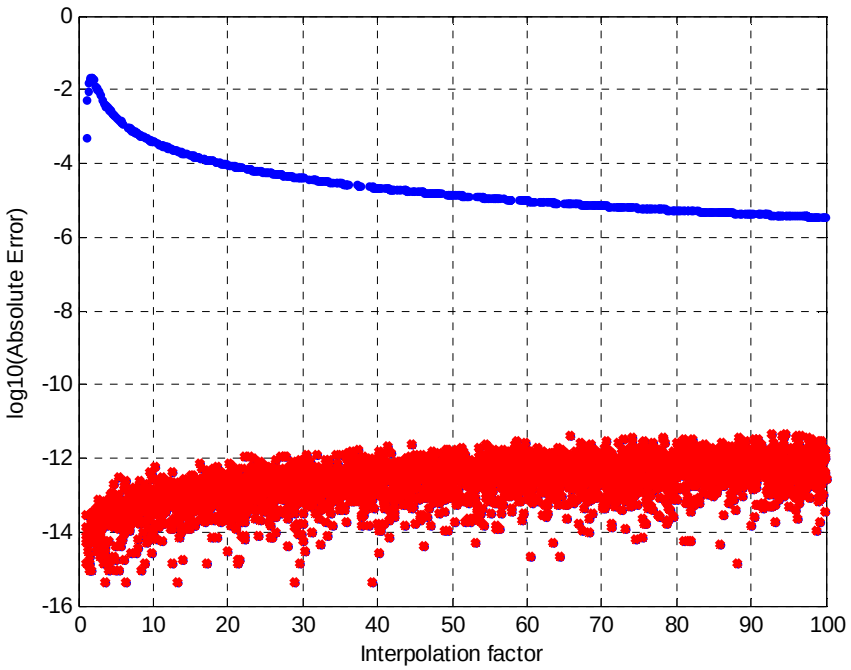

Figure 4– Sub-pixel Measurement: Error vs. Super-sampling Density. Ceiling is shown in dark blue while the "accidentally exact" values are shown in the lighter red.

An interpolated graph of coordinate transformed inverse error, the information, is shown in Figure 5. For all interpolation factors above about 8.78, and within the transformed coordinates, the improvement in error approximately follows a power law. A linear function in a log-log plot transforms into a power-law in linear-linear coordinates. A resampling factor of 20 reduces the error by more than a factor of 5.5 while resampling by a factor of 100 improves this to a factor over 145. The peak resolution improvement that can be achieved will be impacted by noise and use in a non-analytic case. This fact does not indicate that there is infinite margin for improvement, but if the noise can be managed it indicates there is some improvement available here. Substantial compute-time costs are incurred by using resampling ratios above 100.

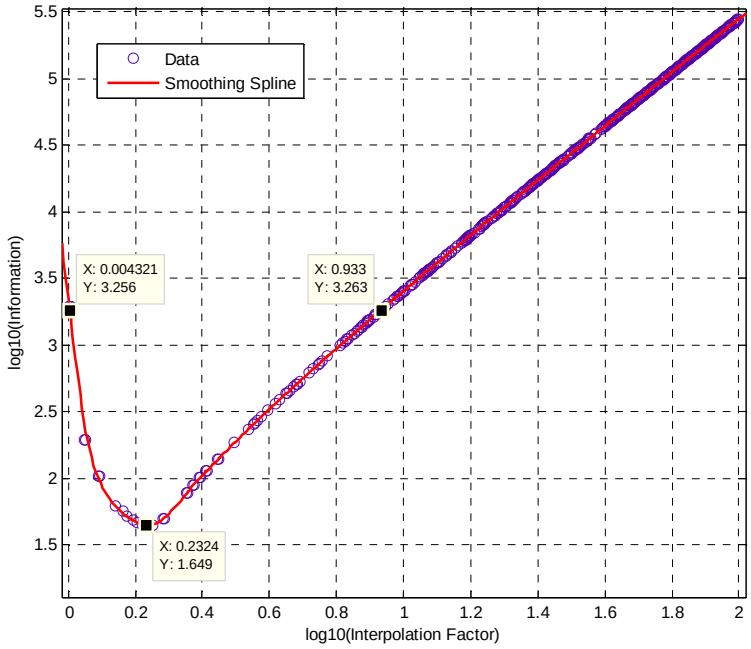

Figure 5 – Sub-pixel Measurement: Information vs. Super-sampling Density

Several values from the smoothing spline interpolant are indicated in Table 1. Resampling ratio values above 100 were not considered due to the high compute time required.

Table 1 –Resampling Versus Error Reduction

Resampling Ratio	Scaling	
	Error	1/ Error
1.00	1.000	1.000
1.71	43.34	0.023
8.78	1.00	0.999
10	0.759	1.317
20	0.179	5.575
50	0.0278	36.014
100	0.00687	145.613

Effects of Noise Energy

Noisy systems have different results each time they are measured. This necessitates the use of an ensemble of repeated measurements to characterize the central tendency and the variation of the system of interest. In this case an ensemble size of 1000 runs, measurement repetitions perturbed by uniform random noise at several sampling densities was used. Figure 6 shows the maximum noise across the ensemble versus a sweep of additive uniform noise factor for different sample sizes. The figure indicates a very clear transition in which noise factor decreases slightly and the error in estimate decreases substantially. As the noise decreases after this point, the variation in the error takes on a very different phenomenology. It can be observed that as the number of samples increases, the error location for this transition becomes smaller.

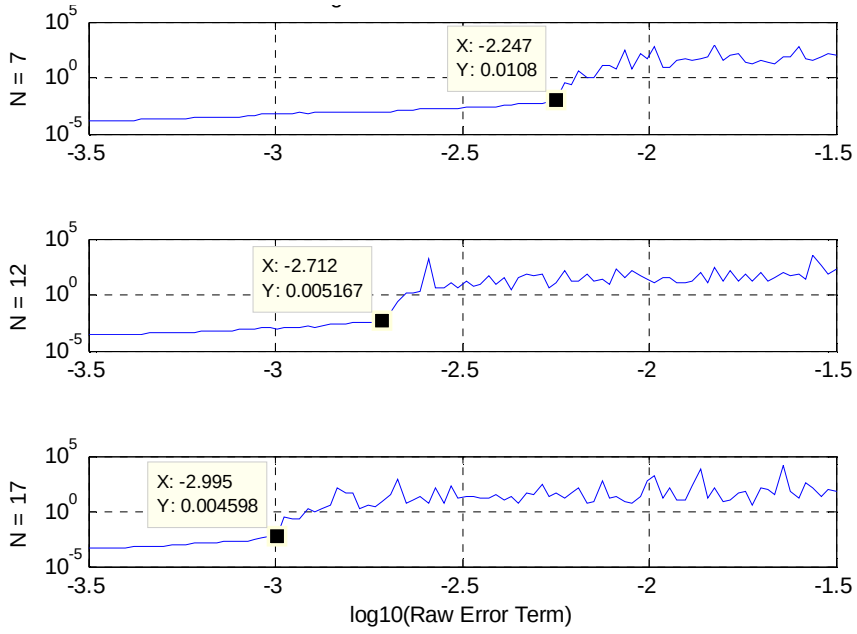

Figure 6 – Sub-pixel Measurement: Error Norm vs. Uniform Noise Scaling Factor for an ensemble of 1000 runs at sampling densities of 7, 12, and 17.

This result motivated a sweep of sampling sizes and noise factors to determine empirically the relationship between the added noise and the transition from very orderly error scaling (on left of transition) to highly random result (on right). The result of this is figure 7, and equation 2. The correlation coefficient of the resultant fit is 99.6%, which is desirably high, but the general form also relates surprisingly well to the energy within the additive noise signal.

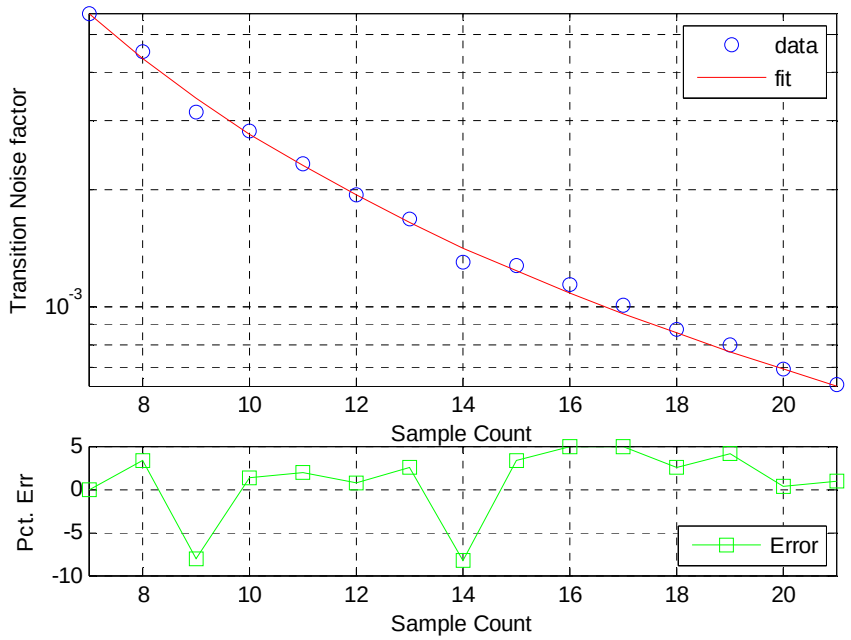

Figure 7 – Sub-pixel Measurement: Characterize Transition Boundary in Error Norm as a Function of Initial Sampling Density

The general form of this transition region, determined using AIC as selection criteria and the Eureqa tool (Schmidt & Lipson, 2009), applied to the numeric results is shown in equation 2.

(2) $$\nu_t = 1.111\left(\frac{1}{2N}\right)^2$$

When taken into account with the facts that the energy in a signal is proportional to the 2-norm of its components, that the mean value for a uniform random function is half the (typically unity) range, and that the measured values were over a logarithmically spaced grid and were the first point within the stable zone equation 2 leads to the conclusion that for a constant template offset and resampling factor the method is capable of withstanding noise with an energy value that is less than or equal to 90% of the energy in the information signal. The signal-energy is the 2-norm of the intensity of the samples divided by the infinity norm of the samples. The noise energy transition is also significant because it informs management of accuracy in registration/measurement by controlling the energy of the error signal in the input. If the energy of the noise is equal to or greater than 90% the energy in the information, then the registration-based measurement becomes problematic. If the energy of the noise is significantly less than the energy of the signal, then the above forms for the analytic case are going to inform the measurement. In image registration, this energy result can be used to inform the spatial scale of the local deformation field.

These analytic results suggest that with proper conditioning, measurement accuracy can be significantly improved compared to conventional methods. Notice again that the maximum relative error for the uniformly noisy case over an ensemble of 1000 elements at N=17 was 0.4%. The median error value left of the transition was about 6% of the max, so over a series of measurements and with any form of reasonable smoothing the expected error term is still around 0.03%. For an additive noise whose mode exists (unlike the uniform random), the method is going to yield lower level errors than given here. The uniform distribution has no true mode, while a large number of more realistic noise probability density functions including the popular Gaussian (or Normal) distribution and the Cauchy distribution, have clear central tendencies in frequency

domain. It is an adverse probability density and demonstrates a boundary which comprises a ceiling for well-behaved distributions.

Smoothing the Discretization

The only "dial" that has not been explored is the impact of numeric discretization on error. For the above material, the representation was performed using MatLab double precision, which conforms to IEEE standard 754 (Moler, 1996) and is represented using 1+11+52 = 64 bits. When comparing the mutual histogram of fit error shown in figure 2, but using 64 bit integers and doubles, this procedure resulted in a straight line whose fit was exact to round-off. Conventional imaging systems yield much smaller bit-length representations – typically 8 to 24 bits for a single color. When comparing the error of 8-bit representations versus 64-bit representations, the discretization error significantly impacts the overall registration accuracy. This is overcome in applied cases by the use of a "Lowess" smooth operated over the template and the data with a small scale parameter of approximately 6%.

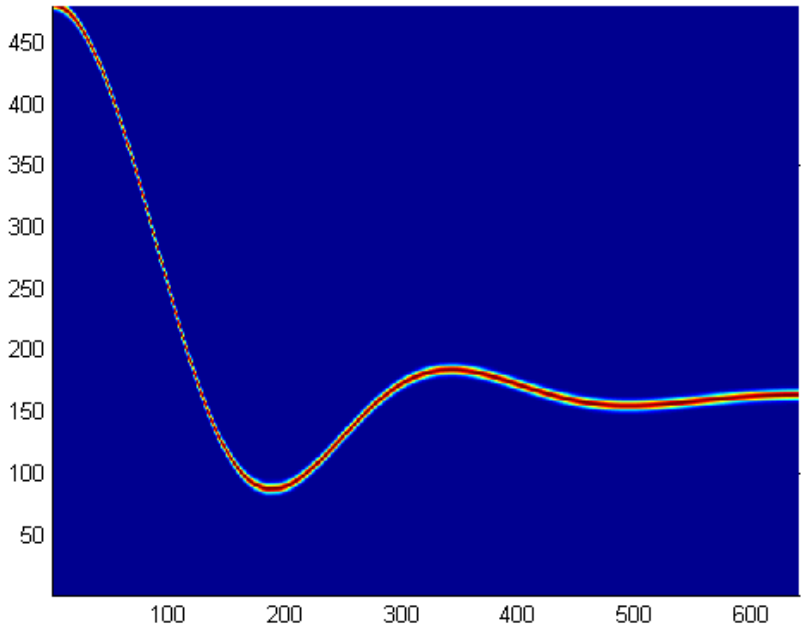

Figure 8 –Nonlinear Pendulum Image. Time is shown on horizontal axis and pendulum position (in pixels, not radians) is approximated on vertical axis.

Results: Synthetic Case

Figure 8 shows the image created by sweeping the mean of a 2d gaussian curve, standard deviation 0.014, along the path generated by following the angle of displacement for a damped nonlinear pendulum. The ordinary differential equation (ODE) for the pendulum is evaluated over a time domain of 0 to 10 (s) which is shown in its equivalent pixel indices (480x640). The 2d reference profile is taken as a subset of the entire range at time = 2.8951 seconds, or x-pixel number 186. It is a vertical slice of the surface, a line of constant time, whose profile is shown in Figure 9. Notice that only the highest 17 values of the Gaussian were used – indicating why the standard deviation was set as it was.

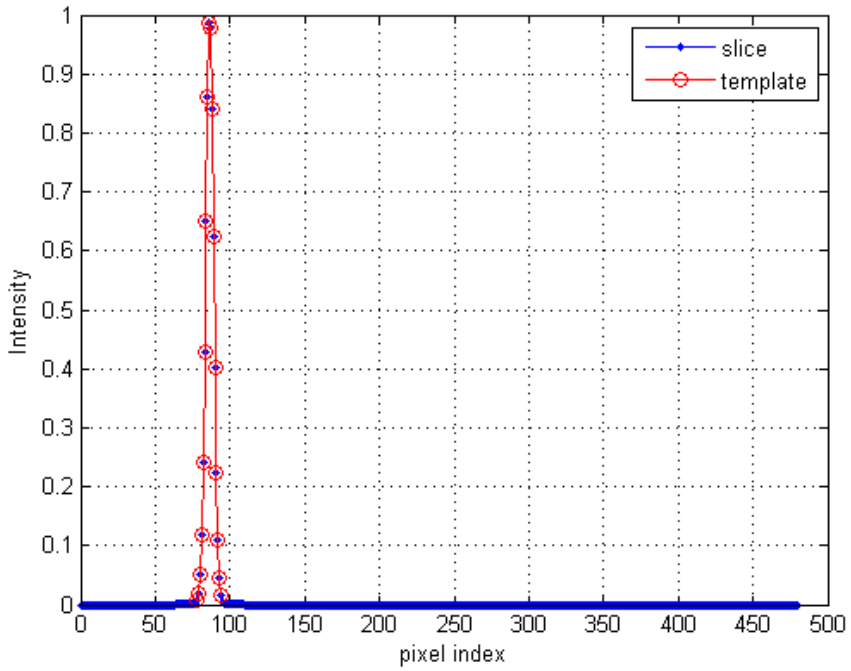

Figure 9 –Nonlinear Pendulum: Section and Template

Notice that the sampling is not truly symmetric around the mean of the Gaussian because the offsets in the Intensity values of points near intensity of 0.88. The impact of this is that without calibrating to a "zero" value, the measurements will have consistent

bias. In the simulation, this means that the template measures a true zero to determine the bias in all other measurements. In physical systems, it means that the measurement system should be calibrated and high precision references can be engineered into the system to allow equivalent calibration. Investment in finding clever ways to achieve real world calibration using the data contained in the image can substantially improve results with limited resources. It is also considered good practice for the measurement system to be shown statistically capable using a Measurement Capability Analysis (MCA) or also as Process Measurement Characterization (PMC). This process is effectively described in (NIST/SEMATECH, 1998). The original, re-sampled, and offset templates are shown in Figure 10. Again, notice the demonstrated zero-padding at the tails. This is critical for accurate determination of position.

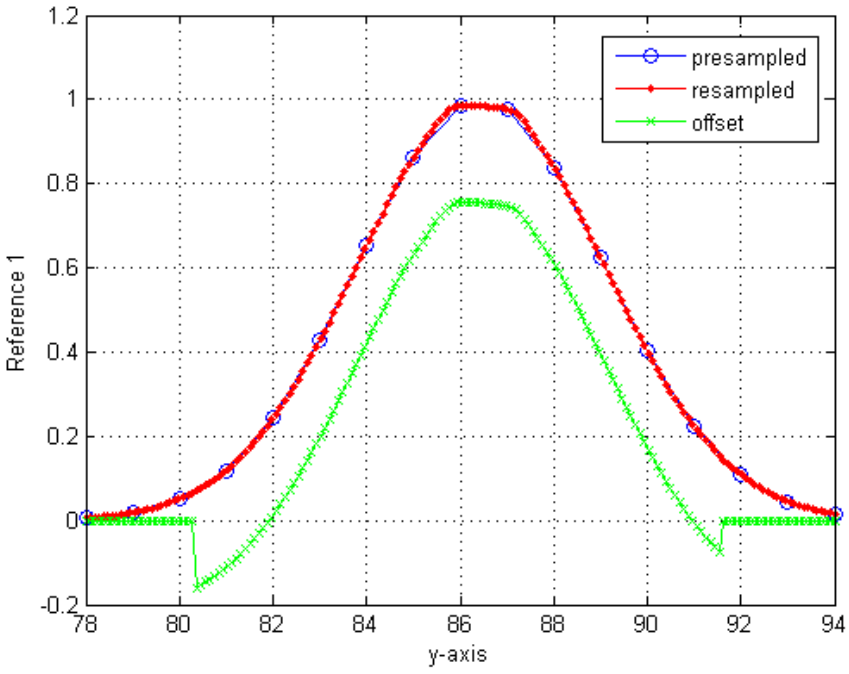

Figure 10 –Nonlinear Pendulum: Reference, Resampled, and Offset Profiles

Next, using the single reference, we iterate through the numeric solution (found using ode45, the MatLab implementation of the Runga-Kutta 4/5 numeric solver) and compare

it with the cubic interpolation of the measured value. The results of this process are shown in Figure 11. The error plot starts around time of 0.3 (s), because the Gaussians in the plot of figure 8 were truncated before this time, as they were displayed with their mean adjacent to the top of the field of view.

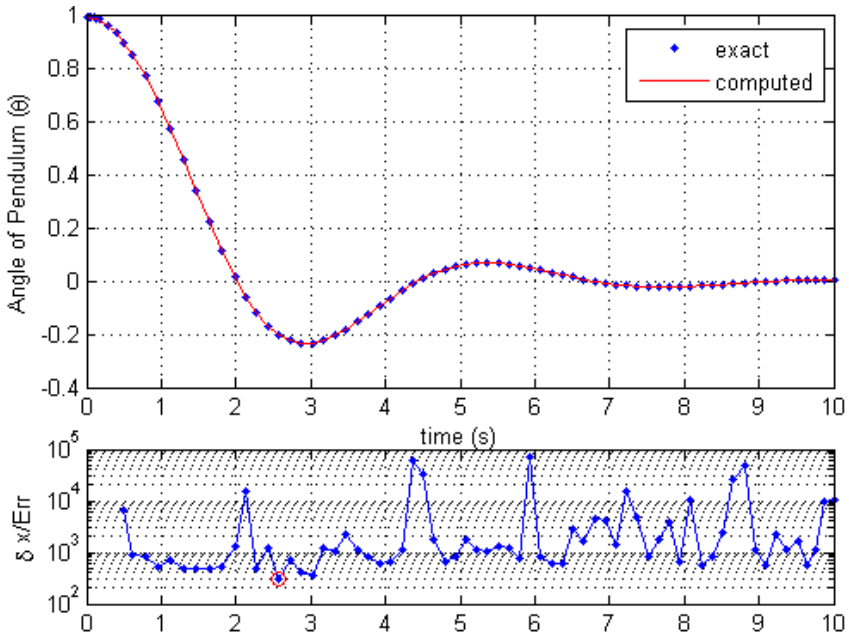

Figure 11 – Nonlinear Pendulum: Synthetic Example Measurement Results. Top subplot shows exact versus computed position while lower shows the log-scale inverse error per pixel, also known as the information per pixel.

The lower subplot is the ratio of the pixel size to the error. This is an estimate of the information per pixel. The smallest value, corresponding with highest error rate, occurs near time of 3(s) and has a value of 311.94. This means that, in this case, the worst case error is about 0.32% of a pixel in size. The error statistics in Table 2 demonstrate close agreement in standard deviation and pseudo-sigma, and some disagreement in mean versus median. The ratio of mean error to maximum error is consistent with the observations of figure 2 at a value of 18.25.

Table 2 – Error statistics for figure 11b

	Statistics (Absolute valued)	
	Error	Δx/err
mean	3.1259E-06	5694.7
median	2.7618E-06	1131.5
std/sqrt(n)	2.8534E-07	1633.9
iqr/1.35	2.5036E-06	1324.6
range	9.9724E-06	68074

In the table, the vertical pixel size is referred to as "Δx", and there are 480 pixels over the rage of $-0.4 \leq y \leq 1$. Pseudo-sigma, indicated by "iqr/1.35", is the interquartile range, the difference between third and first quartiles of the data, divided by 1.35. This fraction of the iqr is a robust estimator for the standard deviation. The disagreement in central tendencies, in the median vs. the mean, suggests there are outliers skewing the mean. In this example, the approach that minimizing the maximum error corresponds to substantially improving the mean and standard deviation errors has been validated. This concludes the synthetic problem section applied to extreme sub-pixel measurement.

OPTIMAL CHARACTERIZATION

It is possible to use a "non-parametric", more properly a model-agnostic or hyper-parametric, smooth with optimal model selection criteria in continuous sum domain to optimally smooth arbitrary continuous systems and to determine model-variation. Although the data domain is most often used when applying this procedure to the data, it can be useful to perform these operations in the continuous sum domain especially in the case of high energy centered additive noise, because this discrete integral transform is noise reducing. After conversion to non-sum domain by performing

an analytic derivative on the fit, the result can again be smoothed with somewhat better results than just smoothing the original form.

Analysis is the process that converts the output of the measurement step, data, into an evaluation of the premise of the experiment. Its purpose is to bring meaningful understanding from raw data. This is accomplished, typically, by summarization of many measurements, often so many that it is too large to be easily handled, manipulated, or even comprehended, into the parameters of an expression which is itself able to be easily handled, manipulated and understood. The simplest non-trivial statistics are those that evaluate to a single scalar, including the moments of the data, of which the mean value and standard deviation are the most popular. The parameters of more complex expressions followed the scalars as summarizations meant to characterize the behavior of the data. A number of great minds performed the early work on the method of Least Squares (LSQ) including Legendre, Gauss, and Markoff (Plackett, 1949) and it is so ubiquitous and important a method that it is now typical of first-year linear algebra. LSQ is an optimization metric that determines parameter values by minimizing the sum of squares of error between values generated by the fit and the data. Measures such as sum of squared errors, sum of absolute errors, maximum error, and correlation coefficient are used to determine the appropriateness of the fit of the data to the analytic expression. If the errors are above what is considered an acceptable threshold or correlation is below what is considered an acceptable threshold, then the model is deemed inappropriate for describing the data. LSQ is used extensively to associate a large number of measurements with relatively few parameters in a model at many levels of work, but it has the problem that it must first have the analytic expression with which to associate parameters before it can be operated to determine what the parameters are. The

formatting and execution are highly model dependent and so this method, while useful, is not model-agnostic.

The idea to use a smoothed polynomial line, or "spline", as an interpolant was originated in the area of descriptive modeling, and not predictive modeling. Drafters over a century ago would elastically bend metal pieces between fixed points to make a smooth line for creating the line of curve for a boat, a building, or other curved surface. Strength of materials gives that the shape of the elastically loaded thin piece of metal comprising these curves is piecewise cubic. The mathematics behind them were developed in the early 1960's (Schoenberg, 1964) and resulted in 1967 in something called "spline smoothing" (Reinsch, 1967), although these techniques did not enter common use for another 30 years because of advancements in video rendering during the 1990's. The spline smooth is a process whereby an optimization is performed whose input is a single "smoothing parameter" and whose output is a piecewise cubic curve whose performance approximates a least squares linear interpolant when the parameter is close to zero, and that approximates a pure cubic spline interpolant when the parameter is close to unity. The spline interpolant to otherwise perfectly linear data can have more parameters than samples and so it is hyper-parametric.

One of the more recent developments in associating data with models are the families of model-selection optimal information criteria. One of the most common of these is the Akaike Information Criterion (AIC) (Akaike, 1976) (MacLachlan & Peel, 2000). As described in (Cavanaugh, 1997) the AIC is an asymptotic estimator of the Kullback-Leibler divergence, or "cross entropy", and indicates the incompatibility between the fitted function and the underlying or "true" model. It is generally applicable to all probability distributions for which consistency and asymptotic normality of the maximum likelihood vector can be established. One of the features driving its popular

use is that it can be expressed very simply in terms of the residual error, the number of samples, and the number of parameters in the model as shown in (Hu, 2007). Sample sizes used here are such that no small-sample correction is needed to adequately determine the AIC.

AIC-Spline Smoothing

In combination with the AIC, however, the smoothing parameter is set so that the smoothed spline is model optimal – it describes the "true model" from only the data and only requires that the regularity conditions are met. It is model-free, model-optimal, model-agnostic smoothing.

Another way to get smoothed estimates of the state underlying some noise corrupted data is by application of the Kalman Filter (Simon, 2006). Kalman filters are used extensively in state estimation, and updating forecasting because they optimally integrate a measurement with a forecast. The biggest challenge for the Kalman filter and its derivatives is how to handle nonlinearity, especially in the state estimate covariance matrix. The novel application presented here is that the Kalman filter equations are used on AIC-spline smoothed data to back out an empirically generated, model-optimal Kalman gain and state covariance update without assuming a model a-priori. This is model-agnostic model determination. In this case the methods are applied, for the sake of simplicity, to scalar functions, but the underlying methods are viable for use in multivariate case.

For the following analytic demonstration, only the non-cumulative domain will be used, but in the case of the tool calibration, a cumulative sum domain was used. The result of applying this method to data is a model-selection-optimal smooth on the data, and an estimate of the state estimate covariance.

Inverse Modeling Procedure:

1. Use optimal model section criterion (AIC) with smoothing spline on data to get smoothed values
    a. For mean state
    b. For state covariance function
2. Use results from step 1 to get measurement covariance
3. Use state covariance and measurement covariance to compose Kalman Gain
    a. If system is simple enough convert to analytic form
    b. For substantial multidimensionality consider leaving underlying equations in smoothed-spline form
4. Validate stability and quality of results
    a. For reduction of error given known/simulated system
    b. For stability of known/simulated system using sensitivity analysis

Spline smoothing used here is consistent with the MatLab implementation in (de Boor, 1978). The smoothing parameter is swept from 0 to 1 in a logistic manner as shown in figure 12. This gives higher density at the "tails" allowing better determination of the optimal smoothing coefficient.

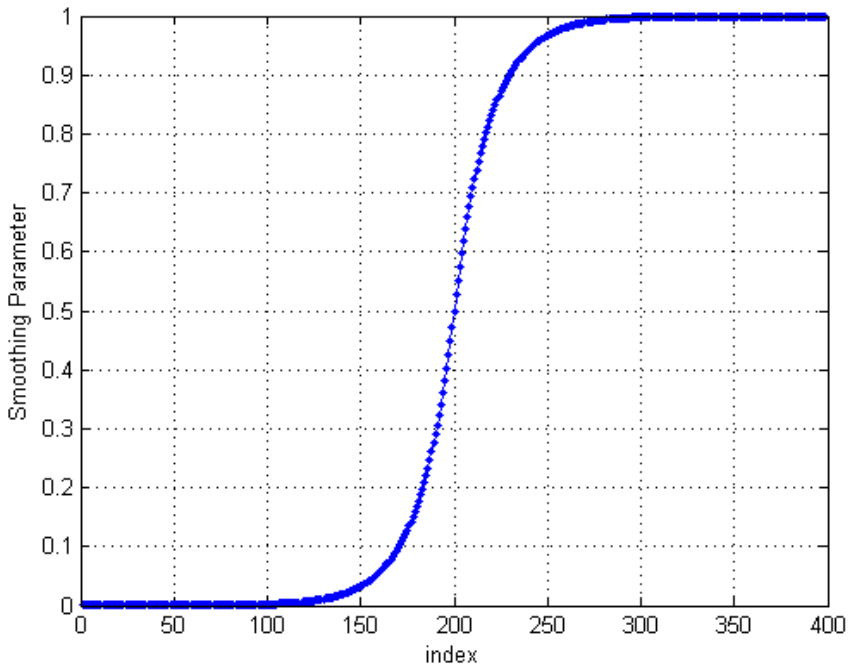

Figure 12 – Optimal Characterization: Smoothing Parameter vs. Index

The Akaike Information Criterion (AIC) is an optimal model selection criterion originally described in (Akaike, 1976). The equation used here to compute the AIC for a particular smoothing parameter is presented in (Hu, 2007):

$$AIC(\nu) = n \cdot \left(\frac{RSS(\nu)}{n} \right) + 2 \cdot k(\nu) \tag{3}$$

This expression indicates that both the residual sum of squares of the fitted model (RSS) and the number of effective model parameters of the effective smooth (k) are functions of the smoothing parameter "ν". The sum of squared residuals is also referred to as sum of squared error (SSE) in mechanical engineering nomenclature. The value for "n" in the above expression is taken as the number of samples.

Applying this AIC-informed smoothing method to the cumulative sum can require more finesse. It many times yields a classic minimum value indicating model-

optimal smoothing parameter, but in cases where the underlying integral-domain model is an indeterminate form, the derivative of the AIC, or its multiplicative inverse must be used, depending on which application of L'Hospital's rule (L'Hospital, 1696) is appropriate. The integral-domain transform, though it reduces the amplitude of centered noise, can confound a mean-noise drift with the estimated model. In both the conventional form, and the integral-domain form, the graph of the AIC, or its derivative, versus the index of the smoothing parameter, often clearly indicates whether there are multiple scales of information in the underlying system, and the approximate scale upon which they operate. Some care is required to determine the appropriate domain within which to formulate this smoothing.

The results for smoothing method to an additive noise corrupted signal follow. The AIC of the smoothing parameter for a sine curve with amplitude 6 and angular frequency 2π that has additive standard normal noise evaluated over a domain of 30 cycles, and that is sampled at 30 samples per period is shown in figure 13. Both the number of cycles and the number of samples per cycle are set at 30 because it is a convention for the minimum sufficient population size to assure statistical significance. The i.i.d. additive noise used here is not applied within the framework of a Kalman filter, so it is not assigned explicitly to measurement or to model variation. Notice the clear interior minima around index of 265. An analysis of the squared error versus smoothing parameter using exact data, for this case, gives the minimum at index 270. The percent difference between these two values is 0.039%, thus the method is considered validated as delivering a good approximation of the parameter that delivers a minimum error approximation to the exact underlying function. It is an assumption in this method that the noise corrupted system is not ideally fit by a cubic interpolant.

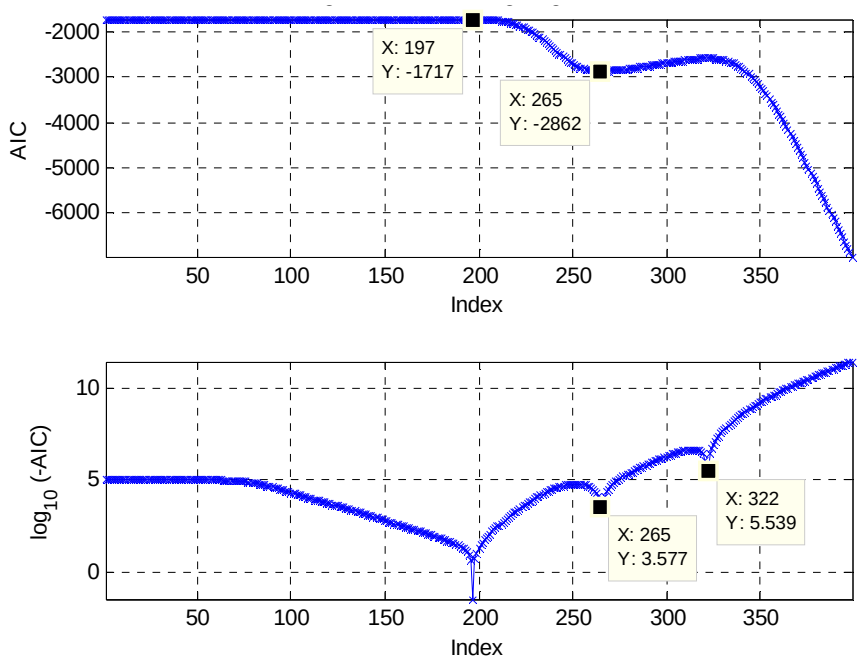

Figure 13 – Optimal Characterization: AIC and Log-Negative Derivative AIC for Smoothing of Sine using Constrained Domain of Smoothing Spline.

The smallest interior AIC indicates the optimal parameter to use. A plot of the corrupted, smoothed, and exact values over a small section of the domain is shown in figure 14. The exact is the spline smoothing that results in the least error compared to the exact value. The improvement due to smoothing is more clearly demonstrated in this smaller portion of the domain than for an image of the entire domain.

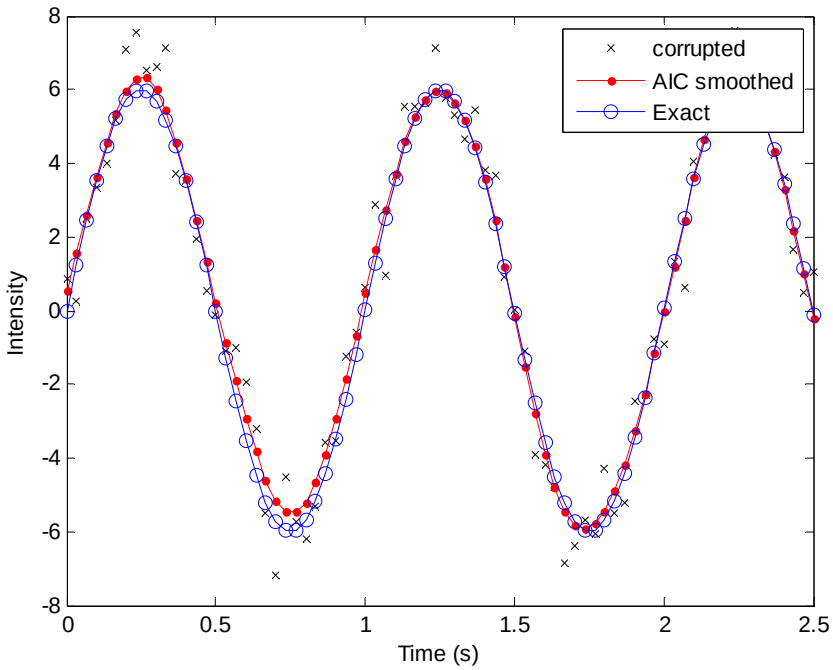

Figure 14 – Optimal Characterization: Validating Smoothing on Corrupted Sine

It can be observed that the smoothed value is interior between the noise-corrupted values and the true values, and that the location of highest divergence is where there is high curvature. The general form of the smoothed function has low frequency deviations from the true value, but the high frequency variations consistent with the additive noise have been reduced. There are also larger scale drifts as can be seen when comparing the AIC smoothed values at the first peak versus the first valley – the first peak overestimates the exact while the first valley slightly underestimates it. For the most part the large scale structure of the underlying system has been recovered. Figure 15 shows the transformation of the noise in CDF-domain with the first and third quartile statistics indicating that the noise was reduced by about 59.3%. This value is consistent with improvement seen with Kalman-derived smoothers (also mentioned in latter sections of this paper) without needing the substantial staging and ephemeris required by them.

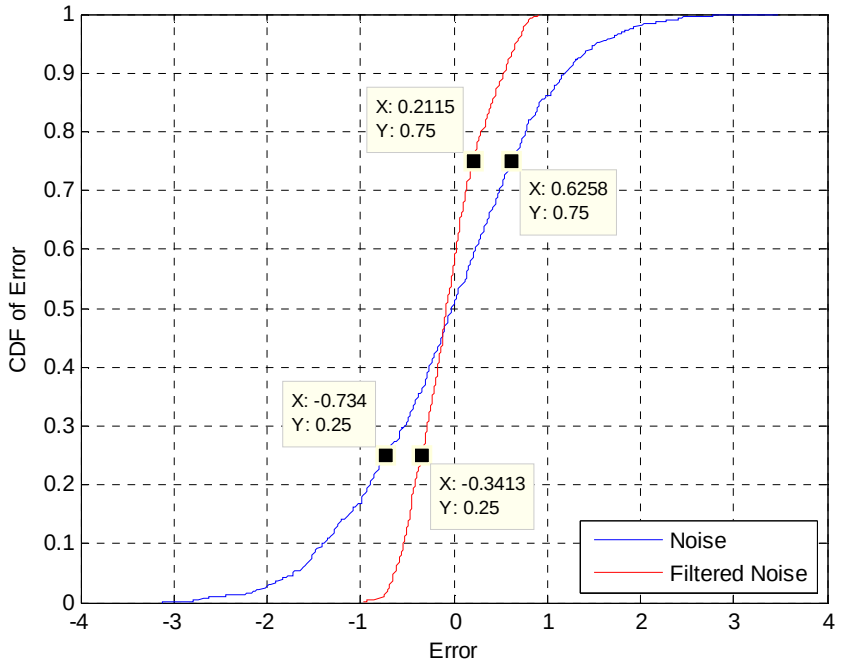

Figure 15 – Optimal Characterization: Comparison of Pre and Post Smooth Error Distributions

It is also asserted in (Hu, 2007) that a family of smoothing parameter values could be used, ranging from the optimal AIC to all functions whose AIC is within a fixed offset of the optimum that can range from a value of AIC+2 for "substantial support" to a value of AIC+10 to envelope all but models with "essentially no support". This is referring to the Bayes Factor described in (MacLachlan & Peel, 2000), so by doing this all candidate models with the selected level of support are accounted for in terms of model variation. Referring again to the AIC plot and its interpretation for "substantial support" a traverse of the smoothing parameter values associated with AIC values within 10 of the minimum by the standard deviation of a uniformly spaced ensemble of 400 smoothing parameter values yielded the modified 4-plot shown in Figure 16. The typical 4-plot displays a histogram in the lower-left quadrant, but the empirical CDF has the information describing the distribution, as well as much lower noise, without the need for the hidden parametric nature of the binning process. The upper-left quadrant shows the

estimated covariance over the domain as calculated using the AIC+10 to account for all models through “essentially no support”.

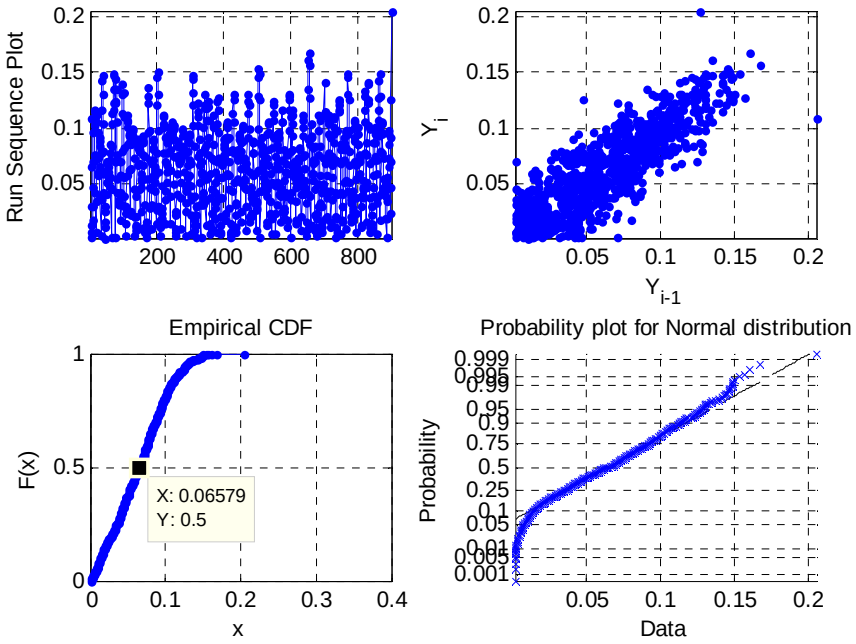

Figure 16 – Optimal Characterization: Estimation of State Variance Function, “P”, using AIC and Akaike Weight Informed Alternative Splines

Although it is tempting to take the median value, shown in the lower right quadrant, and apply it as a constant value for the state-estimate covariance, a comparison of the mean state to the mean variation indicates the constant value is insufficient to describe the estimate uncertainty. The lag-plot (upper right) has a high enough correlation to suggest that the relationship between values is causal and non-constant. Another application of the previously mentioned smooth applied to the plot of the estimated versus the estimated variation in the state yields the relationship shown in figure 17. The solid red line is central tendency of the variation of the state estimate – when the state is at its extreme values, the expected variation is larger, but when it is zero there is a nonzero floor to the variation. This is a very clear and error free relationship.

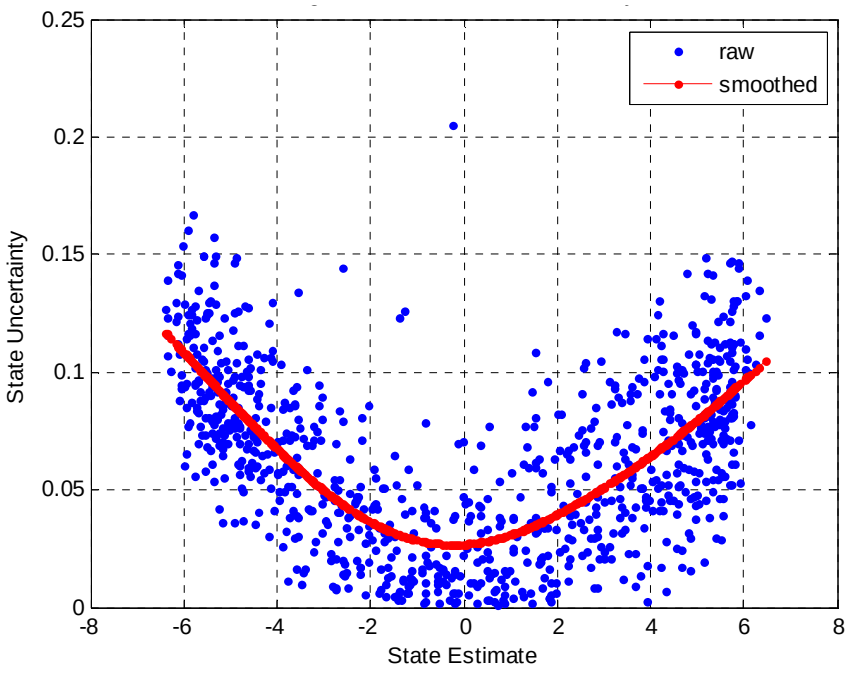

Figure 17 – Optimal Characterization: State Estimate Uncertainty

The smoothed relationship, shown in red, is another implementation of the previously mentioned model-optimal smoothing. Because the smoothing parameter was selected using the AIC we can know that the behavior is characteristic, and is neither constant nor linear. Now that the central tendency of the state update function and the central tendency of uncertainty in state estimate have been approximated using model-optimal smooth, the next step is to transform these high parameter-count formulations into more accessible analytic expressions. The method through which the smoothed signal is used to determine the analytic form of the underlying system requires the description of the Kalman Filter as background.

Empirical Kalman Filter

The Kalman Filter has been a useful tool for state estimation, smoothing, and forecasting since Kalman's original paper was published in 1960 (Kalman, 1960). An

accessible introduction is provided in (Welch & Bishop, 2006) and more completely and deeply described in (Simon, 2006) and (Andrews & Grewal, 2008). Nomenclature used here are, as far as is reasonable, consistent with the work of (Welch & Bishop, 2006).

For the system whose state is described by:

$$x_k = \mathbf{A}x_{k-1} + \mathbf{B}u_{k-1} + w_{k-1} \tag{4}$$

$$z_k = \mathbf{H}x_k + v_k \tag{5}$$

Where the following is true:

$$Q = \mathrm{cov}(w) \tag{6}$$

$$R = \mathrm{cov}(v) \tag{7}$$

$$P_k = E\left(\left[\hat{x}_k - x_k\right]\left[\hat{x}_k - x_k\right]^T\right) \tag{8}$$

The equations describing Discrete-Time Linear Kalman Filter, neglecting control-input terms, are:

$$x_{k-1}^{(+)} = E(x_1) \tag{9}$$

$$x_k^{(-)} = Ax_{k-1}^{(+)} + Bu_{k-1} \tag{10}$$

$$P_k^{(-)} = AP_{k-1}A^T + Q \tag{11}$$

$$K_k = P_k^- H^T \left(HP_k^- H^T + R\right)^{-1} \tag{12}$$

(13) $$P_k = \left(I - K_k H\right) P_k^-$$

(14) $$\hat{x}_k^+ = \hat{x}_k^- + K_k\left(z_k - H\hat{x}_k^-\right)$$

In the extended Discrete-Time Kalman filter the analog equations for (9), (11), (12) and (14), from (Simon, 2006) are:

(15) $$x_k^{(-)} = f_{k-1}\left(x_{k-1}^{(+)}\right)$$

(16) $$P_k^{(-)} = \left(\left.\frac{\partial f_{k-1}}{\partial x}\right|_{\hat{x}_{k-1}^+}\right) P_{k-1} \left(\left.\frac{\partial f_{k-1}}{\partial x}\right|_{\hat{x}_{k-1}^+}\right)^T + \left(\left.\frac{\partial f_{k-1}}{\partial w}\right|_{\hat{x}_{k-1}^+}\right) Q_{k-1} \left(\left.\frac{\partial f_{k-1}}{\partial w}\right|_{\hat{x}_{k-1}^+}\right)^T$$

(17) $$K_k = P_k^- \left.\frac{\partial h_k^T}{\partial x}\right|_{\hat{x}_k^-} \left(\left.\frac{\partial h_k}{\partial x}\right|_{\hat{x}_k^-} P_k^- \left.\frac{\partial h_k^T}{\partial x}\right|_{\hat{x}_k^-} + \left.\frac{\partial h_k}{\partial \upsilon}\right|_{\hat{x}_k^-} R_k \left.\frac{\partial h_k}{\partial \upsilon}\right|_{\hat{x}_k^-}^T\right)^{-1}$$

(18) $$\hat{x}_k^+ = \hat{x}_k^- + K_k\left(z_k - h_k\left(\hat{x}_k^-\right)\right)$$

These provide the general forms for the prior and posterior state estimate. The state uncertainty estimate and the Kalman gain can be explored in terms of these variables. Equations 15-18 indicate the best order for determining the desired variables. Any process that transforms numeric information from one form to another can introduce artifacts into the data that reduce or misdirect downstream treatments. It is therefore a good practice to express the critical downstream forms, such as covariance estimate, or Kalman gain, in terms of the least transformed upstream variables – in terms of the highest quality upstream information. Although this would be prohibitive if performed on nonlinear analytic functions, it is much easier when using numeric data to supply input and output values and let the GA tool determine the analytic expression.

A lag-plot of the state is shown in figure 18. The vertical line test indicates that the previous state estimate is insufficient to uniquely specify the current state, and more information is required.

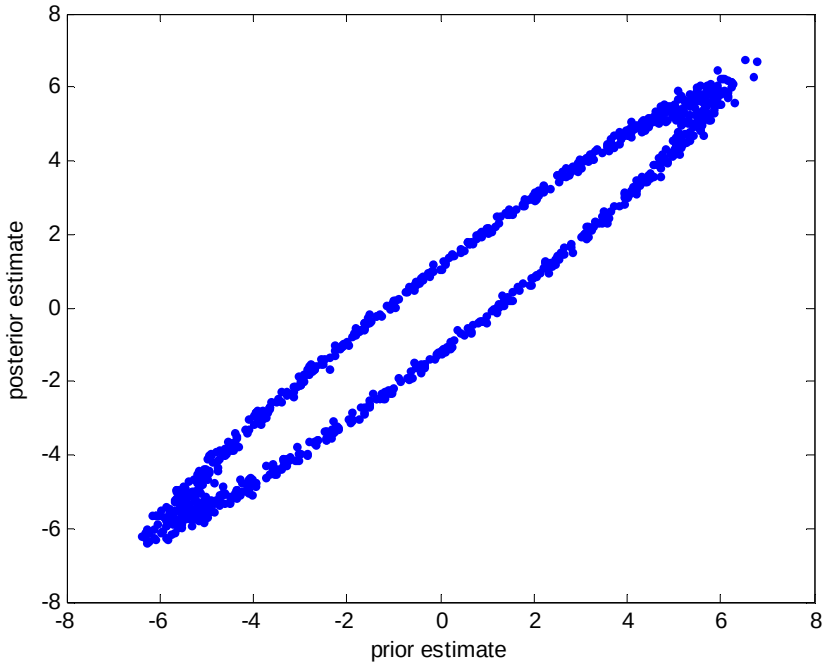

Figure 18 – Optimal Characterization: Lag Plot of State Estimate

There are a number of ways to determine the underlying relationship, and mileage will vary. Simplicity and stability are to be given preference, but heuristics like the number of terms in the preferred Taylor series approximation, or finite difference approximation are alternatives to be considered. In this case, EDA gives good results. When using the Genetic Algorithm (GA) tool a comparison of the AIC or a Cattell "scree" plot analysis (Cattell, 1966) on the SSE terms gives functional results. One of the data segregation modes used in the GA tool is model-generalization, and in this mode one part of the data is used for the fit and a segregated part is used for the validation. It must be kept in mind that not only must each expression fit the data, but together they must

operate stably and well. Equation 19 is the analytic expression determined to relate the current state to the prior state.

$$(19) \qquad \hat{x}_k^- = 1.95621 \cdot \hat{x}_{k-1}^+ - \hat{x}_{k-2}^+, \quad R^2 = 99.993\%$$

Using the model suggested by figure 17, a non-discrete function of the state estimate was used to describe the state covariance. The relationship was found using the previously mentioned Genetic Algorithm tool, but instead of operating only on the current state estimate, the previous 4 states were supplied as candidates, in keeping with giving non-transformed data to get the best fit expressions. The state estimation covariance found is indicated in expression (20) – this comprises the model prior model covariance without the need to account for "Q": everything in here is model related variation, and it is only model related variation. This includes the very good but imperfect modeling of the underlying system using the smoothing spline.

$$(20) \qquad P_k^- = 0.02897 - 0.00475 \cdot \hat{x}_{k-1}^+ \cdot \hat{x}_{k-2}^+ + 0.006851 \cdot \left(\hat{x}_{k-1}^+\right)^2,$$

The R2 statistic for this fit is 99.39% - acceptably high for descriptive use.

Using expression (8) and (19), the measurement covariance "R", not to be confused with the fit statistic, can be estimated. Given the nature of the noise, this is a constant value, so the expectation is computed over the entire domain. The Kalman gain expression, including that the transform from measurement to state domains, "H" is unity, becomes:

$$(21) \qquad K_k = P_K^- \left(P_K^- + 0.85214\right)^{-1}$$

A graph that compares the noise-corrupted data, the AIC-Spline smoothed data, and the updated Kalman filtered data is shown in figure 19. The filtered value is shown in blue.

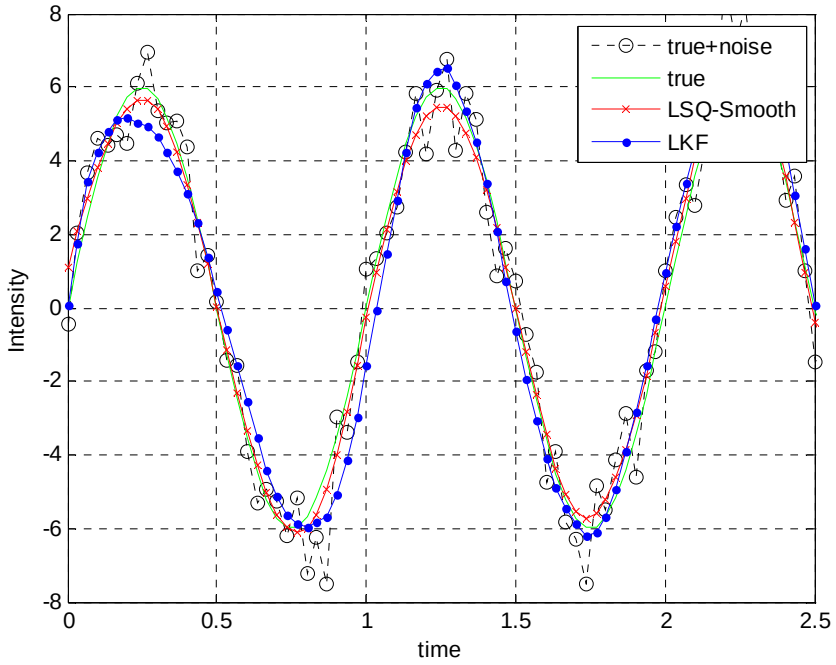

Figure 19 – Optimal Characterization: Graphic Evaluation of Derived Kalman Filter Performance vs. Globally Informed Smooth and True Model.

It can be seen that both the exact-informed AIC smooth and the KF work to reduce the effect of the noise. The Kalman filter is seeded with some basic averaging of the corrupted values. Over an ensemble of 10 runs, this discrete Kalman filter reduces the norm of the noise by 34.59% ± 0.0239% while the AIC-smooth reduces the norm of the noise by 60.73% ± 0.0124%. The system was stable under the assumptions that neither the nature of the noise nor the nature of the underlying system was changing. The Kalman derived slightly overshoots the true value at some points (i.e. just after time 1.25 in figure 19), and this is a substantial improvement in nature over the AIC-Spline smooth which will always undershoot in a case like this. The first run discrete Kalman filter

reduced the noise only by about 57% as much as the AIC-smooth but it compensates for that by more effectively generalizing the underlying system. Although the Kalman-filtered values are compatible with the AIC-informed smooth, they enable utility of the Kalman filter including Rausch-Tung-Streibel (RTS) smoothing, described in (Simon, 2006) and (Andrews & Grewal, 2008), which can improve the signal quality over the raw Kalman filter. This concludes the synthetic problem section applied to empiric model-optimal inverse system modeling.

RESULTS:

The presentation of method and results is as follows. Although the methods described were developed with the use of relatively simple analytic test cases, they must be applicable in real-world circumstances. The methods for measurement and smoothing are shown to be valid in the case of finding the calibration values for the x-y stage of a Panasonic laser scribe, and also in application to the video-based measurement to the 2d position of the indicator for a King model 2-32-G-042 rotameter (fluid velocity measurement device).

Laser Scribe Calibration

When applied to the determination of y-position variation for the calibration of the XY stage of a Panasonic Laser scribe, the measurement method yielded 500,000 position measurements across the wafer that were accurate to ± ~2nm using pixels that were ~700nm in size. The laser scribe is a device that uses an x-y table to move a wafer under a laser beam to cut a channel that a die-saw will then follow to singulate the dies out. The variation in the position of the path between the dies, called a “street” and shown in figure 20, must be characterized as part of tool qualification so that this can be

robustly accounted for in the engineering of die-sizes, and thus make the most of die material and the very expensive upstream processing of the wafer.

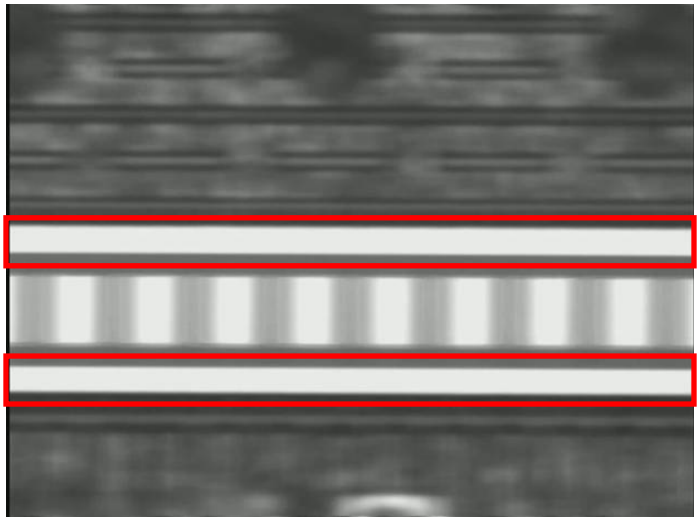

Figure 20 – Test Case 1: Image from Calibration of Laser Scribe. (“Streets” indicated by red boxes)

These calibration values are used to set these variations into the xy-trajectory of the wafer under the beam. The data available to calibrate the tool was a supplied grayscale XGA video that showed the x-y table moving under the laser fixture. The wafer was rotated slightly so that a perfect line on the wafer took up 3-4 multiple rows of pixels in the image. The video was deconstructed into individual files using the K Multimedia Player, a freeware video player/decoder provided by Pandora TV, a Korean company. The files were loaded into MatLab and the non-informative regions were culled. A reasonable section of “street” was selected as template, smoothed, and super-sampled. The number of frames required for a particular feature to cross the field of view of the image was counted, and it was determined that it took about 20 frames for a feature shown on the extreme upstream (right) edge of the field of to no longer be displayed. Using the Nyquist sampling criterion, it was determined that if all of the pixels would be measured every tenth image, then the streets would be measured for y-position at every

point possible. An image smoothing operation that had a sigma of 3 gv (of 255) and a spread of 5 pixels was operated on the images of the reference and the template to remove gv noise (room temperature cameras have 3-8 gv of thermal noise). The super-sampling value selected was 10x. The method was operated as described above on features indicating either side of the street, and then averaged (achieving calibration) to indicate pixel-location of the center of the street. The raw y-position data was then smoothed using an AIC informed spline smooth operated over the ~4.5e5 measurements. The effect of this smooth was evaluated by comparing it to the numeric mean of the y-values for each image over the run – the agreement supports the AIC as characterizing the phenomena and superior to a simple average. An intensity indicator was used – it summed the columns, to indicate when the "intersection" of a horizontal street and a vertical street was crossed. In this way, the data was able to be processed both in terms of performance across the entire wafer and on a die-by-die basis. Figure 21 shows the y-position across the entire wafer.

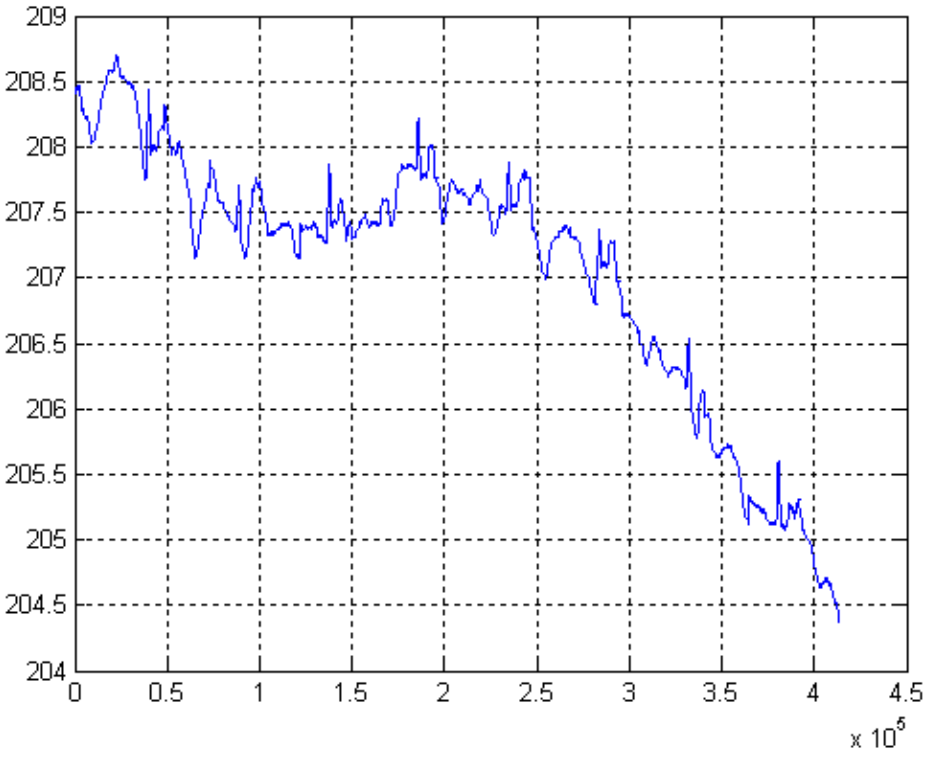

Figure 21 – Test Case 1: Measured Lane Position over Wafer. Horizontal axis is measurement index, and vertical value is measured position of center of "street" in pixels.

The non-linear trend over the wafer can be seen by comparing the general line of transportation between indices 0 and 200,000 with that from around 200,000 to the end. The motion of one is generally horizontal while the other is, in this representation, at a nearly 45 degree angle from the horizontal – it nosedives. If the engineer were to assume that the center value was more characteristic, or that the trajectory over the first half of the path were characteristic, the consequence could be expensive and hard to diagnose.

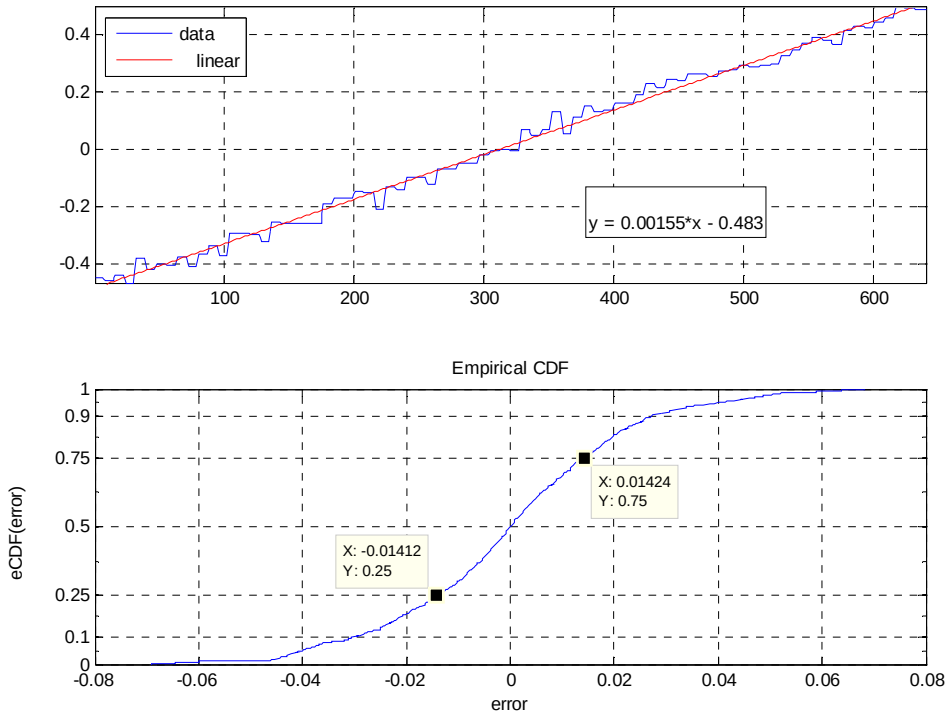

Figure 22 – Test Case 1: Centered Lane Position over Single Frame. The top subplot shows the data and a linear fit of that data and the lower subplot shows a CDF of the error between the linear fit and the data.

The variation over a single frame is interesting. Each of the "flat" regions measured (in blue) in the upper portion of figure 22 is comprised of 10's or 50's of independent measurements varying by a very small amount. Given the production process used to create the template, these are clearly regions of constant position. The linear fit of the measured position is shown by the red line. The CDF of the error between the fit and the measurements is shown in the lower portion. It gives an interquartile distance of 0.0285 pixels, which suggests sample standard deviation in values over the image is about 0.0211 pixels. Inspection clearly reveals portions of near-constant height, for example, around index 150 of figure 2.2.

A comparison of two lanes from the same image is shown in figure 23. Notice the correlations between the lanes that occur at indices of approximately 30, 130, 430, and 600.

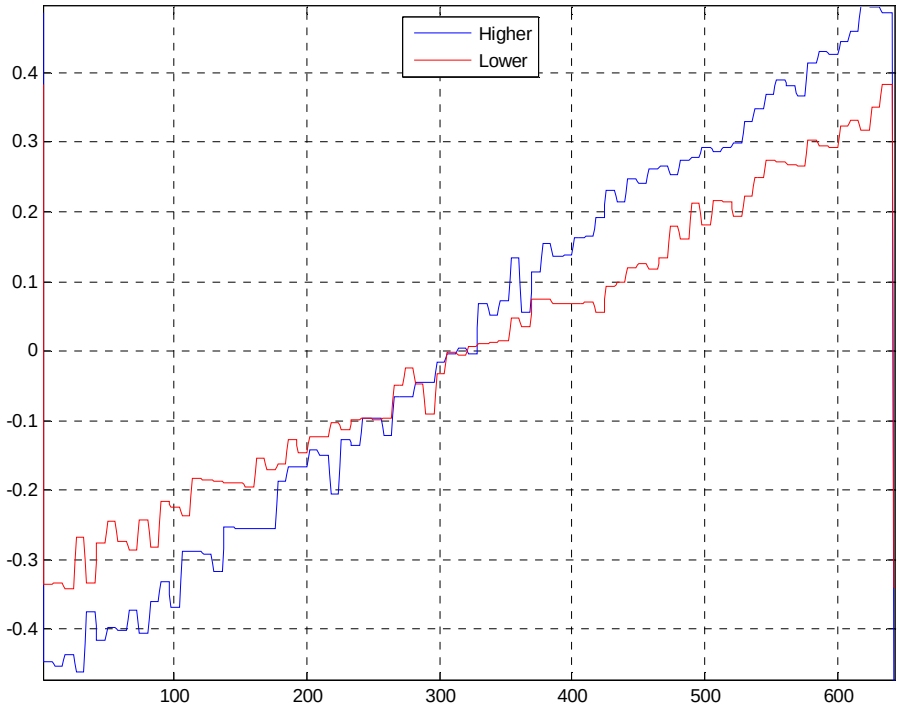

Figure 23 – Test Case 1: Centered Single-image Values for Higher and Lower Lanes. The x axis is pixel index while the y axis indicates the frame-centered estimate of position. The darker blue is the higher lane and the lighter red is the lower lane.

When the variations in lane position are adjusted for the angular misalignment, and sections of corresponding consistent geometry are evaluated for variation from their mean it was found that they vary by much smaller values than the large-scale structure, on the order of 0.0025 pixels, or about 1.8nm.

A die-scale plot of the lane positions is shown in figure 24. There is a “bump” in the trajectory at around pixel 1.5e4. The consistent location along the dies, but only along 8 of them, suggests this is a systematic defect in the calibrating wafer that is engineered into the upstream processes, and not a defect of the xy-stage.

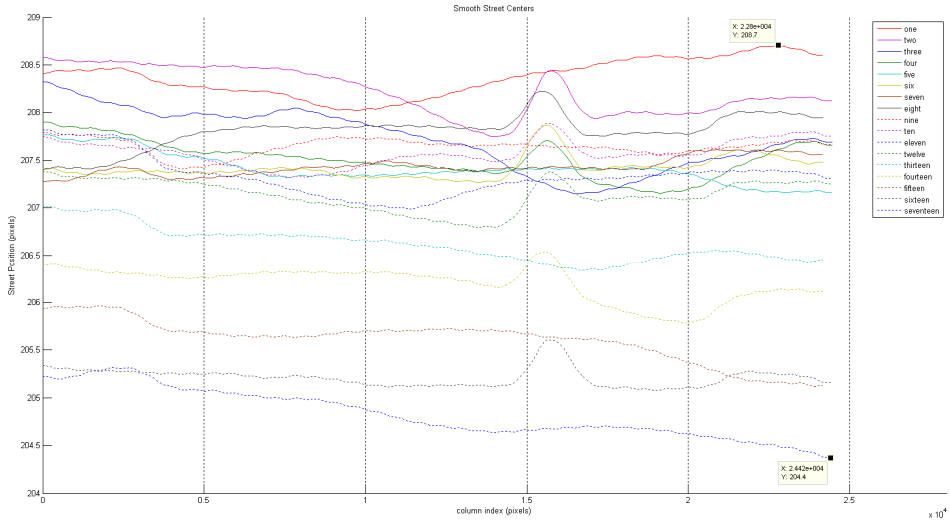

Figure 24 – Test Case 1: Measured Lane Position over each "Die". The x-axis is the column index ranging from 0 through about 2.5e6 pixels. The y-axis is the measured center of the lane, also in pixels ranging from values of 204 through 209.

The very slight ripples are noise artifacts from the smoothing process: the data was smoothed in integral domain, so it is under-smoothed by a small amount and allows the user to see the scale of the pictures supporting the information as well as the information itself. One “ripple” is the result of measuring one entire image. Table 3 indicates the by-die and overall variation statistics that are then used to set the street position calibration of the tool. It can be seen that lane y-position drifts by a range of nearly 2.5 pixels and that the die-to-die variation in lane height can be as much as 67%.

Table 3 – Test Case 1: Lane Measurement Statistics

	Raw Value		Centered Value	
idx	mean	std	mean	std
1	208.368	0.19646	1.30682	0.01075
2	208.223	0.24195	1.16193	0.05624
3	207.722	0.31254	0.66050	0.12683
4	207.509	0.18756	0.44806	0.00186
5	207.410	0.15419	0.34869	-0.03152
6	207.421	0.11575	0.35974	-0.06996
7	207.433	0.08994	0.37175	-0.09577
8	207.809	0.17354	0.74749	-0.01216
9	207.641	0.08621	0.58004	-0.09950
10	207.574	0.13431	0.51299	-0.05139
11	207.329	0.20517	0.26731	0.01947
12	207.136	0.16071	0.07414	-0.02500
13	206.616	0.18015	-0.44570	-0.00556
14	206.200	0.17747	-0.86188	-0.00823
15	205.610	0.21940	-1.45152	0.03369
16	205.213	0.10062	-1.84807	-0.08508
17	204.829	0.25554	-2.23229	0.06983
grand	207.061	0.18571		0.06013
grand abs	206.980	0.17597	0.80464	0.04723

Using a demonstrably linear section of the part that spanned several images that had been model optimally smoothed it was again shown that this method, using pixels that were on the order of 700 nm in size was able to effectively locate the center of the lane such that variation in the mean estimate was consistent with the prior measurement of 1.8 nm. Using the described method to measure the calibration offset of the xy-stage, the tool was able to be set such that the variation between actual scribe position and intended position was less than one micron without risk of harm to the product.

Velocimeter Tracking

When applied to the measurement of the indicator position for a King Rotameter (velocimeter) more preprocessing was required. The method was intended to be used with LabVIEW and a webcam to allow online (non-supervised) flow speed measurement of a Propylene-Glycol/water solution flowing through a solar-thermal collector. The supplied image was intended to be an aggressive benchmark of the method. The initial form of the video is shown in figure 25.

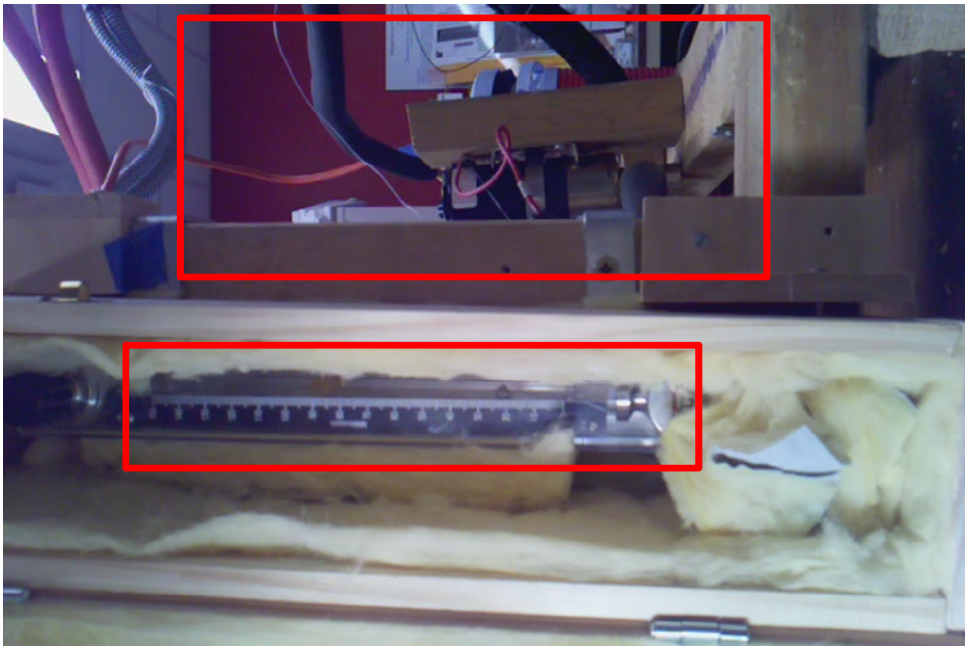

Figure 25 – Test Case 2: Image of Flowmeter Setup.

The operation of this method upon the flowmeter images was ideal for a number of reasons that, on the surface, make the work more challenging, but also show the strength of the method. Inspection of Figure 25 shows that the upper rectangle indicates the portion of the field of view that was within the focal region of the camera, while the lower rectangle indicates the region of interest outside the region of focus. The illumination was not uniform over the scale indicating flow-speed, so the method would

have to be made resistant to this sort of variation in order to operate. The indicator ticks are so blurred that they cannot be clearly differentiated. This shows the great degree of information loss, but it also allows all results to be communicated in terms of pixel-indices, which works reasonably well and enables less ambiguous evaluations of error terms. The indicator was 13 pixels across of a 640 pixel wide image. This is a very small 2% of the field of view, so there could be substantial improvement realized by making the indicator larger and thus encoding more information about its position into the image. The shape of the indicator changed over the field of view due to lighting – it cast a shadow that was different at one end of the scale from the other. The edge of the indicator is not clearly demarked against the background, again resulting in a poor signal – it is expected that in more controlled environment the method would have higher quality results. The axis of motion of the indicator is not aligned with the camera, so 2d motion must be accounted for – the method must be applied in 2 dimensions instead of one.

The preprocessing, like mentioned earlier, involved first segmenting the areas of interest from the total image to reduce the compute time, and then subtracting from the entire image another image in which the ball was out of its typical area (at the extreme value for flow rate and thus outside expected behavior of the system under any type of normal operation, and also blurred by motion resulting in substantially lower intensity values). This substantially reduced the impact of noise, but there was still small-value variation across the field of view. The intensities were mapped from the 8-bit range of 0 to 255 gv to a double-precision ranging from 0 to 1. An empirical CDF of the intensity values revealed that the region around the ball was at substantially higher intensities than the rest of the region of interest, so intensities below the critically determined threshold were set to zero. A template region around the ball was selected and smoothed. A 2-

dimensional convolution was performed between the truncated template of the ball (appropriately super-sampled, and offset in intensity) which resulted in a 2-dimensional array of convolution intensities. A 2-dimensional quadratic surface was fit to the region surrounding the peak of this field, and an analytic root indicating the location of maximum convolution was found. This was performed over all available images.

Figure 26, below, shows the difference between the original and the smoothed and super-sampled template of the ball. The intensities reflect the rescaling, and the low-brightness of the illumination. The template at its peak only traverses 25% of the intensity spectrum. This image has the same information content as if it were represented using 6-bit colors.

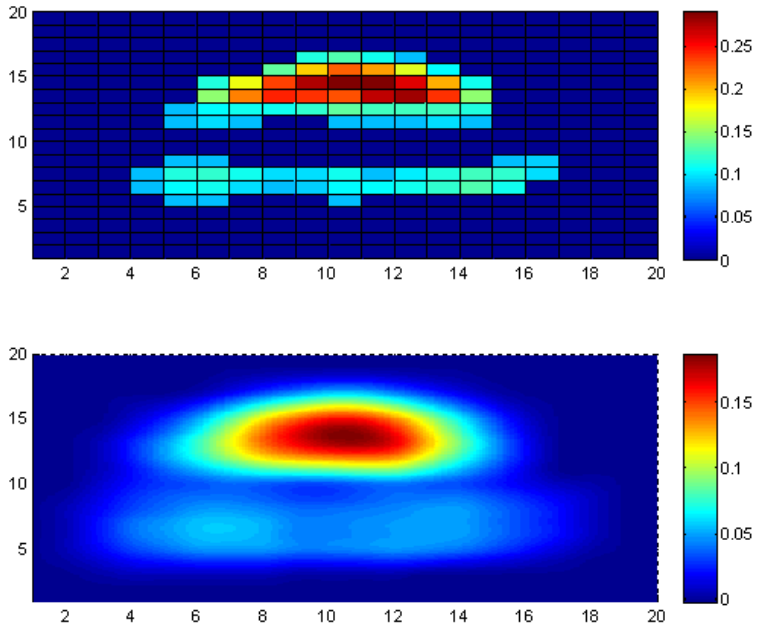

Figure 26 – Test Case 2: Indicator Reference (raw pixel values above, smoothed and super-sampled below)

This shows the effect of the over-sampling rate of 10 that was used here. The smoothing was set consistent with that described above in that it was below the standard deviation of the information and acted to allow clean alignment of the target with the image. Although the inputs to the velocimeter were not supplied, a region within the video was found where the ball responds to what appears to be a step input. The values for this region are displayed in figure 27.

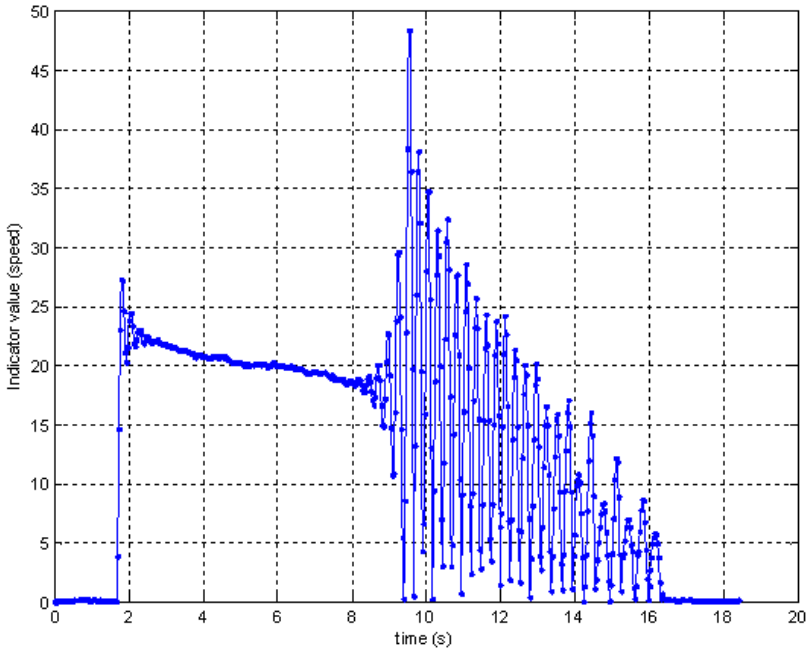

Figure 27 – Test Case 2: Useful Rotameter Indicator Values

The behavior of the indicator is too erratic within the supplied video/images to allow characterization of the system kinematics, and there was no indicator of imposed control inputs to allow formulation as a Kalman Filter. The region of slowly descending constant flow between about 2 (s) and 8(s) looks like a close analog of the step response of an under-damped second order system. Between 10(s) and 16(s), it appears to be governed by a resonance phenomenon. If the system parameters were known, then the control input as a function of indicator position history could be determined even with

phenomena of overshoot or resonance as was demonstrated in the analytic case. This exercise staged the operator of the rotameter to be able to perform online automatic measurement through a video camera observing the rotameter.

CONCLUSIONS

The results demonstrated heuristic method for achieving extreme sub-pixel measurement accuracy on the order of 0.3% of a pixel in controlled but repeatable real-world applications, including model-optimal noise removal from the raw data. The error achieved in the first case is two orders of magnitude better than the performance ceiling indicated by Reed and Adrian. This quality of result was confirmed in the calibration of the laser-scribe, allowed measurement of features to within 2nm using pixels that were 700nm in size, and it allowed the tool to achieve operational variation below one micron. The error found in the second demonstrates the applicability of this method to adverse inputs. It was shown in the velocimeter tracking results that even with substantial accuracy settings, if the system was not set up to reduce the noise sources in a highly controlled manner, then there is a very solid floor to the quality of the estimate.

A framework was demonstrated that provided two model-optimal forms for the Kalman filter, one in spline equations and the other in human-tractable analytic expressions. These did not require prior analytic knowledge of the underlying system or the measurement system. Their results worked “out the gate” and are substantially different in nature than the “textbook” results of conventional linearization methods. A system thus created using the guidelines presented in this work would be expected to have repeatable performance that consistent with those demonstrated. The Kalman filter, with two prior state estimates, gave results compatible with a global optimal smooth comprised of on the order of 1000 samples.

BIBLIOGRAPHY

Adrian, R. (1991). Particle-Imaging Techniques for Experimental Fluid Mechanics. *Annu. Rev. Fluid Mech. Vol 23* , 261-304.

Akaike, H. (1976). An Information Criterion (AIC). *Math. Sci. 14(153)* , 5-9.

Anand, F. S. (2010). *Bayesian Framework for Improved R&D Decisions (Doctoral Dissertation. Georgia Institute of Technology, 2010).* Retrieved June 2012, from http://hdl.handle.net/1853/39530.

Andrews, A., & Grewal, M. (2008). *Kalman Filtering: Theory and Practice using MATLAB.* New York: Wiley-IEEE Press.

Bishop, C. (1996). *Neural Networks for Pattern Recognition.* USA: Oxford University Press.

Candes, E., & Tao, T. (2006, Dec). Near-Optimal Signal Recovery From Random Projections: Universal Encoding Strategies? *IEEE Transactions on Information Theory, Vol.52, no.12* , 5406-5425.

Cattell, R. B. (1966). The Scree Test for the Number of Factors. *Multivariate Behavioral Research* , 245-276.

Cavanaugh, J. (1997). Unifying the Derivations for the Akaike and Corrected Akaike Information Criteria. *Statistics & Probability Letters Vol 33* , 201-208.

de Boor, C. (1978). *A Practical Guide to Splines.* New York: Springer-Verlag.

Dohoho, D. L. (2006, April). Compressed Sensing. *IEEE Transactions on Information Theory, Vol. 32, No. 4* , 1289-1306.

Hu, S. (2007, Mar 15). *Akaike Information Criterion*. Retrieved June 2012, from Center for Research in Computation, North Carolina State University, Raleigh NC: www4.ncsu.edu/~shu3/Presentation/AIC.pdf

Hughes, A. (2008). Ethnic Differences in Microvascular Structure. *Diabetologia* , 1719-1722.

Kalman, R. E. (1960). A New Approach to Linear Filtering and Prediction Problems. *Transactions of the ASME-Journal of Basic Engineering* , 35-45.

Kelly, D. J., Azeloglu, E. U., Kochpura, P. V., Sharma, S. G., & Gaudette, G. R. (2007). Accuracy and Reproducibility of a subpixel extended phase correlation method to determine micron level displacement in the heart. *Med. Eng. Phys. Vol 29* , 154-162.

Laval, B. (2008). Sub-pixel Water Temperature Estimation from Thermal-Infrared Imagery using vectorized lake features. *Remote Sens. Environ. Vol 112* , 1678-1688.

L'Hospital, G. (1696). L'analyse des infiniment petits pour l'intelligence des lignes courbes.

Liebgott, H. (2008). A Method for Vector Displacement Estimation with Ultrasound Imaging and its Application for Thyroid Nodular Disease. 259-274.

MacLachlan, G., & Peel, D. (2000). *Finite Mixture Models.* New York: Wiley Interscience.

Moler, C. (1996, Fall). *Floating Points.* Retrieved from MatLab News and Notes, Fall 1996: http://www.mathworks.com/company/newsletters/news_notes/pdf/Fall96Cleve.pdf

NIST/SEMATECH. (1998). e-Handbook of Statistical Methods. online.

Plackett, R. L. (1949). A Historical Note on the Method of Least Squares. *Biometrika Vol. 36 No. 3 of 4* , 458-460.

Reed, A. R. (2010). Comparison of Subpixel Phase Correlation Methods for Image Registration. *ARNOLD ENGINEERING DEVLEOPMENT CENTER, AROLD AFS TN* .

Reinsch, C. H. (1967). Smoothing by Spline Functions. *Numerische Mathematic Vol. 10, Num. 3* , 177-183.

Schmidt, M., & Lipson, H. (2009). Distilling Free-Form Natural Laws from Experimental Data. *Science Vol. 324, No. 5923* , 81-85.

Schoenberg, I. J. (1964). On Interpolation by Spline Functions and its Minimal Properties. *roceedings of the Conference held in the Mathematical Research Institute at Oberwolfach, Black Forest, August 4-10,1963.* Basel-Stuttgart: Birkhauser.

Shannon, C. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal, 27 (3)* , 379-423.

Shelepin, Y. E., & Bondarko, V. M. (2004). Resolving Ability and Image Discretization in the Visual System. *Neuroci. Behav. Physiol* , 147-157.

Simon, D. (2006). *Optimal State Estimation: Kalman, H-infinity, and Nonlinear Approaches.* New York: Wiley Interscience.

Wang, G. (2007). An Improved Multi-fractal Method for Pavement Cracks Extraction. *Eng. Comput. Vol 24* , 712-722.

Welch, G., & Bishop, G. (2006, July 24). *An Introduction to the Kalman Filter.* Retrieved November 2011, from Web Site of Greg Welch, University of North Carolina at Chapel Hill, Department of Computer Science: http://www.cs.unc.edu/~welch/media/pdf/kalman_intro.pdf

Yamamoto, Y. (2009). Robust particle image velocimetry using gradient methods with upstream difference and downstream difference. *Exp. Fluids Vol. 46* , 659-670.

Yu, W. (2008, November). Development of a Three-dimensional Anthropometry System for Human Body Composition Assessment. *(Doctoral Dissertation. University of Texas at Austin, .*