

High-Speed Microscale Optical Tracking Using Digital Frequency-Domain Multiplexing

Robert A. MacLachlan, *Member, IEEE*, and Cameron N. Riviere, *Member, IEEE*

Abstract—Position-sensitive detectors (PSDs), or lateral-effect photodiodes, are commonly used for high-speed, high-resolution optical position measurement. This paper describes the instrument design for multidimensional position and orientation measurement based on the simultaneous position measurement of multiple modulated sources using frequency-domain-multiplexed (FDM) PSDs. The important advantages of this optical configuration in comparison with laser/mirror combinations are that it has a large angular measurement range and allows the use of a probe that is small in comparison with the measurement volume. We review PSD characteristics and quantitative resolution limits, consider the lock-in amplifier measurement system as a communication link, discuss the application of FDM to PSDs, and make comparisons with time-domain techniques. We consider the phase-sensitive detector as a multirate DSP problem, explore parallels with Fourier spectral estimation and filter banks, discuss how to choose the modulation frequencies and sample rates that maximize channel isolation under design constraints, and describe efficient digital implementation. We also discuss hardware design considerations, sensor calibration, probe construction and calibration, and 3-D measurement by triangulation using two sensors. As an example, we characterize the resolution, speed, and accuracy of an instrument that measures the position and orientation of a 10 mm × 5 mm probe in 5 degrees of freedom (DOF) over a 30-mm cube with 4- μ m peak-to-peak resolution at 1-kHz sampling.

Index Terms—Computer-aided surgery, lock-in amplification, medical robotics, optical position measurement, phase-sensitive detector, position-sensitive detectors (PSDs).

I. INTRODUCTION

POSITION-SENSITIVE detectors (PSDs), or lateral-effect photodiodes, are commonly used for high-speed high-resolution optical position measurement. PSDs that measure in 2-D are particularly attractive for multidimensional position measurement [1]. When more than 2 degrees of freedom (DOF) are to be measured, instruments generally have multiple PSDs and laser/mirror combinations with distinct optical paths so that each PSD still measures one spot with 2 DOF [2]. This is difficult to arrange when the measurement probe must be much smaller than the measurement volume and the angular

Manuscript received October 16, 2007; revised January 22, 2008. First published October 24, 2008; current version published May 13, 2009. This work was supported in part by the U.S. National Institutes of Health under Grant R21 EY016359, by the U.S. National Science Foundation under Grant EEC-9731748, and by the American Society for Laser Medicine and Surgery. The Associate Editor coordinating the review process for this paper was Dr. Juha Kostamovaara.

The authors are with the Robotics Institute, Carnegie Mellon University, Pittsburgh, PA 15213 USA (e-mail: camr@ri.cmu.edu).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TIM.2008.2006132

range is large. We address this problem by using a probe with multiple modulated sources, where each PSD simultaneously tracks multiple sources, enabled by frequency-domain multiplexing (FDM).

Analog lock-in amplifiers have often been used with PSDs to measure the position of a modulated source in the presence of ambient light [3]. Multiple lock-in amplifiers can be used to implement FDM, but the large amount of hardware required has deterred the use of this technique with PSDs. The lock-in concept is widely used with other sensors, and many designers have found that when the modulation frequencies are low enough, it is more economical and more flexible to implement the phase-sensitive detector in software [4], [5]. However, due to the assumption of static signals, the frequency response of the phase-sensitive detector is often neglected. This paper demonstrates how to efficiently implement a digital phase-sensitive detector with any desired frequency response and evaluate the response tradeoffs in the context of dynamic signal measurement.

In Section II, we review the PSD characteristics and quantitative resolution limits, consider the lock-in measurement system as a communication link, discuss the application of FDM to PSDs, and compare this approach with alternative time-domain techniques in widespread use. As a reference design, Section III describes an instrument that performs 5-DOF tracking of microsurgical tools. We discuss in detail several difficulties that are widely encountered in applications of PSDs. Section IV considers the lock-in phase-sensitive detector as a multirate digital signal processing problem, explores the connections with Fourier spectral estimation and filter banks, describes the optimal choice of sampling rate and modulation frequencies under design constraints, and discusses the tradeoffs in window design as relevant to PSD demodulation. Section V experimentally characterizes the performance of the reference design (Section III) in terms of the noise spectrum, resolution, bandwidth, drift, effects of multiplexing, workspace size, relative error, probe error, and rotation/translation coupling. The Appendix describes calibration procedures for the PSD camera and probe.

II. BACKGROUND

A. PSDs

A PSD is a nonimaging analog optical position sensor. It is a large-area planar photodiode with high-resistivity contacts. A spot of light creates a photocurrent, and the contact resistance forms a current divider that splits the current between two

terminals, allowing the position of the spot to be determined from the ratio of the currents. If the currents along an axis are x_0 and x_1 , then the position normalized to the range $[-1, 1]$ is

$$x = \frac{x_1 - x_0}{x_1 + x_0}. \quad (1)$$

In a 2-D duo-lateral PSD, the front and back contacts are high-resistivity sheets, and the front and rear contact pairs are oriented at right angles to each other. This allows the position of the spot to independently be measured in 2-D.

Because of the analog spatial averaging performed by the resistive contact, the PSD responds to the mean center of illumination (or centroid). The size and shape of the spot have little effect on position measurement, so precise focusing is not required. The disadvantage is that, having no spatial selectivity, the PSD is easily disturbed by ambient lighting and unintended reflections.

PSD resolution is limited by noise. The noise analysis is complex because some noise components are correlated between x_0 and x_1 , and so are either cancelled or doubled by (1). If I_s is the detected RMS signal current, R_{ie} is the PSD interelectrode resistance, and I_b is the mean dc sum of the dark current and uniform background illumination, then the standard deviation of the normalized x position is

$$\sigma_n \approx \frac{\sqrt{B}}{I_s} \sqrt{\frac{16 kT}{R_{ie}} + \frac{2qI_b}{3}}$$

and the signal-to-noise ratio (SNR) is $20 \log_{10}(2/\sigma_n)$ dB [3]. If l is the PSD size, then the standard deviation of the spot position on the PSD is

$$\sigma \approx \frac{l\sigma_n}{2}.$$

We have neglected the contribution of transimpedance amplifier noise, which can be made small in comparison to the R_{ie} thermal noise (see Section III-D). The shot noise in I_s does not affect the measurement because fluctuations in the signal amplitude are normalized out. The I_b shot noise does degrade the resolution because it is distributed across the PSD and thus is not perfectly correlated at the two terminals.

Using typical values $I_s = 1.4 \mu\text{A}$ RMS, $R_{ie} = 10 \text{ k}\Omega$, $I_b = 1 \mu\text{A}$, and $B = 1 \text{ kHz}$, then $\sigma_n \approx 29 \times 10^{-6}$, and SNR = 97 dB. If $l = 10 \text{ mm}$, then $\sigma \approx 290 \text{ nm}$ RMS, giving a 1.9- μm 99.9% confidence interval (equivalent resolution peak-to-peak).

The PSD bandwidth is established by R_{ie} and the junction capacitance C_j . These are both distributed, so precise analysis requires a position-dependent transmission line model [6]; however, a lower bound on the bandwidth is

$$f_{3\text{dB}} \geq 1/\pi R_{ie} C_j.$$

For a typical 1-cm 2-D PSD with $R_{ie} = 10 \text{ k}\Omega$ and $C_j = 175 \text{ pF}$, $f_{3\text{dB}} \geq 180 \text{ kHz}$. To use more than a fraction of this bandwidth, it is necessary to consider or even exploit the effects of position-dependent phase shifts [7].

The effect of PSD size on system performance is worth considering. If the system incorporates optical magnification, we can measure any size displacement with any size PSD, so the

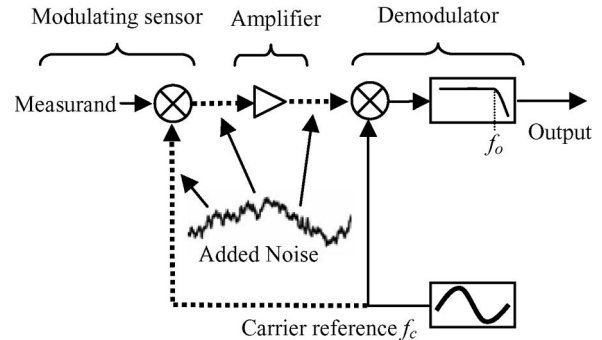


Fig. 1. Lock-in amplification application.

resolution at the sensor (σ) is relatively unimportant, and our primary concern is the SNR. I_s is determined by the spot power, and PSD saturation limits the power density (spot brightness). If d is the spot diameter, then the SNR is proportional to d^2 at the saturation limit. A larger PSD will tend to allow a larger spot size, so the SNR limit also approximately increases as l^2 . However, C_j is also proportional to l^2 , whereas R_{ie} remains constant (at best), so larger devices have a lower bandwidth.

B. Lock-In Amplification

The function of the lock-in amplifier (or phase-sensitive detector) can be described as a narrow tracking bandpass filter with capability to measure amplitude or phase [8]. Although correct, this completely obscures the reason for using lock-in amplification to make measurements near dc, which is to suppress the $1/f$ noise and other low-frequency interference.

To apply lock-in amplification to a measurement problem (Fig. 1), we must somehow incorporate a modulator in the sensor, either by modulating the measurand (e.g., with a chopper) or by using a modulating sensor. A ratiometric sensor like the PSD multiplies the measurand by the excitation amplitude, so it can serve as a modulating sensor. In communications terminology, observe that we have arranged for amplitude modulation (AM) of a carrier f_c by the measurand. If the measurand has a fixed dc value, then the output of the sensor is the carrier scaled by this value. The demodulator recovers the measurand value by multiplying the sensor output by the same carrier reference that originally modulated the measurand and then low-pass filtering the result. The principle of operation of the demodulator is that multiplying by the reference signal converts the AM signal spectrum centered at f_c into two components at $f_c \pm f_c$. The $2f_c$ component is above the low-pass filter cutoff f_o and so is suppressed. With a dc measurand, the difference frequency is zero, so the dc level is reconstructed at the modulator output.

By viewing the lock-in measurement system as an AM communication link, we can see that the frequency response from the measurand to the output is simply that of the low-pass filter in the demodulator; signals below f_o pass through unchanged. The dashed signal paths in Fig. 1 carry AM signals; interference on these paths is rejected unless it falls within the bandwidth $f_c \pm f_o$. This means that the measurement is insensitive to offsets, drift, $1/f$ noise, or any other low-frequency interference introduced at these points.

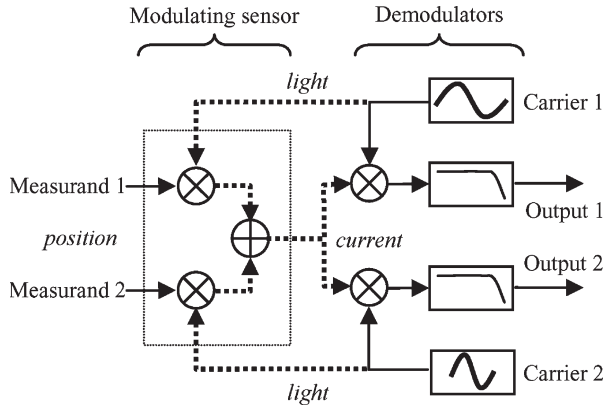


Fig. 2. FDM sensor.

Because these low-frequency disturbances almost invariably do exist in dc measurements, it is somewhat surprising to realize that if the noise were either white or increased at high frequency ($> f_o$), then the signal-to-noise improvement would be at least as good if we discarded the rest of the lock-in system and used the low-pass filter alone. So the removal of low-frequency ($< f_o$) noise is the only reason to use lock-in amplification for dc measurements. Phase-sensitive detectors also see use in inherently ac applications (such as impedance measurement), where $1/f$ noise is generally not a problem, and the phase measurement capability is often valuable.

C. FDM PSDs

The communications concept of frequency domain multiplexing (FDM) can be implemented by a bank of lock-in amplifiers. We can then simultaneously transmit multiple signals over the same medium as long as they are sufficiently separated in frequency to avoid any significant interference. Fig. 2 shows the concept of FDM as applied to PSD measurement, which has been done in practice [9], [10] and has theoretically been analyzed and experimentally validated [7], [11], but remains a little-used technique—perhaps undeservedly so. There is negligible loss of accuracy or resolution, and the hardware complexity is minimal when software demodulation is used.

Note that FDM assumes that the sensor output is the sum of the two modulated measurands. If there were nonlinear interaction between the two carriers, then the two measurements would become coupled, degrading the accuracy of both. However, the PSD centroid measurement property implicitly requires a sum response, so we can expect the nonlinear interactions to be small. To characterize this problem in the frequency domain, observe that any nonlinear interaction between carriers at frequencies f_{c1} , f_{c2} will cause intermodulation distortion (IMD), creating output components at $f_{c1} \pm f_{c2}$. We have experimentally measured these IMD products and found that they are at least 80 dB below the carrier.

In any multiplexing scheme, if the SNR (or resolution) is to be held constant, then the power per channel must also remain constant as the multiplexing is increased (total power increases proportional to multiplexing). As for the feasibility of doing so with FDM PSDs, Qian *et al.* [11] are overly pessimistic. Since PSD saturation is a localized phenomenon, it is possible to add

nonoverlapping spots of the same brightness without causing saturation. In addition, the power may be limited by the source brightness rather than PSD saturation, in which case adding more sources proportionally does add more power. In theory, there is also some resolution degradation due to the shot noise of the other spots, but this is position dependent and likely to be submerged in the R_{ie} noise as long as all the spots are close together. The degree of multiplexing is ultimately limited by the PSD bandwidth, but in the applications we are currently considering, the total bandwidth is less than 10 kHz, so the PSD bandwidth is not a limiting factor.

D. Time-Domain Techniques

An alternative way to achieve the interference-rejection function of lock-in amplification, which has also been used with PSDs, is to modulate the signal on/off, and then, in the receiver, sample the signal during on and off periods, subtracting the two values

$$y(n) = x(2n + 1) - x(2n). \quad (2)$$

Variously known as dc restoration, correlated double sampling, or background subtraction, this approach has the advantage of conceptual and implementation simplicity and can be implemented using either analog hardware or a data acquisition system and software.

Note that if we define a reference signal

$$r = \{-1, +1, -1, \dots\},$$

then (2) becomes

$$y(n) = r(2n + 1)x(2n + 1) + r(2n)x(2n).$$

This allows us to reinterpret background subtraction as a discrete-time implementation of lock-in amplification with decimation ratio 2 and a rectangular window (Section IV). We will see that this results in a frequency response that droops considerably, making the usefully flat signal bandwidth narrower than the noise bandwidth. The actual noise bandwidth depends on the details of the sampler and any antialias filter. In the (likely) case that the sampler is not an integrator, the noise bandwidth will be wider, perhaps considerably so if antialias filtering is neglected.

Background subtraction has been generalized into time-domain-multiplexed systems, where each LED is flashed in turn [12]. Since only one LED is lit at a time, it is not possible to proportionally increase the average power as multiplexing is increased because of both LED characteristics and PSD saturation [13], so the SNR (resolution) is worse than with FDM.

III. REFERENCE DESIGN

To provide a concrete context for design discussion, we will describe an instrument we have developed to support research in the quantification of hand tremor during microsurgery and development of devices for active tremor compensation [14], [15]. This instrument measures the 5 DOF position

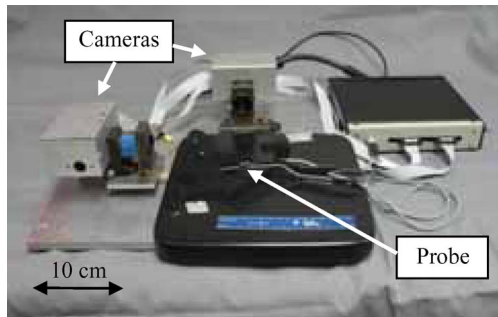


Fig. 3. Five-degree-of-freedom instrument.

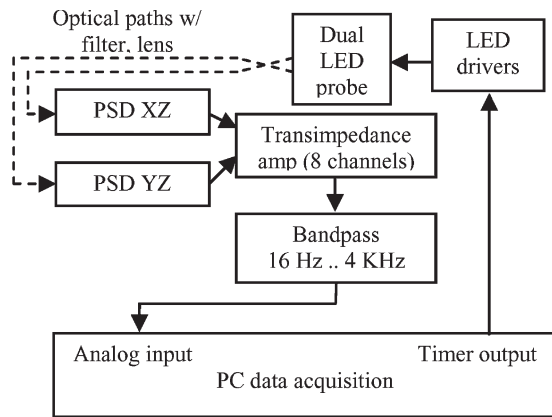


Fig. 4. System architecture.

and orientation of a probe by simultaneously measuring the 3-D position of two LEDs.

The major functional components of the instrument are two 2-D PSD cameras and a probe with two lights on it (Fig. 3). The basic measurement principle is to triangulate the positions of the lights by tracing the rays from each camera. The cameras are arranged at right angles to each other, which provides roughly equal measurement resolution along all three axes. However, most of the design considerations subsequently discussed generally apply to systems using any number of PSDs and LEDs in any geometric arrangement.

The requirements are micrometer resolution over a range of centimeters, 5% accuracy, a compact probe, and kilohertz measurement rate. Although tremor does not contain significant energy above 20 Hz, one of the uses of this instrument is as part of a closed-loop tremor-canceling control system, and to exhibit significant cancellation, the loop bandwidth must substantially exceed the highest tremor frequencies, requiring a high measurement bandwidth. The requirement for position accuracy is modest for human performance measurement because individual variation is great, and when used in a feedback system, the inaccuracy causes only a small variation in open-loop gain, which is largely cancelled by feedback.

A. System Architecture

The instrument only requires a small amount of custom electronics (Fig. 4). A commercial data acquisition board digitizes the sensor signals and generates the excitation signal for the

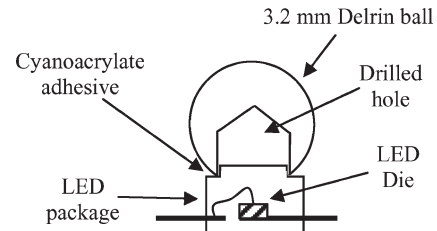


Fig. 5. LED emitter construction.

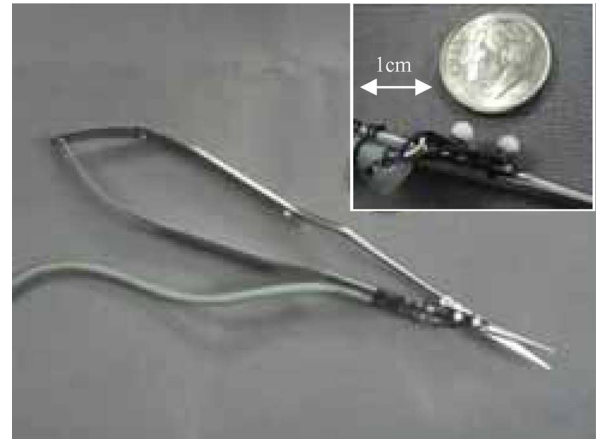


Fig. 6. Probe with two LEDs.

lights. The remainder of the signal processing and the instrument control interface are implemented in Labview software.

The previous version of this instrument measured one 3-D point using a reflective probe, a single modulated illuminator, and analog lock-in amplification [16]. Due to the need to demodulate eight channels, this required a significant amount of hardware, and the thought of further multiplying this hardware two or more times to support FDM was daunting. Instead, we opted to move as much of the signal processing into software as possible. Only current-to-voltage conversion and antialias filtering are done in hardware. All demodulation is performed in software. This does increase the demands on the analog-to-digital conversion in terms of conversion rate and dynamic range as well as the demands on the CPU, but given the low modulation frequency (a few kilohertz), this is well within the capability of standard desktop PCs and data acquisition hardware.

B. LED Probe

An ideal emitter would have rotational symmetry and uniform illumination such that the 2-D centroid of the image coincides with the 3-D center of the emitter when viewed over a wide angle (130°). Any departure from this ideal causes a position error that varies with viewing angle. Our emitter is an illuminated sphere created by attaching a Delrin ball to a packaged LED (Fig. 5). The drilled hole serves two purposes: positive location of the ball on the LED and increased illumination uniformity. To maximize the symmetry of the emitter, light leakage from the clear body of the LED is masked by painting it silver and then black. The LED is an HDSL-4400 (Avago Technologies) with peak emission at 875 nm. The complete probe has two emitters along its axis that are spaced 7 mm apart (Fig. 6).

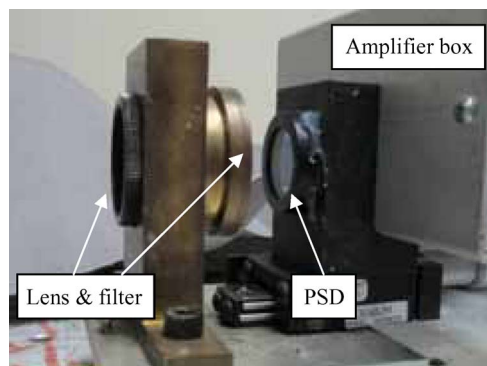


Fig. 7. PSD camera detail.

C. LED Drivers

The LEDs are pulsed using a 150-mA on/off square wave, which is the maximum drive specified for this LED. A square-wave drive was primarily chosen so that the excitation could be generated using a timer output [as opposed to a digital-to-analog converter (DAC)]. A square-wave drive also allows the driver to be a simple switch with a current-limiting resistor. It should be noted, however, that the switch can create RF harmonics that may be coupled into sensor circuits by electromagnetic radiation from the probe cable or through the power supply. These signals are phase coherent with the excitation, so they will not be rejected by the demodulator. It is best to avoid overly fast switching, use a separate power supply, and wrap the probe leads several turns around a ferrite toroid.

D. PSD Camera

Light is focused on the PSD by a single-element double-convex lens (Fig. 7). We have found that for use with PSDs, simple single-element lenses outperform much larger and more expensive camera lenses. This is because camera lenses are designed to minimize spherical and chromatic aberrations, which are irrelevant for our purposes, and in doing so introduce geometric distortion. In fact, spherical aberration is beneficial because it creates a soft-focus effect that increases the spot size, reducing the brightness and, therefore, avoiding PSD saturation. Some camera lenses are also unsuited for near-infrared use because they absorb 75% or more of the illumination.

In our PSD camera, the lens is 12.5 mm in diameter with 30-mm focal length and has a near-infrared anti-reflective coating (Melles Griot). Before reaching the PSD, the light passes through an infrared-pass filter that blocks much of the ambient light. This is important not only because ambient light might overload the sensor and may be modulated (causing hum) but also because even a steady illumination degrades resolution by increasing the shot noise.

Using a thin-lens model, the front-to-back focus ratio of 120–35 mm would be expected to give an optical demagnification of 1 : 3.4. In fact, the calibrated demagnification (see Appendix A) is 1 : 4. The difference may be due to the filter shortening the effective back focus and due to coma resulting from the lens's spherical aberration.

Focusing the PSD camera is more difficult than with an imaging camera, but fortunately, the focus is far less critical (defocusing may even be desirable). We removed the infrared filter, substituted a screen for the sensor, and focused using visible light. Note that in the presence of spherical aberration, the minimum spot size is achieved at a shorter back focus distance than in the thin lens model.

The PSD is a 1-cm² OSI Optoelectronics DL-10 duo-lateral device with nonlinearity of < 1% and drift of < 0.6 $\mu\text{m}/^\circ\text{C}$. The PSD is reverse biased by 5 V to inhibit saturation and increase bandwidth.

Each of the four PSD terminals requires a transimpedance amplifier to convert the photocurrent to a voltage that can be digitized. These amplifiers are located in the camera to minimize noise pickup and gain peaking due to wiring capacitance. The PSD resolution is determined by the SNR and ultimately is limited by PSD intrinsic noise (Section II-A), but attaining this limit requires a low-noise transimpedance amplifier [3]. Typically, 2-D PSDs have an interelectrode resistance (R_{ie}) of 10 k Ω . Two key principles are that the op-amp must be selected for this source resistance [17] and that the gain-setting feedback resistance should be much larger than R_{ie} (we use 200 k Ω). In [13], amplifier noise was dominant because a noisy op-amp (TL071) and a small feedback resistor were used. Good op-amp choices are the Texas Instruments OPA2228, the industry standard OP37, or a decompensated low-voltage-noise FET input op-amp ($e_n < 6 \text{ nV}/\sqrt{\text{Hz}}$).

E. Channel Matching

Because the analog signals are all ac, the only way that the signal conditioning can contribute to drift is by changes in frequency response. Furthermore, because of the normalization done by position computation, gain changes have little effect as long as the pairs of channels for each axis remain well matched. If the gains of the two axis channels differ, then the position sensitivity differs from one side of the PSD to the other, creating nonlinearity.

F. Antialias Filtering

The voltage signal from each transimpedance amplifier is ac coupled through a single-pole 16-Hz high-pass filter, primarily to remove the dc bias. Then, the signal is low-pass filtered through a two-pole Butterworth antialias filter. The low filter order was chosen both for simplicity and to promote channel matching. To ensure matching, the antialias filters were constructed with 1% capacitors and 0.1% resistors, and the capacitors for the axis channel pairs were matched to 0.1%. It would also be possible to use a gain calibration procedure to match the channels in software. Channel matching the high-pass filters helps the rejection of low-frequency interference such as hum.

G. Analog-to-Digital Conversion

The sensor signal has a dynamic range extending from the peak illumination when all lights are on ($\sim 10 \mu\text{A}$ p-p) down to the noise floor ($\sim 100 \text{ pA}$ RMS). It is important that the

least significant bit is well dithered because the modulation scheme causes the quantization noise to be concentrated at the very same frequencies to which the demodulator responds. This defeats the purpose of modulation by allowing low-frequency noise to pass through the demodulator.

A high performing but quite inexpensive solution to digitization would be to use one sigma-delta audio analog-to-digital converter (ADC) per channel either in a custom design or in a high-end PC audio card. This approach provides both dithering and antialias filtering. In our instrument, we chose to ease interfacing to Labview by selecting an 18-bit data acquisition card (National Instruments PCI-6289), which has enough resolution for the signal to be self-dithering.

H. Ray Intersection

Once the four signals from each PSD are demodulated as described in Section IV, they are converted into two axis positions using (1) and then into rays in instrument coordinates using the sensor calibration information (Appendix A).

We find the 3-D light position by intersecting the rays from the two PSD cameras. Since there are four measurements and three unknowns, the problem is overconstrained. The intersection is approximated by the midpoint between the two rays at their point of closest approach. The distance of closest approach (ideally 0) is an indicator of measurement quality. Although it is possible to compute the 3-D position in real time using the geometric intersection formulas, the runtime has been considerably reduced by precomputing a 4-D lookup table that maps sensor axis positions to the corresponding 3-D position and matching error. Multilinear interpolation is then used to compute the output position.

IV. DIGITAL LOCK-IN AMPLIFICATION

We can straightforwardly implement a digital lock-in amplifier by taking Fig. 1, adding an antialias filter chosen to pass the carrier frequency, digitizing the signal at a sampling frequency f_s chosen to minimize noise aliasing, and replacing the reference generator, multiplier, and low-pass filter with digital implementations [4]. Note, however, that we can discard most of the output samples without loss of information because the measurement bandwidth is much less than $f_s/2$.

Efficiency can be improved by defining the demodulator as a multirate function that has an output sampling rate f_d chosen according to the bandwidth requirements and an integer decimation ratio $M = f_s/f_d$ chosen according to the needs of the demodulation process. Let the frequency of each carrier c be $k_c f_d$ a harmonic of the output sample rate chosen such that the channel index k_c satisfies $k_c \leq M/2$. Since the multiplexed superposition of all the carriers repeats every M samples, it makes sense to process the signals in M sample blocks. There is a fixed phase relationship between the carriers and block boundaries, so the demodulator for channel index k_c multiplies each block by the same fixed reference signal

$$r_c(n) = \cos\left(\frac{2\pi}{M}k_c n\right) = \text{Real}\left(e^{-\frac{2\pi j}{M}k_c n}\right). \quad (3)$$

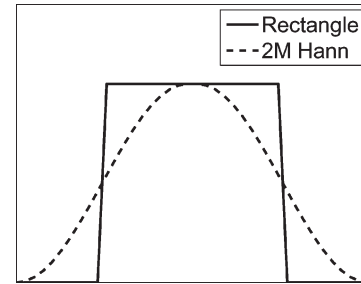


Fig. 8. Rectangular and Hann window functions.

Let $x(n)$ be the n th input sample ($0 \leq n < M$). The demodulating multiplier output is then $x(n)r_c(n)$. If (as in [5]) we decimate by summing each block, then

$$X_{k_c} = \sum_{n=0}^{M-1} x(n)r_c(n). \quad (4)$$

Then, from (3)

$$X_{k_c} = \sum_{n=0}^{M-1} x(n)\text{Real}\left\{e^{-\frac{2\pi j}{M}k_c n}\right\}.$$

This is precisely the real part of the discrete Fourier transform (DFT). The result of demodulation of the carrier k_c is X_{k_c} , which is the k_c th component of the M -point DFT of $x(n)$. Although this suggests the possibility of efficient implementations using fast Fourier transform (FFT), the immediate value of this DFT interpretation is to make the connection with a body of theory that is likely to be familiar to instrumentation designers, namely, spectral estimation via FFT, windowing, and short-time Fourier transform (STFT).

A. Windowed DFT

Generally, spectral estimation first requires multiplying the data by a window function $w(n)$, such as the Hann window (Fig. 8). This reduces the spectral leakage, i.e., a spurious broadening of spectral peaks that degrades the frequency resolution. Using no window is equivalent to a rectangular window, so we can call (4) the rectangular window demodulator. By adding a window to the DFT demodulator, we can reduce the amount of nonharmonic noise and interference that leak into the output

$$X_{k_c} = \sum_{n=0}^{M-1} w(n)x(n)r_c(n). \quad (5)$$

With the windowed DFT, the response of each DFT component to nonharmonic signals is a frequency-shifted version of the response that the window function would have if used as an finite-impulse-response (FIR) filter. Fig. 9 shows the frequency response of the rectangular window. Since it is a low-pass filter, the response peaks at 0 (dc). The response at negative frequencies represents the response near f_s (which aliases to dc).

For our purposes, a desirable property of the rectangular window of length M is that it notches out all the harmonics

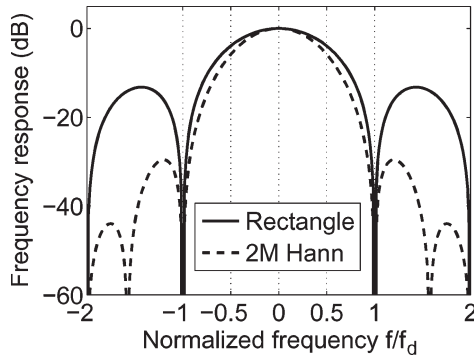


Fig. 9. Rectangular and Hann window frequency responses.

of f_d . We have ensured coherent sampling by the choice of harmonic carrier frequencies $k_c f_d$ and sampling rate $M f_d$; so when using a rectangular window, there is no leakage at any harmonic of f_d . This means that the carrier isolation is perfect (in the absence of distortion). Many other windows do not have this property. However, the Hann window of length OM does have it for $O \geq 2$. Furthermore, as Fig. 9 shows, the response of the Hann window decays much faster as the frequency moves away from dc, which greatly helps the rejection of nonharmonic noise and interference.

B. Demodulation of Time-Varying Measurands

We have shown that given M samples, we can use the DFT to measure the static carrier amplitude X_{k_c} , and that by using a window function, we can minimize the leakage of interfering nonharmonic signals into the amplitude measurement. Recall that we demodulate each M -sample block of x in turn to generate a discrete-time signal $X_{k_c}(t), X_{k_c}(t+1), \dots$. Equation (5) becomes

$$X_{k_c}(t) = \sum_{n=0}^{M-1} w(n)x(tM+n)r_c(n). \quad (6)$$

This is the windowed STFT.

In Section II-B, we observed that the frequency response of the analog lock-in measurement system from measurand to output is simply that of the low-pass filter in the demodulator. However, what is the frequency response using the windowed DFT demodulator? We will see that except for the aliasing introduced during the original digitization or the decimation, the windowed DFT demodulator gives the same overall low-pass characteristic, with the window taking the role of the response-defining low-pass filter.

The windowed DFT demodulator was constructed to be a multirate operation with decimation factor M . For decimation to not cause aliasing, the demodulator bandwidth must be limited to $f_d/2$ before it is downsampled. Remove the product with the reference signal $r_c(n)$ from (5), and consider the effect of the window alone as

$$y(t) = \sum_{n=0}^{M-1} w(n)x(tM+n).$$

This is a polyphase FIR decimation filter with impulse response $w(n)$, whose low-pass response is the needed antialias filter [18]. The same $w(n)$ can be interpreted either as a DFT window or a decimation filter—spectral leakage causes aliasing—and the criteria for a leakage minimizing window are the same as for an antialias filter at sampling rate f_d .

To do better than the rectangular window, we must use a window longer than M . If we choose an overlap factor O , then the demodulator becomes the windowed overlapped STFT

$$X_{k_c}(t) = \sum_{n=0}^{OM-1} w(n)x(tM+n)r_c(n).$$

We can halve the number of multiplications by precomputing

$$d_c(n) = w(n)r_c(n).$$

This converts the entire demodulator into a single modulated polyphase FIR decimation filter

$$X_{k_c}(t) = \sum_{n=0}^{OM-1} d_c(n)x(tM+n). \quad (7)$$

In practice, the filters are not symmetrical because each r_c must be phase adjusted to compensate for phase lag in the antialias filter, nonsimultaneous sampling due to ADC multiplexing, etc.

If we were to demodulate all of the $M/2$ channels, we would have a *filter bank*. There are FFT-based algorithms, such as the polyphase DFT filter bank [18], that are likely to be more efficient than direct implementation by (7) when more than two carriers are demodulated, and are also more efficient than the windowed overlapped STFT. However, FFT requires M to be a power of 2, which is incompatible with the (frequently prime) M selected by the procedure described next.

C. Parameter Optimization

For the determination of f_d , M , and the assignments of carriers to channels, the design constraints include the number of carriers, the maximum f_s , the minimum f_d , and the minimum and maximum $k_c f_d$. Generally, $k_c \geq 2$ is needed to increase the low-frequency interference rejection, and $k_c f_d$ must be less than the analog antialias filter cutoff.

An implementation constraint is that f_s and all the carrier half-periods $2 k_c f_d$ must be integer divisors of a single timebase. In addition, we chose to use a square-wave drive for the LEDs (high distortion of odd harmonics) and a low-order analog antialias filter that allows some aliasing of these harmonics. These aliases may land on the channel of another carrier so the isolation between carriers depends on both the channel assignments and M .

As observed in [11], with a different implementation of FDM, for two square-wave carriers, a high isolation can be achieved by choosing a 1:2 frequency ratio. We have found similar solutions for larger numbers of carriers.

We wrote a program that exhaustively compares the carrier isolation of all possible channel assignments over all feasible

values of f_d and M . This takes only seconds, even with six carriers. Using a 10-MHz timebase, the result for four carriers with $f_d \geq 1$ kHz, $2 \text{ kHz} \leq k_c f_d \leq 8 \text{ kHz}$, and $f_s \leq 62.5 \text{ kHz}$ is $f_d = 1082.2511$, $M = 55$, and $k_c = \{2, 3, 4, 7\}$. The isolation model considers carrier harmonic distortion, first-order IMD, and aliasing. Analysis is simplified by the observation that, due to the coherent sampling, the aliases and IMD products are also harmonics of f_d , and at these frequencies, the response is always either unity or zero when using rectangular or Hann windows. The actual amount of intermodulation is quite low (-80 dB), so aliasing of carrier harmonic distortion is the dominant cause of interference between the channels.

D. Summary of Digital Lock-In Amplification

First, we proposed digitally implementing the analog lock-in amplifier as a multirate decimation operation. Then, we interpreted the multiplexed demodulator as the DFT, investigated interference rejection using the windowed DFT, and then observed that the windowed STFT is a bank of modulated polyphase decimation filters, thus establishing that the window frequency response is the demodulator frequency response, and that spectral leakage causes aliasing. None of these connections is novel. The purpose has been to explain the operation of an efficient digital phase-sensitive detector and to emphasize the importance of the window function.

As the connection with FIR filter design and DFT window selection should make clear, the window for the lock-in demodulation of time-varying signals must be chosen based on application-specific requirements for passband flatness and interference rejection on one hand, and the conflicting requirements for efficient implementation and minimal group delay on the other. By not using a window, Smith *et al.* [5] implicitly claim optimality for the rectangular window, but this is only true for static signals with white noise. If a short window is required ($O \leq 5$), then standard DFT window functions will probably give better results than an equiripple low-pass FIR filter, but do not expect passband flatness or any great attenuation at $f_d/2$. If group delay and efficiency are not a problem, then large FIR filters can be designed with any desired response. For use inside a feedback loop, group delay is crucial, but passband flatness is relatively unimportant. The rectangular window works well, except for interference rejection. If $f_d = 1 \text{ kHz}$ and $k_c = 2$, then interference rejection at 120 Hz is 30 dB better for the Hann window of length $2M$ than for the rectangular window at the cost of an additional $1/2$ sample group delay.

V. PERFORMANCE OF THE REFERENCE DESIGN

A. Noise Spectrum

Resolution is limited by the 3-D measurement noise, which is substantially white (Fig. 10). The peak at 120 Hz is due to interference from ambient lighting.

At low frequencies, the noise density is higher (Fig. 11). The $1/f$ noise corner is about 2 Hz when the instrument is exposed to the lab environment and about 0.1 Hz when enclosed in a

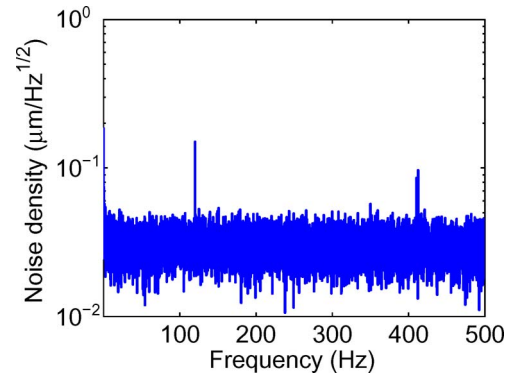


Fig. 10. High-frequency 3-D noise spectrum.

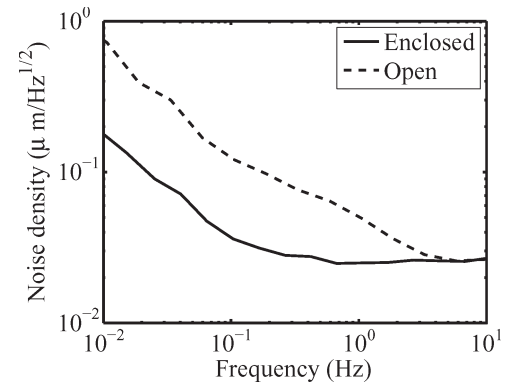


Fig. 11. Low-frequency 3-D noise spectra.

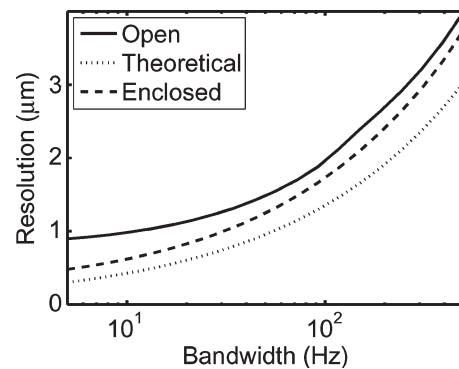


Fig. 12. Three-dimensional position resolution versus bandwidth.

box to shield the optical paths from air currents. When operated in the normal (open) environment, the low-frequency noise is clearly dominated by the effects of air turbulence inherent to the optical measurement principle and is not due to the PSD or signal conditioning.

B. Resolution Versus Bandwidth

To find the resolution, we measured the 3-D RMS amplitude of 1000 s of data that was bandpass filtered between 0.01 Hz and the desired measurement bandwidth. This RMS amplitude was then multiplied by 6.6 to convert it to a 99.9% confidence interval. Fig. 12 shows the resolution versus bandwidth, with the instrument in both open and enclosed environments.

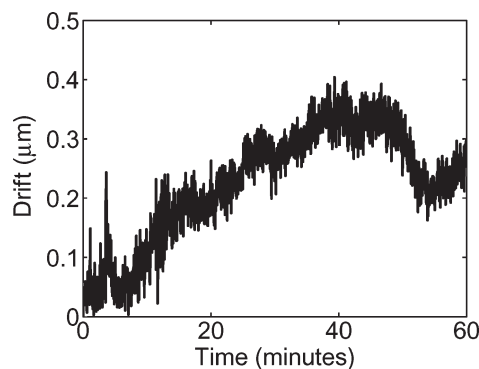


Fig. 13. Three-dimensional position drift.

For reference, we also plot a 3-D resolution limit derived from the ideal 1-D PSD resolution (Section II-A) using the actual PSD current $I_s = 4.5 \mu\text{A RMS}$.

The resolution when enclosed is close to the theoretical limit, and at typical measurement bandwidths of $> 50 \text{ Hz}$, the broadband white noise dominates the low-frequency noise due to air turbulence, so in practice, enclosure gains little. To convert 3-D position resolution to the equivalent angular resolution, multiply by $0.01^\circ/\mu\text{m}$. To convert to 1-D resolution at the PSD, multiply by 0.16.

C. Drift

In practice, drift is dominated by the thermal expansion of the probe test fixture in response to the 300 mW dissipated by the probe. We ran the instrument overnight to thermally stabilize it and then measured the nominally fixed probe position for an hour. Fig. 13 shows the magnitude of the 3-D position drift. The 300-nm drift is 10 ppm of the full scale, showing the excellent drift performance resulting from lock-in amplification.

D. Effects of Multiplexing

In Section II-C, it was claimed that frequency-division multiplexing a PSD would not significantly degrade accuracy or resolution. This was experimentally verified by shutting off one light and observing the effect on the measurement of the other light. The isolation model (Section IV-C) predicts 73-dB channel isolation, whereas we observed approximately -80-dB change in the PSD signal amplitude. The corresponding 3-D position change was $0.7 \mu\text{m}$, which is 24 ppm of the full scale, well below the measurement resolution at full bandwidth. There was no detectable change in the measurement resolution (noise).

E. Workspace

The workspace is approximately a 3-cm cube, but both lights must remain inside the workspace, so the probe motion is limited to about $\pm 1 \text{ cm}$. Typically, rotations of $\pm 20^\circ$ can be measured. Rotation is limited by the constraint that both LEDs must be visible to both sensors, with neither being occluded by the other or by the LED body or probe tip.

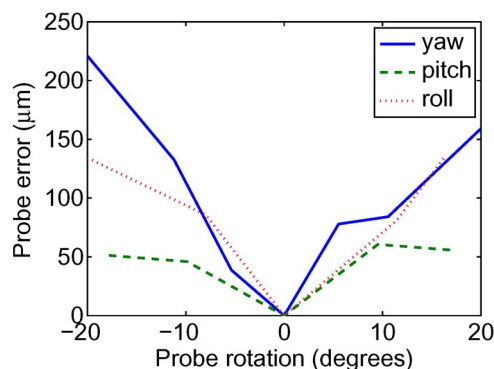


Fig. 14. Probe position error versus rotation.

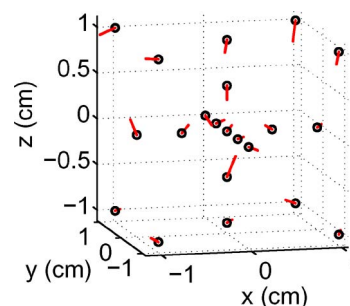


Fig. 15. Translation error (30× exaggeration).

F. Probe Error

As discussed in Section III-B, accuracy depends on the degree to which the 2-D centroid of the light coincides with the 3-D center when light is viewed from different angles. Imperfections in the probe cause the light centroid (and the measured position) to change depending on the angle from which the probe is viewed. Probe error is a cross coupling from angular to translational motion, but it significantly affects the measurement even of pure translational motion, as the sensor viewing angles still slightly change.

We used the probe calibration procedure (Appendix B) to measure probe error by repeatedly calibrating the probe over $\pm 20^\circ$ rotation. Fig. 14 shows the probe position error versus rotation angle for each axis.

G. Relative Error

Fig. 15 shows the position error when the probe undergoes $\pm 1\text{-cm}$ translation. These data were collected during a PSD camera calibration procedure (Appendix A) but are from the other light, which was *not* used to generate the calibration. The error is 1.6% of translation and is exaggerated 30× for clarity. When $\pm 20^\circ$ rotation is introduced, the worst-case relative error measured throughout the workspace is 6.5% of angle and 2.7% of translation.

The UDT DL-10 PSD nonlinearity is specified at $< 1\%$ (Section III-D), so there are other significant error sources. We believe that the reduction in accuracy when probe rotation is introduced is due to probe error. Probe error somewhat degrades the translation accuracy but has an even more dramatic effect on the angular error due to larger change in probe orientation.

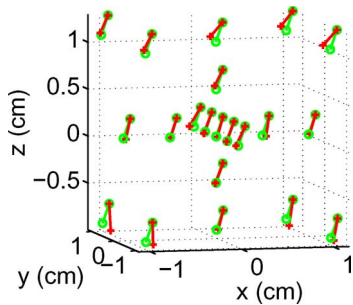


Fig. 16. Angular error during translation ($30\times$ angular exaggeration).

H. Coupling Between Rotation and Translation

There is a cross coupling between rotation and translation: pure rotation causes apparent translation, and vice versa. Probe error is by definition the cause of rotation-to-translation coupling. Since angle is derived from light positions, translation-to-rotation coupling is a consequence of overall position nonlinearity, which is determined by PSD camera nonlinearity as well as probe error. We measured the apparent translation due to a $\pm 1^\circ$ rotation and the apparent rotation due to a ± 3 -mm translation. The worst-case values across the workspace are $21 \mu\text{m}/^\circ$ and $0.36^\circ/\text{mm}$. Fig. 16 shows the angular error as the probe undergoes ± 1 -cm pure translation. At each point, the angular error is represented by the difference in direction of two vectors: one unchanging (circles), representing the actual light positions, and the other rotated in the direction of the angular error but exaggerated $30\times$. The peak error is 1.2° .

VI. DISCUSSION

We have described an approach to multidimensional position measurement based on PSD cameras and FDM LEDs that is well suited to applications where a compact probe with high angular range is needed and where bandwidth is more important than linearity. We have shown that instruments using FDM and software demodulation can approach the theoretical PSD performance limit for a nonmultiplexed PSD. We have described a software implementation of lock-in amplification that requires only the bare minimum of hardware to acquire the PSD data.

We have characterized a 5 DOF instrument with $4\text{-}\mu\text{m}$ resolution and 3% nonlinearity over a 3-cm 40° workspace at 1-kHz sample rate and are currently developing a four-LED probe pair for 9 DOF tracking of a handheld surgical instrument with 3 DOF articulation. Our experience with this system confirms the advantages of PSDs for high-bandwidth position tracking, since a high-performance system can be implemented using a quite simple hardware. The only additional cost of measuring more points is adding more LEDs and drivers. This simple implementation is well suited to use in embedded feedback systems or in cost-critical applications, even when there is no multiplexing. By optimizing the choice of modulation frequencies and sampling rate, we have shown how to minimize interference while also enabling a particularly efficient implementation of the demodulation signal processing and have shown how to choose the demodulator frequency response based on application considerations.

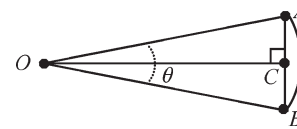


Fig. 17. Light motion due to rotation about O .

APPENDIX CALIBRATION PROCEDURES

Since this instrument is used for relative measurement, its purpose is to basically establish an arbitrary fixed coordinate system where the change in probe position over time can be measured. Because the coordinate system is arbitrary, when calibrating or testing performance, we first align the measurements with the apparent calibration fixture coordinates by applying a 6 DOF motion that minimizes the matching error. This procedure implicitly discards any long-term drift. Our primary reference for testing and calibration is a 6 DOF positioning system with resolution of $0.1 \mu\text{m}$ (Hexapod, Model F-206, Physik Instrumente, Karlsruhe Germany) that has the capability to perform rotations about a software-configured center of rotation. We also used a manual 3-D linear stage with range of 25 mm (Model H-RH, Line Tool Co, Allentown, PA) to test the accuracy under pure translation over the entire workspace.

A. PSD Camera Calibration

Although it does not record an image, the PSD camera can be calibrated using standard camera calibration techniques. To measure a position, we must convert the two normalized axis positions (1) into a ray in the instrument coordinates. This requires precise knowledge of the position and orientation of the camera (external calibration) and camera characteristics such as focal length (internal calibration). We move the probe to known positions across the workspace, recording the axis positions at each point. This is used as an input to a version of [19] that we have adapted to PSD camera calibration.

B. Probe Calibration

We have developed a procedure to determine the light positions relative to the coordinate system of the probe, allowing the motion of a reference point on the probe (such as the probe tip) to be computed from the light position data. By making $\pm 1^\circ$ rotations about each axis, we can determine the apparent center of rotation from the resulting motion of the lights. First, consider the 2-D problem (Fig. 17). Let A and B be the positions of one light, measured at the ends of the $\pm 1^\circ$ sweep. Then, given $C = (A + B)/2$, we can solve for the center of rotation O knowing that

$$|OC| = |AC|/\tan(\theta/2).$$

The 3-D problem is solved by decomposing the localization of the 3-D center of rotation O into three 2-D problems. After rotating the light position data into alignment with the Hexapod coordinate system, as described in Appendix A, the motion due to the rotation about each axis should lie in the plane defined

by the other two axes. The three 2-D results give two independent estimates of each coordinate. The mean is used as the 3-D result.

To make this measurement insensitive to any other sources of nonlinearity, we use the instrument only as a null detector. We adjust the Hexapod center of rotation (in software) to the apparent light center and then repeat the measurement. Ideally, when the probe is rotated at the light center, there will be no motion detected. This calibration procedure is repeated until the light position converges.

REFERENCES

[1] W. Ma, A. Rizzi, and R. Hollis, "Optical coordination sensor for precision cooperating robots," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2000, pp. 1621–1626.

[2] K. Kwon, J. Park, and N. Cho, "A highly sensitive multi-dimensional motion measurement system using a spherical reflector," *Meas. Sci. Technol.*, vol. 17, no. 9, pp. 2421–2429, Sep. 2006.

[3] A. Mäkynen, "Position-sensitive devices and sensor systems for optical tracking and displacement sensing applications," Ph.D. dissertation, Univ. Oulu, Oulu, Finland, Sep. 2000.

[4] L. A. Barragan, J. I. Artigas, R. Alonso, and F. Villuendas, "A modular, low-cost, digital signal processor-based lock-in card for measuring optical attenuation," *Rev. Sci. Instrum.*, vol. 72, no. 1, pp. 247–251, Jan. 2001.

[5] R. W. M. Smith, I. L. Freeston, B. H. Brown, and A. M. Sinton, "Design of a phase-sensitive detector to maximize signal-to-noise ratio in the presence of Gaussian wideband noise," *Meas. Sci. Technol.*, vol. 3, no. 11, pp. 1054–1062, Nov. 1992.

[6] C. Narayanan, A. B. Buckman, I. J. Busch-Vishniac, and W. Wang, "Position dependence of the transient response of a position-sensitive detector under periodic pulsed light modulation," *IEEE Trans. Electron Devices*, vol. 40, no. 9, pp. 1688–1694, Sep. 1993.

[7] C. Narayanan, A. B. Buckman, and I. Busch-Vishniac, "Position detection of multiple light beams using phase detection," *IEEE Trans. Instrum. Meas.*, vol. 43, no. 6, pp. 830–836, Dec. 1994.

[8] D. P. Blair and P. H. Sydenham, "Phase sensitive detection as a means to recover signals buried in noise," *J. Phys. E, Sci. Instrum.*, vol. 8, no. 8, pp. 621–627, Aug. 1975.

[9] K. Gunnam, D. C. Hughes, J. L. Junkins, and N. Kehtornavaz, "A DSP embedded optical navigation system," in *Proc. 6th Int. Conf. Signal Process.*, 2002, pp. 1735–1739.

[10] R. Lutz, "Measuring device for location and attitude determination of an object," U.S. Patent 4 923 303, May 8, 1990.

[11] D. Qian, W. Wang, and I. J. Busch-Vishniac, "A method for measurement of multiple light spot positions on one position-sensitive detector (PSD)," *IEEE Trans. Instrum. Meas.*, vol. 42, no. 1, pp. 14–18, Feb. 1993.

[12] G. Welch, G. Bishop, L. Vicci *et al.*, "The HiBall tracker: High-performance wide-area tracking for virtual and augmented environments," in *Proc. ACM Symp. Virtual Reality Softw. Technol.*, London, U.K., 1999, pp. 1–11.

[13] C. Narayanan, A. B. Buckman, and I. J. Busch-Vishniac, "Noise analysis for position-sensitive detectors," *IEEE Trans. Instrum. Meas.*, vol. 46, no. 5, pp. 1137–1144, Oct. 1997.

[14] C. N. Riviere, J. Gangloff, and M. de Mathelin, "Robotic compensation of biological motion to enhance surgical accuracy," *Proc. IEEE*, vol. 94, no. 9, pp. 1705–1716, Sep. 2006.

[15] C. N. Riviere, W. T. Ang, and P. K. Khosla, "Toward active tremor canceling in handheld microsurgical instruments," *IEEE Trans. Robot. Autom.*, vol. 19, no. 5, pp. 793–800, Oct. 2003.

[16] L. Hotraphinyo and C. N. Riviere, "Precision measurement for microsurgical instrument evaluation," in *Proc. 23rd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Istanbul, Turkey, 2001, pp. 3454–3457.

[17] P. Horowitz and W. Hill, *The Art of Electronics*. New York: Cambridge Univ. Press, 1989.

[18] P. P. Vaidyanathan, "Multirate digital filters, filter banks, polyphase networks, and applications: A tutorial," *Proc. IEEE*, vol. 78, no. 1, pp. 56–93, Jan. 1990.

[19] R. Willson, *Tsai Camera Calibration C Code*. [Online]. Available: <http://www.cs.cmu.edu/~rgw/TsaiCode.html>



Robert A. MacLachlan (A'01–M'02) received the B.S. degree in applied mathematics from Carnegie Mellon University, Pittsburgh, PA, in 1987.

Since 1983, he has been a Research Software Developer and Engineer with Carnegie Mellon University, where he has been involved in electronic design since 1993. In 1999, he joined the Robotics Institute, Carnegie Mellon University, where he designed and developed sensor and control hardware and software for mobile robots and currently develops intelligent medical instruments at the Medical Instrumentation

Laboratory. His research interests include medical robotics, sensor design, signal processing, and control systems, with particular emphasis on exploiting the combination of sensitive measurements and effective data-interpretation algorithms to enable the construction of novel mechatronic systems that solve real-world problems.



Cameron N. Riviere (S'94–M'96) received the B.S. degree in aerospace engineering and ocean engineering from the Virginia Polytechnic Institute and State University, Blacksburg, in 1989 and the Ph.D. degree in mechanical engineering from The Johns Hopkins University, Baltimore, MD, in 1995.

Since 1995, he has been with the Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, where he is currently an Associate Research Professor and the Director of the Medical Instrumentation Laboratory. He is also an adjunct faculty member with the Department of Rehabilitation Science and Technology, University of Pittsburgh. His research interests include medical robotics, control systems, signal processing, learning algorithms, and biomedical applications of human-machine interfaces.

Dr. Riviere served as one of the guest editors of the Special Issue on Medical Robotics of the PROCEEDINGS OF THE IEEE in September 2006.