

**Active Tremor Compensation in Handheld Instrument for
Microsurgery**

Wei Tech Ang

CMU-RI-TR-04-28

*Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Robotics*

The Robotics Institute
Carnegie Mellon University
Pittsburgh, Pennsylvania 15213

May 2004

Thesis Committee:
Pradeep Khosla, Co-chair
Cameron N. Riviere, Co-chair
Gary Fedder
Russell Taylor, Johns Hopkins University

Copyright © 2004 by Wei Tech Ang. All rights reserved.

Abstract

Human's ability to perform precise micromanipulation is limited by small involuntary movements inherent to normal hand motion. Microsurgery is one area where the surgeons' performance is hampered by this manual imprecision. Not only that it complicates many delicate surgical procedures, it also makes certain types of intervention impossible. The most familiar type of erroneous movement affecting a healthy person is physiological tremor.

Instead of going with the more familiar approach of using a teleoperated robotic system, we adopted a less obtrusive and much cheaper approach of implementing accuracy enhancement within a completely handheld tool. This device senses its own motion, distinguishes the erroneous motion from the intended motion, and manipulates its own tip in real time to compensate the erroneous motion.

This dissertation focuses on the sensing and compensation of the erroneous motion, while tremor modeling and estimation is performed by a previously developed weighted-frequency Fourier linear combiner (WFLC) algorithm.

Instantaneous motion of the instrument is sensed by a new magnetometer-aided all-accelerometer inertial measurement unit (IMU). The sensing system consists of three dual-axis miniature accelerometers and a three-axis magnetometer. The redundancy in sensing provides two sources of orientation and position information. The angular motion information derived from the differential sensing kinematics algorithm has very high sensing resolution but suffers from integration drift; on the other hand, the orientation obtained from the gravity and magnetic North vectors is noisy but non-drifting. These two complementary sensing sources are fused via an augmented state quaternion-based Kalman filter to yield high quality sensing.

The instrument tip is manipulated by a three DOF parallel manipulator driven by piezoelectric actuators. The hysteretic non-linearity of the piezoelectric actuator is modeled and linearized by a rate-dependent Prandtl-Ishlinskii operator. Based on the dynamic hysteresis model, an open-loop inverse feedforward controller is implemented to accurately track dynamic motion profiles.

While the targeted application of this dissertation is in microsurgery, the principles of the approach is universal and can be extended to other micromanipulation tasks, such as cell manipulation in the biotech industry, gun-sights or handheld military tracking equipment, and handheld video photography etc.

Acknowledgement

I would like to express my first and foremost gratitude to my thesis co-advisor Cameron Riviere, for his tireless guidance, and most importantly the opportunity to work on the Micron project. The years that we spent building the instrument from ground up, screw by screw and code by code, will always be deeply remembered for many years to come. Cam's brilliance in providing creative solutions to problem still continue to awe me, and reminds me of how much I have not learnt from him.

I also want to thank my other thesis co-advisor, Pradeep Khosla, for his invaluable advice. Pradeep's stellar career achievement is dwarfed by his selfless advice and his willingness to help his students. It is an honor to understudy one of greatest roboticists of the present time.

I am also grateful to Russ Taylor, who has been an inspiration throughout the later part of my doctoral study. I have learnt and benefited so much from his unrivalled expertise and profound knowledge in the field of medical robotics. The capacity of his help in my dissertation work has made him all but in name my third thesis advisor, for which I feel truly fortunate.

I also wish to show my appreciation to Drs. Gary Fedder, Jim Osborne, and Yoky Matsuoka for the interest that they have shown in my Ph.D. career. Thanks also go to Stephanie Matvey, Suzanne Lyon Muth, Monica Hopes, and Debbie Scappatora, for the administrative help they rendered in all these years.

I owe a large part of experiments presented in this thesis to the guys in the Medical Instrumentation Lab. Special thanks to our research engineer Si Yi Khoo, who has been such a wonderful team player and relentless worker; and to my fellow Ph.D. student David Choi, who I wish my future students would be like; and to Nick Patronik, Anirudha Jagtap, Anna Liao, Jeremy Ng, Al Brzeczko, and Alan Guisewise, who have all contributed greatly to advance the project to the current stage.

My stay in the Robotics Institute would not be half as enjoyable without the friendship of Chieh-Chih (Bob) Wang, Jing Xiao, Wen-Chieh (Steve) Lin, Anthony Gallagher, Fernando Alfaro, Curt Bererton, Gabe Brisson, Stella Yu, and Jinxiang Chai. I

also want to mention my Singaporean friends in Carnegie Mellon, the social and self-help network has been a very important and integral part of my life in the last few years.

Last but not least, my ever supportive family back home, for all their love and the belief that their son and brother can accomplish something great. And none of the above truly matters without the love and understanding of my wife Vivian. The final few months leading to my defense have been agonizing and unbelievably tough, I cannot imagine how things would be without her faith and care, for which I feel exceedingly indebted.

Table of Content

Abstract.....	iii
Acknowledgement	v
Table of Content	vii
List of Figures.....	xi
List of Tables	xvii
List of Tables	xvii
Chapter 1 Introduction	1
1.1 Background.....	1
1.2 Approaches to Enhance Accuracy in Microsurgery	3
1.3 Scope.....	5
1.4 Technical Challenges.....	5
1.5 Objectives	6
1.6 Organization.....	7
References.....	7
Chapter 2 Sensing System Design.....	9
2.1 Introduction.....	9
2.1.1 Performance Specifications	9
2.1.2 Externally vs Internally Referenced Sensors	10
2.1.3 Conventional vs All-Accelerometer Inertial Measurement Unit.....	14
2.2 Magnetometer-Aided All-Accelerometer Inertial Measurement Unit.....	15
2.3 Sensing Kinematics.....	18
2.3.1 Differential Sensing Kinematics.....	18
2.3.2 Orientation from Vectors Observation.....	23
2.3.3 Translational Displacement	25
2.4 Error Variance Analysis of IMU Designs.....	26
2.4.1 Angular Velocity Error Variance.....	27
2.4.2 Translational Displacement Error Variance.....	29
2.4.3 Discussion.....	30
References.....	32
Chapter 3 Measurement Model.....	33
3.1 Introduction.....	33
3.2 Physical Model.....	34
3.2.1 Accelerometer.....	34
3.2.2 Magnetometer	50
3.3 Stochastic Model.....	52
3.3.1 Accelerometer.....	52
3.3.2 Magnetometer	55

References.....	57
Chapter 4 Motion Tracking.....	59
4.1 Introduction.....	59
4.2 Deterministic Error Compensation.....	60
4.2.1 Motion Sensing Experiment - Translation.....	61
4.2.2 Motion Sensing Experiment – Static Orientation.....	63
4.2.3 Discussion.....	66
4.3 Sensor Fusion with Kalman Filtering.....	68
4.3.1 Kalman Filter vs Complementary Filter.....	68
4.3.2 Human Motion Tracking.....	70
4.4 Augmented State Quaternion-Based Kalman Filtering.....	72
4.4.1 Orientation Representation.....	72
4.4.2 Angular Velocity Bias Vector.....	74
4.4.3 The Process Model.....	75
4.4.4 The State Measurement.....	78
4.4.5 Kalman Filtering.....	80
References.....	81
Chapter 5 Modeling and Filtering of Erroneous Motion.....	83
5.1 Introduction.....	83
5.2 Tremor Filtering.....	85
5.2.1 Weighted-Frequency Fourier Linear Combiner (WFLC).....	85
5.2.2 WFLC and Inertial Sensor Drift.....	89
5.3 Non-Tremulous Error Filtering.....	90
5.3.1 Introduction.....	90
5.3.2 Cascade Neural Network with Extended Kalman Filtering.....	91
5.3.3 Experimental Methods.....	92
5.3.4 Results.....	93
5.3.5 Discussion.....	96
References.....	97
Chapter 6 Manipulator Design.....	99
6.1 Introduction.....	99
6.1.1 System Requirements.....	99
6.2 Design of Mechanism.....	100
6.3 Inverse Kinematics.....	104
6.4 Forward Kinematics & Workspace Analysis.....	112
References.....	114
Chapter 7 Control of Manipulator System.....	115
7.1 Piezoelectric Actuator Hysteresis.....	115
7.2 Prandtl-Ishlinskii (PI) Hysteresis Model.....	119
7.2.1 Prandtl-Ishlinskii (PI) Operator.....	120
7.2.2 Modified Prandtl-Ishlinskii (PI) Operator.....	122

7.2.3	Parameter Identification.....	123
7.2.4	Inverse Modified Prandtl-Ishlinskii (PI) Operator.....	124
7.3	Rate-Dependent Prandtl-Ishlinskii (PI) Hysteresis Model	126
7.3.1	Rate-dependent Hysteresis Slope.....	126
7.3.2	Rate-dependent Model Identification	127
7.3.3	Rate-dependent Modified Prandtl-Ishlinskii Operator.....	128
7.4	Motion Tracking Experiments – Piezoelectric Actuator	129
7.5	Discussions	136
	References.....	138
Chapter 8 Micron Instrument.....		141
8.1	Introduction.....	141
8.2	System Integration	141
8.2.1	Micron Instrument	141
8.2.2	System Hardware and Software.....	143
	References.....	145
Chapter 9 Conclusion		147
9.1	Conclusion	147
9.2	Contributions.....	149
9.3	Future Work.....	149

List of Figures

Fig. 1.1 Error compensation in microsurgery with Micron, an active handheld instrument that senses its own motion, distinguish between the desired and undesired component of the sensed motion, and manipulate its own tip in an equal but opposite motion to cancel the erroneous motion.	4
Fig. 2.1 Design of the magnetometer-aided inertial measurement unit.	16
Fig. 2.2 Pictures of the implemented magnetometer-aided inertial measurement unit. ..	17
Fig. 2.3 Kinematic representation of the all-accelerometer IMU.	19
Fig. 2.4 Comparison of the joint distributions of ω_x and ω_y estimates of an all-gyro ($P(\omega_{Gx}, \omega_{Gy})$) and an all-accelerometer ($P(\omega_{Ax}, \omega_{Ay})$) design. Higher density regions at the centers represent higher probability.	31
Fig. 3.1 Integration drift of an accelerometer. The plots show the output of an accelerometer at rest. A small offset error in the acceleration sensing grows linearly after the first integration and becomes quadratic in the displacement domain after the second integration.	35
Fig. 3.2 Construction schematic of ADXL-203 dual-axis accelerometer.	37
Fig. 3.3 Definition of angles α and β	39
Fig. 3.4 Hysteretic paths of the accelerometer output when rotated in the gravity plane. When rotated clockwise and anticlockwise through 360° , it follows paths $A^+ - B^-$ and $B^+ - A^-$ respectively; when rotated clockwise and anticlockwise through 180° and back to 0° , it follows paths $A^+ - A^-$ and $B^+ - B^-$ respectively.	40
Fig. 3.5 Hysteretic paths of the accelerometer rotated through 360° in a plane inclined at angle $\alpha = \pm 30^\circ, \pm 90^\circ, \text{ and } \pm 150^\circ$	40
Fig. 3.6 The residual x-accelerometer output versus the y-accelerometer output. The relationship is quadratic.	42
Fig. 3.7 The measured x-accelerometer output, the linear model $f_x(A_x)$, and the model $f_x(A_x) + g_x(V_y)$	43
Fig. 3.8 The residual x-accelerometer output versus the z-acceleration. The relationship is quadratic.	44

Fig. 3.9 The measured x-accelerometer output at $\alpha = \pm 30^\circ$ and the proposed physical model $f_x(A_x) + g_x(V_y) + h_x(V_z)$. The $Z_{\pm 90}$ plot is added for comparison.	44
Fig. 3.10 Scale Factor versus z-acceleration. The relationship is quadratic.	45
Fig. 3.11 Bias and scale factor temperature drift. Steady state is reached at approximately the 7 th hour in this experiment.	46
Fig. 3.12 Orientation error of an accelerometer.	48
Fig. 3.13 Experimental setup for physical model parameters identification. The two orthogonal precision motorized rotary stages produce α and β rotations to the accelerometers and the triad of magnetometers. The extension arm is to place the sensors a safety distance away from the magnetic field generated by the motors.	50
Fig. 3.14 Output of a motionless accelerometer collected over 12 hours at 1 Hz. The black circles are the mean of voltage output every 30 min.	53
Fig. 3.15 Root Allan variance plot of output of a motionless accelerometer collected over 12 hours at 1 Hz. The dominant noise components are a strong velocity random walk (slope $-\frac{1}{2}$), a weak acceleration random walk (slope $+\frac{1}{2}$), and trend (slope +1).	53
Fig. 3.16 Root Allan variance plot of the final 4 hours recordings of a motionless accelerometer collected over 12 hours at 1 Hz. The dominant noise components are a strong velocity random walk (slope $-\frac{1}{2}$) and a weak acceleration random walk (slope $+\frac{1}{2}$). The trend component (slope +1) disappears.	54
Fig. 3.17 Root Allan variance plot of recordings of a motionless accelerometer collected over 30 min at 1000 Hz, after steady state has been reached. The dominant noise component is only velocity random walk (slope $-\frac{1}{2}$).	55
Fig. 3.18 Output of a motionless magnetometer collected over 12 hours at 1 Hz. The black circles are the mean of voltage output every 30 min.	56
Fig. 3.19 Root Allan variance plot of output of a motionless magnetometer collected over 12 hours at 1 Hz. The dominant noise components are a strong velocity random walk (slope $-\frac{1}{2}$), a weak acceleration random walk (slope $+\frac{1}{2}$), and trend (slope +1).	56
Fig. 3.20 Root Allan variance plot of recordings of a motionless magnetometer collected over 30 min at 1000 Hz, after steady state has been reached. The dominant noise component is only velocity random walk (slope $-\frac{1}{2}$).	57

Fig. 4.1 Real-time motion tracking of the intraocular tip position with the magnetometer-aided all-accelerometer inertial measurement unit.	60
Fig. 4.2 Translational motion sensing experimental setup.	61
Fig. 4.3 Results of (a) the manufacturer’s recommended linear model vs (b) the proposed physical model. The dotted line is the computed acceleration from the interferometer. The solid line is the measured acceleration from the x-accelerometer.	63
Fig. 4.4 Static orientation experimental setup.	64
Fig. 4.5 Errors in tilt angle estimation using the proposed physical model and the manufacturer. In this experiment, $\beta = 0^\circ$ and α is rotated from -180° to 180° at 30° interval.	65
Fig. 4.6 Errors in tilt angle estimation using the proposed physical model and the manufacturer. In this experiment, $\alpha = 0^\circ$ and β is rotated from -180° to 180° at 30° interval.	65
Fig. 4.7 Errors in tilt angle estimation using the proposed physical model and the manufacturer. In this experiment, α and β are rotated simultaneously from -180° to 180° at 30° interval.	66
Fig. 4.8 The action of each component of the physical model on the accelerometer response plot. The line ‘Linear Model’ is obtained from the manufacturer’s recommended calibration method. The bias model $B_x(V_y, V_z)$ shifts the linear model parallel towards the actual response and the model $SF_x(V_z)$ alters the gradient and result in the line ‘ $B_x + SF_x$ ’. The misalignment model makes the final shift to the line ‘Actual Response’	67
Fig. 4.9 A complementary Kalman filter.	69
Fig. 4.10 X and Y are the error-free orthogonal axes. An erroneous rotation δ is introduced by numerical integration. (a) Drift error. The X' and Y' axes remain mutually orthogonal and unit length; (b) Scale and skew errors. The X' and Y' axes are not unit length and not orthogonal.	73
Fig. 5.1 An adaptive noise canceller.	85
Fig. 5.2 The Fourier linear combiner. The FLC adaptively creates a dynamic Fourier series model of any periodic input. The reference vector consists of harmonic sines and cosines at a fixed fundamental frequency. These are weighted by the Fourier coefficient vector w_k and summed to obtain a truncated Fourier series model of the input s_k	86

Fig. 5.3 The weighted-frequency Fourier linear combiner (WFLC). The system maintains a running sum of the frequency weight ω_0 . Harmonic sines and cosines of this quantity are taken (vector x_k), weighted by the Fourier coefficient vector w_k , and summed to provide a truncated Fourier series model of the input s_k	88
Fig. 5.4 The WFLC-FLC combination. The highpass or bandpass IIR prefilter filters the input signal s_k to attenuate the voluntary motion. WFLC operates on the filtered signal s_k' , estimates the tremor frequency ω_0 and feed it into FLC. With known frequency, FLC adapts and locks-on to the tremor amplitude and phase.....	88
Fig. 5.5 Result of one dimensional active physiological tremor canceling experiment. The dotted line is a recording of the physiological tremor of a microsurgeon. The solid line shows an offline estimate of voluntary motion, obtained via zero-phase lowpass postfiltering of the uncompensated motion. The fine dotted line shows the motion during WFLC compensation.	89
Fig. 5.6 Diagram of the cascade neural network architecture. The diagram shows a network with three hidden nodes. As each hidden node is added, it is connected to the input and output nodes, as well as each of the preceding hidden nodes.	91
Fig. 5.7 The line 'Target' is the pseudo-voluntary motion, generated by lowpass filtering white noise at 1 Hz cutoff frequency. The line 'Input' represents the network input for the test, obtained by adding one of the recorded hand motion error files to the pseudo-voluntary motion. The line 'Output' indicates the filtered version of the data.	95
Fig. 6.1 The intraocular shaft manipulator is located in the front end of the instrument.....	101
Fig. 6.2 The manipulator mechanism.	102
Fig. 6.3 The three stroke amplification levers crisscross one another at different heights to avoid collision of the linkages.....	103
Fig. 6.4 Kinematic chain of the 3 DOF intraocular shaft parallel manipulator.	105
Fig. 6.5 Kinematic chain of the 3 DOF intraocular shaft parallel manipulator, represented in Lee and Shah [Lee 88] convention.	105
Fig. 6.6 Relationship between the link length L_i and the piezoelectric actuator length λ_i	111

Fig. 6.7 Comparison of the intraocular shaft manipulator workspace versus the tremor space. Tremor space is approximated by a $\varnothing 50 \mu\text{m}$ sphere.....	113
Fig. 7.1 Hysteresis of piezoelectric material is rate-dependent. The plots show the response of a piezoelectric actuator at two different driving frequencies.....	116
Fig. 7.2 The response of the Polytec PI NanoCube™ P-611 three-axis micropositioner (Polytec PI, Inc., Germany) with 10 Hz p-p sinusoidal input. The closed-loop controller exhibits a system response that resembles that of a low-pass filter, with attenuated amplitude and phase lag.....	117
Fig. 7.3 Hysteresis is evident in the response of the Polytec PI NanoCube™ P-611 three-axis micropositioner (Polytec PI, Inc., Germany) with 10 Hz p-p sinusoidal input.	117
Fig. 7.4 Piezoelectric actuator linearization with inverse feedforward control strategy.....	118
Fig. 7.5 The rate-independent generalized backlash operator is characterized by the threshold or backlash magnitude, r , and the weight or backlash gain, w_h	121
Fig. 7.6 The PI hysteresis model with $n = 4$. The hysteresis model is characterized by the initial loading curve. The piecewise linear curve is defined by the equally spaced threshold values \bar{r} and the sum of the weight values \bar{w}_h	122
Fig. 7.7 (a) The one-sided dead-zone operator is characterized by the threshold, d , and the gain w_s	123
Fig. 7.8 The lighter solid lines are the measured piezoelectric actuator response, subjected to a 10 Hz, $12.5 \mu\text{m}$ p-p sinusoidal control input. The dark dotted line is the identified modified PI hysteresis model.	124
Fig. 7.9 An inverse feedforward controller. For a desired periodic actuator output $x(t)$, the inverse modified PI operator Γ^{-1} transforms it into a control input $u(t)$ that would produce a response $z(t)$ in the hysteretic system that closely resembles $x(t)$. This produces an equivalent control system with identity mapping between the desired output and the actual actuator response.....	125
Fig. 7.10 The darker thick line is the modified PI hysteresis model, Γ . The inverse modified PI hysteresis model Γ^{-1} , represented by the lighter thin line, is the mirror image of the hysteresis model about the 45° line.	126
Fig. 7.11 Plot of the hysteresis slopes W_{hi} , $n = 9$, vs. actuation rate, $\dot{x}(t)$. Since the actuation rate is always slow at the turning points of a sinusoid, the first two sum of weights, W_{h0} and W_{h1} are modeled only up till $200 \mu\text{m/s}$	128

Fig. 7.12 Rmse and maximum errors of the rate-independent and rate-dependent controllers in tracking 12.5 μ m p-p stationary sinusoids at different frequencies. The rate-independent controller is based on the modified PI hysteresis model identified at the same 10Hz, 12.5 μ m p-p sinusoid.....	131
Fig. 7.13 Experimental open-loop tracking results of stationary 12.5 μ m p-p sinusoids at 10 Hz. The rate-independent controller is based on the modified PI hysteresis model identified at the same 10Hz, 12.5 μ m p-p sinusoid.	132
Fig. 7.14 Experimental open-loop tracking results of a multi-frequency, non-stationary dynamic motion profile. The motion profile is made up of superimposed modulated 1, 10, and 19 Hz sinusoids with time-varying amplitudes. The rate-independent controller is based on the modified PI hysteresis model identified at the same 10Hz, 12.5 μ m p-p sinusoid. Transient error is observed for the rate-independent controller in the first 2 seconds.(a) Without compensation. (b) Rate-independent controller. (c) Rate-dependent controller.	134
Fig. 7.15 Experimental open-loop tracking results of a multi-frequency, non-stationary dynamic motion profile. The motion profile is made up of superimposed modulated 1, 10, and 19 Hz sinusoids with time-varying amplitudes. The rate-independent controller is based on the modified PI hysteresis model identified at the same 10Hz, 12.5 μ m p-p sinusoid. Transient error is observed for the rate-independent controller in the first 2 seconds.	135
Fig. 8.1 The Micron instrument.	142
Fig. 8.2 The manipulator system is designed to shift the CG of the instrument to within the circle formed by the grasping index finger and the thumb.	143
Fig. 8.3 Overview of Micron system. Micron subsystems are denoted by bold font, peripheral hardware and software systems are denoted by italic and regular font respectively.	145

List of Tables

Table 2.1 Summary of selected commercial externally referenced motion sensing system.....	11
Table 2.2 Summary of a few selected commercial rate gyroscopes.....	12
Table 2.3 Summary of a few selected low-g commercial accelerometers.....	12
Table 2.4 Comparison of error variances and standard deviations of the angular velocity and linear acceleration sensed by a conventional three gyros three accelerometers (3G-3A) IMU versus the proposed all-accelerometer (6A) IMU.....	30
Table 4.1 Rmse and maximum error of the tilt estimation experiments.....	66
Table 5.1 Performance of the neural network and the network architecture that produces the best result.	94
Table 5.2 Percentage rmse reduction comparison between ours (NN_{vm}) and Riviere/Khosla's (NN_{err}) method	94
Table 5.3 Numbers in parenthesis are the test result of the file using the neural network (NN) trained on the same surgeon, to provide comparison. Numbers in bold are tests that equal to or outperform the % rmse reduction of the neural network trained on the same surgeon. Numbers in italics are tests that produce a rmse worse than the raw rmse.....	96
Table 6.1 Specifications of Micron manipulator system.	104
Table 7.1 Measured performance of the Rate-independent and Rate-dependent Inverse Feedforward Controllers in Tracking 12.5 μ m p-p Stationary Sinusoids.	130
Table 7.2 Measured Performance of the Rate-independent and Rate-dependent Inverse Feedforward Controllers in Tracking Multi-Frequency (1, 10, and 19 Hz) Non-Stationary Sinusoids.	133
Table 7.3 Measured Performance of the Rate-independent and Rate-dependent Inverse Feedforward Controllers in Tracking Recorded Real Tremor of Ophthalmologic Surgeons.....	136

Chapter 1

Introduction

1.1 Background

Humans have intrinsic limitations in manual positioning accuracy. These limitations are consequences of small involuntary movements that are inherent in normal hand motion. This manual imprecision hinders micromanipulation tasks, where the intended motion and the erroneous movements have the same order of magnitude, or in engineering terms: the signal-to-noise ratio is low. The influence it has on the performance of such tasks may range from significantly hampered, e.g. microsurgery [Patkin 77], cell manipulation in bio-tech industry, to moderate or slight degradation, e.g. gun-sight or military handheld tracking equipment and consumer video cameras.

The most familiar type of erroneous movement affecting micromanipulation is hand tremor [Harwell 83]. Tremor is defined as any involuntary, approximately rhythmic, and roughly sinusoidal movement [Elble 90]. Physiological tremor is a type of tremor that is inherent in the movement of healthy subjects. Hand tremor contains components that arise from neuromuscular activity, cardioballistic thrust, and the passive resonances in the tissues of the hand that are initiated by both of these activities [Wade 82]. Because of its biological origins, it is believed that some components of hand tremor are stochastic in nature [Elble 90], and analysis of the dynamics of hand tremor time series reveals that

physiological tremor can be best described by a linear stochastic process [Gantert 91]. The effect of tremor on the overall motion is additive, i.e. it is superimposed upon the voluntary motion [Riley 87]. Amplitude modulation of tremor is often quite rapid [Riviere 95], while tremor frequency changes in a slower rate.

In ophthalmological microsurgery, the significant component of physiological tremor is found to be an oscillation at 8-12 Hz whose frequency is independent of the mechanical properties of the hand and arm [Elble 90]. The measured tool tip oscillation during vitreoretinal microsurgery can be as large as 50 μm peak-to-peak (p-p) in each of the principal axes [Hunter 93]. It has also been reported that the vector magnitude of physiological tremor during the most delicate part of the procedure is measured to be 38 μm rms [Singh 02].

Measurements of the hand motion of surgeons have also shown the existence of other sources of non-tremulous erroneous motion such as jerk (i.e., normal myoclonus) and drift. These components are often larger than physiological tremor [Riviere 97].

These inherent limitations complicate many delicate surgical procedures, and make some types of intervention impossible. One of the common procedures in vitreoretinal microsurgery includes removal of membranes as thin as 20 μm from the retina. Another example is the treatment of retinal vein occlusions by injection of anticoagulants using intraocular cannulation [Tang 00; Weiss 98]. This procedure holds the potential of curing a major cause of blindness, estimated to affect 1.6% of persons aged 49 years or older [Mitchell 96]. However, it is generally considered infeasible as it is practically impossible to insert a micropipette into a $\sim\text{Ø}100$ μm retinal vein and deliver the drug without risking tearing the vessel apart. Other examples include procedures in neurological surgery and microvascular surgery.

The high level of manual accuracy demanded by microsurgery restricts the number of qualified surgeons. The fact that human hand stability deteriorates with age further exacerbates the situation. Even for microsurgeons in their prime years, fatigue, alcohol and caffeine consumption, and other factors affect the manual stability [Humayun 97; Lakie 94]. Besides governing their sleep, and alcohol and caffeine consumptions prior to the surgery, some surgeons take beta blockers, which have been shown to have an

attenuating effect [Elmar 98]. There is some degree of consensus among vitreoretinal microsurgeons that instrument-tip positioning accuracy of 10 μm is desired [Charles 96].

1.2 Approaches to Enhance Accuracy in Microsurgery

Efforts to provide solutions to the problem of increasing accuracy in microsurgery have included the use of telerobotic technology [Hunter 93; Schenker 95], where a robotic arm replaces the unstable human hand. Commercially available telerobotic surgery systems include the Zeus® surgical system from Computer Motion, Inc. and the da Vinci® surgical system from Intuitive Surgical, Inc.¹. This approach allows filtering of erroneous motion between master and slave manipulators via motion scaling, which effectively increase the signal-to-noise ratio. This approach, though effective, is costly and obtrusive from the point of view of the surgeon, who is accustomed to treating the patient with his own hands. A robotic arm with the requisite workspace for the application also introduces significant safety and liability issues.

Taylor *et al.* have used a "steady hand" approach, in which a robot and a surgeon directly manipulate the same tool [Taylor 99], with the robot having high stiffness, and moving along with only those components of the manual input force that are deemed desirable. While this system cannot scale input motion, it has advantages in terms of cost and likelihood of user acceptance. Moreover, it lends the surgeon a "third hand," holding a tool in position while the surgeon performs other tasks with his own two hands.

In order to further reduce cost, and to maximize ease of use, user acceptance, and compatibility with current surgical practice, we are implementing accuracy enhancement within a completely hand-held tool, keeping the instrument size and weight as close as possible to those of existing passive instruments. This device should sense its own motion, estimate the undesired component of the sensed motion, and manipulate its own tip in real time to nullify the erroneous motion as shown in Fig. 1.1.

¹ Intuitive Surgical, Inc. and Computer Motion, Inc. have merged in June 30, 2003. The new company continues to use the name Intuitive Surgical, Inc.

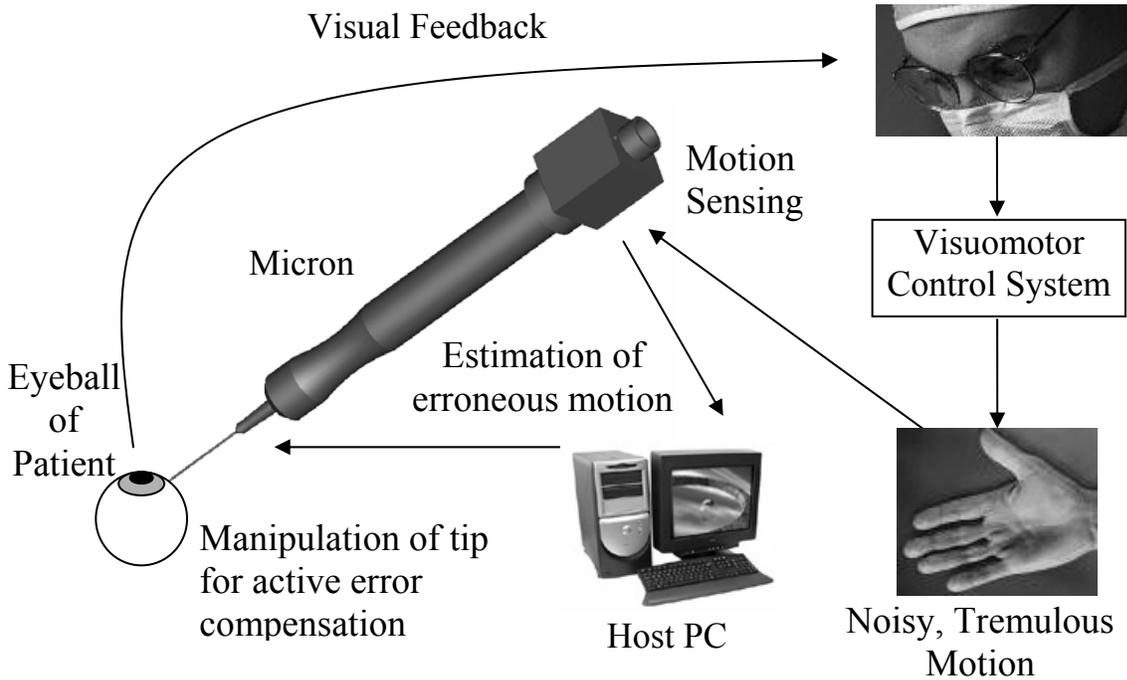


Fig. 1.1 Error compensation in microsurgery with Micron, an active handheld instrument that senses its own motion, distinguishes between the desired and undesired component of the sensed motion, and manipulates its own tip in an equal but opposite motion to cancel the erroneous motion.

Besides the cost and non-obtrusiveness advantage, one important motivation of this approach is its innate ‘fail-safe’ characteristic. From a safety standpoint, the telerobotic and ‘steady hand’ solutions are capable of doing catastrophic damages should any malfunction happen. Micron, on the other hand, with its limited manipulator workspace is not much more dangerous than an existing passive microsurgical tool. Furthermore, in the event of extraordinary medical situations or system glitches where surgeon manual intervention is necessary, the bulky robotic systems in the other approaches might have to be removed to get out of the way of the surgery team. There is no such problem in our approach since the surgery is being performed as it usually does; and even if system failure does occur, one may simply switch off Micron to reduce it to a normal passive instrument.

1.3 Scope

Micron is a complete robotic system that is made up of three distinct sub-systems: the sensing system, filtering system, and manipulation system.

The main focus of this dissertation is on the research and development of the sensing and the manipulation system. It describes in detail the hardware and software components of both the systems, and the implementation of the complete Micron system with laboratory and clinical experiments to test the performance of the instrument.

The filtering system is a software system residing in the host computer that distinguishes between desired and undesired component of the sensed motion. There are, in broad classification, two types of undesired motion components in micromanipulation tasks, the well studied physiological tremor and lesser understood non-tremulous error component such as myoclonic jerk and drift etc. The tremor filtering algorithm used in Micron is based on the dissertation work of Riviere [Riviere 95]. No new research has been performed on this part. While this dissertation focuses on tremor compensation, an algorithm to model and distinguish the non-tremulous errors is proposed. Simulation results of this algorithm are presented, but it is not incorporated into the current version of the Micron system.

While the targeted application of this dissertation is in microsurgery, the principles of the approach and the theory of operation is universal and can be extended for other micromanipulation tasks like cell manipulation in biotech industry and gun-sights or handheld military equipment etc.

1.4 Technical Challenges

Before the objectives of this dissertation work can be succinctly and sufficiently defined, it is important to examine the specific surgical needs and the technical challenges associated with this approach of implementing real-time error compensation in a completely handheld instrument:

- (i) Real-time Issues

Physiological tremor has a distinctive higher frequency band than voluntary motion, which is typically less than 1 Hz in most micromanipulation tasks. However, all classical frequency-selective filters create a phase change in the filtered signal, and the output of the filter would be a time delayed version of the actual physical tremulous motion. In other words, any attempt to cancel this motion might actually worsen the situation.

(ii) Accuracy Issues

To accurately sense and cancel physiological tremor is pushing the limit of the current state of technology. Firstly, sensing motion as small as tens of micrometer is approaching the sensing resolution limit or the noise floor of most motion sensors. High order filtering to reduce the noise floor is not feasible for reasons discussed in the previous point. Secondly, controlling the manipulator tooltip to accurately track a motion of about 10 Hz is beyond the system bandwidth of many actuators.

(iii) Implementation Issues

One of the major motivations of this approach of error compensation is its non-obtrusiveness to the surgeon performing the surgery. It is then paramount to design the functionalities into a handheld instrument without compromising the grasping ergonomic in terms of size and weight. The size and weight constraints greatly restrict the design options and give rise to many engineering difficulties.

1.5 Objectives

It is apparent that the real-time issues are associated with the filtering system, while the accuracy and implementation issues are the major technical challenges of the sensing and manipulation systems. Translating the technical challenges into specific engineering and research goals, the objectives of the dissertation are as follow:

- (i) To develop a six DOF motion sensing system that would enable accurate real-time sensing of the tremulous component in the instrument motion during micromanipulation tasks;

- (ii) To develop a intraocular shaft manipulator system that has adequate bandwidth and tracking precision to cancel physiological hand tremor effectively ;
- (iii) To implement the sensing and the manipulator systems within a completely handheld instrument.

1.6 Organization

The remaining chapters of the dissertation are organized as follow. Chapter 2 describes the features of the proposed magnetometer aided all-accelerometer inertial measurement unit (IMU) and demonstrates the superiority of this novel design by comparing the error variance to that of a conventional IMU design. Chapter 3 develops a detailed measurement model to account for the sensor errors. Based on the sensor characteristics modeled in Chapter 3, Chapter 4 discusses error compensation and sensor fusion via Kalman filtering. Chapter 5 reviews the weighted-frequency Fourier linear combiner (WFLC) algorithm for tremor estimation, and proposes a neural networks approach to estimate and compensate the non-tremulous component of the undesired motion. Chapter 6 presents the design of the manipulator mechanism, the forward and inverse kinematics, and a workspace analysis of the manipulator tool-tip. Chapter 7 discusses the hysteresis problem of the piezoelectric actuators, and the theory and implementation of a controller based on the inverse hysteresis model. Chapter 8 puts together the three sub-systems and discusses the issues related to the integrated Micron handheld instrument. And finally Chapter 9 summarizes the contributions of this dissertation and gives an account on the future work.

References

- [Charles 96] S. Charles, Dexterity enhancement for surgery. In: R.H. Taylor et al. (eds.): *Computer Integrated Surgery: Technology and Clinical Applications*. MIT Press, Cambridge'96 467-471.
- [Elble 90] R. J. Elble and W. C. Koller, Tremor. Baltimore: Johns Hopkins, 1990.
- [Elman 98] Elman, M. J., Sugar, J., Fiscella, R., Deutsch, T. A., Noth J., Nyberg, M., Packo, K., and Anderson, R. J., "The effect of propranolol versus placebo on

- resident surgical performance," *Trans. Am. Ophthalmol. Soc.*, vol. 96, pp. 283-291, 1998.
- [Gantert 92] C. Gantert, J. Honerkamp, J. Timmer, "Analyzing the dynamics of hand tremor time series," *Biol. Cybern.* Vol. 66, pp. 479-484, 1992
- [Humayun 97] Humayun, M. U., Rader, R. S., Pieramici, D. J., Awh, C. C., de Juan, E., Jr., "Quantitative measurement of the effects of caffeine and propranolol on surgeon hand tremor", *Archives of Ophthalmology*, Vol. 115, pp. 371, 1997
- [Hunter 93] I. W. Hunter, T. D. Doukoglou, S. R. Lafontaine, P. G. Charette, L. A. Jones, M. A. Sagar, G. D. Mallinson, and P. J. Hunter, "A teleoperated microsurgical robot and associated virtual environment for eye surgery," *Presence*, 2:265-280, 1993.
- [Lakie 94] Lakie, M., Frymann, K., Villagra, F., Jakeman, P., "The effect of alcohol on physiological tremor," *Experimental Physiology*, Vol. 79, pp. 273-276, 1994.
- [Mitchell 96] Mitchell, P., Smith, W., and Chang, A., "Prevalence and associations of retinal vein occlusion in Australia: the Blue Mountains Eye Study," *Arch. Ophthalmol.*, vol. 114, pp. 1243-1247, 1996.
- [Patkin 77] M. Patkin, "Ergonomics applied to the practice of microsurgery," *Aust. N. Z. J. Surg.*, vol. 47, pp. 320-329, June 1977.
- [Riley 87] P.O. Riley and M. Rosen, "Evaluating manual control devices for those with tremor disability," *J. Rehabilitation Research Dev.*, vol24, pp.99-110, 1987
- [Riviere 95] C. N. Riviere, "Adaptive suppression of tremor for improved human-machine control," *Ph.D. dissertation*, Johns Hopkins Univ, Baltimore, MD, 1995.
- [Riviere 97] C. N. Riviere, R. S. Rader, and P. K. Khosla, "Characteristics of hand motion of eye surgeons," *Proc. 19th Annu. Conf. IEEE Eng. Med. Biol. Soc.*, Chicago, 1997.
- [Schenker 95] P. S. Schenker, E. C. Barlow, C. D. Boswell, H. Das, S. Lee, T. R. Ohm, E. D. Paljug, G. Rodriguez, and S. T. 2, "Development of a telemanipulator for dexterity enhanced microsurgery," *Proc. 2nd Intl. Symp. Medical. Robotics and Comp. Assisted Surgery*, pp.81-88, 1995.
- [Singh 02] S. P. N. Singh and C. N. Riviere, "Physiological tremor amplitude during retinal microsurgery," in *Proc. IEEE 28th Annu. Northeast Bioeng. Conf.*, pp. 171-172, 2002.
- [Tang 00] Tang, W. M. and Han, D. P., "A study of surgical approaches to retinal vascular occlusions." *Arch. Ophthalmol.* Vol. 118, pp.138-143, 2000.
- [Taylor 99] R. Taylor, P. Jensen, L. Whitcomb, A. Barnes, R. Kumar, D. Stoianovici, P. Gupta, Z. Wang, E. de Juan, and L. Kavoussi, "A steady-hand robotic system for microsurgical augmentation," in: C. Taylor, A. Colchester (eds.), *Medical Image Computing and Computer-Assisted Intervention - MICCAI'99*, Springer, Berlin, pp. 1031-1041.
- [Wade 82] P. Wade, M. A. Gresty, L.J. Findley, "A Normative Study of Postural Tremor of the Hand," *Arch Neurol*, Vol. 39, pp.358-362, 1982
- [Weiss 98] Weiss, J. N., "Treatment of central retinal vein occlusion by injection of tissue plasminogen activator into a retinal vein," *Am. J. Ophthalmol.* Vol. 126, pp.142-144, 1998.

Chapter 2

Sensing System Design

2.1 Introduction

This chapter presents the design of the intelligent handheld microsurgical instrument's on-board sensing system. A survey of several state-of-the-art sensing systems with different sensing modalities is first conducted; follow by a description of the proposed magnetometer-aided all-accelerometer inertial measurement unit (IMU) with justifications of the design choices. Sections 2.3 and 2.4 present the sensing kinematics to retrieve the complete six degrees of freedom motion information of the instrument. The last section shows analytically the superiority of the proposed design to other conventional designs using error variance analysis.

2.1.1 Performance Specifications

Physiological hand tremor of trained surgeons during vitreoretinal microsurgery has been reported to be in the frequency band of 8 – 12 Hz and measure up to 50 μm p-p in each principal direction. To effectively reduce the erroneous tremulous motion down to the accuracy level that the micronsurgeons desired, i.e. $\sim 10 \mu\text{m}$, the sensing system would have to meet the required performance specifications.

First of all, the sensing system should have a sensing bandwidth higher than the tremor frequency band. It is also desirable to have the sensing resolution and the rms sensing accuracy to be about an order smaller than the tremor amplitude. Also, we need the system to run at a sampling rate fast enough to detect a change equivalent to the resolution of the system. Based on an assumed sinusoidal tremor profile of 10 Hz, 50 μm p-p, we define the following performance specifications:

- (i) Sensing resolution $\sim 5 \mu\text{m}$, $\sim 10^{-3^\circ}$;
- (ii) Sensing accuracy $\sim 5 \mu\text{m}$ rms, $\sim 10^{-3^\circ}$ rms;
- (iii) Bandwidth ≥ 13 Hz;
- (iv) Sampling rate ≥ 500 Hz.

For the sensing system to be deployed in an operating room, its sensing modality must be non-obtrusive to the surgeon and free from interference with other electronic equipment. Other critical implementation considerations would be the size and weight of the sensing system, since it is to be fitted into a handheld device.

2.1.2 Externally vs Internally Referenced Sensors

In broad classification, there are two different categories of motion sensing technology, externally and internally referenced sensors. Externally referenced sensors require interactions between on-board components and external active sources or receivers, whereas internally referenced sensors are self-contained sensing modalities.

Sensing modalities of externally referenced sensors include infrared, radio frequency, visible light, ultrasound, electromagnetic wave, etc. Examples of externally referenced motion sensors for indoor applications include the Optotrak, Polaris, and Aurora systems from Northern Digital Inc., the Fastrak and Liberty systems from Polhemus Inc., and the IS-900 system based on the Constellation technology from Intersense, Inc. The sensing modality, size, weight, and performance specifications of some of the selected commercial externally referenced motion sensing systems are summarized in Table 2.1.

Table 2.1 Summary of selected commercial externally referenced motion sensing system.

System	Optotrak (Northern Digital, Inc.)	Polaris (Northern Digital, Inc.)	Aurora (Northern Digital, Inc.)	Fastrak (Polhemus, Inc.)	Liberty (Polhemus, Inc.)	Constellation IS-900 (Intersense, Inc.)	
Modality	Infrared	Infrared	Electro- magnetic	Electro- magnetic	Electro- magnetic	Ultrasound + Inertial	
Sampling Rate (Hz)	145	60	22	120	240	180	
Accuracy (rms)	deg	NA	NA	0.8	0.15	0.15	2.5
	μm	100	350	900	760	760	2000
Resolution	deg	NA	NA	NA	0.3 (@ 1')	0.00012 (@ 1')	0.05
	μm	10	NA	NA	60 (@ 1')	38 (@ 1')	750
Weight ¹ (g)	< 5	< 10	0.7	NA	23	44	
Size ¹ (mm)	$\varnothing 4 \times 2$ per unit	$\varnothing 6$ per unit	8 x $\varnothing 0.8$	22.9 x 27.9 x 15.2	22.9 x 27.9 x 15.2	100 x 80 x 44	

¹The on-board unit.

Internally referenced sensors sense the inertial quantities of a moving body, such as acceleration, angular velocity, and angular acceleration. The motion of a body in three dimensional Euclidean space is usually measured by a six degrees-of-freedom (DOF) inertial measurement unit (IMU) consists of three accelerometers and three rate gyroscopes. With well-defined initial conditions, an IMU measures the basic quantities that define the motion of a body – velocity, position, angular rate, and orientation (or attitude in navigation terminology). Table 2.2 and 2.3 show the cost, size, weight, and performance specifications of a few selected commercially available rate gyroscopes and accelerometers respectively.

Table 2.2 Summary of a few selected commercial rate gyroscopes.

Make/Model	Honeywell / GG-1342	KVH / DSP-3000	Token / CG-L43D	Analog Devices / ADXRS150
Class	Inertial Grade	Tactical/Navigation Grade	Rate Grade	Rate Grade
Type	Ring Laser	Interferometric Fiber Optic	Vibrating Piezoelectric	MEMS
Cost Range	> \$10000	>\$5000	\$50-100	\$ 30-40
Output Noise (@ 15 Hz)	1.3 deg/h	60 deg/h	1.4 deg/s	0.75 deg/s
Bias Stability (deg/h)	0.0035	20	NA	NA
3 dB Bandwidth (Hz)	NA	50	20	40
Sampling Rate	Limited by ADC or microprocessor, usually > 1 kHz.			
Input Range (deg/s)	NA	± 375	± 90	± 150
Size (mm)	190.5 x 114.3 x 45.7	88.9 x 58.4 x 33.0	8 x 16 x 5	7 x 7 x 3.5
Weight (g)	1360	270	1	< 1

Table 2.3 Summary of a few selected low-g commercial accelerometers.

Make/Model	Crossbow / CXL02LF3	Entran / EGAS3	Kistler / 8393A2	Analog Devices / ADXL203	IC Sensors / 3150
Type	MEMS	Piezoresistive	Capacitive	Differential Capacitive	Piezoresistive
No. of Axes	3	3	3	2	1
Cost Range	\$800	> \$500	> \$500	\$20	> \$100
Output Noise (mg^1)	1.0 rms	NA	0.57 rms (@ 15 Hz)	2.25 rms (@ 15 Hz)	0.5 p-p
3 dB Bandwidth (Hz)	50	> 150	> 10	15 (User select)	200
Sampling Rate	Limited by ADC or microprocessor, usually > 1 kHz.				
Input Range (g^1)	± 2	± 5	± 2	± 1.5	± 2
Size (mm)	19.8 x 44.5 x 27.2	15 x 15 x 17	31.8 x 31.8 x 38.9	5 x 5 x 2	22.9 x 22.9 x 11.4
Weight (g)	46	NA	60	1	13

¹g is the gravitational acceleration.

It can be seen from Table 2.1 that all of the externally referenced sensing systems do not meet most of the performance specifications for our application, as specified in Section 2.1.1. Furthermore, the operation of all external referenced sensing technologies requires a line of sight or other similar provision between the on-board unit and the external stationary unit. In order to sense the motion of the instrument during microsurgery, it would inadvertently be obtrusive to the surgeon to maintain this line of sight, given the tight and constraint workspace the surgeon has to work with.

All inertial sensors listed in Table 2.2 and 2.3 satisfy the sampling rate and sensing bandwidth requirements. The technical specifications of inertial sensors are rated differently, to compare the sensing resolution and accuracy, we need to transform the required specifications to the equivalent engineering units.

Based on a sinusoidal tremor profile of 10 Hz, 50 μm p-p, we have

$$\begin{aligned}x(t) &= A \sin \omega t, \\ \dot{x}(t) &= A\omega \cos \omega t, \\ \ddot{x}(t) &= -A\omega^2 \sin \omega t.\end{aligned}\tag{2.1}$$

We further assume the distance L between the intraocular shaft tip and the instrument center of mass is known, the equivalent tremor angular velocity is thus

$$\omega(t) = \frac{\dot{x}(t)}{L}.\tag{2.2}$$

With $A = 0.025$ mm, $\omega = 2\pi f$, and $L = 100$ mm,

$$\begin{aligned}\dot{x}_{\max} &= 2A\omega = 3.14 \text{ mm/s}, \\ \ddot{x}_{\max} &= 2A\omega^2 = 197.4 \text{ mm/s}^2 = 20.1 \text{ mg} \text{ (} g = 9801 \text{ mm/s}^2\text{)}, \\ \omega_{\max} &= 1.8 \text{ }^\circ\text{/s}.\end{aligned}$$

The sensing resolution of an inertial sensor is limited by its output noise floor. All the accelerometers listed in Table 2.3 have a rms output noise that is either very close to or smaller than a tenth of \ddot{x}_{\max} . To meet the resolution criteria of 0.18 $^\circ\text{/s}$, we would need gyros with performance between rate-grade and tactical/navigation grade.

Sensing accuracy of inertial sensors is greatly handicapped by the notorious integration drift. The fast growing integration drift arises from the accumulation of integrated sensing errors. Errors in the sensed angular rate cause orientation estimate

errors to grow linearly over time ($\propto t$) after integration. On top of this, the greatest repercussion of orientation error is the fact that accelerometer also senses gravity and the attainment of effective body acceleration relies on the orientation information to remove gravity from the sensed quantity. As a result the compounded error growth rate of position estimate is cubic over time ($\propto t^3$) after two integrations. This is made worse by the fact that tremor acceleration is typically two orders of magnitude smaller than gravity, gravity residual errors considered small for other applications would therefore be intolerable here.

This is also the reason inertial measurement units are seldom used alone in high performance motion tracking applications. Externally referenced sensors are usually employed to bind the long term inertial sensor drift errors.

However, in our attempt to sense and compensate hand tremor, the signal of interest becomes the higher frequency hand tremor, while the intended motion becomes the “noise,” as it were. When the filtering algorithm separates the slower varying voluntary motion from the tremor, the low-frequency sensor drifts are removed at the same time. As a result, precise tracking of the absolute pose of the instrument is not of importance, as long as we can bind the accumulation of drift errors and remove gravity effectively from the kinematic model to recover the tremor motion.

Therefore, from both technical and end-user perspectives, internally referenced sensors emerge to be a better suited sensing technology than externally referenced sensors.

2.1.3 Conventional vs All-Accelerometer Inertial Measurement Unit

A conventional three accelerometers and three gyroscopes (3A-3G) IMU has been a popular choice of motion sensor in aerospace navigation applications since the 1950s. It is sometimes referred to as strapdown IMU or inertial navigation system (INS) when combined with other electronics and sensor fusion software.

We have seen from the previous section and other literatures [Barbour 01; Weston 00; Verplaetse 94] that high quality gyros with the precision needed for our application are

too bulky to be fitted into a handheld device, and they are almost always too costly. On the other hand, inexpensive, batch-processed gyros, although having suitable size and weight, currently lack the level of required sensing resolution and accuracy. Technology breakthrough in the gyro manufacturing process is not likely in the near future due to challenges associated with micro-miniaturization of gyros [Tan 01].

Exploiting the more affordable and mature micro-machined accelerometer technology, researchers have proposed gyroscope-free or all-accelerometer sensor designs. Chen *et al.* [Chen 94] proposes an original cube configuration design with six accelerometers, the minimum number of sensors to recover all the six kinematic parameters. The sensing direction of each accelerometer is along the respective cube face diagonal, in such a way that these diagonals form a regular tetrahedron. Tan *et al.* [Tan 01] show that, except for a ‘measure zero’ set of configurations, any arbitrary configuration of six or more accelerometers are also feasible, although only the cube design yield a closed-form solution. The proposed closed-form solution solves for the angular acceleration and hence error growth rate for the orientation and position estimates are an order higher than a conventional IMU, i.e. $\propto t^2$ and $\propto t^4$ respectively.

2.2 Magnetometer-Aided All-Accelerometer Inertial Measurement Unit

The proposed sensing system for the handheld microsurgical instrument consists of an all-accelerometers inertial measurement unit (IMU) and a tri-axial magnetometer. The IMU is made up of three low-cost miniature Analog Devices ADXL-203 dual-axis accelerometers. These sensors are housed in two locations in the instrument handle, as shown in Fig. 2.1. The back sensor suite located at the distal end from the intraocular shaft tip houses two dual-axis accelerometers (one X -, one Y -, and two Z -sensing directions) and the Honeywell HMC-2003 tri-axial magnetometer. One dual-axis accelerometer is located at the front sensor suite at the proximal end to the tool tip, measuring motion in X - and Y -directions. The back sensor suite measures $26 \text{ mm} \times 26$

mm \times 30 mm and the cylindrical handle body measures $\varnothing 20$ mm \times 90 mm. Pictures of the magnetometer-aided all-accelerometer IMU is shown in Fig. 2.2.

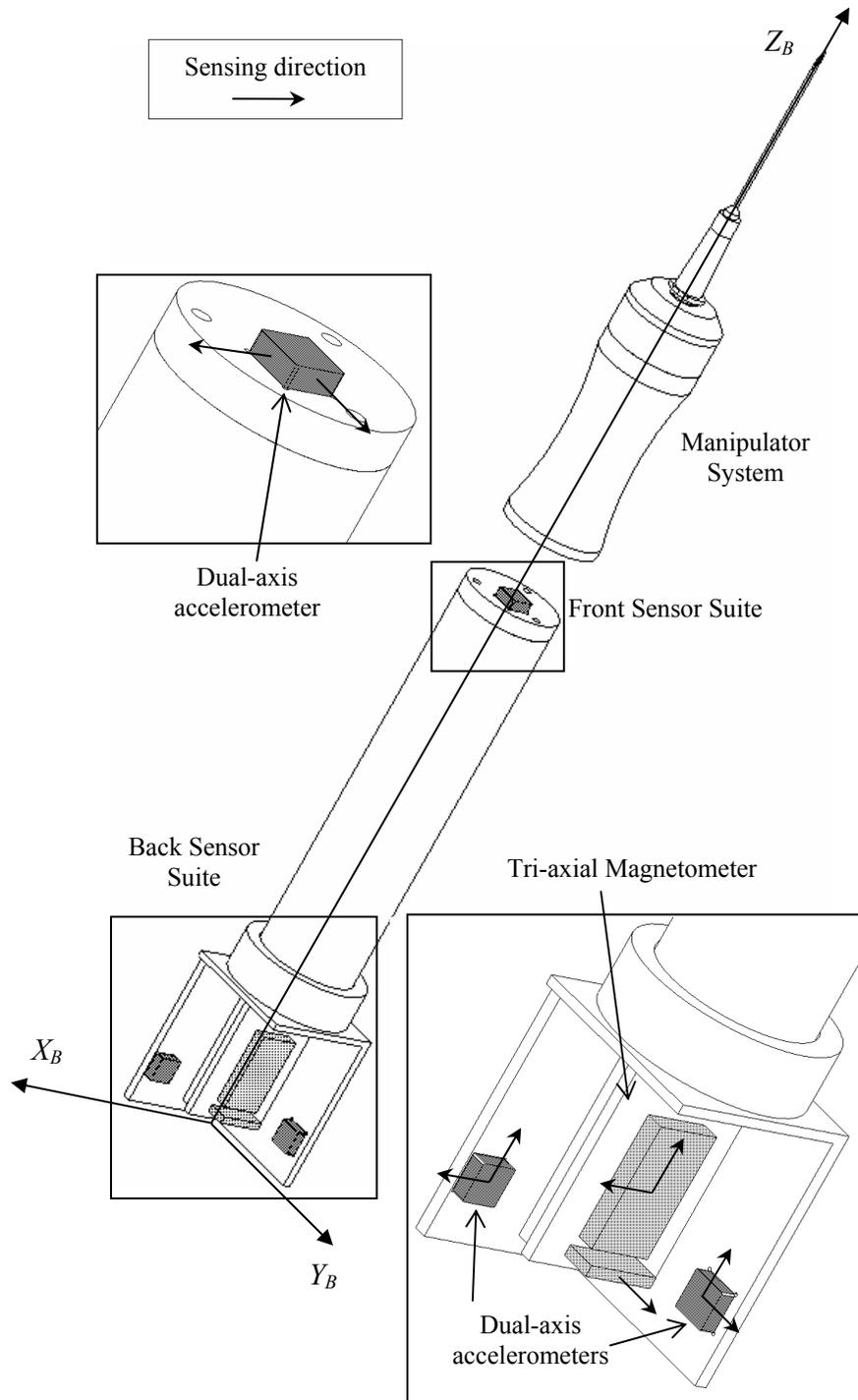


Fig. 2.1 Design of the magnetometer-aided inertial measurement unit.



Fig. 2.2 Pictures of the implemented magnetometer-aided inertial measurement unit.

An all-accelerometer IMU design is proposed as the primary sensing modality because of the superiority of low cost miniature accelerometers to low cost miniature gyros, in terms of sensing resolution and accuracy. Compared to other existing all-accelerometer IMUs in a cube configuration, the proposed design can use off-the-shelf dual-axis accelerometers, since the sensing directions can be orthogonal. This results in a more compact, simpler, and cheaper design. The most compelling reason to have the proposed all-accelerometer IMU design is that this configuration enables high quality linear and angular inertial sensing. This is to be shown in Section 2.4.

Despite the apparent cost, size and resolution superiority, the attractiveness of all-accelerometers IMU over a conventional IMU is marred by its faster error growth rate. To address this problem, a three-axis magnetometer is added to aid the IMU. In an indoor environment such as an operating room, a magnetometer senses the direction and intensity of the combined magnetic field of the earth and that generated by the electrical components in its vicinity. Since a magnetometer does not need to maintain a light of sight with its external reference source, the self-contained advantage of the IMU is preserved. If the direction of the magnetic field remains constant throughout the operation period, which is a valid assumption in a microsurgery setting, the magnetometer's North-seeking information would provide an external reference to bind the sensor drift in the pan orientation. Furthermore, as long as the effective body acceleration of the instrument is not in the same order as the gravity for an extended time, which is a valid assumption for microsurgery, the downward pointing gravity sensed by the accelerometers can be combined with the non-drifting North-seeking information from the magnetometer to provide orthogonal bounds to the IMU orientation errors via sensor fusion technique such as Kalman filtering. With bounded orientation error, the error growth rate of the all-accelerometer IMU position estimates reduces by two orders, i.e. $\propto t^2$. It is to be discussed in Chapter 4.

With the addition of the magnetometer, we have sensing redundancy in both angular and translational motion sensing. The next section describes the sensing kinematics to compute the orientation of the instrument and the position of the tool-tip.

2.3 Sensing Kinematics

2.3.1 Differential Sensing Kinematics

The primary outputs of the proposed all-accelerometer IMU are a set of three pairs of linear acceleration measurements at three different locations. This section presents the differential sensing kinematics to recover the 3 DOF angular velocity of the instrument from the sensed acceleration information.

We define a body frame $\{B\}$ at the base of back sensor suite, with its Z_B -axis collinear with the instrument tip as shown in Fig 2.1. In the back sensor suite, there are two dual axis accelerometers, ${}^B A_1 = [{}^B a_{1x} \ \bullet \ {}^B a_{1z}]^T$ and ${}^B A_2 = [\bullet \ {}^B a_{2y} \ {}^B a_{2z}]^T$. The third dual axis accelerometers, ${}^B A_3 = [{}^B a_{3x} \ {}^B a_{3y} \ \bullet]^T$ is located at the front sensor suite close to the instrument tip. The symbol \bullet denotes undefined quantity. The locations of the accelerometers with respect to the body frame $\{B\}$ are known from the mechanical design and through sensor calibrations. Without loss of generality, we assume that the proof masses of each of the dual axis accelerometer are coincident at $\{1\}$, $\{2\}$ and $\{3\}$ respectively, and all the sensing axes are aligned with the body principle axes: X_B , Y_B , and Z_B . Fig. 2.3 shows the accelerometer locations, the sensing axes and kinematic frames of the instrument.

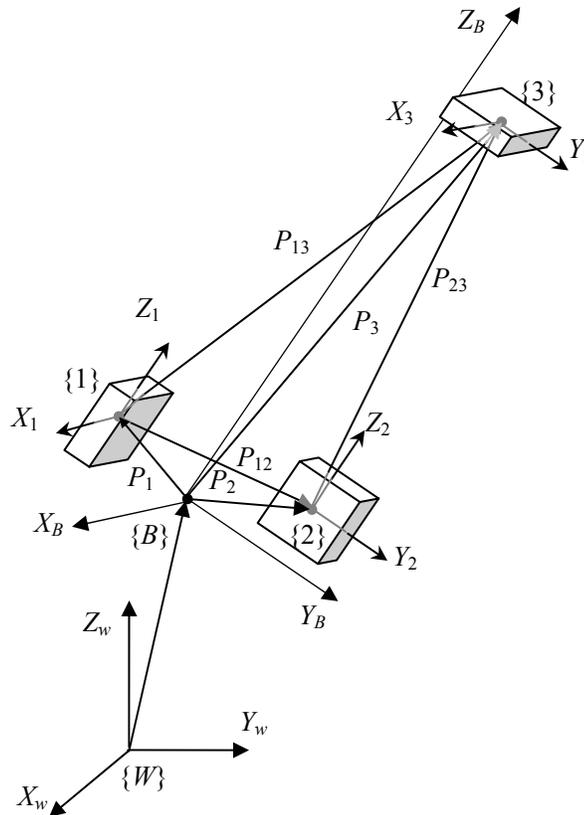


Fig. 2.3 Kinematic representation of the all-accelerometer IMU.

The X_B , Y_B and Z_B axes of frame $\{B\}$ are initialized to coincide with X_w , Y_w and Z_w axes of the world coordinate system $\{W\}$ at time $t = 0$. To derive the angular velocity

vector, ${}^B\Omega = [{}^B\omega_x \ {}^B\omega_y \ {}^B\omega_z]^T$ from the accelerometers, we start from the motion equation of a rigid body in \mathbf{E}^3 space. The total accelerations, ${}^B A_i$, each accelerometer at $\{i\}$ senses include the inertial acceleration of the body, ${}^B A_{CG}$, the gravity, ${}^B g$, and the rotation-induced accelerations: the centripetal acceleration, ${}^B A_{i/C}$, and the tangential acceleration, ${}^B A_{i/T}$,

$${}^B A_i = {}^B A_{CG} + {}^B g + {}^B A_{i/C} + {}^B A_{i/T}, \quad i = 1, 2, 3; \quad (2.3a)$$

$${}^B A_i = {}^B A_{CG} + {}^B g + \underbrace{{}^B\Omega \times {}^B\Omega \times R + {}^B\alpha \times R}_{\text{rotation-induced}} \quad (2.3b)$$

where the R is the vector from the unknown instantaneous center of rotation to the point of sensing.

Taking the difference between the accelerations readings at $\{1\}$, $\{2\}$ and $\{3\}$, the non-rotation induced acceleration components ${}^B A_{CG}$ and ${}^B g$ are eliminated, since the linear inertial acceleration of the body and the gravity should be identical at different locations,

$${}^B A_{13} = {}^B A_3 - {}^B A_1 = {}^B\Omega \times {}^B\Omega \times {}^B P_{13} + {}^B\dot{\Omega} \times {}^B P_{13} = \begin{bmatrix} {}^B a_{13x} & \bullet & \bullet \end{bmatrix}^T \quad (2.4a)$$

$${}^B A_{23} = {}^B A_3 - {}^B A_2 = {}^B\Omega \times {}^B\Omega \times {}^B P_{23} + {}^B\dot{\Omega} \times {}^B P_{23} = \begin{bmatrix} \bullet & {}^B a_{23y} & \bullet \end{bmatrix}^T \quad (2.4b)$$

$${}^B A_{12} = {}^B A_2 - {}^B A_1 = {}^B\Omega \times {}^B\Omega \times {}^B P_{12} + {}^B\dot{\Omega} \times {}^B P_{12} = \begin{bmatrix} \bullet & \bullet & {}^B a_{12z} \end{bmatrix}^T \quad (2.4c)$$

where the symbol \bullet denoted undefined quantity and ${}^B P_{ij} = [{}^B p_{ijx} \ {}^B p_{ijy} \ {}^B p_{ijz}]^T$, $i, j = 1, 2, 3$, is position vector from $\{i\}$ to $\{j\}$ with respect to $\{B\}$, which are known values from the hardware design and calibrations. Equations (2.4a-c) provide three independent constraint equations to solve for angular velocity vector ${}^B\Omega = [{}^B\omega_x \ {}^B\omega_y \ {}^B\omega_z]^T$.

2.3.1.1 Numerical Solution

If the sensors are initialized at rest at time, $t = 0$, the angular acceleration ${}^B\dot{\Omega}$ can be expressed in terms of angular velocity,

$${}^B\dot{\Omega}(t) = \frac{{}^B\Omega(t) - {}^B\Omega(t-T)}{T}, \quad {}^B\Omega(0) = \mathbf{0}_{3 \times 1} \quad (2.5)$$

where T is the sampling interval. Thus, equations (2.4a-c) can be rewritten as

$$f_1 = -{}^B p_{13x} ({}^B\omega_y^2 + {}^B\omega_z^2) + {}^B p_{13y} {}^B\omega_x {}^B\omega_y + {}^B p_{13z} {}^B\omega_x {}^B\omega_z$$

$$- {}^B p_{13y} \frac{{}^B \omega_z - {}^B \omega_z(t-T)}{T} + {}^B p_{13z} \frac{{}^B \omega_y - {}^B \omega_y(t-T)}{T} - {}^B a_{13x}, \quad (2.6a)$$

$$f_2 = {}^B p_{23x} {}^B \omega_x {}^B \omega_y - {}^B p_{23y} ({}^B \omega_x^2 + {}^B \omega_z^2) + {}^B p_{23z} {}^B \omega_y {}^B \omega_z \\ + {}^B p_{23x} \frac{{}^B \omega_z - {}^B \omega_z(t-T)}{T} - {}^B p_{13z} \frac{{}^B \omega_x - {}^B \omega_x(t-T)}{T} - {}^B a_{13y}, \quad (2.6b)$$

$$f_3 = {}^B p_{12x} {}^B \omega_x {}^B \omega_z + {}^B p_{12y} {}^B \omega_y {}^B \omega_z - {}^B p_{12z} ({}^B \omega_x^2 + {}^B \omega_y^2) \\ + {}^B p_{12x} \frac{{}^B \omega_y - {}^B \omega_y(t-T)}{T} - {}^B p_{12y} \frac{{}^B \omega_x - {}^B \omega_x(t-T)}{T} - {}^B a_{13z}. \quad (2.6c)$$

The time indexes t for ${}^B \omega_x$, ${}^B \omega_y$, and ${}^B \omega_z$ are omitted for ease of representation.

The angular velocity vector ${}^B \Omega = [{}^B \omega_x \quad {}^B \omega_y \quad {}^B \omega_z]^T$ may be found by solving the nonlinear system of equations (2.6a-c) by nonlinear least-squares solver such as the Gauss-Newton method or the Levenberg-Marquart method,

$$\min \{ r({}^B \Omega) : {}^B \Omega \in \mathfrak{R}^3 \}, r({}^B \Omega) = \frac{1}{2} \| F({}^B \Omega) \|_2^2, F({}^B \Omega) = [f_1 \quad f_2 \quad f_3]^T. \quad (2.7)$$

The starting values for the nonlinear solver are chosen to be

$${}^B \tilde{\Omega}(t) = {}^B \Omega(t-T) + ({}^B \Omega(t-T) - {}^B \Omega(t-2T)). \quad (2.8)$$

The angular displacement vector is then given by

$${}^B \Theta(t) = \int_t^{t-T} {}^B \Omega(t) dt + {}^B \Theta(t-T) \quad (2.9)$$

The directional cosine matrix that relates the body frame $\{B\}$ to the world coordinate system $\{W\}$ at each sampling interval is updated by first updating the quaternions,

$$\dot{q}(t) = \tilde{\Omega}(t)q(t) \quad (2.10)$$

$$q(t) = \int_t^{t-T} \dot{q}(t) dt + q(t-T) \quad (2.11)$$

$$\tilde{\Omega} = \frac{1}{2} \left[\begin{array}{ccc|c} 0 & {}^B \omega_z & -{}^B \omega_y & {}^B \omega_x \\ -{}^B \omega_z & 0 & {}^B \omega_x & {}^B \omega_y \\ {}^B \omega_y & -{}^B \omega_x & 0 & {}^B \omega_z \\ \hline -{}^B \omega_x & -{}^B \omega_y & -{}^B \omega_z & 0 \end{array} \right] \quad (2.12)$$

where $\tilde{\Omega}$ is the augmented cross product matrix. In terms of quaternions, the directional cosine matrix is given by

$${}^W C_B = \begin{bmatrix} q_0^2 + q_1^2 - q_2^2 - q_3^2 & 2(q_1 q_2 - q_0 q_3) & 2(q_1 q_3 + q_0 q_2) \\ 2(q_1 q_2 + q_0 q_3) & q_0^2 - q_1^2 + q_2^2 - q_3^2 & 2(q_2 q_3 - q_0 q_1) \\ 2(q_1 q_3 - q_0 q_2) & 2(q_2 q_3 + q_0 q_1) & q_0^2 - q_1^2 - q_2^2 + q_3^2 \end{bmatrix}. \quad (2.13)$$

2.3.1.2 Analytical Solution

Equation (2.7) might run into numerical instability or converge to local minima in the presence of excessive sensor noise due to external disturbance, such as impact of sensor housing with a hard surface. In micromanipulation tasks, the angular velocity of the instrument is almost always very small. Thus the ${}^B \Omega^2$ centripetal acceleration terms in (2.4a-c) may be ignored,

$${}^B A_{13} = {}^B \dot{\Omega} \times {}^B P_{13} = \begin{bmatrix} {}^B a_{13x} & \bullet & \bullet \end{bmatrix}^T \quad (2.14a)$$

$${}^B A_{23} = {}^B \dot{\Omega} \times {}^B P_{23} = \begin{bmatrix} \bullet & {}^B a_{23y} & \bullet \end{bmatrix}^T \quad (2.14b)$$

$${}^B A_{12} = {}^B \dot{\Omega} \times {}^B P_{12} = \begin{bmatrix} \bullet & \bullet & {}^B a_{12z} \end{bmatrix}^T. \quad (2.14c)$$

The angular acceleration vector is then solved by

$${}^B \dot{\Omega} = \begin{bmatrix} {}^B \dot{\omega}_x \\ {}^B \dot{\omega}_y \\ {}^B \dot{\omega}_z \end{bmatrix} = P_D^{-1} A_D, \quad (2.15)$$

where the relative position cross product matrix P_D and the differential acceleration vector A_D are defined as

$$P_D = \begin{bmatrix} 0 & {}^B p_{13z} & -{}^B p_{13y} \\ -{}^B p_{23z} & 0 & {}^B p_{23x} \\ {}^B p_{12y} & -{}^B p_{13x} & 0 \end{bmatrix}, \quad (2.16)$$

$$A_D = \begin{bmatrix} {}^B a_{13x} \\ {}^B a_{23y} \\ {}^B a_{12z} \end{bmatrix}. \quad (2.17)$$

The angular velocity vector is found by integrating (2.15)

$${}^B \Omega(t) = \int_t^{t-T} {}^B \dot{\Omega}(t) dt + {}^B \Omega(t-T). \quad (2.18)$$

2.3.2 Orientation from Vectors Observation

2.3.2.1 The Wahba's Problem

The determination of orientation of a rigid body from vector observations is also known as the Wahba's problem.

The original problem proposed by Wahba in 1965 was on the estimation of a satellite's attitude based on direction cosines of objects as observed in the satellite body frame and direction cosines of the same objects in a known frame of reference [Wahba 65]. In other words, it is to find the orthonormal direction cosine or rotation matrix M that minimizes in the least squares sense the error function,

$$E(M) = \sum_{j=1}^n \|v_j^* - Mv_j\|^2, \quad (2.19)$$

where $\{v_j^*\}$ is a set of unit vectors measured in the body frame and $\{v_j\}$ are the corresponding unit vectors in a reference frame. Examples of these vectors may be the unit vectors pointing to the direction of a specific star such as the Sun, the Earth's magnetic field or gravity, etc. The solution of the Wahba's problem yields non-drifting attitude estimate. Equation (2.19) is presented in Wahba's original notations.

Recasting the Wahba's problem in the context of orientation tracking with magnetometer-aided inertial measurement unit, the sensors measure in the body frame the gravity vector and the magnetic North vector, the solution to the problem gives the orientation of the body with respect to the world frame. Two non-collinear and non-zero length vectors is the minimum requirement to solve the Wahba's problem unambiguously.

Gebre-Egziabher [Gebre-Egziabher 00] formulates an iterated least squares algorithm to solve for the error quaternion, and the error quaternion is cast into a standard linearized measurement equation of a complementary Kalman filter. In the recent development of the MARG sensor, Yun *et al.* [Yun 03] follow Gebre-Egziabher's approach to the Wahba's problem but compound the error quaternion with the gyros updated quaternion and express this quantity as the measurement states of a Kalman filter. The authors state computational efficiency as the primary reason of adopting this algorithm over their previous implementation [Marins 01].

2.3.2.2 The TRIAD Algorithm

Among the proposed solutions, the TRIAD algorithm [Shuster 81] is the simplest and best suited for a two-vector Wahba's problem, despite its inefficiency in solving the general case with three or more observed vectors when compare to other methods [Farrell 66; Bar-Itzhack 85; Markley 00].

We define the world coordinate system $\{W\}$ such that the z_W axis aligns with the negated gravity vector ${}^W G$, the x_W axis forms a plane with z_W that contains the magnetic North vector ${}^W N$, and the y_W axis completes the right-handed coordinate system. Hence, the gravity and the unit magnetic North vector in $\{W\}$ are defined as

$${}^W G = g[0 \ 0 \ -1]^T, \quad {}^W N = [{}^w n_x \ 0 \ {}^w n_z]^T, \quad (2.20)$$

where g is the local gravity constant.

The accelerometers and magnetometers sense the gravity and the unit magnetic North vector in the body frame $\{B\}$,

$${}^B G = [{}^B g_x \ {}^B g_y \ {}^B g_z]^T, \quad {}^B N = [{}^B n_x \ {}^B n_y \ {}^B n_z]^T. \quad (2.21)$$

The z_W axis in the body frame is simply

$${}^B z_W = -\frac{{}^B G}{\|{}^B G\|}. \quad (2.22)$$

Normalizing the cross product between ${}^B z_W$ and ${}^B N$ will give the y_W axis in the body frame,

$${}^B y_W = \frac{{}^B z_W \times {}^B N}{\|{}^B z_W \times {}^B N\|}. \quad (2.23)$$

A second cross product between ${}^B y_W$ and ${}^B z_W$ will yield x_W in the body frame,

$${}^B x_W = {}^B y_W \times {}^B z_W. \quad (2.24)$$

The direction cosines of the world frame $\{W\}$ with respect to the body frame $\{B\}$ is

$${}^W C_B = \begin{bmatrix} {}^B x_W & {}^B y_W & {}^B z_W \end{bmatrix}^T. \quad (2.25)$$

2.3.3 Translational Displacement

With complete orientation information of the instrument body frame, the instantaneous position of the instrument tip with respect to a fixed frame may be determined by simple kinematic computation.

The effective translational body motion with respect to $\{W\}$ is obtained from removing the gravity component from the accelerometer readings,

$${}^W A_{iE} = {}^W C_B {}^B A_i - {}^W G \quad (2.26)$$

where $i = 1, 2$, or 3 . Integrating (2.26), we get the effective velocities at the sensor location $\{i\}$,

$${}^W U_{iE}(t) = \int_{t-T}^t {}^B A_{iE}(t) dt + {}^W U_{iE}(t-T) \quad (2.27)$$

The effective velocity at the origin of $\{B\}$ with respect to $\{W\}$, ${}^W U_E = [{}^W u_{Ex} \ {}^W u_{Ey} \ {}^W u_{Ez}]^T$ is found by transforming the sensors' position, ${}^B P_{iB}$, $i = 1, 2$, or 3 and body angular velocity ${}^B \Omega$ vectors into $\{W\}$,

$${}^W P_{iB} = {}^W C_B {}^B P_{iB}, \quad (2.28)$$

$${}^W \Omega = {}^W C_B {}^B \Omega, \quad (2.29)$$

and then compound the results derived from different accelerometers by their respective error variances,

$${}^W u_{Ex} = k_{13x} ({}^W u_{1x} + {}^W p_{1Bz} {}^W \omega_y - {}^W p_{1By} {}^W \omega_z) + k_{31x} ({}^W u_{3x} + {}^W p_{3Bz} {}^W \omega_y - {}^W p_{3By} {}^W \omega_z) \quad (2.30a)$$

$${}^W u_{Ey} = k_{23y} ({}^W u_{2y} + {}^W p_{2Bx} {}^W \omega_z - {}^W p_{2Bz} {}^W \omega_x) + k_{32y} ({}^W u_{3y} + {}^W p_{3Bx} {}^W \omega_z - {}^W p_{3Bz} {}^W \omega_x) \quad (2.30b)$$

$${}^W u_{Ez} = k_{12z} ({}^W u_{1z} + {}^W p_{1By} {}^W \omega_x - {}^W p_{1Bx} {}^W \omega_y) + k_{21z} ({}^W u_{2z} + {}^W p_{2By} {}^W \omega_x - {}^W p_{2Bx} {}^W \omega_y) \quad (2.30c)$$

where

$$k_{jid} = \frac{\sigma_{Aid}^2}{\sigma_{Aid}^2 + \sigma_{Ajd}^2}, \quad i, j = 1, 2, \text{ or } 3, \text{ and } d = x, y, \text{ or } z. \quad (2.31)$$

To obtain the instantaneous velocity of the instrument tip with respect $\{W\}$,

$${}^W U_{tip} = {}^W U_E + {}^W \Omega \times {}^W C_B {}^B P_{tip} \quad (2.32)$$

where the tip position with respect to the body frame $\{B\}$ is ${}^B P_{tip} = [0 \ 0 \ l]^T$. The length l is determined from calibration. Integrating (2.32) we get the instantaneous tip displacement

$${}^W P_{tip}(t) = \int_{t-T}^t {}^B U_{tip}(t) dt + {}^W P_{tip}(t-T). \quad (2.33)$$

2.1 Error Variance Analysis of IMU Designs

The strongest motivation of having the proposed all-accelerometer IMU design is its ability to achieve higher quality linear and angular inertial measurements. The rms output noise floor of a band-limited inertial sensor determines its sensing resolution, and indirectly impacts the sensing accuracy. One way to evaluate the performance of an IMU design analytically is to compare the error variance or standard deviation of the sensed or derived motion quantities. In this section, we compare the proposed all-accelerometer IMU with a conventional three gyros, three accelerometers (3G-3A) IMU in terms of error variance per sampling time of the estimated translational accelerations and angular velocities. We assume that the fictitious conventional IMU is made up of three Analog Devices ADXRS150 MEMS gyroscopes and three accelerometers with identical performance as the Analog Devices ADXL-203. The ADXS150 gyros are chosen because they are the smallest and exhibit the best noise performance among the low cost miniature gyros surveyed in Table 2.2.

To analyze the error variance of the translational and angular displacement estimations, we first define the acceleration vector to be

$$A = [a_{1x} \ a_{1z} \ a_{2y} \ a_{2z} \ a_{3x} \ a_{3y}]^T, \quad (2.34)$$

and the angular velocity vector from the gyros and that derived from the accelerometers to be

$$\Omega_G = [\omega_{Gx} \ \omega_{Gy} \ \omega_{Gz}]^T, \quad (2.35)$$

$$\Omega_A = [\omega_{Ax} \ \omega_{Ay} \ \omega_{Az}]^T. \quad (2.36)$$

Taking into account the stochasticity of the sensors, the sensor vectors become

$$A = \bar{A} + \delta A, \quad (2.37)$$

$$\Omega_G = \bar{\Omega}_G + \delta \Omega_G, \quad (2.38)$$

$$\Omega_A = \bar{\Omega}_A + \delta \Omega_A, \quad (2.39)$$

where \bar{A} , $\bar{\Omega}_G$, and $\bar{\Omega}_A$ are the means, δA , $\delta\Omega_G$, and $\delta\Omega_A$ are zero mean random vectors with known variances. Assuming the sensors' noise are not correlated, the covariance matrices and A and Ω_G may be computed simply as

$$\Xi_A = E[(A - \bar{A})(A - \bar{A})^T] = \begin{bmatrix} \sigma_{a1x}^2 & & & & & \\ & \sigma_{a1z}^2 & & & & \\ & & \sigma_{a2y}^2 & & & \\ & & & \sigma_{a2z}^2 & & \\ & & & & \sigma_{a3x}^2 & \\ & & & & & \sigma_{a3y}^2 \end{bmatrix}, \quad (2.40)$$

$$\Xi_{\Omega_G} = E[(\Omega_G - \bar{\Omega}_G)(\Omega_G - \bar{\Omega}_G)^T] = \begin{bmatrix} \sigma_{\omega Gx}^2 & 0 & 0 \\ 0 & \sigma_{\omega Gy}^2 & 0 \\ 0 & 0 & \sigma_{\omega Gz}^2 \end{bmatrix}, \quad (2.41)$$

where σ_{aid}^2 and $\sigma_{\omega Gd}^2$, $i = 1, 2$, or 3 and $d = x, y$, or z , are the error variances of the accelerometers and the gyros. From (2.29), we may define the differential acceleration vector,

$$A_D = [a_{13x} \ a_{23y} \ a_{12z}]^T = MA, \quad (2.42)$$

where

$$M = \begin{bmatrix} -1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & -1 & 0 & 0 & 1 \\ 0 & -1 & 0 & 1 & 0 & 0 \end{bmatrix}. \quad (2.43)$$

Hence the covariance of the differential acceleration vector is

$$\Xi_{AD} = MC(A)M^T = \begin{bmatrix} \sigma_{a1x}^2 + \sigma_{a3x}^2 & 0 & 0 \\ 0 & \sigma_{a2y}^2 + \sigma_{a3y}^2 & 0 \\ 0 & 0 & \sigma_{a1z}^2 + \sigma_{a2z}^2 \end{bmatrix}. \quad (2.44)$$

2.3.4 Angular Velocity Error Variance

2.3.4.1 Gyroscope Angular Velocity Error Variance

The error variance of the gyro may be computed from the square of the rms output noise, since the zero mean rmse is by definition one error standard deviation. Assuming

all the Analog Devices ADXRS150 gyros are identical and band-limited at 15 Hz, from Table 2.2, the error variance or the diagonal elements of Ξ_{Ω_G} may be computed as

$$\sigma_{\omega_G}^2 = (0.75 \text{ deg/s})^2 = 0.563 \text{ deg}^2/\text{s}^2. \quad (2.45)$$

2.3.4.2 Accelerometer Derived Angular Velocity Error Variance

From Section 2.3.1, the differential accelerometer vector is related to the accelerometer-derived angular velocity vector by a non-linear function,

$$A = F(\Omega_A), \quad (2.46)$$

To compute the accelerometer-derived angular velocity covariance Ξ_{Ω_A} , we need to find the inverse mapping function

$$\Omega_A = F^{-1}(A) \quad (2.47)$$

which has no closed-form solution and need to be computed by a nonlinear least-squares solver such as the Gauss-Newton method or the Levenberg-Marquart method. Since we cannot obtain the noise covariance Ξ_{Ω_A} by computing the Jacobian to approximate F^{-1} with the multidimensional Taylor Series, we adopt the scaled unscented transformation – a nonlinear variance transformation technique proposed by Julier [Julier 00].

The scaled unscented transformation is motivated by the notion of it is easier to approximate a probability distribution than it is to approximate a nonlinear function. For the three-dimensional ($n = 3$) angular velocity vector, a set of at least seven ($2n + 1$) weighted samples, called the sigma points, are generated. This sample set is zero mean and has covariance equal to the noise covariance of the differential acceleration vector Ξ_{AD} . The nonlinear transformation F^{-1} is applied to each of the sigma points, and the covariance of the body angular velocity vector is approximated by the calculated covariance of the transformed sample set.

The sigma points S are selected by the algorithm:

$$\begin{aligned} S_0 &= 0, \quad c_0 = \kappa/(n + \kappa); \\ S_i &= \left(\sqrt{(n + \kappa)\Xi_{AD}} \right)_i, \quad c_i = \frac{1}{2(n + \kappa)}; \\ S_{i+n} &= -\left(\sqrt{(n + \kappa)\Xi_{AD}} \right)_i, \quad c_{i+n} = \frac{1}{2(n + \kappa)}; \end{aligned} \quad (2.48)$$

where $\kappa \in \mathfrak{R}$ and $(n + \kappa) \neq 0$, $(\sqrt{(n + \kappa)\Xi_{AD}})_i$ is the i^{th} row or column of the matrix square root of $(n + \kappa)\Xi_{AD}$, and c_i is the weight associated with the i^{th} sigma point.

The mean and noise covariance of Ω_A is computed as

$$\bar{\Omega}_A = \sum_{i=0}^{2n} c_i \hat{S}_i, \quad (2.49)$$

$$\Xi_{\Omega_A} = \sum_{i=0}^{2n} c_i (\hat{S}_i - \bar{\Omega}_A)(\hat{S}_i - \bar{\Omega}_A)^T, \quad (2.50)$$

where

$$\hat{S}_i = F^{-1}(S_i). \quad (2.51)$$

Thus assuming all the Analog Devices ADXL-203 accelerometers are identical and band-limited at 15 Hz, from Table 2.3, the error variances of the accelerometer derived angular velocity may be computed as

$$\sigma_{\omega Ax}^2 = \xi_{11} + \xi_{12} + \xi_{13}, \quad (2.52a)$$

$$\sigma_{\omega Ay}^2 = \xi_{21} + \xi_{22} + \xi_{23}, \quad (2.52b)$$

$$\sigma_{\omega Az}^2 = \xi_{31} + \xi_{32} + \xi_{33}, \quad (2.52c)$$

where ξ_{ij} $i, j = 1, 2, 3$, are the elements of covariance matrix Ξ_{Ω_A} .

2.3.5 Translational Displacement Error Variance

In the proposed all-accelerometer design, since we have two accelerometers in each sensing axis, we may compound the redundant sensing information,

$$A_d = \frac{\sigma_{Ajd}^2}{\sigma_{Aid}^2 + \sigma_{Ajd}^2} A_{id} + \frac{\sigma_{Aid}^2}{\sigma_{Aid}^2 + \sigma_{Ajd}^2} A_{jd}. \quad (2.53)$$

Assuming all accelerometers have identical noise characteristics, the error variance and standard deviation become

$$\frac{1}{\sigma_{Ad}^2} = \frac{1}{\sigma_{Aid}^2} + \frac{1}{\sigma_{Ajd}^2}, \quad (2.54)$$

with $i = 1, 2$, or 3 and $d = x, y$, or z , and $\sigma_{Aid}^2 = \sigma_{Ajd}^2$,

$$\sigma_{Ad}^2 = \frac{1}{2} \sigma_{Aid}^2, \quad \sigma_{Ad} = \frac{\sigma_{Ai}}{\sqrt{2}}. \quad (2.55)$$

This results in an improvement of a factor of $\sqrt{2}$ or 29.3% in the standard deviations of the translational displacement measurements.

2.3.6 Discussion

The error variances and standard deviations per sampling time computed in the previous sections are tabulated in Table 2.4.

The standard deviation of the ω_{Ax} , ω_{Ay} , and ω_{Az} are reduced by 99.1%, 99.1% and 97.5% over those of ω_{Gx} , ω_{Gy} , and ω_{Gz} respectively. The intuition of this result is depicted in Fig. 2.4, taking ω_{Gx} and ω_{Gy} as an example. The discs labeled $P(\omega_{Gx}, \omega_{Gy})$ and $P(\omega_{Ax}, \omega_{Ay})$ are the joint distributions of the $\Omega_{Gxy} = [\omega_{Gx} \ \omega_{Gy}]^T$ and $\Omega_{Axy} = [\omega_{Ax} \ \omega_{Ay}]^T$ estimates from gyros and accelerometers respectively. Higher density regions at the centers represent higher probability. With the same uncertainty distribution at $\{3\}$, the larger $|{}^B P_{13}|$ ($= |{}^B P_{23}|$) gets, the smaller the angle sustaining the cone will become. This angle may be viewed as being the standard deviation of ω_{Ax} or ω_{Ay} , since $\sigma_{\omega_{Ax}}$ and $\sigma_{\omega_{Ay}}$ are identical. The noise in the ω_{Az} measurement is the worst as length $|{}^B P_{12}|$ is constraint by the slenderness of the handheld instrument.

Table 2.4 Comparison of error variances and standard deviations of the angular velocity and linear acceleration sensed by a conventional three gyros three accelerometers (3G-3A) IMU versus the proposed all-accelerometer (6A) IMU.

Measurement	Error Variance, σ^2		Error Standard Deviation, σ		
	3G-3A	6A	3G-3A (deg)	6A (deg)	% Reduction
ω_x, ω_y	0.563 deg ² /s ²	4.23×10^{-5} deg ² /s ²	0.75 deg/s	0.65×10^{-2} deg/s	99.1
ω_z	0.563 deg ² /s ²	3.55×10^{-4} deg ² /s ²	0.75 deg/s	1.88×10^{-2} deg/s	97.5
a_x, a_z, a_z	486.3 mm ² /s ⁴	243.1 mm ² /s ⁴	22.1 mm/s ²	15.6 mm/s ²	29.3

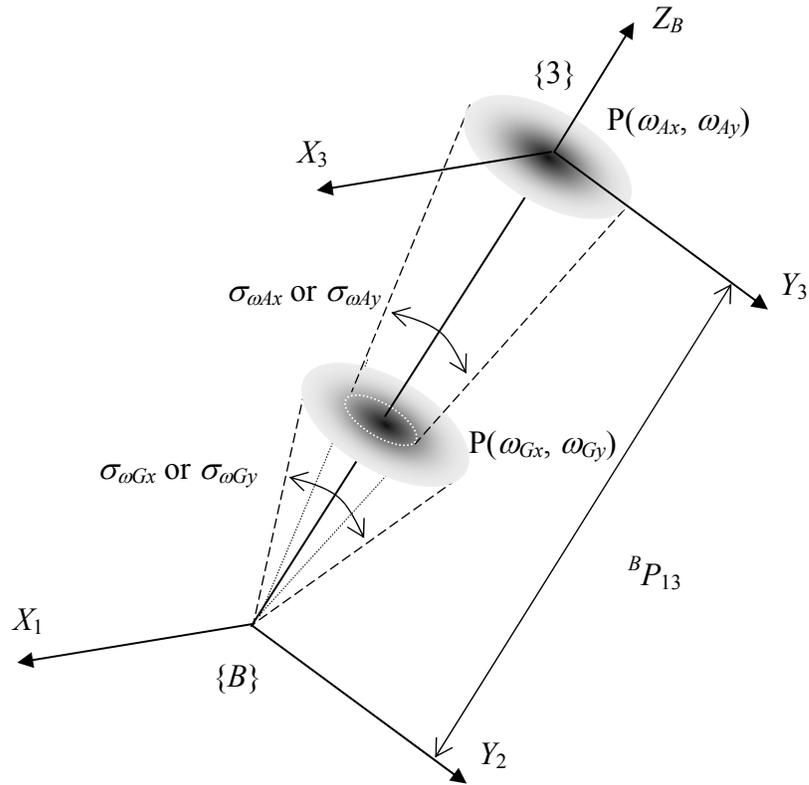


Fig. 2.4 Comparison of the joint distributions of ω_x and ω_y estimates of an all-gyro ($P(\omega_{Gx}, \omega_{Gy})$) and an all-accelerometer ($P(\omega_{Ax}, \omega_{Ay})$) design. Higher density regions at the centers represent higher probability.

With this argument, the angular sensing resolution of the sensing system will be maximized if the distances between the accelerometers having the same sensing direction are maximized. Within the physical constraint of a slender handheld instrument, the sensor locations are designed to be

$${}^B P_{12} = \begin{bmatrix} -d_1 \\ d_1 \\ 0 \end{bmatrix}, \quad {}^B P_{23} = \begin{bmatrix} d_2 \\ d_2 - d_1 \\ d_3 \end{bmatrix}, \quad \text{and} \quad {}^B P_{13} = \begin{bmatrix} d_2 - d_1 \\ d_2 \\ d_3 \end{bmatrix}, \quad (2.54)$$

where $d_1 = 20.2$ mm, $d_2 = 14.3$ mm, and $d_3 = 116.0$ mm.

On top of the 29.3% direct improvement in the translation sensing, the improvement in the orientation estimation will result in a much complete removal of the gravity, which in turn has a great impact on the quality of the translation sensing since gravity is typically two orders of magnitude larger than hand accelerations.

References

- [Barbour 01] N. Barbour and G. Schmidt, "Inertial Sensor Technology Trends," *IEEE Sensors Journal*, vol.1. No. 4, pp. 332-339, Dec 2001.
- [Bar-Itzhack 85] I. Y. Bar-Itzhack and Y. Oshman, "Attitude Determination from Vector Observations: Quaternion Estimation," *IEEE Trans. Aerospace & Electronic Syst.*, Vol. 21. No.1, pp.128-135, Jan. 1985.
- [Chen 94] J. H. Chen, S. C. Lee, D. B. DeBra, "Gyroscope Free Strapdown Inertial Measurement Unit by Six Linear Accelerometers", *Journal of Guidance, Control, and Dynamics*, Vol.17, No.2, March-April 1994, pp.286-290.
- [Elble 90] R. J. Elble and W. C. Koller, *Tremor*. Baltimore: Johns Hopkins, 1990.
- [Farrell 66] J. L. Farrell, J. C. Stuelpnagel, R. H. Wessner, and J. R. Velman, "Solution to problem 65-1: A least squares estimate of satellite attitude," *SIAM Review*, Vol. 8, No. 3, pp. 384-386, July 1966.
- [Gebre-Egziabher 00] D. Gebre-Egziabher, G. H. Elkaim, J. D. Powell, and B. W. Parkinson, "A Gyro-Free Quaternion-Based Attitude Determination System Suitable for Implementation using Low Cost Sensors," *Proc. IEEE Position, Location, and Navigation Symp.*, pp. 185-192, March 2000.
- [Hunter 93] I. W. Hunter, T. D. Doukoglou, S. R. Lafontaine, P. G. Charette, L. A. Jones, M. A. Sagar, G. D. Mallinson, and P. J. Hunter, "A teleoperated microsurgical robot and associated virtual environment for eye surgery," *Presence*, 2:265-280, 1993.
- [Julier 00] S. Julier, J. Uhlmann, and H. F. Durrant-Whyte, "A New Method for the Nonlinear Transformation of Means and Covariances in Filters and Estimators," *IEEE Trans. Automatic Control*, Vol. 45, No.3, pp. 477-482, Mar 2000.
- [Marins 01] J. L. Marins, X. Yun, E. R. Bachmann, R. B. McGhee, and M. J. Zyda, "An Extended Kalman Filter for Quaternion-Based Orientation Estimation Using MARG Sensors", *Proc. IEEE/RSJ Int. Conf. Intelligent Robots and Systems*, Maui, pp. 2003-2011, 2001.
- [Markley 00] F. L. Markley and D. Mortari, "Quaternion attitude estimation using vector observations," *J. Astronautical Sciences*, Vol. 48, No. 2&3, pp. 369-380, Apr-Sep 2000.
- [Shuster 81] M. D. Shuster and S. D. Oh, "Three-Axis Attitude Determination from Vector Observations," *J. Guidance, Control, and Dynamics*, Vol. 4, No. 1, pp. 70-77, Jan-Feb 1981.
- [Smith 90] R. Smith, M. Self, and P. Cheeseman, "Estimating Uncertain Spatial Relationships in Robotics", in *Autonomous Robot Vehicles*, Eds. I.J. Cox and G.T. Wilfong, Springer-Verlag, 1990, pp.167-193.
- [Tan 01] C-W. Tan, A. Park, K. Mostov, and P. Varaiya, "Design of Gyroscope-Free Navigation Systems," *IEEE Intelligent Transportation Systems Conference Proceedings*, pp.286-291, Aug 2001.
- [Verplaetse 94] C. J. Verplaetse, "Inertial-Optical Motion-Estimating Camera for Electronic Cinematography," *M.Sc. Thesis, MIT*, Ch.2, June 1994.
- [Wahba 65] G. Wahba, "Problem 65-1, A Least Squares Estimate of Satellite Attitude," *SIAM Review*, Vol. 7, No. 3, pp. 409, July 1965.
- [Weston 00] J. L. Weston and D. H. Titterton, "Modern Inertial Navigation Technology and its Application," *Electronics and Communication Engineering Journal*, pp.49-64, April 2000.
- [Yun 03] X. Yun, M. Lizarraaga, E. R. Bachmann, and R. B. McGhee, "An Improved Quaternion-Based Kalman Filter for Real-Time Tracking of Rigid Body Orientation," *Proc. IEEE/RSJ Int. Conf. Intelligent Robots & Systems*, Las Vegas, pp.1074-1079, Oct 2003.

Chapter 3

Measurement Model

3.1 Introduction

Management of modeling errors is one of the most important factors in the performance of a motion tracking system. Understanding the nature and dynamics of the sensors' errors enables compensation methods and filtering algorithms to be designed and hence improves the sensing accuracy. Error analysis is especially vital in the effectiveness of using inertial sensing technology to track rigid body motion.

A measurement model of a motion sensing system is a mathematical representation that relates the sensed motion input to the final computed pose with the sensor errors accounted for, i.e. $Pose(t) = f(Motion, Errors, t)$.

A measurement model consists of a physical model and a stochastic model. The physical model describes the physics of the sensors' operation and their physical relationships with respect to one another and the environment, including geographic location. It is the deterministic part of the sensing system. In the proposed model, we only attempt to model the system response to an external excitation (i.e. motion) at the sensor level, rather than at a component level that involves low-level circuitry and material properties. The stochastic model includes the random drift of the inertial and magnetic sensors observed under otherwise benign operating conditions.

The objective of this chapter is to develop a measurement model of the proposed magnetometer aided all-accelerometers IMU based on a thorough error analysis. The classification, nature and characteristics of the errors, and outlines the experimental methods to identify these error parameters will be discussed. The mathematical models of the identified errors are presented in a form where error compensations can be readily implemented.

3.2 Physical Model

The sensing system physical model maps the inputs (in terms of motion) to the outputs (in terms of the engineering quantities). The physical errors are stemmed from the intrinsic material properties and imperfect manufacturing processes. These deterministic errors in most cases may be compensated by calibrations.

3.2.1 Accelerometer

3.2.1.1 Accelerometer Errors

With companies like Analog Devices leading the way, low-cost micro-electromechanical systems-based (MEMS) accelerometers have gained substantial ground in inertial navigation applications in the past several years, especially in non-military and consumer markets [Barbour 01; Weston 00]. However, size and cost advantages notwithstanding, the performance of MEMS silicon accelerometers has not reached tactical or navigation grade. A test of MEMS accelerometers on rolling artillery projectiles conducted by Davis [Davis 98] show that they yield an average tracking error of about 0.1g throughout a 28-second test. Though the report claims that such performance is acceptable, double integration of an acceleration error of 0.1g would mean a position error of more than 350m at the end of the test.

It is this notorious integration drift that is responsible for the fact that inertial measurement technology is seldom used alone in high precision navigation or motion tracking applications. Any seemingly small error in the acceleration measurement would grow quadratically over time in the position measurement after the double integration, as

shown in Fig. 3.1. Therefore, to employ accelerometers effectively in high precision tracking applications, it is imperative to obtain a comprehensive model to account for the errors.

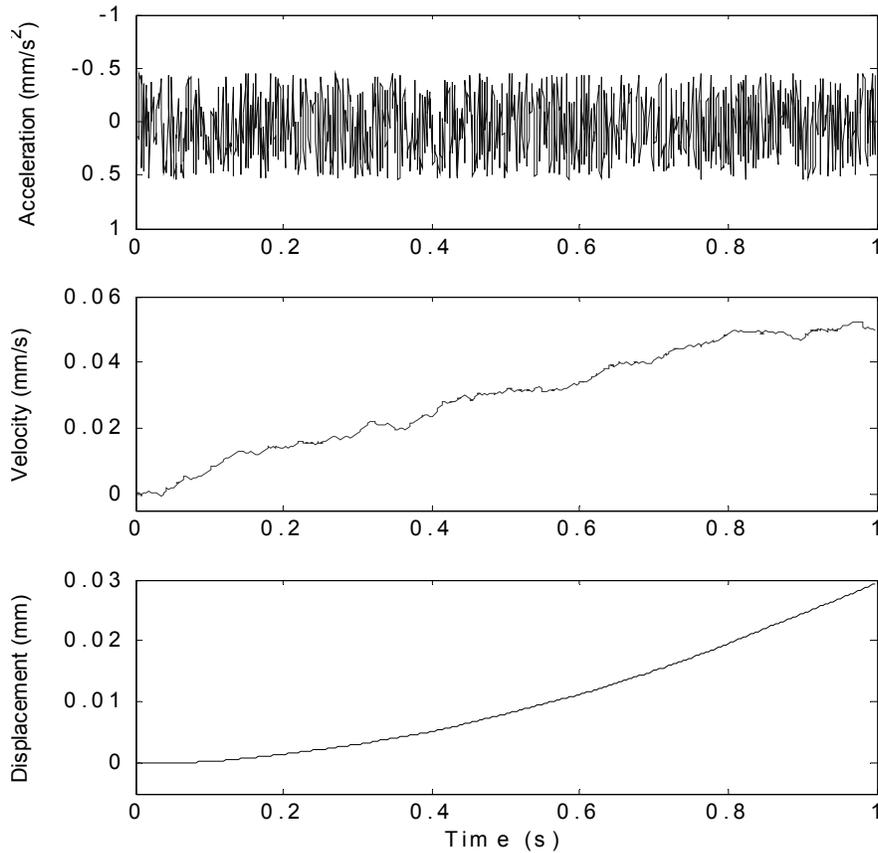


Fig. 3.1 Integration drift of an accelerometer. The plots show the output of an accelerometer at rest. A small offset error in the acceleration sensing grows linearly after the first integration and becomes quadratic in the displacement domain after the second integration.

There have been some modeling efforts of MEMS capacitive accelerometers on equivalent electrical circuit models that represent the physics of operation [Lee 99; Bourgeois 97]. The main problem of this type of modeling is that it is not useful to a user that is merely concerned with how to use the accelerometers effectively and does not care so much about the underlying physics and low level circuitry of the sensors. More important and useful physical parameters by which accelerometers are usually rated, such

as scale factor, bias, nonlinearity, cross-axis sensitivity, misalignment, noise characteristics, and temperature sensitivity are missing from these models.

For many low- g ($< 2g$) inertial sensing applications, such as indoor robotic navigation [Barshan 94], entertainment [Verplaetse 96], and our application in microsurgery, the signal-to-noise ratio is low and thus any unmodeled error in the aforementioned physical parameters would undermine the effectiveness of the intended application over time.

The goal of this section is to develop a physical model of an accelerometer in order to use the accelerometer effectively, in terms of the common parameters published in the manufacturer's datasheets. Simple experiments to reveal the behavior of these parameters are described. We adopt a phenomenological modeling method to relate the experimental observations to mathematical representations of these parameters without requiring complete understanding of the underlying physics. While the physical model developed is specific to the accelerometer make and model tested, the experiments and modeling methodology are generic and may be used with other types of accelerometers.

3.2.1.2 Accelerometer Physics and Construction

The Analog Devices ADXL-203 is a dual-axis miniature MEMS-based capacitive accelerometer. It measures $5\text{mm} \times 5\text{mm} \times 2\text{mm}$, weighs less than 1 gram, and has an effective sensing range of $\pm 1.5g$.

It is a complete acceleration measurement system on a single monolithic integrated circuit. It contains a surface micro-machined polysilicon structure built on top of a silicon wafer, with signal conditioning circuitry to implement an open-loop acceleration measurement architecture. Polysilicon springs at the four corners suspend the square proof mass over the surface of the wafer and provide a resistance against acceleration forces. A schematic illustration is shown in Fig. 3.2.

Deflection of the structure due to externally applied acceleration is measured using a differential capacitor. The differential capacitor involves the interdigitation of moving fingers attached to the proof mass with independent fixed fingers attached to the base. The differential capacitance is proportional to the overlapping area and the distance between the moving and fixed fingers. With the capacitor area fixed by the

manufacturing process, the differential capacitance and hence the circuit output voltage is proportional to the difference between distance d_1 and d_2 , as shown in the inset of Fig. 3.2. The movement of the proof mass is controlled by the suspension springs, and thus Hooke's Law applies in relating the displacement to the applied specific force. Since the modeling approach here is phenomenological, no equations of the physics of operation will be presented here. The reader is requested to refer to the datasheets and application notes from the manufacturer's website (www.analog.com) for more details. The ADXL-203 provides two output modes: analog and digital pulse wave modulation (PWM). The former is used in our implementation, but the approaches described are also applicable to the latter.

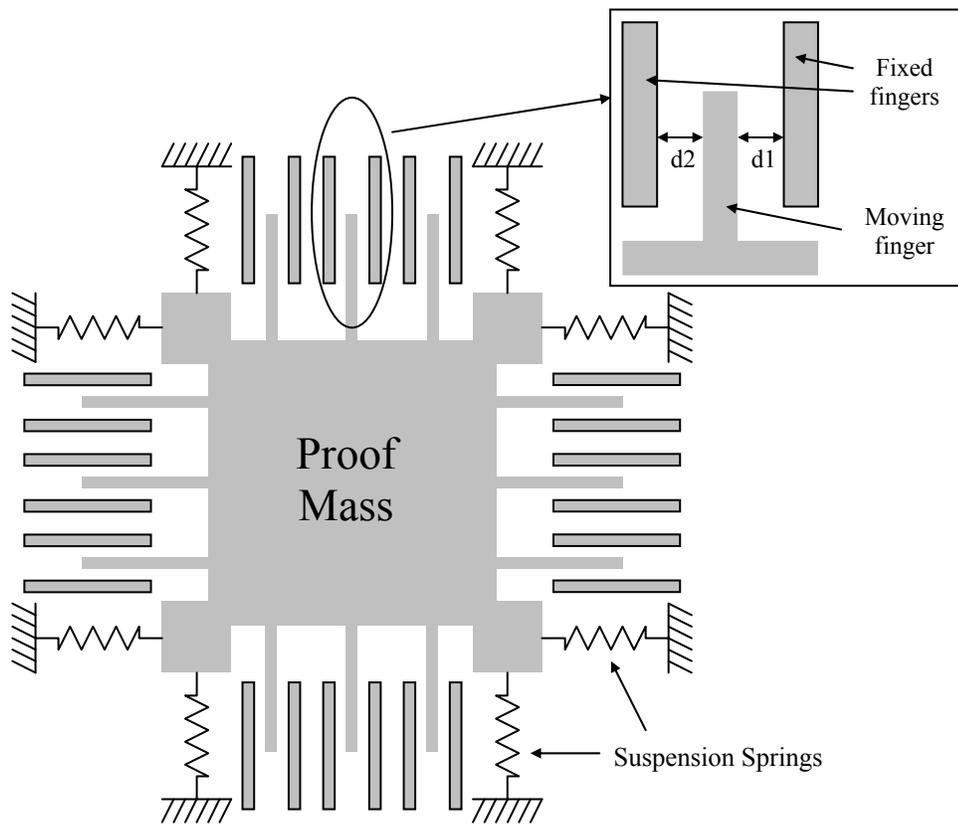


Fig. 3.2 Construction schematic of ADXL-203 dual-axis accelerometer.

3.2.1.3 Scale Factor and Bias

Scale factor is defined as the ratio of a change in the output to a change in the input intended to be measured. It is sometimes referred to as sensitivity. Bias or zero offset is the average sensor output over a specific time measured at specified operating conditions that has no correlation with the input. The scale factor and bias are affected by factors related to material and construction, e.g., hysteresis, nonlinearity, cross-axis effects, etc., as well as factors related to operation, such as temperature and pressure.

Scale factor of an accelerometer is usually expressed in volts (V) per g , where g is the gravitational acceleration. Bias is expressed in volts (V) or in terms of g if the scale factor is known. To determine the scale factor and bias, the manufacturer specification sheets recommend the following:

$$\text{Scale factor, } SF = (V_{+g} - V_{-g})/2 \quad \text{V/g;} \quad (3.1)$$

$$\text{Bias, } B = (V_{+g} + V_{-g})/2 \quad \text{V} \quad (3.2)$$

where V_{+g} and V_{-g} are the output of the accelerometer in volts when aligned with gravity and when aligned opposite to the direction of gravity, respectively. Thus, with accelerometer output V_o , the sensed acceleration (in terms of g) is computed as,

$$\text{Acceleration, } A = (V_o - B)/SF \quad g. \quad (3.3)$$

However, there are two problems to this approach: composite nonlinearity and temperature drift.

3.2.1.4 Composite Nonlinearity

The input–output relationship of the ADXL-203 accelerometer is nonlinear. The nonlinear behavior is a result of the material properties and the construction of the sensor. Two simple experiments will be performed to reveal the behavior of the accelerometer voltage output at orientations defined by two angles, α and β (see Fig. 3.3). The definition of the angle α is such that 0° and 180° are when the plane that contains both x - and y -sensing axes is perpendicular to the gravity plane and $\pm 90^\circ$ is where the two planes are parallel. When $\alpha = 90^\circ$, the angle β is 0° when the x -accelerometer is aligned with the gravity vector.

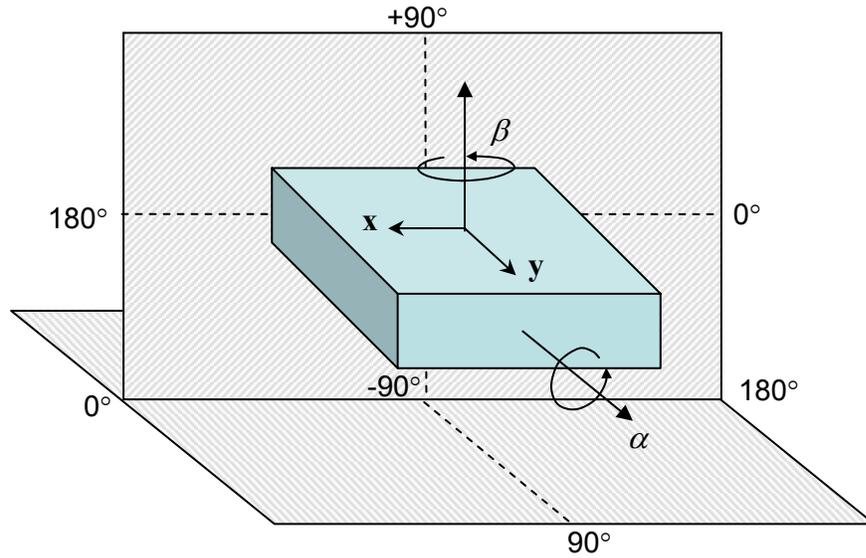


Fig. 3.3 Definition of angles α and β .

In the first experiment, starting with $\alpha = 90^\circ$ (see Fig. 3.4),

- when the accelerometer is rotated from $\beta = 0^\circ$ to 360° clockwise and anticlockwise in the gravity plane, hysteretic paths $A^+ - B^-$ and $B^+ - A^-$, respectively, are observed;
- when it is rotated clockwise and anticlockwise from $\beta = 0^\circ$ to 180° and then back to 0° in the gravity plane, paths $A^+ - A^-$ and $B^+ - B^-$, respectively, are observed.

In the second experiment, the accelerometer is inclined such that both the x- and y-sensing axes lie in a plane offset from the direction of gravity by the angle α . It is observed that

- when the accelerometer is rotated from $\beta = 0^\circ$ to 360° in the inclined plane α , there is a non-parallel upward shift of the entire hysteretic path with increasing α ;
- experiments at $\pm\alpha^\circ$ yield the same hysteretic paths.

There is no change in the accelerometer output at $\alpha = 0^\circ$ and 180° , since the sensing plane is orthogonal to gravity. For ease of illustration, only hysteretic paths $Z_{\pm 30}$, $Z_{\pm 90}$, and $Z_{\pm 150}$ are shown in Fig. 3.5. Note that $Z_{\pm 90}$ is the same as Fig. 3.4.

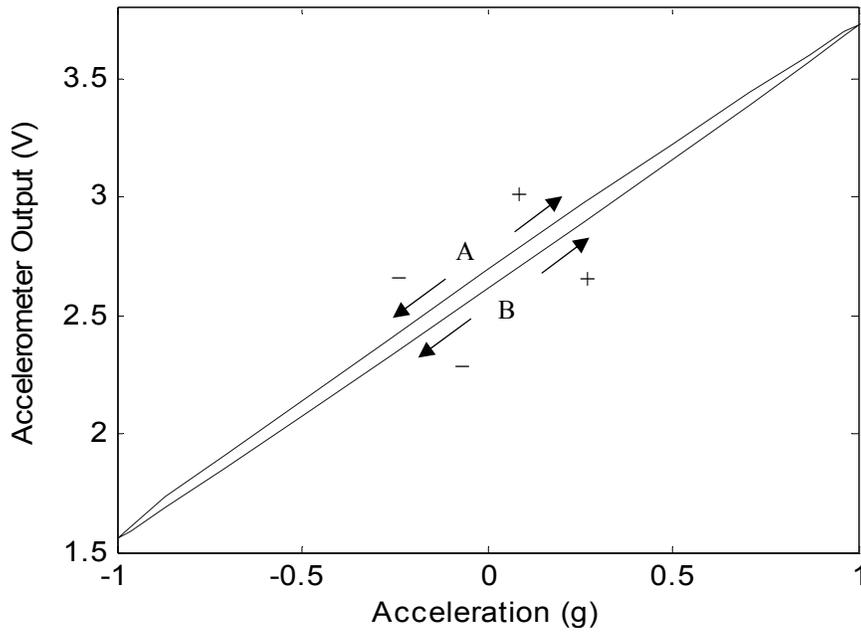


Fig. 3.4 Hysteretic paths of the accelerometer output when rotated in the gravity plane. When rotated clockwise and anticlockwise through 360° , it follows paths $A^+ - B^-$ and $B^+ - A^-$ respectively; when rotated clockwise and anticlockwise through 180° and back to 0° , it follows paths $A^+ - A^-$ and $B^+ - B^-$ respectively.

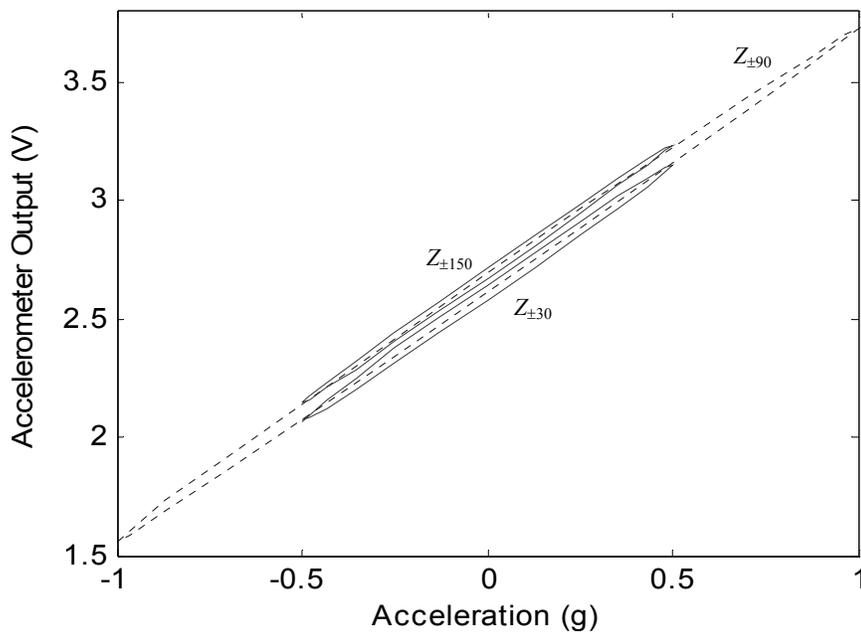


Fig. 3.5 Hysteretic paths of the accelerometer rotated through 360° in a plane inclined at angle $\alpha = \pm 30^\circ$, $\pm 90^\circ$, and $\pm 150^\circ$.

These observations are the result of composite nonlinearity errors, including material nonlinearity, hysteresis, and cross-axis effects. It is almost impossible to tell which factor contributes to which observation without specialized instrumentation. However, it is not necessary to understand the underlying physics, nor to model these errors explicitly; there is sufficient information for us to model and compensate these errors phenomenologically.

The hysteretic behavior of the ADXL-203 accelerometer is different from the classical hysteresis found in magnetism and other shape memory alloys. One possible explanation of this observation is the cross-axis effect from the acceleration in y-axis. This is evident from the fact that, since acceleration in x- and y-sensing directions are 90° out of phase, the hysteresis error of the x-axis is largest at the point where the y-direction acceleration is the largest, and vice versa.

One probable cause of the shift of the entire hysteresis path at different inclination α is the out-of-plane cross-axis effect from the acceleration in the z direction. This is again evident from the fact that the hysteretic curves shift upward as the out-of-plane z-axis acceleration increases.

Thus, the acceleration sensed by the x-accelerometer is the combined effect of the real acceleration in the x-axis plus cross-axis accelerations from the other two orthogonal principal directions, which alter the bias and the scale factor,

$$A_x = (V_x - B_x(V_y, V_z)) / SF_x(V_z) \quad (3.4)$$

where V_y and V_z are respectively the output of the y-accelerometer and an external z-accelerometer from another ADXL-203 in the sensor suite.

We define a linear model $f_x(A_x)$ by rearranging (3.3),

$$f_x(A_x) = A_x \cdot SF_x^{90} + B_x^{90} \quad (3.5)$$

where SF_x^{90} and B_x^{90} are the scale factor and bias obtained from the manufacturer's recommended calibration method. The superscript "90" signifies $\alpha = 90^\circ$.

The shift in bias is a linear superposition of the cross-axis effects from the y- and z-acceleration,

$$B_x(V_y, V_z) = B_x^{90} + g_x(V_y) + h_x(V_z). \quad (3.6)$$

To model $g_x(V_y)$, the corresponding data generated by the linear model $f_x(A_x)$ is subtracted from the measurement data of plot $Z_{\pm 90}$ (where $h_x(V_z) = 0$),

$$R_x^y = V_{ox} - V_x \quad (3.7)$$

where R_x^y is the residual x-accelerometer output attributed to the effect of acceleration in y-direction, V_{ox} is the measured x-accelerometer output and V_x is the data generated by the linear model $f_x(A_x)$. R_x^y , V_{ox} , and V_x are all $n \times 1$ vectors, where n is the number of data points. Plotting the residue vector versus the measured y-accelerometer output vector reveals that the relationship is quadratic as shown in Fig. 3.6. Hence the cross-axis effect of y-acceleration on the x-axis may be modeled by the least-squares fitting of a second-order polynomial to the data,

$$g_x(V_y) = p_{x2}V_y^2 + p_{x1}V_y + p_{x0}. \quad (3.8)$$

The result of this model is shown in Fig. 3.7.

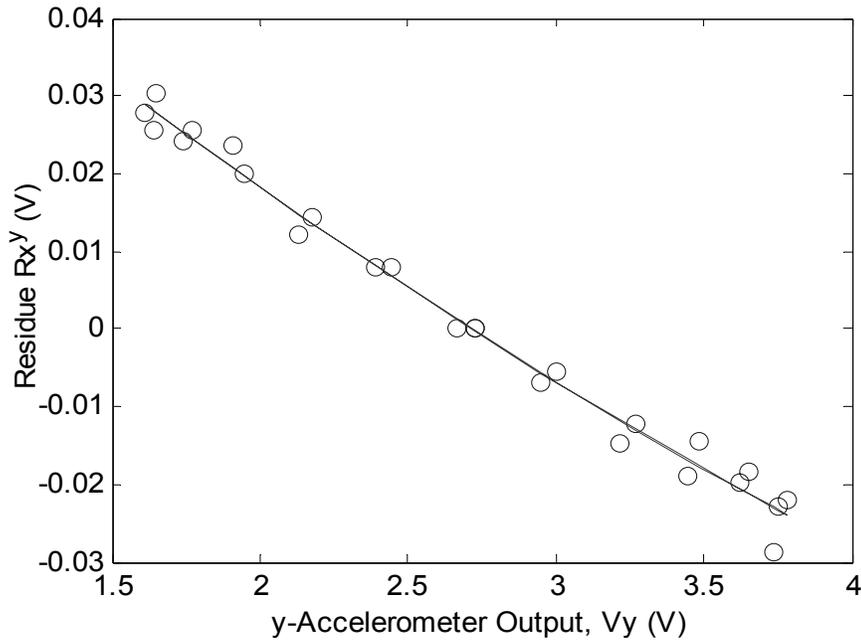


Fig. 3.6 The residual x-accelerometer output versus the y-accelerometer output. The relationship is quadratic.

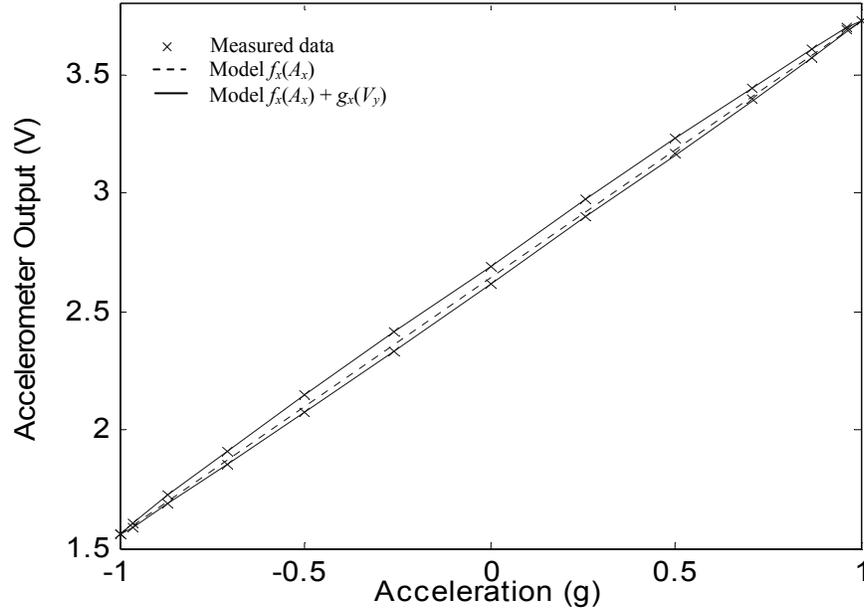


Fig. 3.7 The measured x-accelerometer output, the linear model $f_x(A_x)$, and the model $f_x(A_x) + g_x(V_y)$.

To model $h_x(V_z)$, we first compute the bias of the hysteretic paths generated at each inclined plane using (2) and then the difference of the bias at each α compared with that at $\alpha = 90^\circ$. Thus

$$R_x^z = B_x^\alpha - B_x^{90}, \quad (3.9)$$

where R_x^z is the residual x-accelerometer output attributed to the effect of acceleration in z-direction, B_x^α is the vector of the bias at each α , and B_x^{90} contains repeated values of the bias at $\alpha = 90^\circ$. They are all $m \times 1$ vectors, where m is the number of inclined planes in which we have conducted the experiment. Plotting R_x^z versus the z-accelerometer output shows that the relationship is quadratic as shown in Fig. 3.8. Hence the cross-axis effect of z-acceleration on x-axis may be modeled by the least squares fitting of a second-order polynomial to the data,

$$h_x(V_z) = q_{x2}V_z^2 + q_{x1}V_z + q_{x0}. \quad (3.10)$$

The result of this model is shown in Fig. 3.9.

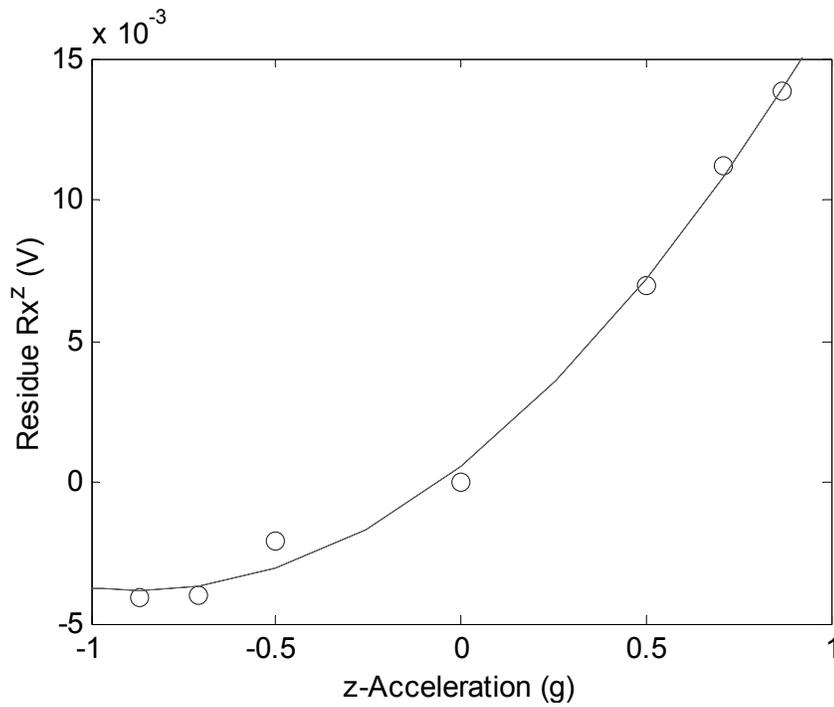


Fig. 3.8 The residual x-accelerometer output versus the z-acceleration. The relationship is quadratic.

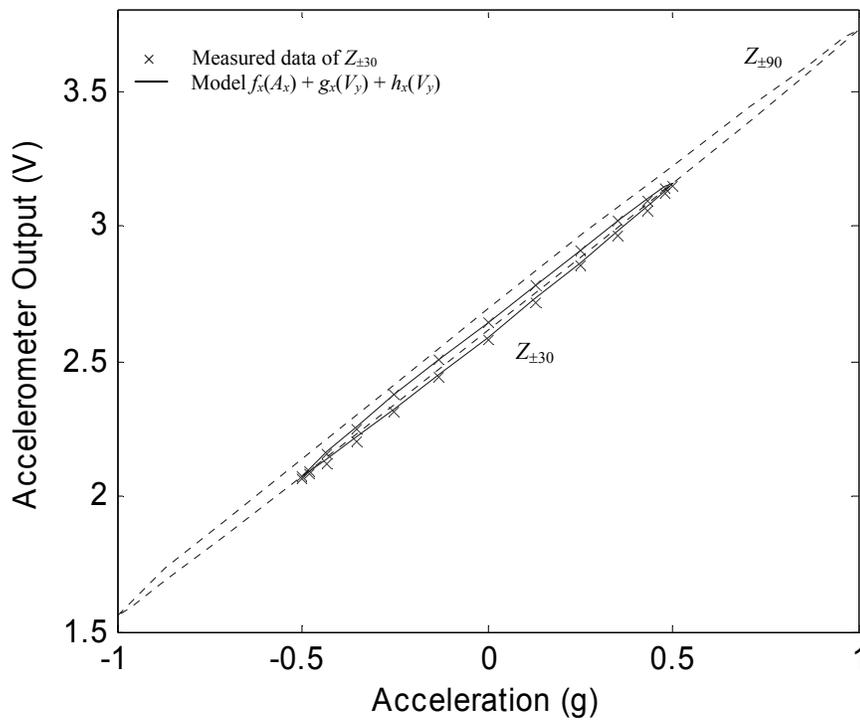


Fig. 3.9 The measured x-accelerometer output at $\alpha = \pm 30^\circ$ and the proposed physical model $f_x(A_x) + g_x(V_y) + h_x(V_z)$. The $Z_{\pm 90}$ plot is added for comparison.

Now we substitute (3.6), (3.8), and (3.10) into (3.4),

$$A_x = (V_x - B_x^{90} - p_{x2}V_y^2 - p_{x1}V_y - p_{x0} - q_{x2}V_z^2 - q_{x1}V_z - q_{x0})/SF_x(V_z), \quad (3.11)$$

where $SF_x(V_z)$ is found from the plot of SF_x versus the z-acceleration V_z as shown in Fig. 3.10. The scale factor model is thus

$$SF_x(V_z) = r_{x2}V_z^2 + r_{x1}V_z + r_{x0}. \quad (3.12)$$

The same approach is repeated to model the y-sensing axis,

$$A_y = (V_y - B_y^{90} - p_{y2}V_x^2 - p_{y1}V_x - p_{y0} - q_{y2}V_z^2 - q_{y1}V_z - q_{y0})/SF_y(V_z), \quad (3.13)$$

where

$$SF_y(V_z) = r_{y2}V_z^2 + r_{y1}V_z + r_{y0}. \quad (3.14)$$

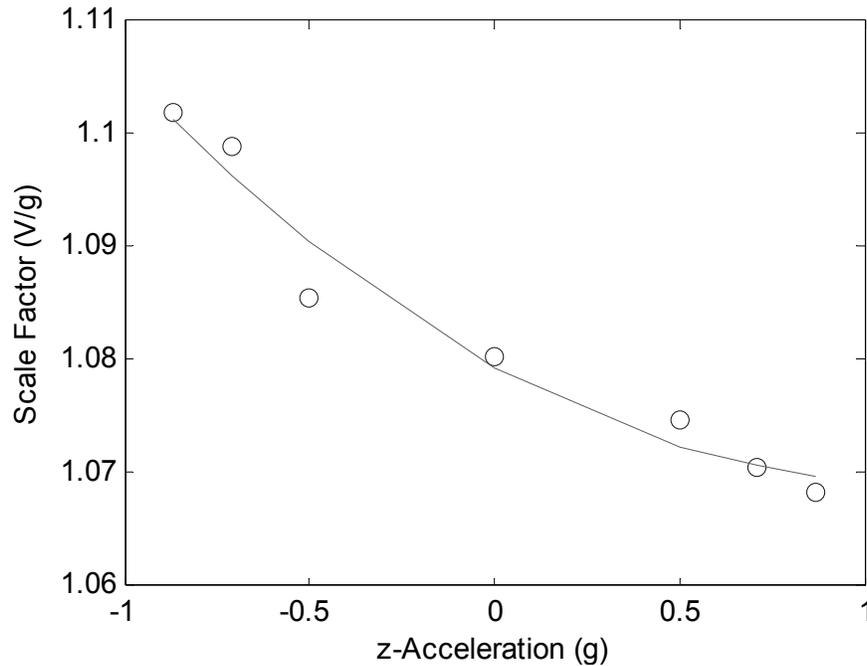


Fig. 3.10 Scale Factor versus z-acceleration. The relationship is quadratic.

3.2.1.5 Temperature Drift

The scale factor and bias are sensitive to temperature change. Since typical low-g motion tracking applications are generally indoors, we may assume the ambient temperature to remain constant over the operation period. However, a change in the operating temperature due to the gradual heating up of the sensor circuitry is inevitable.

Fig. 3.11 shows the drift in the zero bias and the scale factor at different time intervals over a period of 12 hours, with a constant ambient temperature of about 22°C. We observe that in this experiment the zero bias and the scale factor drift to a steady state after about seven hours.

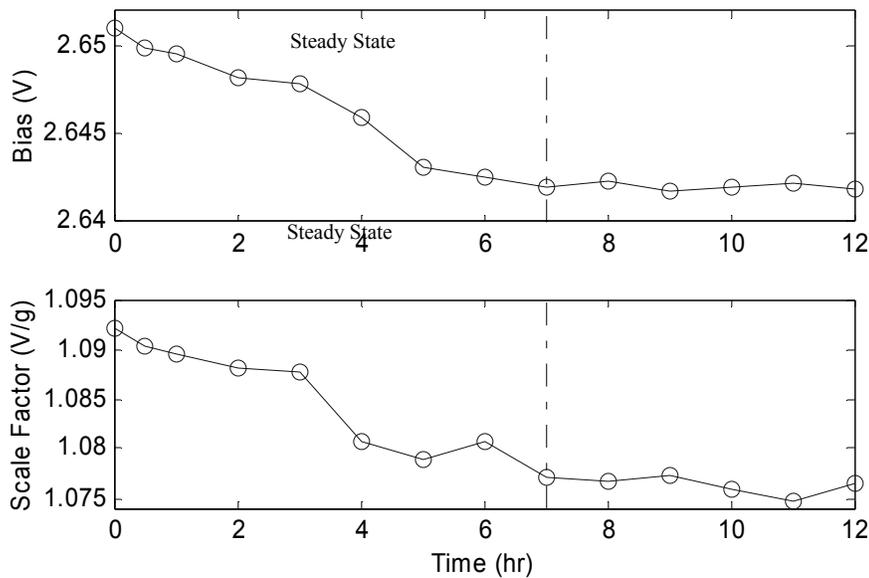


Fig. 3.11 Bias and scale factor temperature drift. Steady state is reached at approximately the 7th hour in this experiment.

Barshan *et al.* [Barshan 94] use an exponential function to model the bias drift of two types of rate gyroscopes, and the drift model is implemented via extended Kalman filtering. To the best of the authors' knowledge there is no literature that models the scale factor drift of any type of inertial sensors. Although Barshan's approach can be easily applied with some modifications to model both the scale factor and bias of an accelerometer, it is subject to potentially large modeling error because the drift characteristics of the sensors are inconsistent between operations. Moreover, quantifying this error may involve impractically great experimental effort.

An alternate straightforward solution is to wait for the sensors' temperature drift to reach steady state, which ranges from 2 to 12 hours among the 3 accelerometers (6 sensing axes) that we have tested.

Another approach that we have tried is ovenization. A resistor is placed closed to the accelerometer to heat up and maintain the accelerometer at a temperature higher than the ambient. Experiments have shown that steady state is reached within the first hour. However, the process changes the behavior of the accelerometer, producing higher nonlinearity and larger hysteresis. More experiments will be needed to investigate the effectiveness of this approach.

3.2.1.6 Misalignment

There are two types of accelerometer misalignment errors, orientation error and position error. Misalignment errors occur at the system level, arise from machining errors of the mechanical structure on which the sensors are mounted.

1) *Orientation Error.* Orientation error is the misalignment between the accelerometer sensing axes and the orthogonal principal body axes $X_B = [1 \ 0 \ 0]^T$, $Y_B = [0 \ 1 \ 0]^T$, and $Z_B = [0 \ 0 \ 1]^T$. The actual sensing axes are defined as

$$\Theta_{ix} = [\theta_{ixx} \ \theta_{ixy} \ \theta_{ixz}]^T, \quad i = 1 \ \& \ 3, \quad (3.15a)$$

$$\Theta_{jy} = [\theta_{jyx} \ \theta_{jyy} \ \theta_{jyz}]^T, \quad j = 2 \ \& \ 3, \quad (3.15b)$$

$$\Theta_{kz} = [\theta_{kzx} \ \theta_{kzy} \ \theta_{kzz}]^T, \quad k = 1 \ \& \ 2, \quad (3.15c)$$

where the subscript i , j , and k denote the sensor locations. Since sensor location does not affect orientation error, we assume the sensors collapse at the origin of the body frame, as shown in Fig. 3.12. For an arbitrary body acceleration ${}^B A = [{}^B a_x \ {}^B a_y \ {}^B a_z]^T$, the misaligned accelerometers sense a dominant component of acceleration in their intended sensing direction plus two other cross-axis acceleration components,

$$A_{ix} = \Theta_{ix}^T {}^B A, \quad i = 1 \ \& \ 3, \quad (3.16a)$$

$$A_{jy} = \Theta_{jy}^T {}^B A, \quad j = 2 \ \& \ 3, \quad (3.16b)$$

$$A_{kz} = \Theta_{kz}^T {}^B A, \quad k = 1 \ \& \ 2, \quad (3.16c)$$

where A_{ix} , A_{jy} , and A_{kz} are the accelerometer outputs. The actual body acceleration may then be found by

$$\begin{bmatrix} \hat{a}_{1x} \\ \hat{a}_{3y} \\ \hat{a}_{1z} \end{bmatrix} = \begin{bmatrix} \Theta_{1x}^T \\ \Theta_{3y}^T \\ \Theta_{1z}^T \end{bmatrix}^{-1} \begin{bmatrix} a_{1x} \\ a_{3y} \\ a_{1z} \end{bmatrix}, \quad (3.17a)$$

$$\begin{bmatrix} \hat{a}_{3x} \\ \hat{a}_{2y} \\ \hat{a}_{2z} \end{bmatrix} = \begin{bmatrix} \Theta_{3x}^T \\ \Theta_{2y}^T \\ \Theta_{2z}^T \end{bmatrix}^{-1} \begin{bmatrix} a_{3x} \\ a_{2y} \\ a_{2z} \end{bmatrix}, \quad (3.17b)$$

where \hat{a}_{id} , $i = 1, 2$, or 3 and $d = x, y$, or z are the misalignment corrected accelerometer outputs. In static operating condition where the accelerometers sense only the gravity, the body acceleration is found by compounding measurements of the same direction according to their respective error variance,

$${}^B A = \begin{bmatrix} {}^B a_x \\ {}^B a_y \\ {}^B a_z \end{bmatrix} = \begin{bmatrix} k_{13}\hat{a}_{1x} + k_{31}\hat{a}_{3x} \\ k_{23}\hat{a}_{2y} + k_{32}\hat{a}_{3y} \\ k_{12}\hat{a}_{1z} + k_{21}\hat{a}_{2z} \end{bmatrix}, \quad (3.18)$$

where the scaling factors k_{ij} are defined in (2.31).

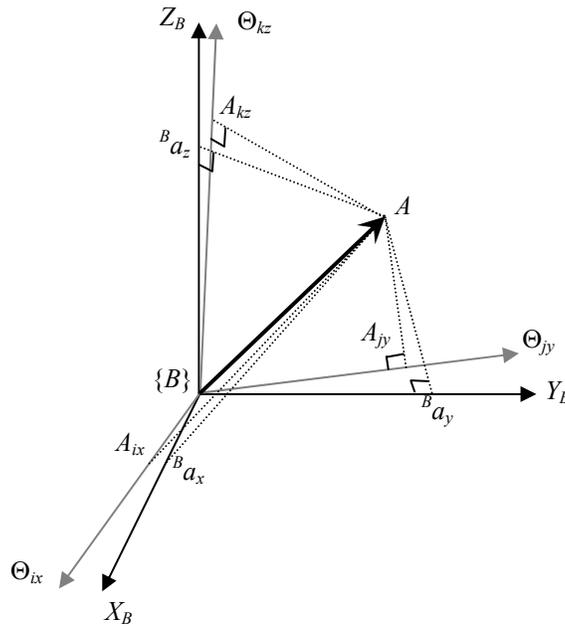


Fig. 3.12 Orientation error of an accelerometer.

The actual sensing axes of the six accelerometers can be determined experimentally. To find $\Theta_{1,x}$, for example, the instrument is mounted with the reference X_B - Z_B surface of the back sensor suite aligned with the gravity plane. The instrument is rotated through the angles α° and β° until a maximum (or minimum if X_B is aligned with $-g$) x -output voltage is registered. The orientation of the real x -sensing axis with respect to the reference body frame is then found by

$$\Theta_x = R_y(\alpha_x)R_z(\beta_x)X_B = [\theta_{xx} \ \theta_{xy} \ \theta_{xz}]^T, \quad (3.19)$$

where $R_y(\alpha_x)$ and $R_z(\beta_x)$ are 3×3 rotation matrix about Y_B and Z_B respectively, with $X_B = [0 \ 0 \ 1]^T$. Actual sensing vector of other accelerometers are found in a similar fashion.

2) *Position error*. Position error refers to the deviation of the accelerometer proof mass location from the nominal position with respect to the body frame. The consequence of position error is inaccuracy in determining the rotation induced acceleration components. To determine the position error, we need to put the instrument in known or accurately measurable angular motion, which provide ground truths on which the position error can be computed. This proves to be difficult to achieve with our current laboratory equipment. Moreover, the erroneous acceleration contributed by the position error is often well within the noise floor of the accelerometer, since the rotation induced acceleration during operation is always small compared to the gravity. Hence, for the current implementation, we ignore the effect of the position error.

3.2.1.7 Experimental Setup

All the experiments described in the previous sections are performed on an experimental setup made of two orthogonal precision motorized rotary stages with resolution of less than 0.1° , as shown in Fig. 3.13. The extension arm is a safety precaution so that the magnetic field generated by the motors will not interfere with the sensor outputs. A capacitor of $0.33\mu\text{F}$ is connected respectively between the x - and y -axis analog outputs and the signal ground, which combines with the sensor output impedance of $32 \text{ k}\Omega$ to become a simple first order R-C low-pass filter with cut-off frequency at around 15.2 Hz . The sensor output is sampled at 1 kHz by a 16-bit analog-to-digital converter. The inclinations of the rotary stages are checked prior to each experiment by a

liquid level. The noise floor of the sensor analog outputs is about 3.1 mV p-p, which is computed from 4 times the measured rms noise, the upper limit of the peak-to-peak noise 95.4% of the time, as specified in the datasheet. This would correspond to a minimum perceivable inclination change of approximately 0.2° . This means that experimental errors due to human inconsistency and parallax error in the order of 0.01° are negligible compared to the noise floor. Because of the stochastic nature of the sensor output, the mean of 5 seconds worth of data (5000 data points) is used in these experiments.

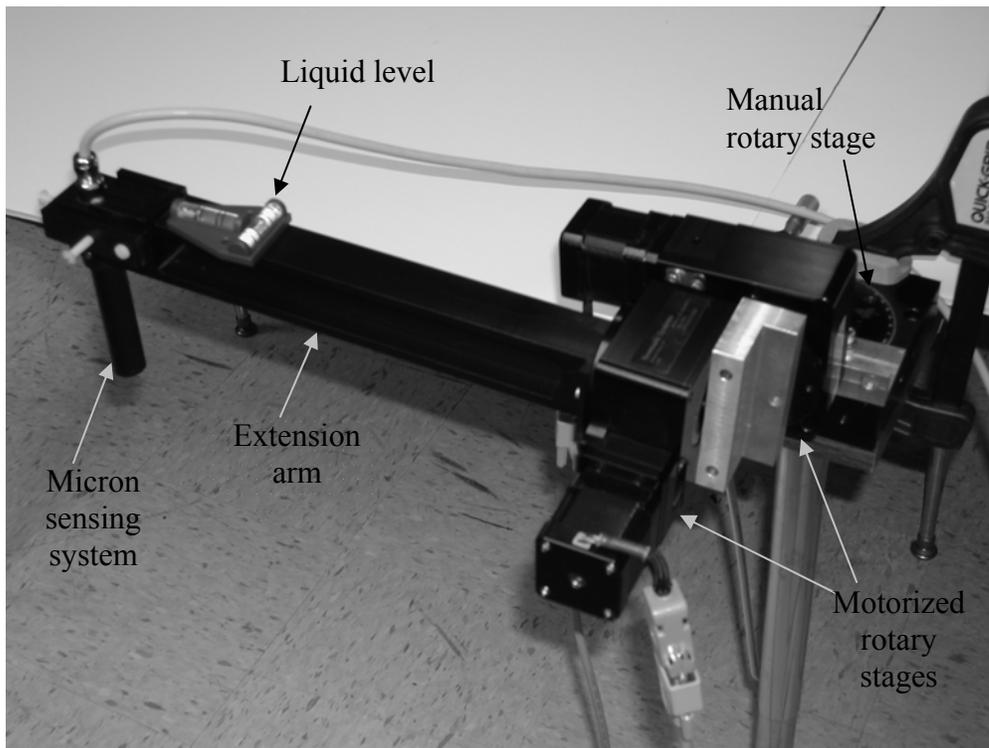


Fig. 3.13 Experimental setup for physical model parameters identification. The two orthogonal precision motorized rotary stages produce α and β rotations to the accelerometers and the triad of magnetometers. The extension arm is to place the sensors a safety distance away from the magnetic field generated by the motors.

3.2.2 Magnetometer

3.2.2.1 Magnetometer Errors

There are typically four types of error associated with a magnetometer: hard iron error, soft iron error, scale factor error, and misalignment [Gebre-Egziabher 01].

Hard iron errors are caused by the presence of time invariant magnetic field other than the Earth's magnetic field. In our application, we do not care about the exact heading of the Earth's magnetic North, as long as the direction and strength of the resultant magnetic field is time invariant throughout the operation. Soft iron errors are magnetic field generated by certain types of material in response to an externally applied field. There is no soft iron material in our instrument and it is pretty safe to assume none present in a clinical setting. Scale factor of the magnetometer is irrelevant our application since the magnetometer measurement is normalized to obtain a directional unit vector and the absolute magnetic field strength is assumed to be constant through out the operation. Thus, in Gebre-Egziabher's classification of magnetometer errors, misalignment is the only one that needs to be compensated in our application.

3.2.2.2 Misalignment

Orientation error of the triad of magnetometers are defined and determined in the same way as the accelerometers,

$$\begin{bmatrix} {}^B \hat{m}_x \\ {}^B \hat{m}_y \\ {}^B \hat{m}_z \end{bmatrix} = \begin{bmatrix} \Theta_{mx}^T \\ \Theta_{my}^T \\ \Theta_{mz}^T \end{bmatrix}^{-1} \begin{bmatrix} m_x \\ m_y \\ m_z \end{bmatrix}, \quad (3.20)$$

where \hat{m}_d and m_d , $d = x, y$, or z are the misalignment corrected and raw magnetometer outputs respectively; Θ_{mx} , Θ_{my} , and Θ_{mz} are the actual sensing axes of the magnetometers.

3.2.2.3 Vector Normalization

To normalize the magnetic field vector, we need also the knowledge of the magnetometer voltage outputs when a sensing axis is aligned with the magnetic North ($V_{N\max}$) and South ($V_{N\min}$). For each axis, we define

$$V_N^d = \frac{1}{2} (V_{N\max}^d - V_{N\min}^d), \quad (3.21)$$

$$V_{Bias}^d = \frac{1}{2} (V_{N\max}^d + V_{N\min}^d), \quad (3.22)$$

where the superscript denotes the sensing direction with respect to the body frame, $d = x, y$, or z . The magnetic North vector is then given by

$${}^B N = \begin{bmatrix} {}^B n_x & {}^B n_y & {}^B n_z \end{bmatrix}^T, \quad (3.23)$$

where

$${}^B n_d = \frac{V_M^d - V_{Bias}^d}{V_N^d}, \quad (3.24)$$

and V_M^d is the magnetometer voltage output of the d axis. For the magnetic North vector to remain unit length, it is normalized again,

$${}^B N = \frac{{}^B N}{\|{}^B N\|}. \quad (3.25)$$

3.2.2.4 Experimental Setup

The experiment to find the magnetometer orientation misalignments share the setup with that of the accelerometers, as described in Section 3.2.1.7.

3.3 Stochastic Model

3.3.1 Accelerometer

This section adopts the stochastic framework established in the IEEE standards for accelerometers [IEEE 1293-1998]. The Allan variance method [Tehrani 83; Ng 97] recommended by the IEEE standards is selected for the noise analysis over other methods [Ford 99; Seong 00] in the same framework which each has merits in its own rights. Allan variance is a characterization of the noise and other processes in the time series of data as a function of averaging time. The method is simple to compute, a more precise noise characterization than the conventional method of representing the drift error by a RMS number, and relatively easy to interpret and understand than the power spectral density method in the frequency domain.

The accelerometers are left motionless and the outputs are recorded for 12 hours at 1 Hz. As shown in Fig. 3.14, the voltage output shows a visible bias drift in the first 8 hours. Allan variance analysis of the recordings reveals that the dominant stochastic noise components of all the six accelerometers are velocity random walk (white noise in

acceleration), acceleration random walk (white noise in jerk or integrated white noise), and trend (a.k.a. bias instability), as shown in Fig. 3.15.

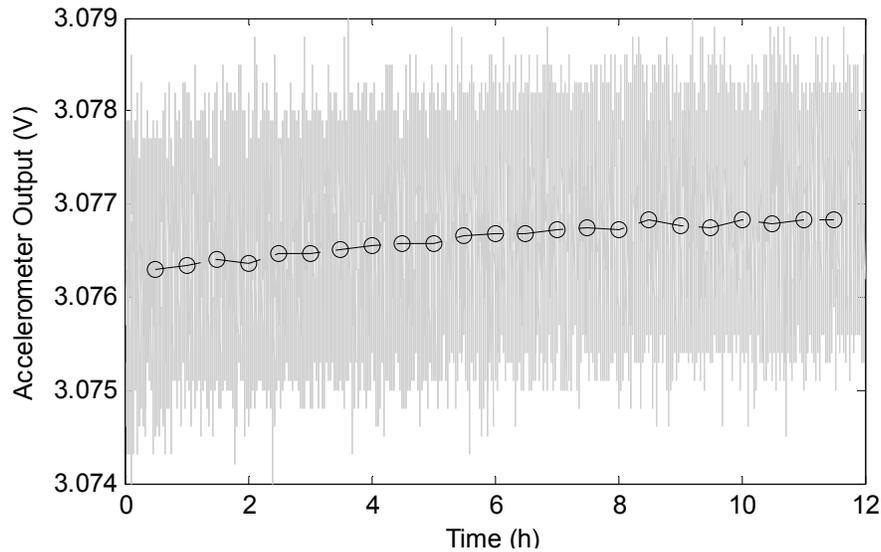


Fig. 3.14 Output of a motionless accelerometer collected over 12 hours at 1 Hz. The black circles are the mean of voltage output every 30 min.

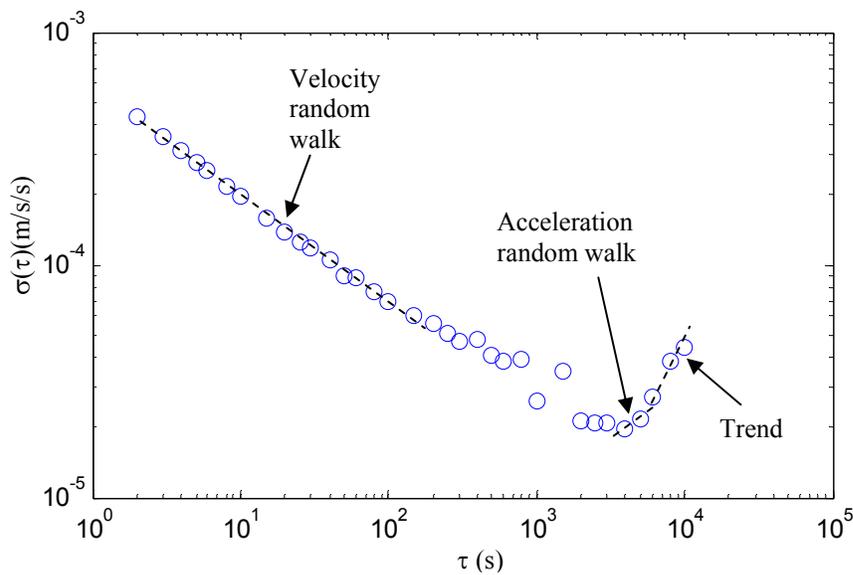


Fig. 3.15 Root Allan variance plot of output of a motionless accelerometer collected over 12 hours at 1 Hz. The dominant noise components are a strong velocity random walk (slope $-1/2$), a weak acceleration random walk (slope $+1/2$), and trend (slope $+1$).

Random walk is a zero-mean Gaussian stochastic process with stationary independent increments and with standard deviation that grows as the root of time. The accelerometer velocity random walk (root Allan variance slope: -0.5 ; units: $(\text{m/s})/\text{h}^{0.5}$) and acceleration random walk (root Allan variance slope: $+0.5$; units: $(\text{m/s}^2)/\text{h}^{0.5}$) are the errors build up with time that are due to white noise in the acceleration and jerk respectively. Trend (root Allan variance slope: $+1$) is a deterministic or systematic error [Tehrani 83], it is the same error as what Barshan and Durrant-Whyte [Barshan 95] referred to as ‘time varying bias’. The value of each of the noise element is found by extrapolating the respective portion of the curve to meet with the vertical line at 1 hour.

Performing the analysis again on the final 4 hours of the same data set, trend disappears from the root Allan variance plots of these four axes, as shown in Fig. 3.16. This finding is consistent with the experimental observation in Section 3.2.1.5, where the accelerometers experience bias and scale factor drifts over time and reach a steady state after a few hours. The drifts are believed to be due to the gradual heating up of the sensor internal circuitry.

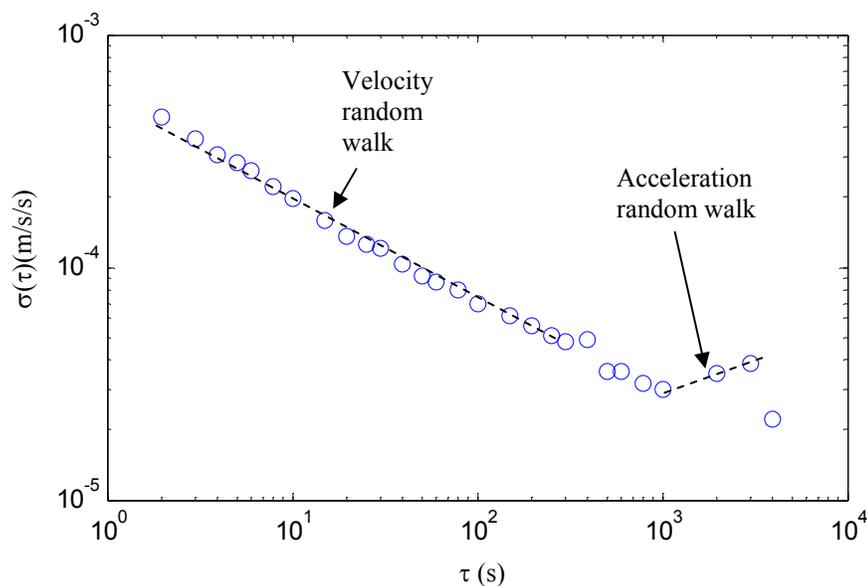


Fig. 3.16 Root Allan variance plot of the final 4 hours recordings of a motionless accelerometer collected over 12 hours at 1 Hz. The dominant noise components are a strong velocity random walk (slope $-1/2$) and a weak acceleration random walk (slope $+1/2$). The trend component (slope $+1$) disappears.

In a subsequent experiment, a set of accelerometer outputs after they reach steady state are recorded for 30 minutes, at 1000 Hz. The data collection period of 30 minutes is the typical duration of a vitreoretinal microsurgery, our targeted application; and 1000 Hz is the sampling rate of our system in operation. The root Allan variance plots in Fig. 3.17 shows that the dominant stochastic noise component of the accelerometers is only velocity random walk.

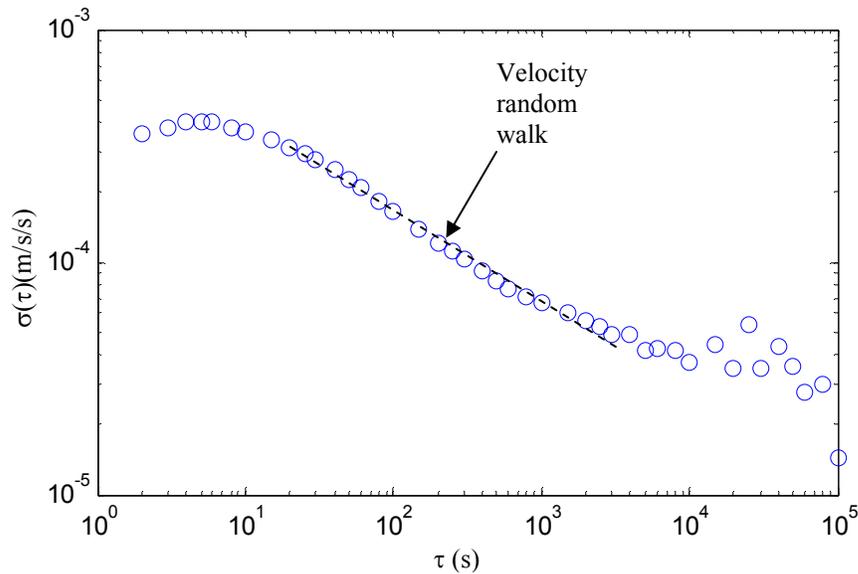


Fig. 3.17 Root Allan variance plot of recordings of a motionless accelerometer collected over 30 min at 1000 Hz, after steady state has been reached. The dominant noise component is only velocity random walk (slope $-1/2$).

3.3.2 Magnetometer

Experiments on the Honeywell HMC-2003 three-axis magnetometer show that it exhibits very similar noise characteristics to the accelerometer. The magnetometer output collected over a 12 hours period show a bias drift in the first four hours (see Fig. 3.18). The root Allan variance plot, as shown in Fig 3.19, reveals that the dominant accelerometer noise sources are a white noise component, an integrated white noise component, and trend. The plot for 30 min, 1000 Hz recording after steady state is shown in Fig. 3.20. Similar to the accelerometers, only white noise component is present.

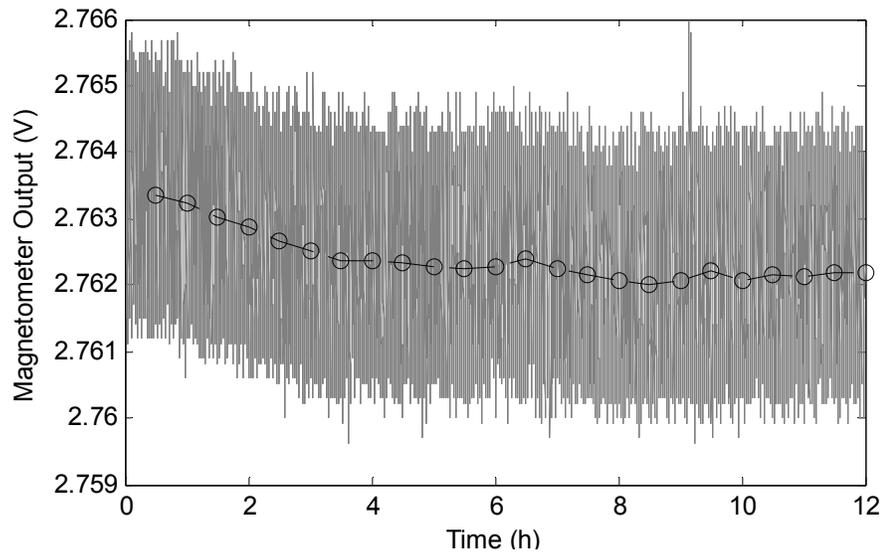


Fig. 3.18 Output of a motionless magnetometer collected over 12 hours at 1 Hz. The black circles are the mean of voltage output every 30 min.

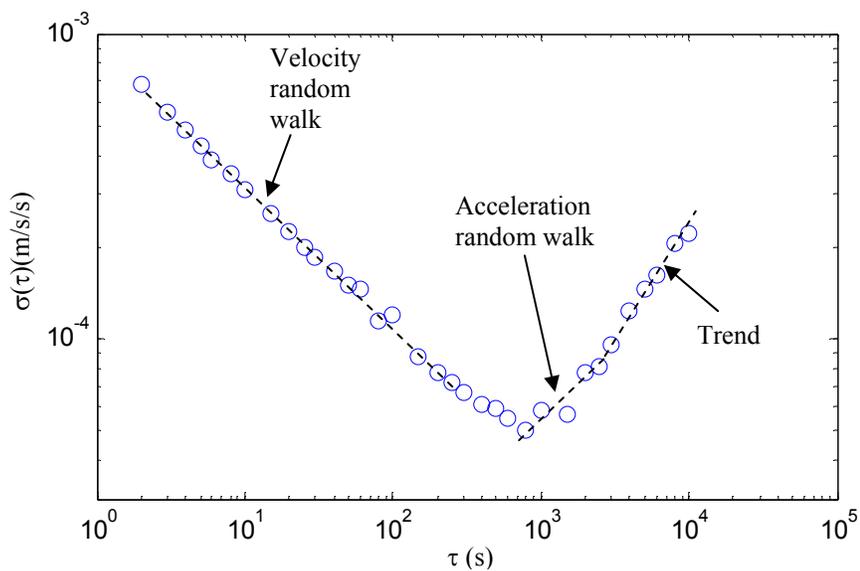


Fig. 3.19 Root Allan variance plot of output of a motionless magnetometer collected over 12 hours at 1 Hz. The dominant noise components are a strong velocity random walk (slope $-1/2$), a weak acceleration random walk (slope $+1/2$), and trend (slope $+1$).

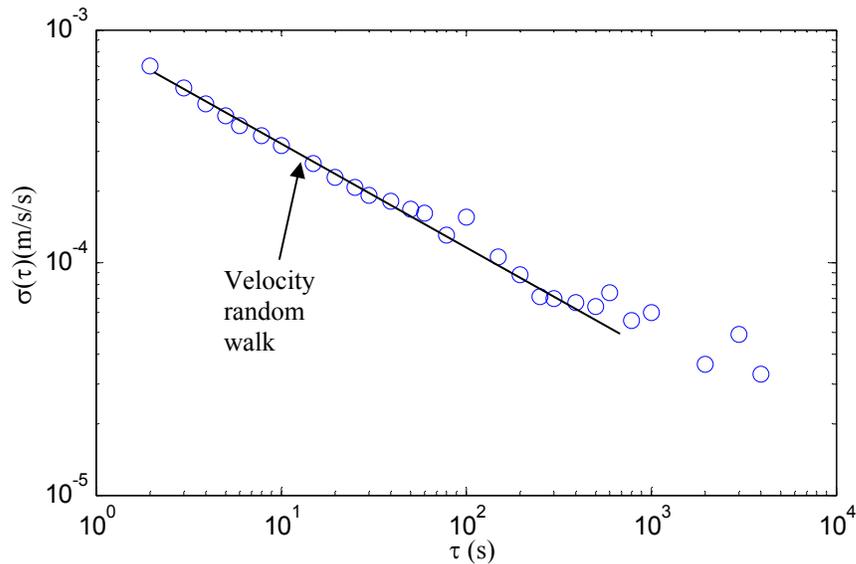


Fig. 3.20 Root Allan variance plot of recordings of a motionless magnetometer collected over 30 min at 1000 Hz, after steady state has been reached. The dominant noise component is only velocity random walk (slope $-\frac{1}{2}$).

References

- [Barbour 01] N. Barbour and G. Schmidt, "Inertial Sensor Technology Trends," *IEEE Sensors Journal*, vol.1. No. 4, pp. 332-339, Dec 2001.
- [Barshan 94] B. Barshan and H. F. Durrant-Whyte, "Inertial Navigation Systems for Mobile Robots", *IEEE Trans. On Robotics and Automation*, Vol.11, No.3, pp. 328-342, Feb 1994.
- [Bourgeois 97] C. Bourgeois, F. Porrent, and A. Hoogerwerf, "Analytical Modeling of Squeeze-Film Damping in Accelerometers", *Proc. IEEE Int'l Conf. Solid-State Sensors and Actuators*, pp. 1117-1120, June 1997.
- [Davis 98] B. S. Davis, "Using Low-Cost MEMS Accelerometers and Gyroscopes as Strapdown IMUs on Rolling Projectiles," *Proc. Position Location and Navigation Symp.*, pp.594-601, Apr. 1998.
- [Ford 99] J. J. Ford, M. E. Evans, "On-line Estimation of Allan Variance Parameters," *Proceedings of Information, Decision and Control IDC 99*, pp.439-444, 1999.
- [Gebre-Egziabher 01] D. Gebre-Egziabher, G. H. Elkaim, J. D. Powell, B. W. Parkinson, "A Non-Linear Two-Step Estimation Algorithm for Calibrating Solid-State Strapdown Magnetometers," *Proc. Int. Conf. on Integrated Navigation Systems*, St. Petersburg, Russia, May 2001.
- [IEEE 1293-1998] IEEE Standard Specification Format Guide and Test Procedure for Linear Single-Axis, Nongyroscopic Accelerometers, Annex I, IEEE Std 1293-1998.
- [Lee 99] W. F. Lee, P. K. Chan, and L. Siek, "Electrical Modeling of MEMS Sensors for Integrated Accelerometer Applications," *Proc. Electron Devices Meeting*, pp.88-91, June 1999.

- [Ng 97] L. C. Ng and D. J. Pines, "Characterization of Ring Laser Gyro Performance Using the Allan Variance Method," *Journal of Guidance, Control, and Dynamics*, Vol.20, No.1, pp.211-214, 1994.
- [Seong 00] S. M. Seong, J. G. Lee, and C. G. Park, "Equivalent ARMA Model Representation for RLG Random Errors," *IEEE Trans. on Aerospace and Electronic Systems*, Vol.36, No.1, pp.286-290, 2000.
- [Tehrani 83] M. Tehrani, "Ring Laser Gyro Data Analysis with Cluster Sampling Technique," *Proceedings of SPIE*, vol.412, pp.207-220, 1983.
- [Verplaetse 96] C. Verplaetse, "Inertial proprioceptive devices: Self-motion-sensing toys and tools," *IBM Systems Journal*, Vol.35, Nos. 3 & 4, pp.639-650, 1996.
- [Weston 00] J. L. Weston and D. H. Titterton, "Modern Inertial Navigation Technology and its Application," *Electronics & Comm. Eng. Journal*, pp.4

Chapter 4

Motion Tracking

4.1 Introduction

In the previous chapter, we have analyzed and developed a measurement model that consists of a deterministic physical model and a stochastic model for the magnetometer-aided all-accelerometer inertial measurement unit. Based on the measurement model, the sensor signal processing steps to track motion in real-time is depicted in Fig. 4.1. The sensor outputs are first corrected for the deterministic errors, then the differential sensing kinematics and the TRIAD algorithms compute the body angular velocity vector and the direction cosine matrix respectively, and they are fused by the augmented state quaternion-based Kalman filter based on their stochastic error characteristics to obtain an orientation estimate. With the updated orientation information, the gravity is subtracted from the vector of six sensed acceleration to find the effective body accelerations. The redundant acceleration information is subsequently fused to get the 3×1 acceleration vector before it is integrated twice to yield the position of the body frame. Finally, the tool-tip position is calculated by simple motion kinematics.

Section 4.2 describes the implementation of the physical model to compensate the deterministic errors. Motion sensing experiments will be presented to demonstrate the effectiveness of the physical model in reducing the sensing errors to the limit of

stochastic noise. Section 4.3 reviews the state of sensor fusion with Kalman filtering to perform orientation tracking using inertial sensing technology. Our implementation of the Kalman filter is presented in Section 4.4, followed by a summary of real-time motion tracking experiments and a discussion in the last two sections.

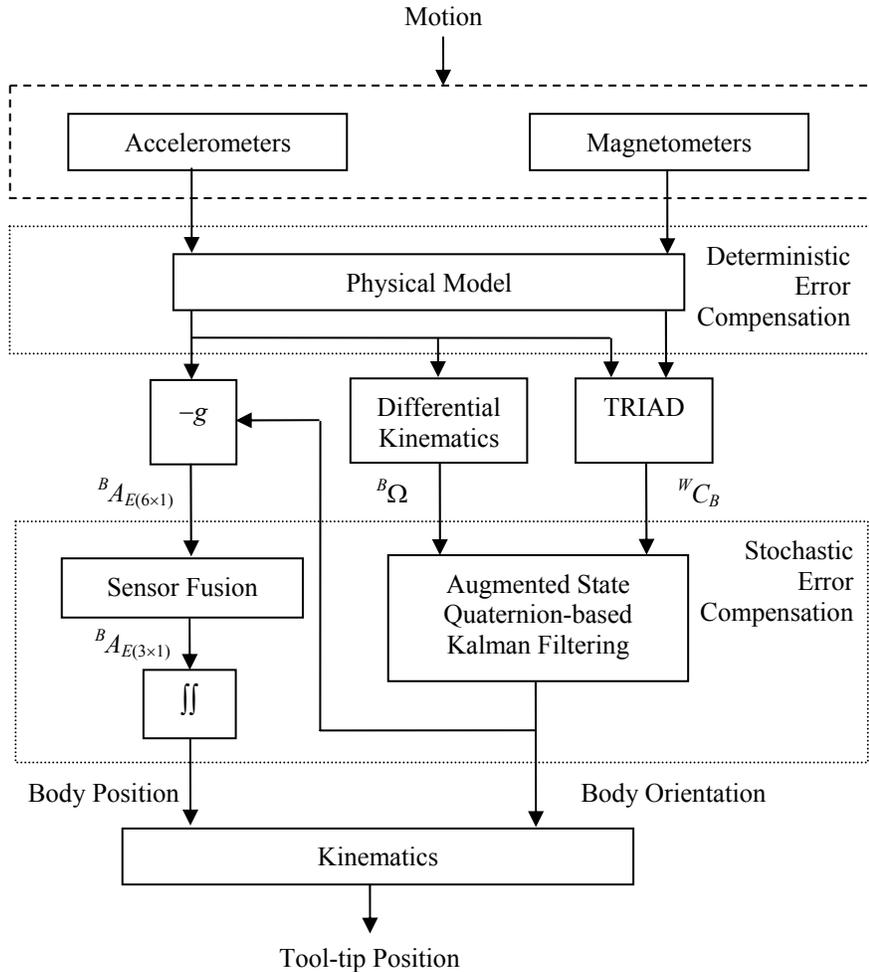


Fig. 4.1 Real-time motion tracking of the intraocular tip position with the magnetometer-aided all-accelerometer inertial measurement unit.

4.2 Deterministic Error Compensation

The physical model is implemented and to be verified in two motion sensing experiments. The first experiment tests the effectiveness of an ADXL-203 accelerometer in sensing tremor like motion. The second experiment uses the accelerometers and the

magnetometers as tilt and pan sensors to compute the static orientation of the instrument using the TRIAD algorithm.

4.2.1 Motion Sensing Experiment - Translation

4.2.1.1 Experimental Setup

The first motion sensing experiment consists of a motion generator and a displacement sensor. The motion generator is a 3-axis piezo-driven linear nano-positioner (Polytec-PI, Inc., Model P-611) and the motion sensor is an infrared interferometer (Philtec, Inc., Model D63) with sub-micrometer resolution. An ADXL-203 accelerometer is mounted on top of the motion generator with its reference edges aligned with the edges of the motion generator. The motion generator is mounted on the rotary stages with an inclination of $\alpha = +45^\circ$, and rotated to $\beta = +90^\circ$ as shown in Fig. 4.2. The definitions of angles α and β are given in Section 3.2. The motion generator generates an oscillation of $60 \mu\text{m}$ p-p at 10 Hz in the x-sensing direction, which is perpendicular to the gravity vector. The interferometer is placed in front of the motion generator aligned with the x-direction to measure the displacement profile of the motion generator. The interferometer is sampled at 1 kHz.

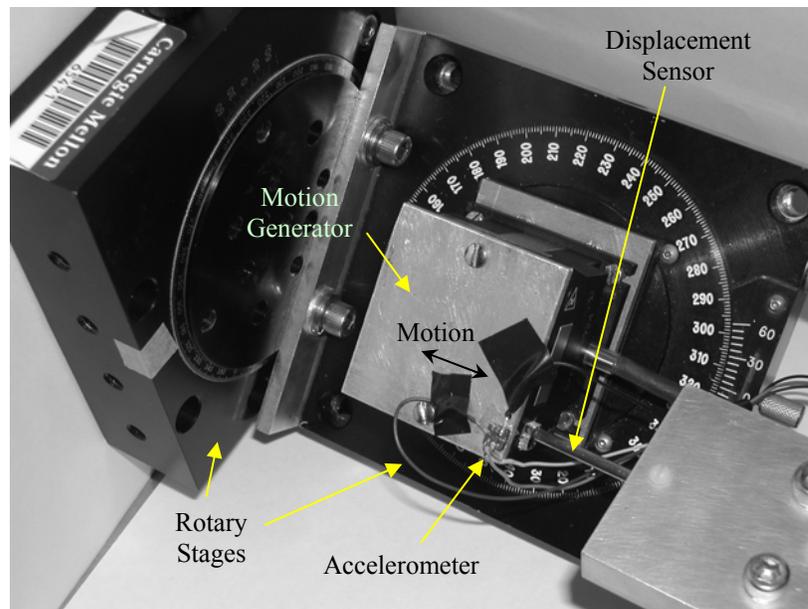
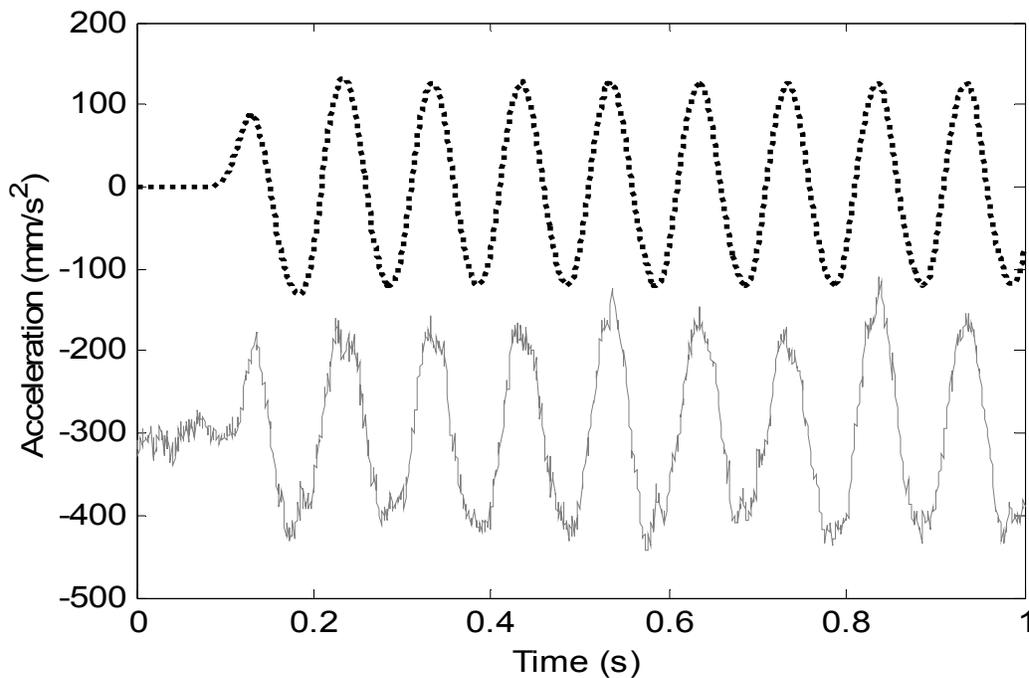


Fig. 4.2 Translational motion sensing experimental setup.

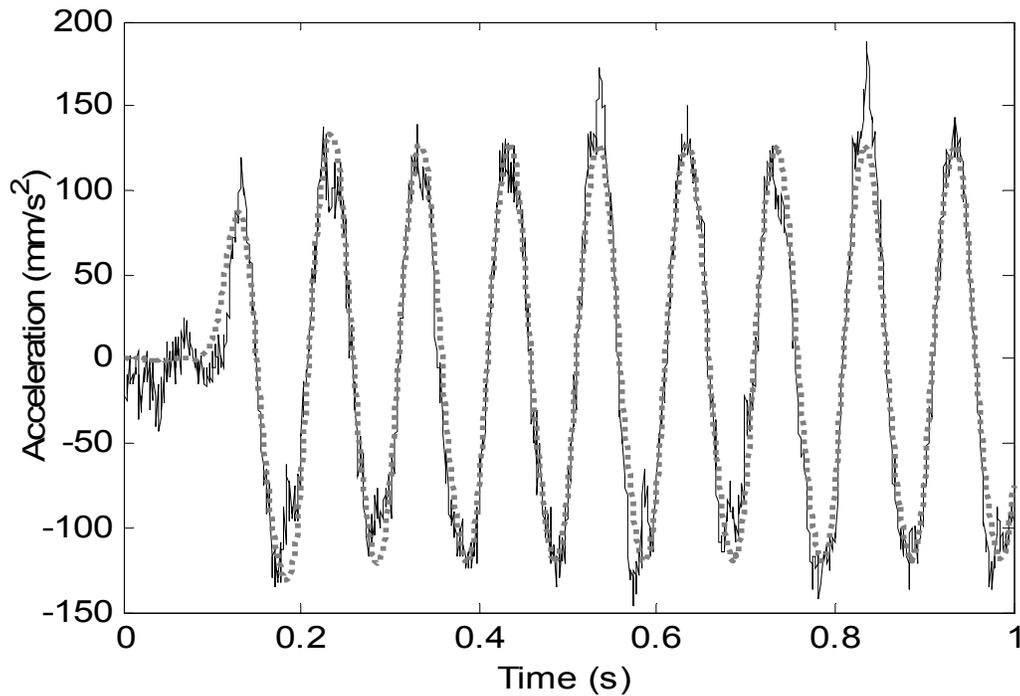
4.2.1.2 Results

The data collected by the interferometer is low-pass filtered with a second-order digital Butterworth filter with cut-off frequency at 20 Hz. It is then corrected for phase shift and attenuation before being differentiated twice to obtain the acceleration profile of the motion generator. The accelerometer voltage output is also corrected for phase shift and attenuation due to the analog R-C low-pass filter and then converted to acceleration in mm/s² based on both the manufacturer's recommended linear model, (3.1) – (3.3), and the proposed physical model, (3.4) – (3.19). The gravity constant g used is 9801.31 mm/s², computed based on the latitude and altitude of our lab.

Fig. 4.3(a) shows the plot of the acceleration profile sensed by the interferometer and that from the accelerometer based on the linear model. The rmse is 299.8 mm/s², most of which is the result of the shift in bias. Fig. 4.3(b) shows the plot of the acceleration profiles sensed by the interferometer and that from the accelerometer based on the proposed physical model. The physical model eliminates most of the bias shift error. The rmse is 20.8 mm/s², an improvement of 93.1% over the linear model.



(a) Manufacturer's recommended linear model



(b) The proposed physical model

Fig. 4.3 Results of (a) the manufacturer's recommended linear model vs (b) the proposed physical model. The dotted line is the computed acceleration from the interferometer. The solid line is the measured acceleration from the x-accelerometer.

4.2.2 Motion Sensing Experiment – Static Orientation

4.2.2.1 Experimental Setup

Three experiments are conducted to evaluate the Micron sensing system in sensing static orientation with the measurement model implemented.

The first experiment is to test the accelerometers in tilt sensing. The Micron sensing system is mounted on an extension arm attached to a roll-pitch-yaw unit made up of one manual and two motorized rotary stages. The motorized and the manual rotary stages all have positioning accuracy of less than 0.1° . The experiment setup is identical to that shown in Fig. 3.13 and is reproduced in Fig. 4.5. The orientation of the extension arm is initialized to be perpendicular with gravity with a liquid level. The two motorized stages are setup to produce pitch and yaw or the two tilt angles. The motorized stages are programmed to move the Micron instrument through a series of orientations. Tilt angles

results obtained from manufacturer recommended linear model are also included for comparison. The second experiment is a pan test for the magnetometer. A similar setup as shown in Fig. 4.5 is used, with minor adjustment to avoid collision with the table edges and the clamp. The Micron sensing system is rotated through a series of pan angles with different tilt configurations. The last experiment is the combination of the pan and tilt information from the magnetometers and accelerometers respectively via the TRIAD algorithm in Section 2.3.2. The results are tabulated in Table 4.3.

To account for the stochasticity of the sensors, the mean of 500 data points is recorded at each orientation.

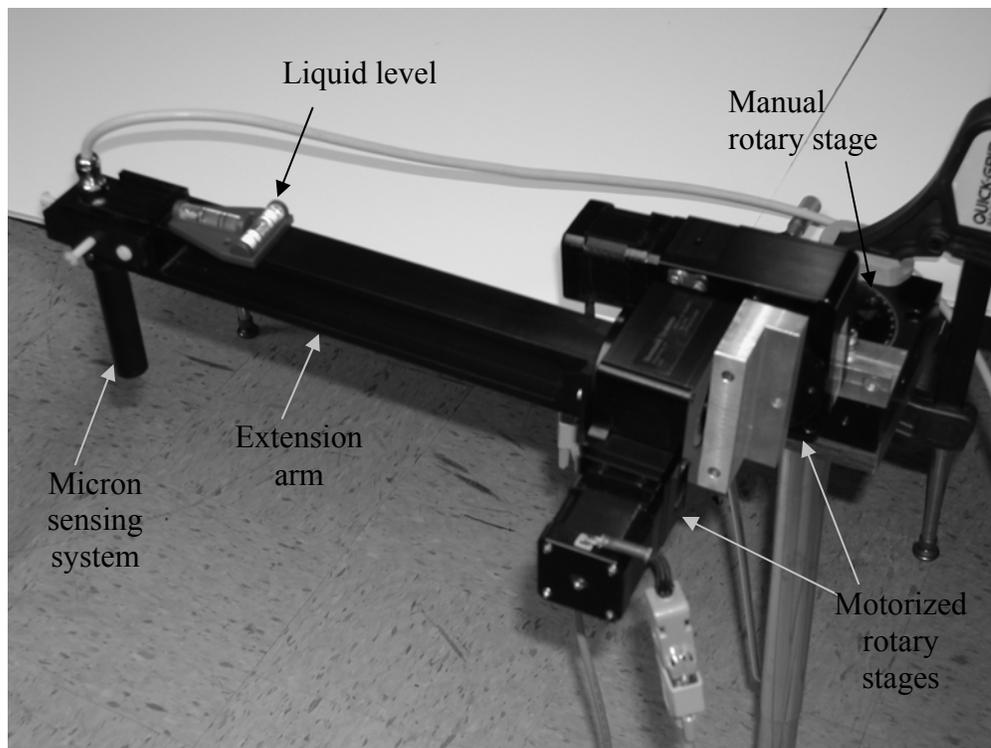


Fig. 4.4 Static orientation experimental setup.

4.2.2.2 Results

Fig. 4.5-4.7 show the errors in tilt estimation of the accelerometers with the proposed physical model and with the manufacturer recommended linear model in the first experiment. The numerical results of the experiment are tabulated in Table 4.1. Fig. 4.8 shows the errors in pan estimation of the magnetometer with the proposed physical model. The numerical results of the experiment are tabulated in Table 4.2.

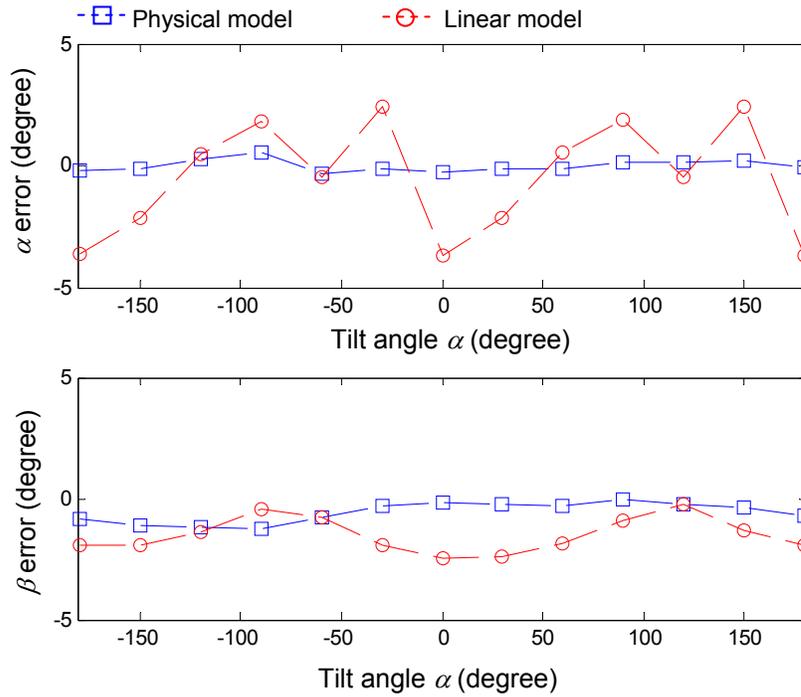


Fig. 4.5 Errors in tilt angle estimation using the proposed physical model and the manufacturer. In this experiment, $\beta = 0^\circ$ and α is rotated from -180° to 180° at 30° interval.

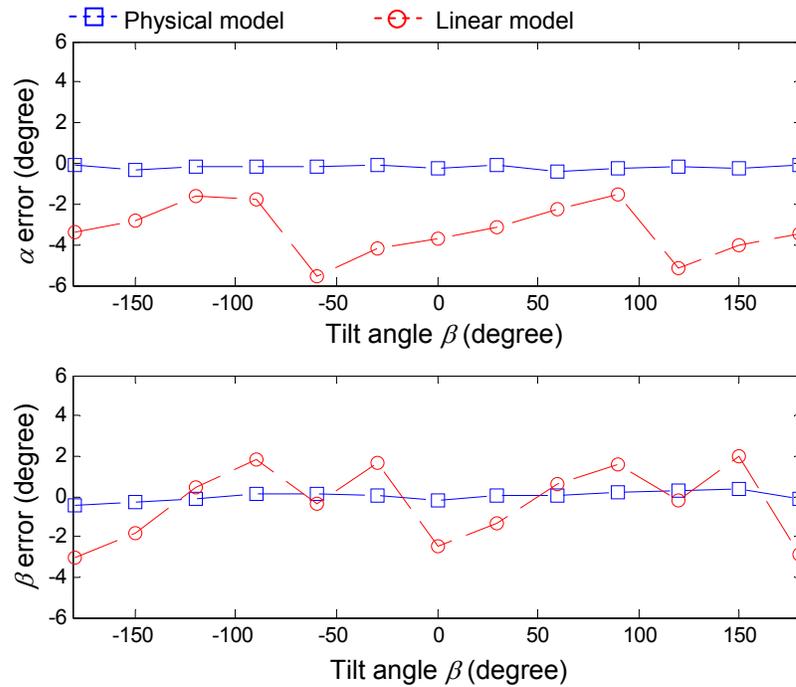


Fig. 4.6 Errors in tilt angle estimation using the proposed physical model and the manufacturer. In this experiment, $\alpha = 0^\circ$ and β is rotated from -180° to 180° at 30° interval.

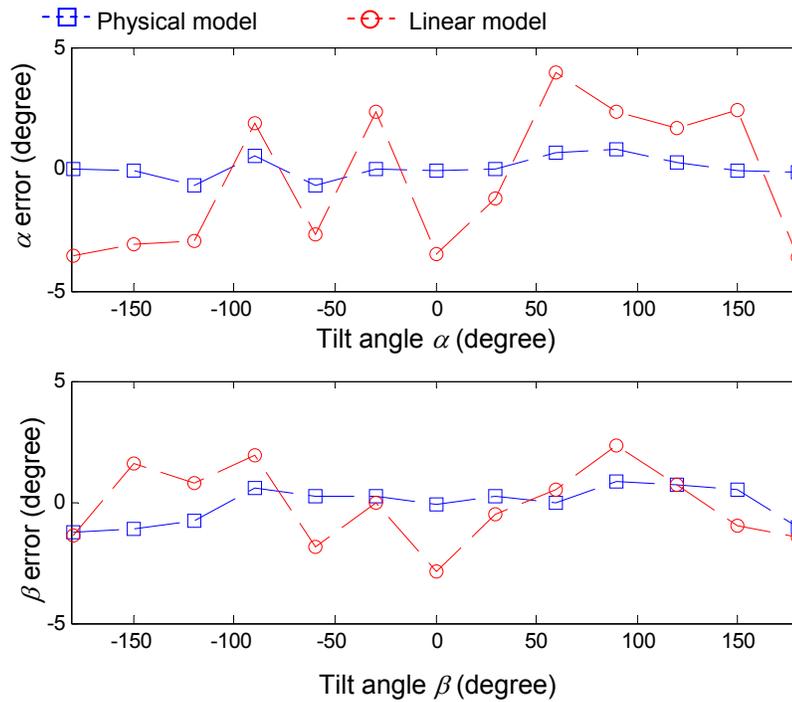


Fig. 4.7 Errors in tilt angle estimation using the proposed physical model and the manufacturer. In this experiment, α and β are rotated simultaneously from -180° to 180° at 30° interval.

Table 4.1 Rmse and maximum error of the tilt estimation experiments.

	Tilt angle α			Tilt angle β		
	Linear Model	Physical Model	% reduction	Linear Model	Physical Model	% reduction
Rmse (deg)	2.88	0.32	88.9	1.64	0.60	63.4
Max error (deg)	3.95	0.82	79.2	2.36	0.82	65.3

4.2.3 Discussion

For the accelerometers, the proposed physical model performed well compared to the manufacturer's recommended linear model as described by (3.3). From Table 4.1, the rmse are reduced by 88.9% and 63.4% and the maximum error by 79.2% and 65.2% for the α and β angles respectively.

The action of each component of the physical model is depicted in Fig. 4.4. The accelerometer is first calibrated at $\alpha = +90^\circ$, which gives the line labeled 'Linear Model'.

The accuracy of the proposed physical model is limited by the stochasticity and inconsistent behavior of the accelerometer between operations. The random noise makes it difficult to model the misalignment accurately. In our application, in which the tremor acceleration to be sensed is about two orders of magnitude smaller than g , small orientation error in the sensing axes will result in a gravity residue comparable to the motion acceleration. The variation of the temperature drift pattern handicaps the usefulness of modeling the drift. While the practicality of ovenization is still under investigation, waiting for the accelerometer to reach its steady state seems to be the most robust solution.

The magnetometer is producing a pretty large error in the pan angle estimation. The Earth's magnetic-North vector in the northern hemisphere has a tilt component, and the tilt angle it is tested to be about 55° from the line of horizon. This means that the magnetometer is more sensitive to a tilt angle change than angle change. In other words, the effective voltage output range in pan sensing is reduced and would thus be more sensitive to noise and disturbance. However, the pan angle estimation is not as important as the tilt angle in our application since the removal of gravity does not depend on it. In addition, the pan angle estimation by the magnetometer is non-drifting, and any errors in pan angle will be removed together with the other low frequency motion components during tremor filtering.

4.3 Sensor Fusion with Kalman Filtering

4.3.1 Kalman Filter vs Complementary Filter

Inertial sensing technology has been used extensively in aerospace navigation and in robotic motion sensing in the past few decades. To treat the notorious integrating drifts of inertial sensors, most of these applications employ a variation of Kalman filters and/or complimentary filters to perform real-time sensor fusion. These techniques typically use other on-board or external sensing sources to provide non-drifting references to bind the inertial sensor drifts.

It is widely recognized that the most important and difficult part of Kalman filtering is the modeling of the state variables to be optimized. A good model has to succinctly capture the dynamics of the system and yet simple enough to be implemented. Even with more than 40 years of development notwithstanding, successful design of a Kalman filter is still part science, part art, and part trial and error.

In aided inertial motion tracking applications, the state variables of a Kalman filter usually take one of the two forms: the sensed engineering quantities (i.e. acceleration, velocity, and attitude etc.) or their errors.

A Kalman filter that operates on the error states is called a complementary Kalman filter. The optimal estimates of the errors are then subtracted from the sensed quantities to obtain the optimal estimates. The complementary Kalman filter, as shown in Fig. 4.1, has become the standard methodology of integrating non-inertial with inertial measurements in aviation and missile navigation since the 1960s. This method requires dynamic models for both the navigation variable states and error states [Britting 71; Benson 75; Weinred 78; Friedland 78; Shibata 86; Goshen-Meskin 92].

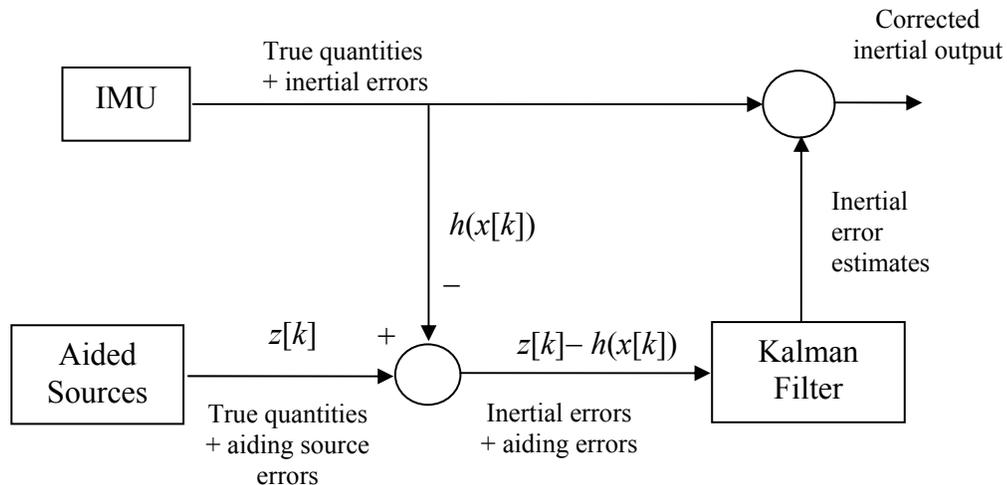


Fig. 4.9 A complementary Kalman filter.

Brown and Hwang [Brown 97] offer three reasons for this popular selection. Firstly, this method provides a degree of generality that readily accepts a wide variety of aiding measurements from different sensors. This is important as the combination of aiding

sensors may vary throughout the flight mission. Secondly, if a nominal trajectory can be pre-computed, which is a reasonable scenario for aviation and missile navigation, the differencing operation $z[k] - h(x[k])$ linearizes the nonlinear process dynamic in the inertial system and the measurement relationships, thus resulting in a linearized Kalman filter. And last but not least, this approach maintains high dynamic response of the sensed engineering quantities while frequency-selective filters operate on the error terms. For example, to fuse velocity measurements from a tachometer and an accelerometer more effectively, one common practice is to reduce the sensors' noise by low-passing the tachometer output and high-passing the integrated accelerometer output. The complementary filter approach allows the important dynamic quantities (in this case the velocity) to pass through the system without any distortion or delay, and hence the nickname distortionless filtering.

For many other inertial sensing applications other than aviation and missile navigation, most or all of the listed beneficial features of a complementary Kalman filter do not exist, therefore there is no compelling reason to choose a complementary Kalman filter over a regular Kalman filter (including extended Kalman filter). On the contrary, in actual implementation it is often more convenient to have the state variables in total quantities than in incremental quantities.

4.3.2 Human Motion Tracking

In more recent years, there is an increasing interest in the application of inertial sensing technology to sense human movement, especially in the entertainment and medical applications.

Foxlin [Foxlin 98] presents two commercially available miniature inertial sensing systems for tracking head-mounted display in virtual reality applications. The core sensing module in both systems is made up of a triad of accelerometers, a triad of rate gyros, and a triad of magnetometers, called InertiaCubeTM. The InertiaCubeTM is a three DOF orientation tracker that uses the gravity and the Earth's magnetic field as stationary references to correct for the drifts in the integrated tilt (roll and pitch) and pan (yaw) estimates from the gyros. Sensor fusion is performed via a separate bias complimentary

Kalman filter [Foxlin 96]. The algorithm adopts a heuristic approach to ignore the accelerometers readings and rely solely on the gyros when the gravity estimation is corrupted by body movement. The process and measurement error covariance matrices and in turn the Kalman gains are computed, in the author's own words, in a 'sloppy' manner. The intrinsic sensor errors such as the bias and scale factor nonlinearity and temperature dependent variations are not explicitly modeled.

Bachmann *et al.* [Bachmann 03] and Yun *et al.* [Yun 03] invented the MARG (Magnetic (3 axes), Angular Rate (3 axes), and Gravity (3 axes)) sensors for tracking human body movements and apply these motion information to control human models in virtual environment. The sensor unit is completed with an on-board microcontroller and signal conditioning circuitry. Taking advantage of the numerical properties of quaternion orientation representation, a quaternion-based Kalman filter is implemented to fuse the information from the vector of nine sensors. To avoid the computational load of solving the nonlinear measurement equations, the authors adopted a numerical approach proposed by Gebre-Egziabher [Gebre-Egziabher 00] to obtain the quaternions using Gauss-Newton method. Estimation of the gravity vector is by averaging the accelerometer outputs over an unspecified period of time, on the assumption that human movements do not produce constant linear acceleration. The gyro error dynamic is assumed to follow a Gauss-Markov model, but the computation and propagation of error covariance matrices has not been discussed and the Kalman gains are determined through experimental 'tweaking'. Bias and scale factor of each sensor are modeled using linear models.

Other Kalman filtering implementations in human motion tracking with a conventional three-gyro, three-accelerometer inertial measurement unit (IMU) include Rehbinder and Hu [Rehbinder 01] and Luinge [Luinge 02]. Without complementary sensors in the pan or yaw DOF, drift in the heading estimate will be inevitable.

Besides the fundamental problem of integration drift associated with the inertial sensing technology, one other major challenge in implementing real-time orientation tracking is the 'true' real-time signal processing requirement of our application. The use of frequency selective filters to reduce the sensor noise would have caused a phase shift

in the processed signal. As a result, the delay between the instantaneous instrument motion and the perceived motion defeats the objective of real-time active erroneous motion compensation. The direct consequence of this is the difficulty in obtaining high quality sensing information from the noisy sensors. This inherent problem in our application motivates the implementation of an effective Kalman filter to decipher high quality sensing information from noisy sensors.

The addition of the three-axis magnetometer to the inertial sensor suite allows us to form two complementary orientation sensing sources. On one hand, we have the all-accelerometer IMU with very high sensing resolution but poor long term sensing accuracy. On the other hand, the magnetic North vector may be combined with the gravity vector to provide non-drifting but low resolution orientation estimates. The poor sensing resolution is a result of the noisy magnetometer and the residual body acceleration of the instrument in estimating the gravity vector. Sensor fusion of these two sensing sources to obtain a high resolution, non-drifting orientation estimate forms the subject matter of the remaining sections of the chapter.

4.4 Augmented State Quaternion-Based Kalman Filtering

4.4.1 Orientation Representation

The lack of physical intuition in the quaternion orientation representation notwithstanding, it has been very well received in the aerospace navigation community because of its numerical advantages. The quaternion eliminates the singularity problem and computational inefficiency associated with solving the transcendental equations in the Euler angles representation. Moreover, quaternion is robust against numerical errors that crept into the direction cosine matrix when updated by numerical integration methods, as a consequence of truncation, rounding off and commutation.

In general, the updating of the orientation orthogonal matrix by numerical integration of a matrix differential equation results in three transformation errors: skew, scale and drift errors. Two-dimensional examples of these errors are shown in Fig. 4.7.

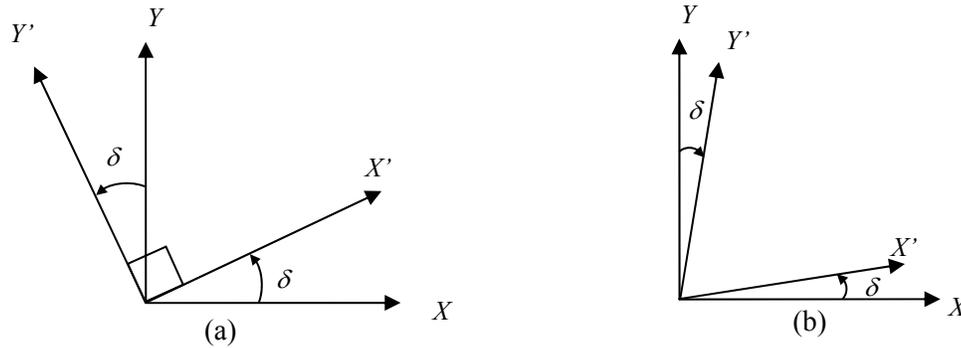


Fig. 4.10 X and Y are the error-free orthogonal axes. An erroneous rotation δ is introduced by numerical integration. (a) Drift error. The X' and Y' axes remain mutually orthogonal and unit length; (b) Scale and skew errors. The X' and Y' axes are not unit length and not orthogonal.

It is well documented [Wilcox 67; Mortensen 74; Friedland 78; Jiang 91; Lovren 98] that the intrinsic property of the quaternions eliminates the skew error. [Lovren 98] concludes through simulations that the drift error introduced by numerical integration (second order approximation) is negligible compared to the analytical approach in the classical coning motion example. The classical coning motion is one where the angular velocity vector has a constant magnitude but the direction is rotating about the nominal forward motion axis with an increasing deviation angle. It is widely known that this motion maximizes the non-commutative vector (a rotation vector times an angular velocity). [Lovren'98] also shows that the scale error introduced by numerical transformation is large comparing to the analytical case in the coning motion example.

From these literatures, we can conclude the following about numerical transformation errors in orientation estimation with quaternions:

- (i) Skew error, $\delta q_{sk} = 0$;
- (ii) Drift error, δq_{dr} is small using a second order approximation numerical integration method.
- (iii) Scale error, δq_{sc} is large comparing to the analytical integration approach. However, this can be easily eliminated by normalizing the quaternion.

Updating of position vectors by numerical integration introduces only drift error, by the same argument we have for orientation estimation, it is small enough to be neglected.

The quaternion is defined as

$$Q = [q_0 \quad \bar{q}]^T, \quad (4.1)$$

where

$$q_0 = \cos\left(\frac{1}{2} \|\Theta\|\right), \quad (4.2)$$

$$\bar{q} = [q_1 \quad q_2 \quad q_3]^T = \sin\left(\frac{1}{2} \|\Theta\|\right) \frac{\Theta}{\|\Theta\|}, \quad (4.3)$$

$$\Theta = \Omega T = [\theta_x \quad \theta_y \quad \theta_z]^T. \quad (4.4)$$

The unit vector $\Theta/\|\Theta\|$ is the axis of rotation and $\|\Theta\|$ is the angle of rotation, in the axis-angle representation. The quaternion has three DOF and satisfies the constraint

$$Q^T Q = 1. \quad (4.5)$$

4.4.2 Angular Velocity Bias Vector

We have discussed in the Section II-B that the body angular velocity vector ${}^B\Omega$ derived from the differential sensing kinematics algorithm is corrupted by an integrated drift. This drift may be treated as a bias. Thus, the true body angular velocity vector Ω is related to ${}^B\Omega$ by

$$\Omega = {}^B\Omega - B - \xi_\Omega, \quad (4.6)$$

where ξ_Ω is a vector of transformed zero mean Gaussian white noise processes that drives the angular velocity vector, and B is angular velocity bias vector assumed to be driven by transformed integrated Gaussian white noise processes,

$$\dot{B}(t) = \xi_B(t), \quad (4.7)$$

Although the noise vectors ξ_Ω and ξ_B are no longer Gaussian after nonlinear transformations, they are assumed to remain Gaussian for ease of implementation, the consequence is that the estimation will be suboptimal. Noise vectors ξ_Ω and ξ_B are assumed to be uncorrelated. The derivation of the covariance matrices of ξ_Ω and ξ_B will be discussed in Section 4.4.3.3.

4.4.3 The Process Model

4.4.3.1 State Vector

The state vector is selected to be the 4-tuple quaternion Q , augmented with the bias vector B of the body angular velocity,

$$x(t) = [Q(t)_{(4 \times 1)} \ B(t)_{(3 \times 1)}]^T. \quad (4.8)$$

It can be shown that the quaternion is related to the angular velocity bias vector by the differential equation

$$\dot{Q}(t) = \frac{1}{2} \tilde{\Omega}_\times \left({}^B \Omega(t) - B(t) - \xi_\Omega(t) \right) Q(t), \quad (4.9)$$

where

$$\tilde{\Omega}_\times = \begin{bmatrix} 0 & -\Omega^T \\ \Omega & \Omega_\times \end{bmatrix}, \quad (4.10)$$

and Ω_\times is the cross product matrix of Ω . Equation (4.9) may be rewritten as

$$\dot{Q}(t) = \frac{1}{2} \tilde{\Omega}_\times \left({}^B \Omega(t) - \xi_1(t) \right) Q(t) - \frac{1}{2} \tilde{Q}(t) B(t), \quad (4.11)$$

since $\tilde{\Omega}_\times$ is linear and homogeneous in its argument,

$$\tilde{\Omega}_\times (B) Q = \tilde{Q} B, \quad (4.12)$$

$$\tilde{Q} = \begin{bmatrix} q_0 & -q_3 & q_2 \\ q_3 & q_0 & -q_1 \\ -q_2 & q_1 & q_0 \\ -q_1 & -q_2 & -q_3 \end{bmatrix}. \quad (4.13)$$

Derivation and properties of (4.13) can be found in [Lefferts 82].

4.4.3.2 Discrete State Equation

Assuming the angular displacement is small and a constant angular acceleration motion model within one sampling period, the discrete state dynamic equation or process model is defined as

$$x[k] = \Phi[k]x[k-1] + \Psi[k]w[k], \quad (4.14)$$

where the state transition matrix is given by

$$\Phi = \begin{bmatrix} \Phi_{Q(4 \times 4)} & \Phi_{B(4 \times 3)} \\ \mathbf{0}_{(3 \times 4)} & I_{(3 \times 3)} \end{bmatrix}; \quad (4.15)$$

$$\Phi_Q = \cos\left(\frac{1}{2} \|\Theta\|\right) I_{(4 \times 4)} + \frac{\sin\left(\frac{1}{2} \|\Theta\|\right)}{\|\Theta\|} \tilde{\Theta}_\times, \quad \tilde{\Theta}_\times = \tilde{\Omega}_\times T; \quad (4.16)$$

$$\Phi_B = -\frac{1}{2} \Phi_Q \tilde{Q}. \quad (4.17)$$

The noise vector and the noise transformation matrices

$$w[k] = [\xi_\Omega[k] \ \xi_B[k]]^T, \quad (4.18)$$

$$\Psi = \begin{bmatrix} -\frac{1}{2} \tilde{Q} & \mathbf{0}_{(4 \times 3)} \\ \mathbf{0}_{(3 \times 3)} & I_{(3 \times 3)} \end{bmatrix}. \quad (4.19)$$

The process noise covariance matrix is thus

$$W[k] = \begin{bmatrix} \frac{1}{4} \tilde{Q}[k] \Xi_\Omega[k] \tilde{Q}^T[k] & \mathbf{0} \\ \mathbf{0} & \Xi_B[k] \end{bmatrix}, \quad (4.20)$$

where $\Xi_\Omega[k]_{(3 \times 3)}$ is the noise covariance matrix of the angular velocity ${}^B\Omega$ to be derived in the next section.

4.4.3.3 State Noise Covariance Matrix

We have known from the Allan variance analysis that the dominant noise component of the ADXL-203 accelerometers is velocity random walk or a zero-mean Gaussian white noise process ξ_A . Without loss of generality, we assume all accelerometer axes have identical noise characteristic and uncorrelated. The noise covariance matrix of the differential acceleration vector $A_D = [a_{13x} \ a_{23y} \ a_{12z}]^T$ from (2.39) is rewritten as

$$\Xi_{AD} = E\{(\xi_{AD})(\xi_{AD})^T\} = 2\sigma_A^2 I_{(3 \times 3)}, \quad (4.21)$$

where σ_A^2 is the variance of ξ_A .

From Chapter 2, the body angular velocity ${}^B\Omega$ is related to the differential acceleration vector A_D and its own time derivative ${}^B\dot{\Omega}$ by a nonlinear function,

$${}^B\Omega(t) = F(A_D(t), {}^B\dot{\Omega}(t)). \quad (4.22)$$

The noise of ${}^B\Omega(t)$ thus has a transformed white noise component $\xi_\Omega(t)$ associated with $A_D(t)$ and a transformed integrated white noise component $\xi_B(t)$ due to ${}^B\dot{\Omega}$.

In the absence of an analytical solution for (4.22), the scaled unscented transformation [Julier 00] is applied to compute the noise covariance of ${}^B\Omega$. The scaled unscented transformation is motivated by the notion of it is easier to approximate a probability distribution than it is to approximate a nonlinear function.

To find the noise covariance of $\xi_\Omega(t)$, we first disregard the effect of ${}^B\dot{\Omega}$. For the three-dimensional ($n = 3$) angular velocity vector, a set of at least seven ($2n + 1$) weighted samples, called the sigma points, are generated. This sample set is zero mean and has covariance equal to the noise covariance of the differential acceleration vector A_D . The nonlinear transformation F is applied to each of the sigma points, and the covariance of the body angular velocity vector is approximated by the calculated covariance of the transformed sample set.

The sigma points S are selected by the algorithm:

$$\begin{aligned} S_0 &= 0, \quad c_0 = \kappa/(n + \kappa); \\ S_i &= \left(\sqrt{(n + \kappa)\Xi_{AD}} \right)_i, \quad c_i = 1/2(n + \kappa); \\ S_{i+n} &= -\left(\sqrt{(n + \kappa)\Xi_{AD}} \right)_i, \quad c_{i+n} = 1/2(n + \kappa); \end{aligned} \quad (4.23)$$

where $\kappa \in \Re$ and $(n + \kappa) \neq 0$, $\left(\sqrt{(n + \kappa)\Xi_{AD}} \right)_i$ is the i^{th} row or column of the matrix square root of $(n + \kappa)\Xi_{AD}$, and c_i is the weight associated with the i^{th} sigma point.

The mean and noise covariance of ${}^B\Omega$ is computed as

$${}^B\bar{\Omega} = \sum_{i=0}^{2n} c_i \hat{S}_i, \quad (4.24)$$

$$\Xi_\Omega = \sum_{i=0}^{2n} c_i \left(\hat{S}_i - {}^B\bar{\Omega} \right) \left(\hat{S}_i - {}^B\bar{\Omega} \right)^T, \quad (4.25)$$

where

$$\hat{S}_i = F(S_i). \quad (4.26)$$

The noise covariance of the angular velocity bias vector Ξ_B due to the noise in ${}^B\dot{\Omega}$ is found by repeating (4.23)–(4.36), with the effect of A_D disregarded.

4.4.4 The State Measurement

4.4.4.1 Gravity Vector Estimation

The TRIAD algorithm assumes a well-defined gravity vector, which is not readily available for a body in continuous motion. However, the body acceleration of the instrument during microsurgery is typically two orders of magnitude smaller than the gravity vector. Therefore, approximating the gravity vector with the raw accelerometer outputs would give a reasonably high quality estimate of about 1% error.

Acceleration content of the instrument in operation includes a high frequency (8-12 Hz) tremulous motion component and a low frequency (< 1 Hz) motion component, made up of primary the voluntary hand movement and other non-tremulous physiological errors. The corruption of the gravity estimate by these components results in a noise process characterized by superimposed white and colored noise sequences. A formal treatment of the colored noise is to model it as a bias and augment the Kalman filter with extra states. For ease of implementation, we assume the gravity noise vector to be white noise processes. The consequence of this assumption is that the state estimates of the Kalman filter will become suboptimal.

4.4.4.2 Measurement Vector

The measurement vector is chosen to be the quaternion derived from the direction cosine matrix from the TRIAD algorithm in (2.20),

$$z = \begin{bmatrix} q_{m0} \\ q_{m1} \\ q_{m2} \\ q_{m3} \end{bmatrix} = \begin{bmatrix} \sqrt{1 + c_{11} + c_{22} + c_{33}}/2 \\ (c_{23} - c_{32})/4q_{m0} \\ (c_{31} - c_{13})/4q_{m0} \\ (c_{12} - c_{21})/4q_{m0} \end{bmatrix}, \quad (4.27)$$

where c_{ij} , $i, j = 1, 2, 3$, are the elements of ${}^W C_B$.

4.4.4.3 Measurement Equation

The measurement equation is

$$z[k] = Hx[k] + v[k], \quad (4.28)$$

$$H = \begin{bmatrix} I_{(4 \times 4)} & 0_{(4 \times 3)} \\ 0_{(3 \times 4)} & 0_{(3 \times 3)} \end{bmatrix}, \quad (4.29)$$

where H is the measurement matrix and $v[k]$ is the measurement noise vector.

4.4.4.4 Measurement Noise Covariance Matrix

The gravity noise vector has two components, a white noise process with variance associated with the accelerometers and a colored noise process due to the body acceleration of the instrument. The noise covariance of the unit gravity vector is given by

$$\Xi_G = \sigma_G^2 I = \frac{\sigma_W^2 + \sigma_C^2}{\|{}^B G\|^2} I, \quad (4.30)$$

where σ_G^2 is the variance of a component of gravity vector ${}^B G$ along a direction normal to the expectation value of the gravity vector, $E\{{}^B G\}$. σ_W^2 is the variance of the white noise process. Since there are two accelerometers in each sensing direction, we may compound the two sensor readings and reduce the noise variance by half,

$$\frac{1}{\sigma_W^2} = \frac{1}{\sigma_A^2} + \frac{1}{\sigma_A^2} \Rightarrow \sigma_W^2 = \frac{1}{2} \sigma_A^2, \quad (4.31)$$

where σ_A^2 is the variance of ξ_A . σ_C^2 is the variance of the color noise process, it is assumed to be equal to 1% of the magnitude of the gravity,

$$\sigma_C = 0.01 \|G\|. \quad (4.32)$$

Although not strictly true, the components of colored noise vector are assumed to be uncorrelated.

We have found from the Allan variance analysis that the dominant noise component of the magnetometers is a vector of zero-mean Gaussian white noise processes ξ_M , where ξ_M and ξ_A are assumed to be uncorrelated. Assuming also that the three magnetometer axes are uncorrelated, the noise covariance of the magnetometer is thus $\sigma_M^2 I$. Therefore, the noise covariance of the unit magnetic North vector is given by

$$\Xi_N = \sigma_N^2 I = \frac{\sigma_M^2}{\|{}^B M\|^2} I, \quad (4.33)$$

where $\|{}^B M\|$ is the magnitude of the magnetic field.

The covariance of the error angle vector associated with the direction cosine matrix determined by the TRIAD algorithm can be shown to be in the form [Shuster 81]

$$\begin{aligned} \Xi_{\theta} = & \sigma_G^2 I + \frac{1}{\|{}^B z_W \times {}^B N\|^2} \left\{ (\sigma_N^2 - \sigma_G^2) {}^B z_W {}^B z_W^T + \right. \\ & \left. \sigma_G^2 ({}^B z_W \bullet {}^B N) ({}^B z_W {}^B N^T + {}^B N {}^B z_W^T) \right\}, \end{aligned} \quad (4.34)$$

where ${}^B z_W$ is the negated normalized gravity vector. The measurement noise covariance is thus

$$V[k] = \frac{1}{4} \tilde{Q} \Xi_{\theta} \tilde{Q}^T. \quad (4.35)$$

4.4.5 Kalman Filtering

4.4.5.1 Prediction

From (4.14), the predicted discrete state equation is

$$x^-[k] = \Phi[k]x[k-1]. \quad (4.36)$$

The projected error covariance is computed as

$$P[k] = \Phi[k]P[k-1]\Phi^T[k] + W[k]. \quad (4.37)$$

4.4.5.2 Filtering

The Kalman gain is obtained by compounding the noise covariance matrices of the state equation and measurement equation,

$$K[k] = P^-[k]H^T(HP^-[k]H^T + V[k])^{-1}. \quad (4.38)$$

The updated state vector with a new measurement is then

$$x[k] = x^-[k] + K[k](z[k] - Hx^-[k]), \quad (4.39)$$

and the updated error covariance is

$$P[k] = (I - K[k]H)P^-[k]. \quad (4.40)$$

References

- [Bachmann 03] E. R. Bachmann, X. Yun, D. McKinnery, R. B. McGhee, and M. J. Zyda, "Design and Implementation of MARG Sensors for 3-DOF Orientation Measurement of Rigid Bodies," *Proc. IEEE Int. Conf. Robotics & Automation*, Taipei, pp. 1171-1178, Sep 2003.
- [Barshan 94] B. Barshan and H. F. Durrant-Whyte, "Inertial Navigation Systems for Mobile Robots", *IEEE Trans. Robotics and Automation*, Vol.11, No.3, pp. 328-342, Feb 1994.
- [Brown 97] R. G. Brown and P. Y. C. Hwang, *Introduction to Random Signals and Applied Kalman Filtering*, John Wiley & Sons, United States, 1997.
- [Foxlin 96] E. Foxlin, "Inertial Head-Tracker Fusion by a Complementary Separate-Bias Kalman Filter," *Proc. VRAIS'96*, IEEE, pp.185-194, 1996.
- [Foxlin 98] E. Foxlin, M. Harrington, and Y. Altshuler, "Miniature 6-DOF inertial system for tracking HMDs," *In SPIE vol. 3362, Helmet and Head-Mounted Displays III, AeroSense*, Orlando, FL, April 13-14, 1998.
- [Friedland 78] Friedland B., "Analysis Strapdown Navigation Using Quaternions", *IEEE Trans. on Aerospace and Electronic Syst*, Vol. AES-14, No.5, pp.764-768, Sep. 1978.
- [IEEE 1293-1998] IEEE Standard Specification Format Guide and Test Procedure for Linear Single-Axis, Nongyroscopic Accelerometers, Annex I, IEEE Std 1293-1998.
- [Jiang 91] Y. F. Jiang and Y. P. Lin, "Error Analysis of Quaternion Transformations," *IEEE Trans. Aerospace and Electronic Systems*, Vol. 27. No.4, pp.634-639, Jul. 1991.
- [Julier 00] S. Julier, J. Uhlmann, and H. F. Durrant-Whyte, "A New Method for the Nonlinear Transformation of Means and Covariances in Filters and Estimators," *IEEE Trans. Automatic Control*, Vol. 45, No.3, pp. 477-482, Mar 2000.
- [Lefferts 82] E. J. Lefferts, F. L. Markley, and M.D. Shuster, "Kalman Filtering for Spacecraft Attitude Estimation," *J. Guidance, Control, and Dynamics*, Vol. 5, No. 5, pp. 417-429, Sep-Oct 1982.
- [Lovren 98] Lovren, N. and Pieper, J. K., "Error Analysis of Direction Cosines and Quaternion Parameters Techniques for Aircraft Attitude Determination", *IEEE Trans. on Aerospace and Electronic Syst*, Vol. AES-34, No.3, pp.983-989, 1998.
- [Luinge 02] H. J. Luinge, "Inertial Sensing of Human Movement," *Ph.D. Thesis*, University of Twente, Twente University Press, Enschede, the Netherlands, 2002.
- [Mortensen 74] Mortensen, R. E. "Strapdown guidance error analysis", *IEEE Trans. on Aerospace and Electronic Systems*, Vol. AES-10, No.4, pp.451-457, Jul. 1974.
- [Newcomb 82] C. Newcomb and I. Filnn, "Improving linearity of piezoelectric ceramic actuators", *Electronics Letters*, Vol.18, No.11, pp.442-444, May 1982.
- [Rehbinder 01] H. Rehbinder and X. Hu, "Drift-Free Attitude Estimation for Accelerated Rigid Bodies," *Proc. IEEE Int. Conf. Robotics & Automation*, Seoul, pp. 4244-4249, May 2001.
- [Riviere 03] C. N. Riviere, W. T. Ang, and P. K. Khosla, "Toward Intelligent Microsurgical Instrument," *IEEE Trans. Robotics and Automation*, Vol. 19, No. 5, pp. 793-800, Oct. 2003.
- [Riviere 04] C. N. Riviere and S. P. N. Singh, "Performance envelope and hand tremor amplitude during manual microsurgery," *IEEE Trans. Robotics*, submitted.
- [Wilcox 67] Wilcox, J. C., "A new algorithm for strapped-down inertial navigation", *IEEE Trans. on Aerospace and Electronic Systems*, Vol. AES-3, No.5, pp.796-802, Sep. 1967.
- [Benson'75] Benson, D., "A Comparison of Two Approaches to Pure-Inertial and Doppler-Inertial Error Analysis", *IEEE Trans. on Aerospace and Electronic Systems*, Vol. AES-11, No.4, pp.447-455, 1975.

- [Britting'71] Britting, K. R., "Inertial Navigation System Analysis", Wiley Interscience, New York 1971.
- [Goshen-Meskin'92] Goshen-Meskin, D. and Bar-Itzhack, I. Y., "Unified Approach to Inertial Navigation System Error Modeling", *Journal of Guidance, Control and Dynamics*, Vol.15, No.3, pp.648-653, May-June, 1992
- [Shibata'86] Shibata, M., "Error Analysis Strpdwn Inertial Navigation Using Quaternions", *Journal of Guidance, Control and Dynamics*, Vol.9, No.3, pp.379-381, 1986.

Chapter 5

Modeling and Filtering of Erroneous Motion

5.1 Introduction

To effectively implement real-time error compensation in a completely handheld microsurgical instrument, it is of paramount importance to accurately model both tremor and various types of non-tremulous involuntary movement. The objective of the erroneous motion filtering algorithms is to estimate the undesired components from the sensed position of the tool-tip, and output a negated erroneous motion to the manipulator system as control input.

Most of the previous works in the area of modeling and suppressing erroneous motion during human-machine control have been in pathological tremor. Tremor is characterized by having a higher frequency band than voluntary motion. Many of the approaches have taken advantage of the frequency separation between voluntary and tremulous motion. Riley and Rosen [Riley 87] have investigated lowpass filtering, while Prochazka *et al.* [Prochazka 92] have shown some success with the approach of closed-loop functional electrical stimulation. Gonzalez *et al.* [Gonzalez 95] has also proposed a signal equalizer technique to suppress pathological tremor. Pledgie *et al.* [Pledgie 00; Pledgie 99] have presented methods that suppress the power of pathological tremors through force feedback and impedance control.

These methods may be extended to physiological tremor notwithstanding, their usefulness in our application is impeded by the real-time motion canceling criterion. In general, classical frequency selective filters and feedback controllers cause a phase change between the input and output signals. Consequently, in time domain, a perceived motion is actually a time-delayed version of a real physical motion. For example, consider a tremor filtering method having the effect of a 2nd order low-pass filter with cut-off frequency at 20 Hz. Assuming the nominal tremor amplitude and frequency is 50 μm at 10 Hz, the resulting phase shift and time delay would be about 0.24 radian and 13 ms respectively. A 13 ms delay between the perceived and actual motion would translate into a maximum canceling error of about 25 μm , which well exceeds our targeted positioning accuracy.

In the current implementation, this dissertation has adopted the weighted-frequency Fourier linear combiner (WFLC) algorithm developed by Riviere *et al.* [Riviere 95] where adaptive noise canceling of tremor is based on a dynamic sinusoidal model. The basic idea and characteristics of this zero-phase adaptive filter will be presented in the next section.

Tremor has been the most studied source of physiological positioning error because of its distinct frequency characteristic which is invariant to mass and length of the limbs. To date, there are very few literature dedicated to study the nature and characteristics of other significant sources of error, e.g. myoclonic jerk and drift, or low-frequency error. Since little is known about these components, and since reference signals for adaptive noise canceling, are unavailable, suppression is difficult. Owing to such, a neural network technique is proposed to perform a black-box modeling and real-time suppression of these non-tremulous errors. Section 5.3 describes the proposed method and results from simulated motion compensation experiments. More in-depth study will need to be conducted before this algorithm can be incorporated into the Micron system for clinical experiments.

5.2 Tremor Filtering

5.2.1 Weighted-Frequency Fourier Linear Combiner (WFLC)

The underlying problem of real-time tremor suppression has lent itself to be an ideal application for adaptive noise canceling techniques. An adaptive noise canceller is a filter whose parameters are adjusted by a learning algorithm based on new observations. An adaptive noise canceller has two parallel sources of input signals, a primary source s_k made up of a desired signal d_k corrupted with an uncorrelated noise n_k ; and a tapped delayed reference input x_k superimposed with a noise n_k' , which is correlated with n_k . Noise y_k is an estimate of n_k obtained from the adaptive filter by minimizing the mean square error between n_k and y_k by typically a gradient descent method such as the least mean square (LMS) algorithm. The output ε_k is computed by subtracting an estimated noise y_k from s_k . An adaptive noise canceling system is depicted in Fig. 5.1.

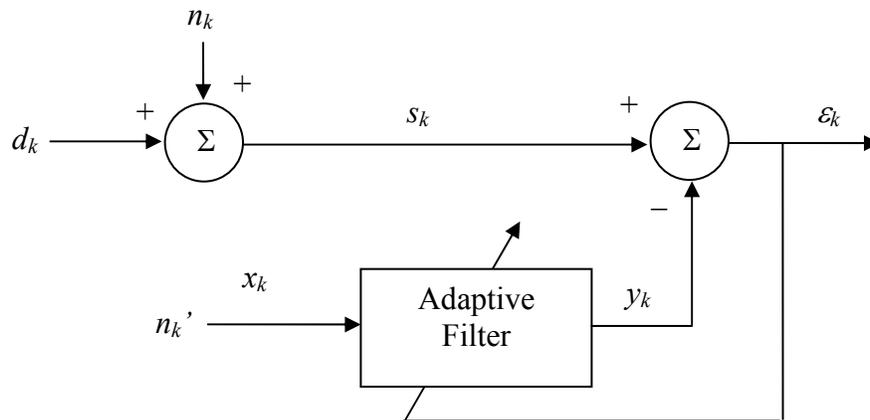


Fig. 5.1 An adaptive noise canceller.

The weighted-frequency Fourier linear combiner (WFLC) algorithm is effectively a zero-phase adaptive notch or band-stop filter with the stop band centered at the dominant incoming frequency. It is an extension of the Fourier Linear Combiner (FLC) [Vaz 94]. The FLC is an adaptive algorithm that uses a dynamic truncated Fourier series model to

estimate incoming quasiperiodic signals of known frequency by adapting the amplitude and phase of an artificially generated reference signal,

$$y_k = \sum_{r=1}^M [a_r \sin(r\omega_0 k) + b_r \cos(r\omega_0 k)] \quad (5.1)$$

The LMS algorithm is used to update the adaptive filter weights, which are the Fourier coefficients. The FLC, as shown in Fig. 5.2, is given as follow

$$x_{rk} = \begin{cases} \sin(r\omega_0 k), & 1 \leq r \leq M, \\ \cos((r-M)\omega_0 k), & M+1 \leq r \leq 2M, \end{cases} \quad (5.2)$$

$$\varepsilon_k = s_k - w_k^T x_k, \quad (5.3)$$

$$w_{k+1} = w_k + 2\mu x_k \varepsilon_k, \quad (5.4)$$

where $w_k = [w_{1k} \ \cdots \ w_{2Mk}]^T$ is the adaptive weight vector, s_k is the input signal, M is the number of harmonics in the model, and μ is an adaptive gain parameter which may be viewed as a parameter to adjust the width of the notch filter created at ω_0 [Widrow 85]. The FLC is computationally inexpensive [Vaz 94], inherently zero phase [Vaz 89], and has an infinite null [Widrow 85]. The time constant for convergence can be shown to be

$$\tau_\varepsilon = \frac{1}{2\mu}. \quad (5.5)$$

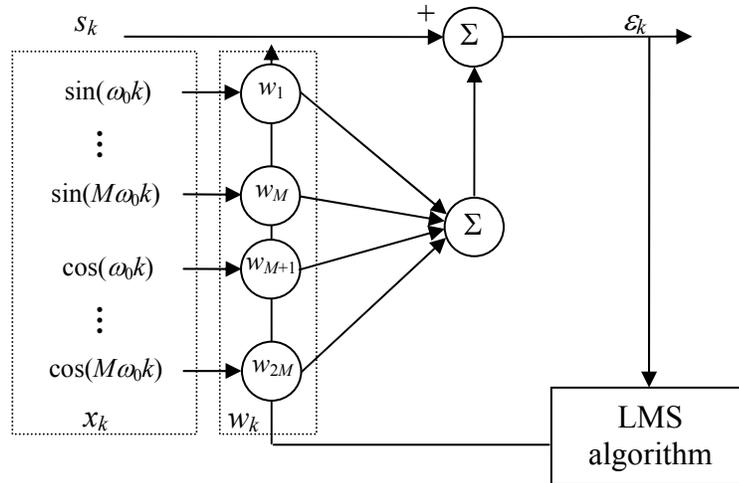


Fig. 5.2 The Fourier linear combiner. The FLC adaptively creates a dynamic Fourier series model of any periodic input. The reference vector consists of harmonic sines and cosines at a fixed fundamental frequency. These are weighted by the Fourier coefficient vector w_k and summed to obtain a truncated Fourier series model of the input s_k .

The capability of FLC to adapt and lock-on to the amplitude and phase of an input periodic signal enable it to be employed in periodic interference cancellation. The WFLC algorithm extends the FLC algorithm to also adapt to the time-varying reference signal frequency, using a modification of the Least Mean Square (LMS) algorithm. The WFLC's explicit zero-phase model of tremor makes it attractive for real-time active tremor suppression. Replacing the fixed reference frequency of the FLC by an adaptive weight w_{0_k} , we have

$$\varepsilon_k = s_k - \sum_{r=1}^M [w_{r_k} \sin(rw_{0_k} k) + w_{r+M_k} \cos(rw_{0_k} k)] \quad (5.6)$$

A modified LMS algorithm is used to construct w_{0_k} recursively from (5.6),

$$w_{0_{k+1}} = w_{0_k} - 2\mu\varepsilon_k \frac{\partial \varepsilon_k}{\partial w_{0_k}}, \quad (5.7)$$

where

$$\frac{\partial \varepsilon_k}{\partial w_{0_k}} = -k \sum_{r=1}^M r \left[w_{r_k} \cos\left(r \sum_{t=1}^k w_{0_t}\right) - w_{r+M_k} \sin\left(r \sum_{t=1}^k w_{0_t}\right) \right] \quad (5.8)$$

Note that the sinusoidal arguments $w_{0_k} k$ in (5.6) are replaced with running sums appropriate to frequency modulation. Making w_{0_k} adaptive with its own gain μ_0 , we arrive at the frequency recursion of WFLC given by

$$x_{r_k} = \begin{cases} \sin\left(r \sum_{t=1}^k w_{0_t}\right), & 1 \leq r \leq M \\ \cos\left((r-M) \sum_{t=1}^k w_{0_t}\right), & M+1 \leq r \leq 2M \end{cases} \quad (5.7)$$

$$\varepsilon_k = s_k - w_k^T x_k, \quad (5.8)$$

$$w_{0_{k+1}} = w_{0_k} + 2\mu_0 \varepsilon_k \sum_{r=1}^M r (w_{r_k} x_{M+r_k} - w_{M+r_k} x_{r_k}), \quad (5.9)$$

$$w_{k+1} = w_k + 2\mu_1 x_k \varepsilon_k. \quad (5.10)$$

The WFLC algorithm is shown in Fig. 5.3. The overall tremor filtering system, as shown in Fig. 5.4, is in fact a WFLC-FLC combination,

$$\hat{\varepsilon}_k = s_k - \hat{w}_k^T x_k, \quad (5.11)$$

$$\hat{w}_{k+1} = \hat{w}_k + 2\hat{\mu} x_k \hat{\varepsilon}_k. \quad (5.12)$$

where $\hat{w}_k = [\hat{w}_{1k} \ \cdots \ \hat{w}_{2M_k}]^T$. Thus, the system operates essentially as a FLC with a time-varying reference frequency.

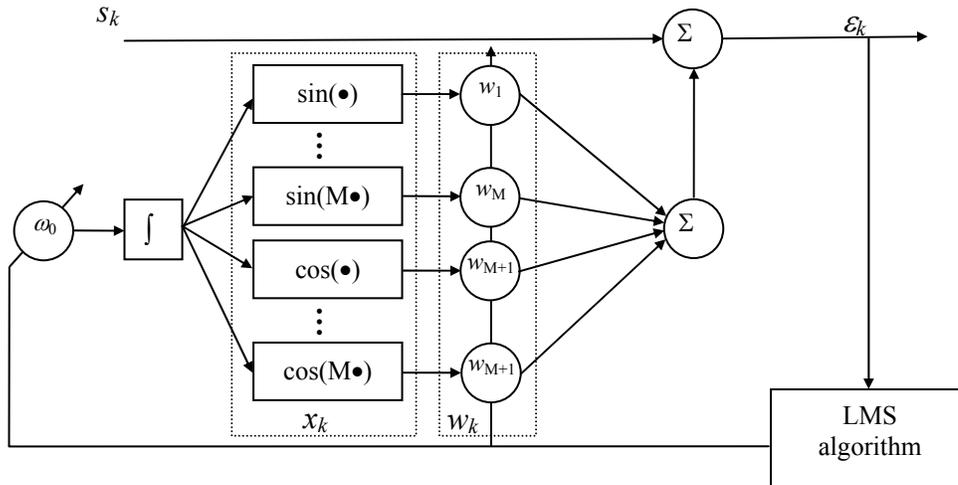


Fig. 5.3 The weighted-frequency Fourier linear combiner (WFLC). The system maintains a running sum of the frequency weight ω_0 . Harmonic sines and cosines of this quantity are taken (vector x_k), weighted by the Fourier coefficient vector w_k , and summed to provide a truncated Fourier series model of the input s_k .

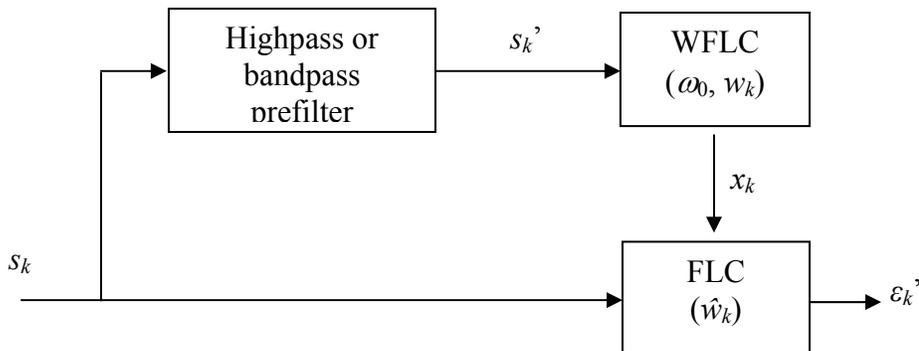


Fig. 5.4 The WFLC-FLC combination. The highpass or bandpass IIR prefilter filters the input signal s_k to attenuate the voluntary motion. WFLC operates on the filtered signal s_k' , estimates the tremor frequency ω_0 and feed it into FLC. With known frequency, FLC adapts and locks-on to the tremor amplitude and phase.

A sixth-order elliptical band-pass prefilter with cut-off corner frequencies set at 7 and 13 Hz is used to reduce the power of the voluntary motion before the input signal is fed into WFLC. This is to prevent w_{0k} from adapting to the low-frequency voluntary component which may have higher signal power than the tremor. This prefilter inevitably

creates a phase lag in the WFLC output. The fact that this affects only the frequency and not the amplitude, and since tremor frequency tends to vary slowly [Gresty 90], the impact of this delay on the overall performance of the system is negligible. Active canceling of physiological tremor using this algorithm has been demonstrated in a 1-dimensional motion canceling experiment [Riviere 95, Riviere 98]. The algorithm achieves an average rms reduction of about 69% in the tremor amplitude. Fig. 5.5 is a reproduction of a result plot published in [Riviere 98].

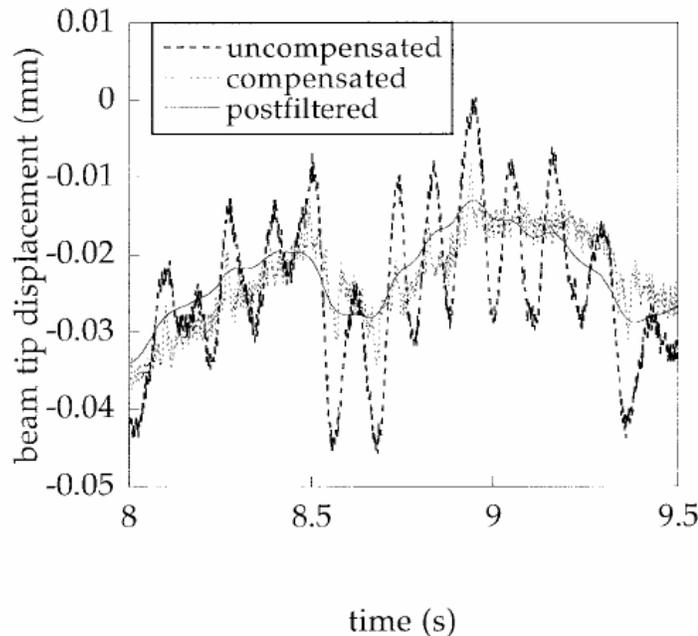


Fig. 5.5 Result of one dimensional active physiological tremor canceling experiment. The dotted line is a recording of the physiological tremor of a microsurgeon. The solid line shows an offline estimate of voluntary motion, obtained via zero-phase lowpass postfiltering of the uncompensated motion. The fine dotted line shows the motion during WFLC compensation.

5.2.2 WFLC and Inertial Sensor Drift

The usefulness of WFLC in practical real-time system to estimate and suppress tremor is limited by its adaptive nature. WFLC being an adaptive system is only suitable to be used in the displacement domain. This is because the transient errors in the higher order Laplace domains (i.e. velocity and acceleration) will become tracking offset errors in the lower order domains after integrations. Since the accelerometer measures acceleration, any erroneous DC bias offset in the measurement will become an error that

increases quadratically over time. As a result, when the power of this integration drift exceeds that of the higher frequency tremor, even with WFLC's the band-pass prefiltering, the WFLC algorithm adapts to the lower frequency sensor drift rather than the tremor. This further exemplifies the importance of the measurement model and the augmented state quaternion-based Kalman filtering proposed in Chapter 3 and 4.

5.3 Non-Tremulous Error Filtering

5.3.1 Introduction

Besides tremor, other types of involuntary motions such as pathological myoclonus can also cause incapacitation of human motor control. Myoclonic movements are twitches or jerks resulting from sudden muscle contractions that can occur alone or in a sequence [Nechyba 95]. Non-tremulous types of erroneous movement are aperiodic, erratic, unpredictable at our current state of understanding, and can overlap in frequency with voluntary motion. Effective online canceling is therefore difficult.

The mapping from human intention to human movement output is nonlinear. Neural networks model nonlinear processes well, and have been used in modeling of human control strategies [Nechyba 95]. The complexity and multiplicity of involuntary hand motion components, and the paucity of knowledge about components such as drift, makes a neural network approach well suited to modeling of human movement error processes. Riviere and Khosla [Riviere 97] used a cascade neural network for noise canceling in human hand motion. Their experiments showed that the neural network successfully modeled and reduced the errors on recorded hand movement files of four surgeons. In this section, we are presenting a different approach in employing the same cascade neural network technique on the same set of data and compare our results with those obtained by Riviere and Khosla. We also go a step further to investigate the effectiveness of the trained networks on surgeons other than those on which they were trained.

Though the experiments presented here focus on surgery, the concepts demonstrated are directly relevant to a wide range of manipulation applications with small signal-to-noise ratio, e.g. helping rehabilitation patients with pathological tremors to control

computer cursor [Ang 02], enhancing manual accuracy in cell manipulation in the biotech industry, etc.

5.3.2 Cascade Neural Network with Extended Kalman Filtering

Instead of the traditional rigid, fixed architecture network with backpropagation, we used a cascade neural network with Kalman filtering proposed by Nechyba [Nechyba 98]. Nechyba combines (i) flexible cascade neural networks, which dynamically adjust the size of the neural network as part of the learning process, and (ii) node-decoupled extended Kalman filtering (NDEKF), a fast converging alternative to backpropagation.

When training starts, the network has no hidden nodes, only linear connections between the input and the output nodes. This enables the network to capture any linear relationship between the inputs and outputs as shown in Fig. 5.6.

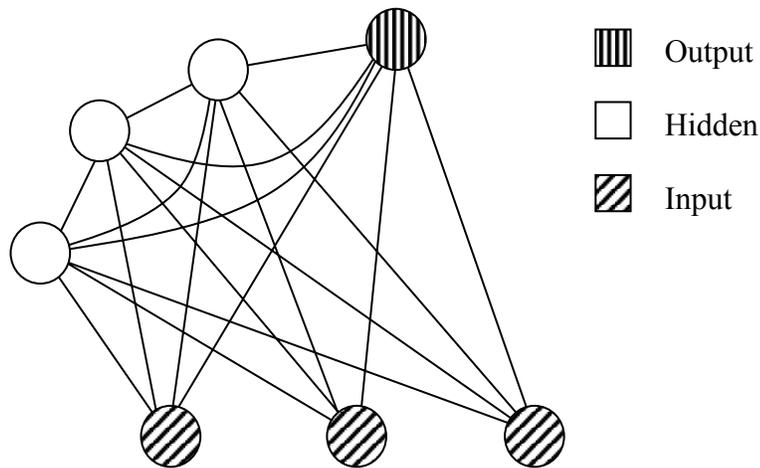


Fig. 5.6 Diagram of the cascade neural network architecture. The diagram shows a network with three hidden nodes. As each hidden node is added, it is connected to the input and output nodes, as well as each of the preceding hidden nodes.

During training, each time the error performance stagnates, a new hidden node is added to the network from a pool of candidate units (transfer functions). In our experiments, these candidates include Sigmoid, Gaussian, Sine, and Bessel. The best candidate unit is selected after being trained independently and in parallel with different

random initial weights using the quickprop algorithm. Once a new hidden unit is installed, the hidden-unit input weights are frozen, while weights to the output units are retrained. The process is repeated until the algorithm succeeds in reducing the root mean square error (rmse) sufficiently for the training set or the number of hidden units reaches a specified maximum number.

Extended Kalman filtering (EKF) is an extension of Kalman filter to deal with non-linear systems via linearization about the current parameter estimates. In neural network training, learning is cast as an identification problem for a non-linear dynamic system. The neural network weights represent the state of the non-linear system. The EKF theory is then used to derive a recursion for the weight updates. This work uses NDEKF, in which the network weights are grouped such that each group contains the input nodes, the output nodes and one hidden node. For each group, elements of the error covariance matrix estimate corresponding to other groups can be ignored, greatly reducing the computational complexity.

5.3.3 Experimental Methods

Hand movement data of surgeons were recorded in Wilmer Eye Institute of Johns Hopkins University. Each surgeon held a microsurgical instrument with the tip inserted in a sclerotomy in the eye of a mannequin face. A Hall effect sensor mounted inside the mannequin eye detected the position, in one dimension, of a 0.26g permanent magnet mounted on the tip of the instrument. Data were recorded for 16s at a sampling rate of 250Hz. The surgeons attempted to hold the instrument motionless for the duration of each test, therefore any motion in these recordings is considered to be error. A total of 15 files were obtained from four surgeons (5, 5, 3, and 2 files, respectively).

To ease disambiguation of erroneous movement from desired movement for purposes of evaluation, surgeons were given fixed targets at which to point, and tried to keep the instrument motionless, thus ensuring that all recorded motion is error. To make the experiments more realistic we generate low frequency pseudo-voluntary motions and add them to the recorded still hand error movement. The pseudo-voluntary motions thus serve as the target motions in our experiments. The magnitude of the randomly generated

pseudo-voluntary motions has a ratio of roughly 1:1 to the mean rms error of the 15 data files. Gaussian white noise sequences are also generated and then low-pass filtered before adding to the randomly generated low frequency motions. Two different pseudo-voluntary motions are generated in this manner, one for the training data files and the other for testing data files.

Separate neural networks were used for each of the four surgeons. For each surgeon, a cascade-architecture neural network, using extended Kalman filtering for learning, was trained using one of the training data files described above. The remaining data files from each surgeon were used for testing of the trained network. The rmse with respect to the pseudo-voluntary motion was calculated for each file, both before and after processing by the neural network.

The input to the neural network was a window of data in the time series, i.e. the number of input nodes depended on the length of the window. The output of the neural network was the error-compensated motion, and since we were looking only at one-dimensional data, there will only be 1 output node. Different combinations of number of input nodes and hidden nodes were tested to obtain the best net architecture for each surgeon. Riviere and Khosla [Riviere 97] used the same set of data but chose the error estimate as the network output, so that the output of the neural network could be used directly by other downstream actions to cancel this error. In addition, Riviere and Khosla fixed the number of input nodes to be 100 and the maximum number of hidden nodes to be 10. The results of our experiment will be compared with those reported in [Riviere 97].

We also investigate the existence of a generalized neural network for all surgeons by testing how well a neural network trained on one surgeon's file perform in predicting the time series of the others.

5.3.4 Results

The neural network reduced the rmse with respect to the randomly generated pseudo-voluntary motion for all the testing data files. Table 5.1 shows for each of the surgeon the mean raw rmse of the data files (with the number of the testing files in parenthesis), the mean rmse of the output of the neural network and their standard deviations, and the

neural network architecture that gives the best result. Table 5.2 shows that our approach (NN_{vm}) of using the pseudo-voluntary motion as the training targets outperforms Riviere and Khosla's method (NN_{err}) that uses the erroneous motion as the training target.

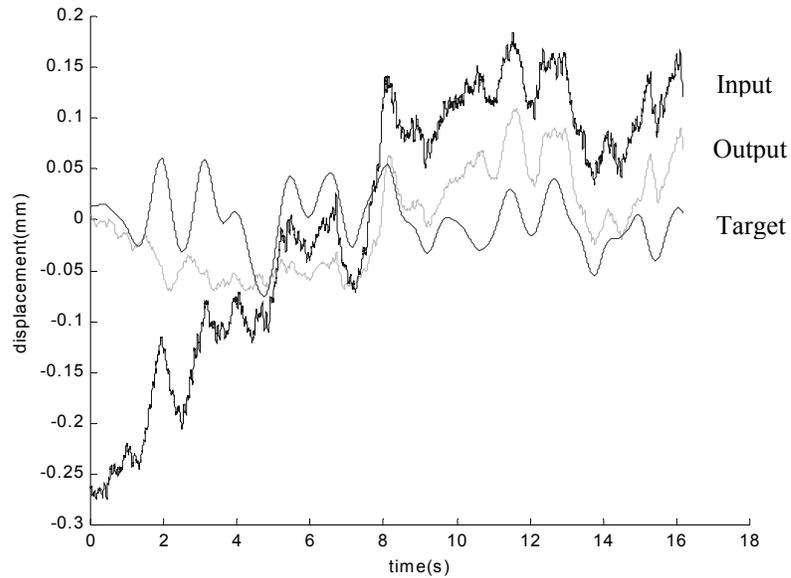
Table 5.1 Performance of the neural network and the network architecture that produces the best result.

Surgeon # (no.of testing files)	Mean raw rmse (mm)	Mean rmse of neural network output (mm)	Standard deviation (mm)	Best Network Architecture
1 (4)	0.112	0.055	0.005	75 input, 3 hidden
2 (4)	0.046	0.033	0.002	60 input, 6 hidden
3 (2)	0.048	0.037	0.001	100 input, 3 hidden
4 (1)	0.127	0.056	-	50 input, 6 hidden

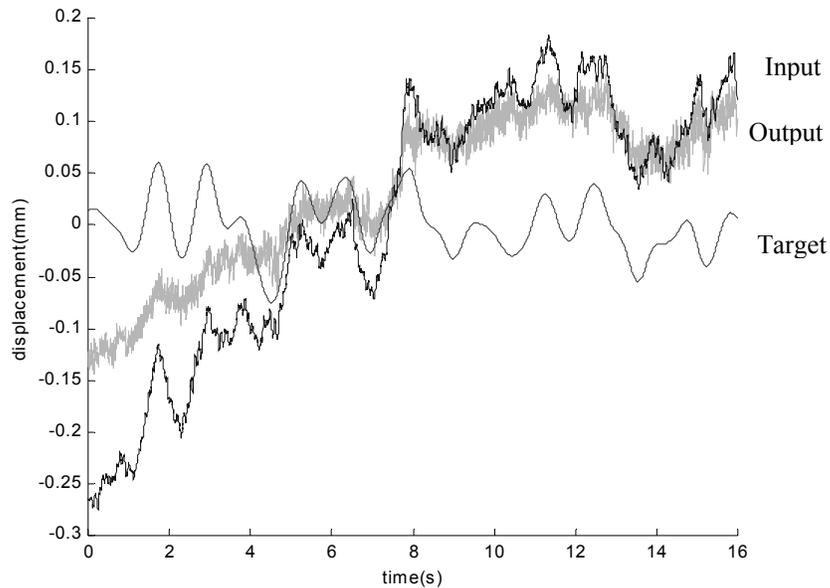
Table 5.2 Percentage rmse reduction comparison between ours (NN_{vm}) and Riviere/Khosla's (NN_{err}) method

Surgeon #	% rmse reduction (NN_{vm})	% rmse reduction (NN_{err})
1	50.9	44.6
2	28.3	26.1
3	22.9	18.8
4	55.9	29.9
Average	39.5	29.9

Fig. 5.7a and 5.7b depict the sample results of our approach and Riviere and Khosla's approach respectively.



(a) Sample result of our approach



(b) Sample result of Riviere/Khosla's approach

Fig. 5.7 The line 'Target' is the pseudo-voluntary motion, generated by lowpass filtering white noise at 1 Hz cutoff frequency. The line 'Input' represents the network input for the test, obtained by adding one of the recorded hand motion error files to the pseudo-voluntary motion. The line 'Output' indicates the filtered version of the data.

Table 5.3 shows how well a neural network trained on one surgeon's file perform in predicting the movement of the others. Only 1 file of each surgeon is cross-tested here.

Table 5.3 Numbers in parenthesis are the test result of the file using the neural network (NN) trained on the same surgeon, to provide comparison. Numbers in bold are tests that equal to or outperform the % rmse reduction of the neural network trained on the same surgeon. Numbers in italics are tests that produce a rmse worse than the raw rmse.

NN trained on Surgeon #	NN test on Surgeon #	% rmse reduction
1	2	26.2 (23.1)
	3	38.1 (28.6)
	4	45.1 (<i>56</i>)
2	1	-85.5
	3	28.6 (28.6)
	4	-243.1
3	1	0
	2	15.4 (23.1)
	4	-0.9
4	1	32.7 (54.5)
	2	-30.8
	3	22.7 (28.6)

5.3.5 Discussion

The results show the feasibility of the basic approach of neural network error canceling in human-machine control. The neural network reduced the rmse of the surgeons' erroneous motion by an average of 39.5%.

NN_{vm} outperforms NN_{err} by 9.6% in rmse reduction. During training, both methods terminate at reaching the maximum number of hidden nodes. Riviere and Khosla used 100 input nodes throughout their experiment while we explore the effect of different input-hidden node combination on the performance of the neural networks. NN_{vm} has better result may be because of the optimization of the network architecture in the number of input and hidden nodes, not necessary the superiority of one method over the other.

We observe that there are distinct differences in the motion profile between the outputs of NN_{err} and those of NN_{vm} . Firstly the NN_{err} amplifies the high frequency noise (tremor) while NN_{vm} attenuates the high frequency noise. This is because NN_{err} is trained to learn the high frequency error, while NN_{vm} is trained to learn the smooth voluntary motion. In addition, qualitatively, NN_{vm} 's estimation of voluntary motion has shown a better preservation of motion profile than NN_{err} .

The experiments on a neural network trained on one surgeon's file in predicting the movement of the others has drawn inconclusive results. Out of the 12 runs, 3 has equal or better predictions than the NN trained on files from the same surgeon, 5 has worse rmse reduction result and 4 has worse rmse than the original error data files. Generally, as one would expect, the results show more worse (9/12) than better (3/12). However, the fact that the neural network of surgeon 1 in predicting the motion of surgeons 2 and 3 outperforms their own NN suggests that there might exist an optimal NN for a group of surgeons. More meaningful conclusion can only be drawn with more in-depth studies.

More extensive testing of the proposed neural network technique will have to be conducted before it can be implemented in the real system for laboratory and clinical experiments. In future, we will be extending the current work to conduct experiments involving three dimensional hand movement data containing both voluntary and erroneous motion. The data collection will include more testing subjects performing tasks like tracing a predetermined path or tracking a predetermined trajectory.

References

- [Ang 02] W. Ang, C. Riviere, and P. Khosla, "Assistive Human-Machine Interfaces via Artificial Neural Networks," *Proc. of the 3rd International Symposium on Robotics and Automation*, September, 2002, pp. 477-480.
- [Charles 96] S. Charles, Dexterity enhancement for surgery. In: R.H. Taylor et al. (eds.): *Computer Integrated Surgery: Technology and Clinical Applications*. MIT Press, Cambridge'96 467-471.
- [Elble 90] R. J. Elble and W. C. Koller, *Tremor*. Baltimore: Johns Hopkins, 1990.
- [Gonzalez 95] J. G. Gonzalez, E.A. Heredia, T. Rahman, K.E.Barner, S.K.Basu, and G.R.Arce, "A new approach to suppressing abnormal tremor through signal equalization," *Proc. RESNA Annu. Conf.*, pp.707-709, 1995.
- [Gresty 90] M. Gresty and D. Buckwell, "Spectral analysis of tremor: Understanding the results," *Electroencephalogr. Clin. Neurophysiol.*, vol. 53, pp. 976-981, 1990.
- [Harwell 83] R. C. Harwell and R. L. Ferguson, "Physiologic tremor and microsurgery," *Microsurgery*, vol. 4, pp. 187-192, 1983.
- [Nechyba 95] M. Nechyba and Y. Xu, "Human skill transfer: neural networks as learner and as teachers," *Proc. IEEE Intl. Conf. Intell. Rob. Sys.*, vol.3, pp.314-319, 1995.
- [Nechyba 98] M. Nechyba, "Learning and Validation of Human Control Strategies", *tech. report CMU-RI-TR-98-06*, Robotics Institute, Carnegie Mellon University, May, 1998.
- [Patkin 77] M. Patkin, "Ergonomics applied to the practice of microsurgery," *Aust. N. Z. J. Surg.*, vol. 47, pp. 320-329, June 1977.
- [Pledge 00] S. Pledge, K. E. Barner, S. K. Agrawal, and T. Rahman, "Tremor Suppression Through Impedance Control", *IEEE Transactions on Rehabilitation Engineering*, Vol. 8, No. 1, 2000, 53-59.

- [Pledge 99] S. Pledge, K. E. Barner, S. K. Agrawal, and T. Rahman, "Tremor Suppression Through Force Feedback", *Proceedings of the Sixth International Conference on Rehabilitation Robotics*, July 1-2, 1999, Stanford, CA, pp. 16-26, 1999.
- [Prochazka 92] A. Prochazka, J. Elek, and M. Javidan, "Attenuation of pathological tremors by functional electrical stimulation I:method", *Ann Biomed Eng* 20:205-24, 1992.
- [Riley 87] P.O. Riley and M. Rosen, "Evaluating manual control devices for those with tremor disability," *J. Rehabilitation Research Dev.*, vol24, pp.99-110, 1987
- [Riviere 00] C. N. Riviere and P. S. Jensen, "A study of instrument motion in retinal microsurgery," *Proc. 22nd Annu. Conf. IEEE Eng. Med. Biol. Soc.*, Jul. 2000.
- [Riviere 95] C. N. Riviere, "Adaptive suppression of tremor for improved humanmachine control," *Ph.D. dissertation*, Johns Hopkins Univ, Baltimore, MD, 1995.
- [Riviere 97a] C. N. Riviere, R. S. Rader, and P. K. Khosla, "Characteristics of hand motion of eye surgeons," *Proc. 19th Annu. Conf. IEEE Eng. Med. Biol. Soc.*, Chicago, 1997.
- [Riviere 97b] C. N. Riviere and P. K. Khosla, "Augmenting the human-machine interface: improving manual accuracy," *Proc. IEEE Intl. Conf. Robot. Autom.*, Albuquerque, pp. 20-25, Apr. 1997.
- [Riviere 98] C. N. Riviere, R. S. Rader, N. V. Thakor, "Adaptive canceling of physiological tremor for improved precision in microsurgery", *IEEE Trans Biomed Eng* 45(7): 839-46, 1998.
- [Schenker 95] P. S. Schenker, E. C. Barlow, C. D. Boswell, H. Das, S. Lee, T. R. Ohm, E. D. Paljug, G. Rodriguez, and S. T. 2, "Development of a telemanipulator for dexterity enhanced microsurgery," *Proc. 2nd Intl. Symp. Medical. Robotics and Comp. Assisted Surgery*, pp.81-88, 1995.
- [Taylor 99] R. Taylor, P. Jensen, L. Whitcomb, A. Barnes, R. Kumar, D. Stoianovici, P. Gupta, Z. Wang, E. de Juan, and L. Kavoussi, "A steady-hand robotic system for microsurgical augmentation," in: C. Taylor, A. Colchester (eds.), *Medical Image Computing and Computer-Assisted Intervention - MICCAI'99*, Springer, Berlin, pp. 1031-1041.
- [Vaz 89] C. Vaz and N. Thakor, "Adaptive Fourier estimation of time-varying evoked potentials," *IEEE Trans. Biomed. Eng.*, vol. 36, pp. 448-455, Apr. 1989.
- [Vaz 94] C. Vaz, X. Kong, and N. Thakor, "An adaptive estimation of periodic signals using a Fourier linear combiner," *IEEE Trans. Signal Processing*, vol. 42, pp. 1-10, Jan. 1994.
- [Widrow 85] B. Widrow and S. D. Stearns, *Adaptive Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1985, ch. 12.

Chapter 6

Manipulator Design

6.1 Introduction

This chapter presents the system design of the tool-tip micro-manipulator for the intelligent handheld microsurgical instrument. The chapter starts off with a brief discussion of the system requirements of the manipulator; it is followed by a detailed description of the design of the manipulator mechanism with explanation of the design choices. The last two sections computed the inverse and forward kinematics of the manipulator, and analyze the workspace of the instrument tip with respect to the physiological tremor.

6.1.1 System Requirements

The primary task of the manipulator system is to deflect the intraocular shaft to cancel physiological tremor in real-time. To interpret the task statement in an alternate perspective, the manipulator is required to manipulate the tool tip to track dynamically the negated motion profile of physiological tremor. To quantify the manipulator system requirements, we consider a tremulous motion profile with nominal 50 μm p-p amplitude and in the frequency band of 8-12 Hz. In addition, it is reported by Gupta *et al.* [Gupta 99]

that roughly 75% of the forces measured during retinal microsurgery are found to be less than 7.5 mN. With this information, the performance specifications are:

- (i) Tracking precision $\leq 1 \mu\text{m rms}$;
- (ii) Tracking resolution $\leq 1 \mu\text{m}$;
- (iii) Actuation bandwidth $> 13 \text{ Hz}$;
- (iv) Actuation force $> 15.0 \text{ mN}$;
- (v) Manipulator workspace $\geq \text{Ø}100 \mu\text{m sphere}$.

Besides these engineering quantities, more subtle but equally important requirements include the size and weight of the manipulator must also be considered. The trade-offs between performance and usability will have to be carefully balanced.

6.2 Design of Mechanism

The proposed design of the intraocular shaft manipulator is a three degrees-of-freedom parallel mechanism driven by piezoelectric actuators. A parallel mechanism is best suited to this application because of its rigidity, compactness, and simplicity in design, as compared to a serial mechanism. The intelligent microsurgical instrument has six degrees-of-freedom sensing but only three degrees-of-freedom manipulation because the tip of the intraocular shaft may be approximated as a point in Euclidean space. We may disregard changes in orientation of the intraocular shaft, since they will be small in any case, given the small workspace of the manipulator. This reduces the dimension of the configuration space of the manipulator to three, and simplifies the mechanical design and the online computation of inverse kinematics.

Piezoelectric actuators emerge as the obvious choice based on primary its high actuation bandwidth. A P-885.50 (Polytec PI, Inc., Germany) piezoelectric stack measures 5 mm x 5 mm x 18 mm, and deflects to a maximum of about 15.0 μm with an applied voltage of +100VDC. It offers good control linearity, an excellent response time of 50 μs and an actuation force of up to about 800 N. Detailed specifications of the piezoelectric actuators may be found in the manufacturer's specification sheet [Polytec-PI]. The response time of the piezoelectric actuator ensures the velocities in the joint

space are more than adequate to map out the trajectory of the instrument tip in the workspace at the speed needed for canceling of tremor.

Fig. 6.1-6.3 depict the mechanism of the intraocular shaft manipulator. The manipulator system is located at the tool-tip end of the instrument. It is attached to the main handle and the sensing system by three screws, as shown in Fig. 6.1. The wiring of the manipulator system runs through a hole close to the circumference of the handle, while that of the front accelerometer runs through the center of the hollow shaft.

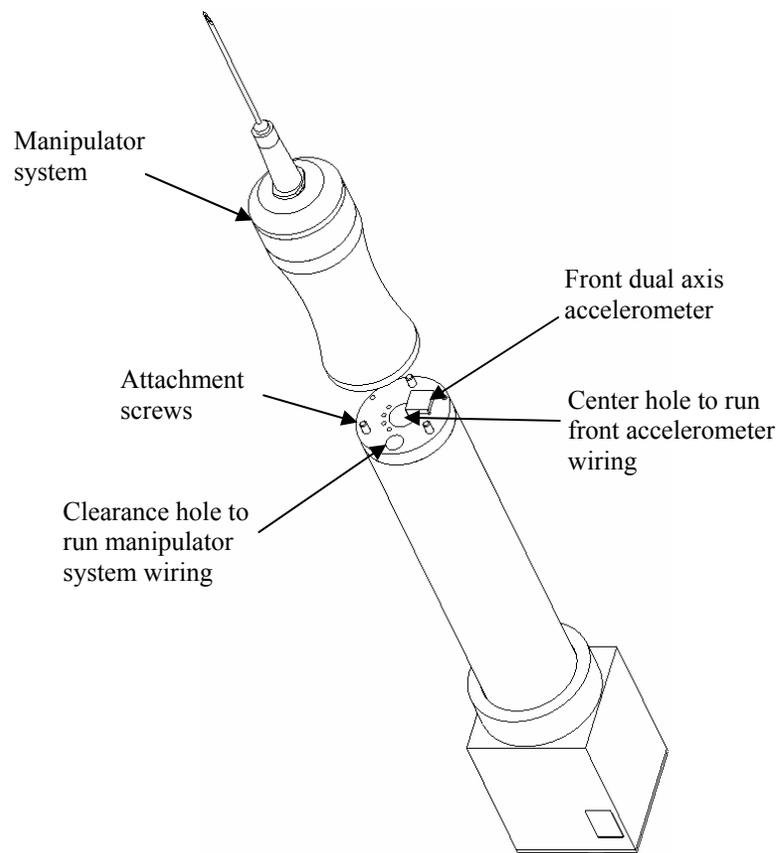


Fig. 6.1 The intraocular shaft manipulator is located in the front end of the instrument.

The manipulator is designed to allow interchanging of disposable surgical needles and other tools with female Luer connectors as the intraocular shaft. The intraocular shaft is attached to a standard Luer male connector fixed at the center of the three-legged rigid star (see Fig. 6.2). The three legs of the rigid star form the apexes of an equilateral triangle. The rigid star is screwed onto the flexi-star, which has the exact same shape, by

a contact pin at each of its legs. The flexi-star is a flexible thin plate made of stainless spring steel. The precision of location and orientation of the star assembly is ensured by three locator pins and is bolted to the center of the star housing by a screw. The locator pins and the screw constrain the star assembly in the three degrees of freedom that are not being driven, namely, translation in the two coordinates transverse to the long axis of the instrument, and rotation about the long axis.

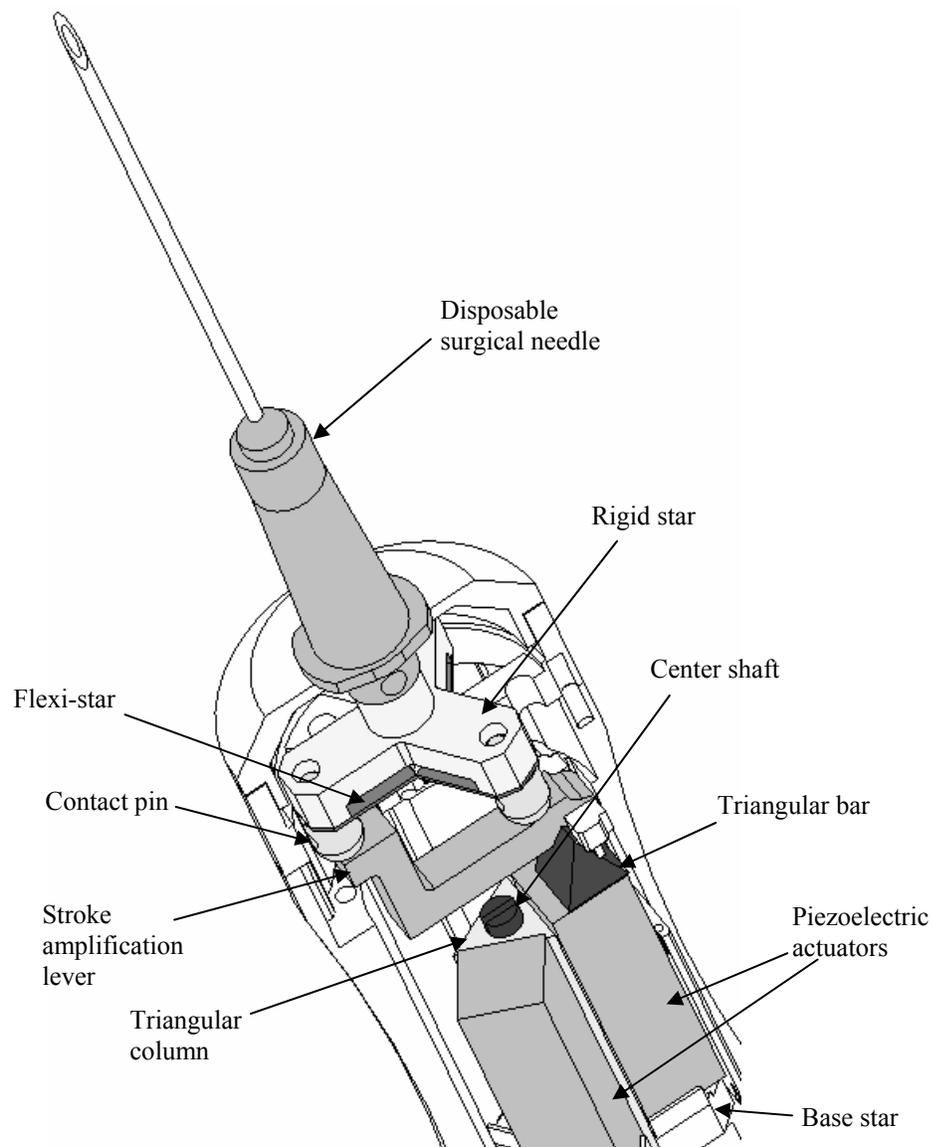


Fig. 6.2 The manipulator mechanism.

The three overhanging legs of the flexi-star is deflected by three stainless steel lever hinged 120° apart near the front opening of the narrowed section of the handle that aids grasping. The three levers serve as mechanical linkages that amplify the stroke of the piezoelectric actuators by 8 times, yielding a total stroke of $> 100 \mu\text{m}$ per axis. Three piezoelectric actuators are located on the three faces of the triangular column, and sandwiched between the base star and the levers pins. When voltage is applied to the piezoelectric stacks, they expand and rotate the levers which in turn push against the contact pins to move the star assembly and the intraocular shaft. To minimize slippage and wear and tear of the actuators at the contact points, a triangular block is epoxyed to each of the actuators with its apex resting in the V-slot on the bottom surface of the levers. The apex of the V-slot is inline with the fulcrum of the lever to eliminate unwanted lateral reaction forces on the piezoelectric actuators. The three levers crisscross one another at different heights to avoid collision of the linkages, as shown in Fig.6.3.

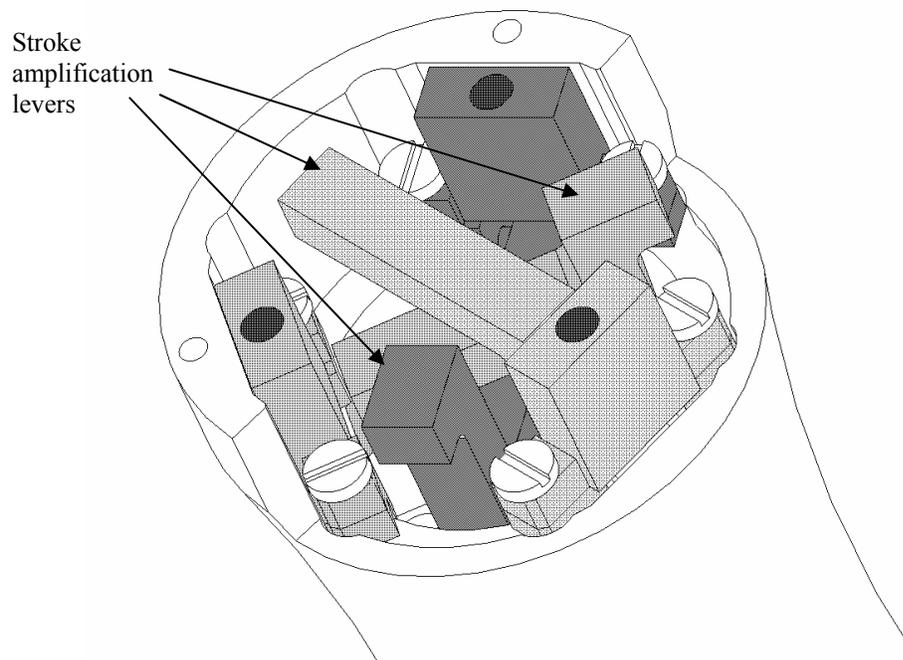


Fig. 6.3 The three stroke amplification levers crisscross one another at different heights to avoid collision of the linkages.

The triangular column is also serving as a mechanism to constrain the lateral movement of the actuators. The center stepped shaft has two tapering sections that push the ball bearing plates on each side of the triangular column radially outward to press the piezoelectric actuators against the wall of the handle. There is also a calibration screw at each of the three legs of the base star to compensate for the manufacturing inconsistencies in the length of the piezoelectric actuators.

The specifications of the manipulator are summarized in Table 6.1.

Table 6.1 Specifications of Micron manipulator system.

	x -axis	y -axis	z -axis
Maximum tip displacement (μm)	> 1000	> 1000	>100
Maximum tip velocity ($\mu\text{m}/\mu\text{s}$)	> 20	> 20	> 2

6.3 Inverse Kinematics

Since we do not care about the orientation of the instrument tip in its workspace, the dexterous workspace of the manipulator is equivalent to its reachable workspace. Moreover, this type of parallel manipulator is homeomorphic, i.e. it has a one-to-one forward and inverse mapping between its joint space and Euclidean space. It therefore has only boundary singularities and no internal singularities.

Within the small workspace for which the manipulator is intended, the constraints imposed by the design features make its kinematics essentially equivalent to those of Lee and Shah [Lee 88]. The base coordinate system $\{0\}$ with axes x_0 - y_0 - z_0 is attached to the centroid of the base star, with x_0 axis point towards pin joint P_1 . Frame $\{1\}$ with axes x_1 - y_1 - z_1 is attached to the centroid of the rigid star such that the intraocular shaft aligns with the z_1 axis and x_1 point towards the contact pin 1, as shown in Fig. 6.4. The equivalent kinematic chain, following the convention of Lee and Shah, is shown in Fig. 6.5.

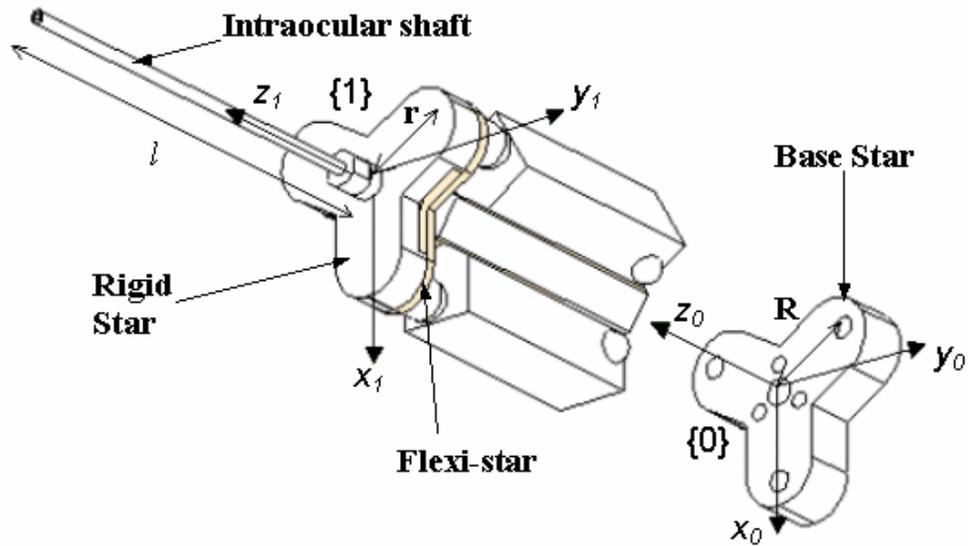


Fig. 6.4 Kinematic chain of the 3 DOF intraocular shaft parallel manipulator.

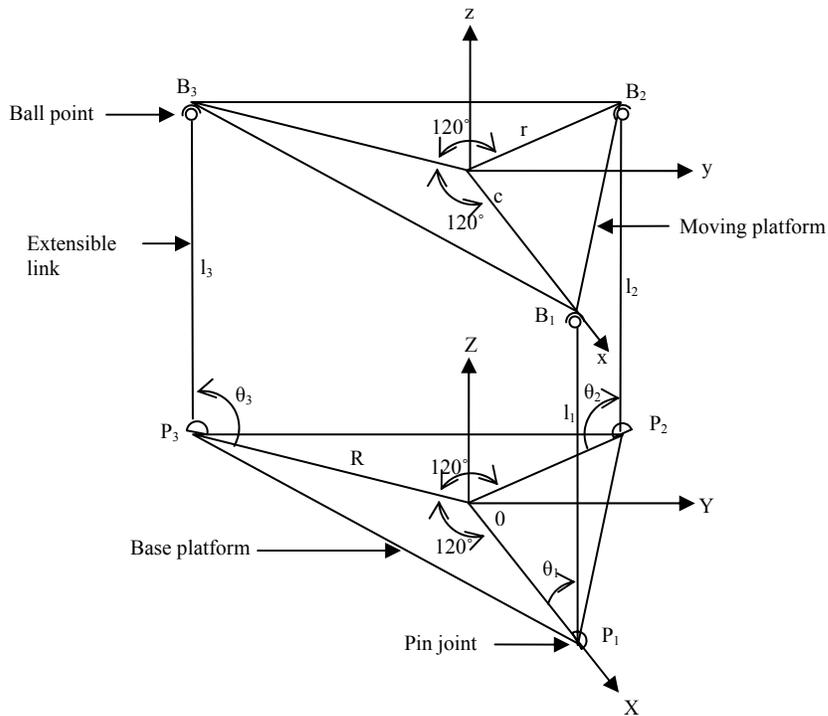


Fig. 6.5 Kinematic chain of the 3 DOF intraocular shaft parallel manipulator, represented in Lee and Shah [Lee 88] convention.

The value R is the distance between the centroid of base star (origin of $\{0\}$) and the vertex of the equilateral triangle form by its three legs. Similarly, the value r is the distance between the centroid of the rigid star (origin of $\{1\}$) and the vertex of the

equilateral triangle formed by its three legs. Hence, the coordinates of the pin joints with respect to $\{0\}$ are

$${}^0P_{P_1} = \begin{bmatrix} R \\ 0 \\ 0 \end{bmatrix}, \quad (6.1a)$$

$${}^0P_{P_2} = \begin{bmatrix} -\frac{1}{2}R \\ \frac{\sqrt{3}}{2}R \\ 0 \end{bmatrix}, \quad (6.1b)$$

$${}^0P_{P_3} = \begin{bmatrix} -\frac{1}{2}R \\ -\frac{\sqrt{3}}{2}R \\ 0 \end{bmatrix}, \quad (6.1c)$$

and coordinated of the ball joints with respect to $\{1\}$ are

$${}^1P_{B_1} = \begin{bmatrix} r \\ 0 \\ 0 \end{bmatrix}, \quad (6.2a)$$

$${}^1P_{B_2} = \begin{bmatrix} -\frac{1}{2}r \\ \frac{\sqrt{3}}{2}r \\ 0 \end{bmatrix}, \quad (6.2b)$$

$${}^1P_{B_3} = \begin{bmatrix} -\frac{1}{2}r \\ -\frac{\sqrt{3}}{2}r \\ 0 \end{bmatrix}. \quad (6.2c)$$

The position and orientation of $\{1\}$ with respect to $\{0\}$ is given by the homogeneous transformation

$${}^0T_1 = \left[\begin{array}{ccc|c} N & O & A & {}^0P_1 \\ \hline 0 & 0 & 0 & 1 \end{array} \right] = \begin{bmatrix} n_x & o_x & a_x & x_c \\ n_y & o_y & a_y & y_c \\ n_z & o_z & a_z & z_c \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad (6.3)$$

where N , O , and A are the directional cosines of the principle axes of Frame $\{1\}$ with respect to those of Frame $\{0\}$. The orthonormality of the unit directional cosine vectors impose six constraint equations on the nine elements,

$$\begin{aligned}
N \cdot N &= 1, \\
O \cdot O &= 1, \\
A \cdot A &= 1, \\
O \cdot A &= 0, \\
O \cdot N &= 0, \\
A \cdot N &= 0.
\end{aligned} \tag{6.4}$$

The Cartesian position of the i^{th} ball joint with respect to $\{0\}$ can be found by

$$\begin{bmatrix} X_{Bi} \\ Y_{Bi} \\ Z_{Bi} \\ 1 \end{bmatrix} = {}^0T_1 \begin{bmatrix} x_{Bi} \\ y_{Bi} \\ z_{Bi} \\ 1 \end{bmatrix}. \tag{6.5}$$

The vector ${}^0P_1 = [x_c \ y_c \ z_c]^T$ in (6.3) is the position of the origin of $\{1\}$ with respect to $\{0\}$ and since Frame $\{1\}$ has the same orientation as $\{0\}$ at initialization,

$${}^0P_1 = [0 \ 0 \ z_{c0}]^T. \tag{6.6}$$

The length of the link, which is equal to the distance between the i^{th} ball joint and i^{th} pin joint is

$$L_1^2 = (n_x \rho + X_c - 1)^2 + (n_y \rho + Y_c)^2 + (n_z \rho + Z_c)^2, \tag{6.7}$$

$$\begin{aligned}
L_2^2 &= \frac{1}{4} \left[(-n_x \rho + \sqrt{3} o_y \rho + 2X_c + 1)^2 \right. \\
&\quad \left. + (-n_y \rho + \sqrt{3} o_y \rho + 2Y_c - \sqrt{3})^2 + (-n_z \rho + \sqrt{3} o_z \rho + 2Z_c)^2 \right], \tag{6.8}
\end{aligned}$$

$$\begin{aligned}
L_3^2 &= \frac{1}{4} \left[(-n_x \rho - \sqrt{3} o_x \rho + 2X_c + 1)^2 \right. \\
&\quad \left. + (-n_y \rho - \sqrt{3} o_y \rho + 2Y_c + \sqrt{3})^2 + (-n_z \rho - \sqrt{3} o_z \rho + 2Z_c)^2 \right], \tag{6.9}
\end{aligned}$$

where

$$L_i = \frac{l_i}{R}, \quad i = 1, 2, 3, \tag{6.10}$$

and in our design the ratio

$$\rho = \frac{r}{R} = 1, \tag{6.11}$$

and

$$X_c = \frac{x_c}{R}, Y_c = \frac{y_c}{R}, Z_c = \frac{z_c}{R}. \quad (6.12)$$

The bolts mounting the flexi-star impose the physical constraints equivalent to the pin joints, which limit the links to move in the planes

$$y = 0, y = \sqrt{3}x, y = -\sqrt{3}x, \quad (6.13)$$

respectively. The constraint equations are thus

$$n_y \rho + Y_c = 0, \quad (6.14)$$

$$-n_y \rho + \sqrt{3}o_y \rho + 2Y_c = -\sqrt{3}[-n_x \rho + \sqrt{3}o_x \rho + 2X_c], \quad (6.15)$$

$$-n_y \rho + \sqrt{3}o_y \rho + 2Y_c = \sqrt{3}[-n_x \rho - \sqrt{3}o_x \rho + 2X_c]. \quad (6.16)$$

Simplifying (6.15) and (6.16), we have

$$n_y = o_x, \quad (6.17)$$

$$X_c = \frac{1}{2}\rho(n_x - o_y). \quad (6.18)$$

Let the displacement of the intraocular shaft caused by tremor to be $(-d_x, -d_y, -d_z)$.

Thus, the canceling displacement of the intraocular shaft (d_x, d_y, d_z) would be:

$$d_x = x_c + la_x, \quad (6.19)$$

$$d_y = y_c + la_y, \quad (6.20)$$

$$d_z = z_c - z_{c0} + l(a_z - 1), \quad (6.21)$$

where l is the length of the intraocular shaft.

Let λ_1, λ_2 and λ_3 be the joint space variables or the extensions of the piezoelectric actuators. With six additional constraints imposed by the orthonormality of N, O , and A , the system of twelve equations (6.22a-1) is then solved for λ_1, λ_2 and λ_3 ,

$$N \cdot N = 1, \quad (6.22a)$$

$$O \cdot O = 1, \quad (6.22b)$$

$$A \cdot A = 1, \quad (6.22c)$$

$$O \cdot A = 0, \quad (6.22d)$$

$$O \cdot N = 0, \quad (6.22e)$$

$$A \cdot N = 0, \quad (6.22f)$$

$$d_x = x_c + la_x, \quad (6.22g)$$

$$d_y = y_c + la_y, \quad (6.22h)$$

$$d_z = z_c - z_{c0} + l(a_z - 1), \quad (6.22i)$$

$$n_y \rho + Y_c = 0, \quad (6.22j)$$

$$n_y = o_x, \quad (6.22k)$$

$$X_c = \frac{\rho}{2}(n_x - o_y). \quad (6.22l)$$

The system is first solved for the directional cosine o_y , in terms of the known constants, using symbolic mathematics software Maple v.5. To solve for o_y , it is necessary to root the following quartic equation:

$$A_{o_y} \xi^4 + B_{o_y} \xi^3 + C_{o_y} \xi^2 + D_{o_y} \xi + E_{o_y} = 0, \quad (6.23)$$

where

$$A_{o_y} = 8r^4 l^2 + r^6 + 16r^2 l^4, \quad (6.24)$$

$$B_{o_y} = 20r^3 d_x l^2 - 16r^2 l^4 + 20r^5 d_x - 4r^6 - 20r^4 l^2 + 16r d_x l^4 \quad (6.25)$$

$$C_{o_y} = 4d_x^2 l^4 - 12r^5 d_x - 12r^2 l^4 - 2r^4 d_y^2 + 8r^2 d_y^2 l^2 - 36r^3 d_x l^2 + 4r^4 d_x^2 \\ + 4d_y^2 l^4 + 20d_y^2 l^2 r^2 + 12r^4 l^2 - 24r d_x l^4 + 6r^6 \quad (6.26)$$

$$D_{o_y} = -8r^4 d_x^2 - 8d_x^2 l^4 + 20r d_y^2 d_x l^2 + 8r^2 l^4 + 12r^5 d_x - 4r^3 d_y^2 d_x \\ + 4r^4 l^2 - 16r^2 d_x^2 l^2 - 4r^6 - 20d_y^2 l^2 r^2 + 12r^3 d_x l^2 + 4r^4 d_y^2 \quad (6.27)$$

$$E_{o_y} = 4d_y^4 l^2 + 4r^3 d_y^2 d_x - 4d_y^2 l^4 - 2r^4 d_y^2 - 20r d_y^2 d_x l^2 - 4r^5 d_x - 4r^4 l^2 + r^6 \\ + 4r^4 d_x^2 + r^2 d_y^4 + 4d_x^2 l^4 + 8r^2 d_x^2 l^2 + 4r^3 d_x l^2 + 8r d_x l^4 + 4r^2 l^4 \quad (6.28)$$

Since the range of motion of the piezoelectric-driven links is small compared to the length of the intraocular shaft l and the radius of the rigid star r , o_y is always close to unity. Thus, the quartic equation (6.23) may be solved by Newton-Raphson or any other root-finding numerical methods [Press 99] by using 1.0 as the starting value and by constraining the root to be less than unity.

Having found o_y , o_z can be computed from

$$o_z = \sqrt{\frac{v_{o_z}}{-2r^2 + l^2}}, \quad (6.29)$$

where

$$v_{o_z} = -r^2 - 2rd_x + d_y^2 - 2o_y r^2 + 2o_y rd_x + 3o_y^2 r^2. \quad (6.30)$$

The sign of o_x is not known, but its square may be calculated from (6.22b),

$$o_x^2 = 1 - o_y^2 - o_z^2. \quad (6.31)$$

Then, from (6.22k),

$$n_y^2 = o_x^2. \quad (6.32)$$

To solve for n_x , it is required to solve for the root of the following quadratic equation,

$$A_{n_x} n_x^2 + B_{n_x} n_x + C_{n_x} = 0, \quad (6.33)$$

where

$$A_{n_x} = \frac{r^2}{4} + l^2, \quad (6.34)$$

$$B_{n_x} = -rd_x - \frac{r^2 o_y}{2}, \quad (6.35)$$

$$C_{n_x} = d_x^2 + rd_x o_y + \frac{1}{4} r^2 o_y^2 - l^2 (o_x^2 - o_y^2). \quad (6.36)$$

The correct root is then

$$n_x = \frac{-B_{n_x} + \sqrt{B_{n_x}^2 - 4A_{n_x} C_{n_x}}}{2A_{n_x}}. \quad (6.37)$$

The rest of the directional cosines are thus given by

$$a_x = \frac{2d_x - rn_x + ro_y}{2l}, \quad (6.38)$$

$$a_z = \sqrt{1 - a_x^2 - a_y^2}, \quad (6.39)$$

$$n_z = -\text{sgn}(d_x) \sqrt{1 - o_z^2 - a_z^2}, \quad (6.40)$$

$$n_y = \frac{-n_z o_z}{n_x o_y} = o_x, \quad (6.41)$$

$$a_y = \frac{d_y - rn_y}{l}. \quad (6.42)$$

The length of the links are related to the negated tremulous displacement (d_x, d_y, d_z) by

$$L_1 = \frac{R}{2} \sqrt{(n_x \rho + X_c - 1)^2 + (n_y \rho + Y_c)^2 + (n_z \rho + Z_c)^2} \quad (6.43)$$

$$L_2 = \frac{R}{2} \sqrt{(-n_x \rho + \sqrt{3} o_x \rho + 2X_c + 1)^2 + (-n_y \rho + \sqrt{3} o_y \rho + 2Y_c - \sqrt{3})^2 + (-n_z \rho + \sqrt{3} o_z \rho + 2Z_c)^2} \quad (6.44)$$

$$L_3 = \frac{R}{2} \sqrt{(-n_x \rho - \sqrt{3} o_x \rho + 2X_c + 1)^2 + (-n_y \rho - \sqrt{3} o_y \rho + 2Y_c + \sqrt{3})^2 + (-n_z \rho - \sqrt{3} o_z \rho + 2Z_c)^2} \quad (6.45)$$

where

$$X_c = \frac{d_x - l a_x}{R}, \quad (6.46)$$

$$Y_c = \frac{d_y - l a_y}{R}, \quad (6.47)$$

$$Z_c = \frac{d_z - l(a_x - 1) - z_{c0}}{R}. \quad (6.48)$$

Since the piezoelectric actuators are always vertically oriented, the lengths of the piezoelectric actuators λ_1 , λ_2 , and λ_3 are related to the length of the links by Pythagoras' Theorem, as depicted in Fig.6.6.

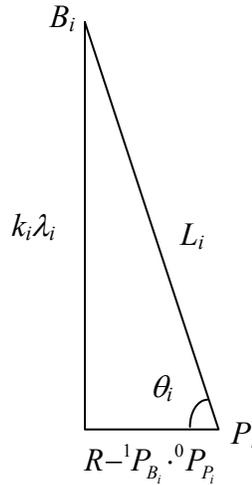


Fig. 6.6 Relationship between the link length L_i and the piezoelectric actuator length λ_i .

Thus,

$$\lambda_1 = \frac{1}{k_1} \sqrt{L_1^2 - (R - r n_x)^2}, \quad (6.49)$$

$$\lambda_2 = \frac{1}{k_2} \sqrt{L_2^2 - \left(R - \sqrt{\left(-\frac{1}{2}rn_x + \frac{\sqrt{3}}{2}ro_x \right)^2 + \left(-\frac{1}{2}rn_y + \frac{\sqrt{3}}{2}ro_y \right)^2} \right)^2}, \quad (6.50)$$

$$\lambda_3 = \frac{1}{k_3} \sqrt{L_3^2 - \left(R - \sqrt{\left(-\frac{1}{2}rn_x - \frac{\sqrt{3}}{2}ro_x \right)^2 + \left(-\frac{1}{2}rn_y + \frac{\sqrt{3}}{2}ro_y \right)^2} \right)^2}, \quad (6.51)$$

where k_i , $i = 1, 2, 3$, are the mechanical amplification gain of the linkages, which are equal to 8 as described in the previous section.

6.4 Forward Kinematics & Workspace Analysis

In order to perform a workspace analysis of the intraocular shaft tip, it is necessary to first compute the forward kinematics of the parallel manipulator, i.e. given the length of the piezoelectric actuators, what is the displacement of the tool-tip in Cartesian space.

Let θ_1 , θ_2 , and θ_3 are defined to be the angles between the links L_1 , L_2 , and L_3 and the base star. It can be shown that θ_i are related to L_i by

$$L_1^2 + L_2^2 + 3 - 3\rho^2 + L_1L_2 \cos \theta_1 \cos \theta_2 - 2L_1L_2 \sin \theta_1 \sin \theta_2 - 3L_1 \cos \theta_1 - 3L_2 \cos \theta_2 = 0, \quad (6.52)$$

$$L_2^2 + L_3^2 + 3 - 3\rho^2 + L_2L_3 \cos \theta_2 \cos \theta_3 - 2L_2L_3 \sin \theta_2 \sin \theta_3 - 3L_2 \cos \theta_2 - 3L_3 \cos \theta_3 = 0, \quad (6.53)$$

$$L_3^2 + L_1^2 + 3 - 3\rho^2 + L_3L_1 \cos \theta_3 \cos \theta_1 - 2L_3L_1 \sin \theta_3 \sin \theta_1 - 3L_3 \cos \theta_3 - 3L_1 \cos \theta_1 = 0. \quad (6.54)$$

The coordinates of ball joints with respect to $\{0\}$ in terms of θ_i and L_i are given by

$$X_{b1} = 1 - L_1 \cos \theta_1, \quad Y_{b1} = 0, \quad Z_{b1} = L_1 \sin \theta_1, \quad (6.55)$$

$$X_{b2} = -\frac{1}{2}(1 - L_2 \cos \theta_2), \quad Y_{b2} = +\frac{\sqrt{3}}{2}(1 - L_2 \cos \theta_2), \quad Z_{b2} = L_2 \sin \theta_2, \quad (6.56)$$

$$X_{b3} = -\frac{1}{2}(1 - L_3 \cos \theta_3), \quad Y_{b3} = -\frac{\sqrt{3}}{2}(1 - L_3 \cos \theta_3), \quad Z_{b3} = L_3 \sin \theta_3. \quad (6.57)$$

The position of the centroid of the upper equivalent triangle or the origin of $\{1\}$ with respect to $\{0\}$ can be found by

$$X_c = \frac{1}{3} \sum_{i=1}^3 \frac{x_{bi}}{R}, \quad (6.58)$$

$$Y_c = \frac{1}{3} \sum_{i=1}^3 \frac{y_{bi}}{R}, \quad (6.59)$$

$$Z_c = \frac{1}{3} \sum_{i=1}^3 \frac{z_{bi}}{R}. \quad (6.60)$$

Having found (X_c, Y_c, Z_c) , the tip position can be computed by taking a cross product between the position vectors (X_c, Y_c, Z_c) to the ball joints, ${}^0P_{1B_i}$ and ${}^0P_{1B_j}$, $i, j = 1, 2, 3$, $i \neq j$, normalize the resultant vector and multiply by the length of the intraocular shaft l ,

$${}^0P_{1B_i} = {}^0P_{B_i} - {}^0P_1, i = 1, 2, 3, \quad (6.61)$$

$${}^0P_{tip} = \frac{{}^0P_{1B_i} \times {}^0P_{1B_j}}{\|{}^0P_{1B_i} \times {}^0P_{1B_j}\|} l. \quad (6.62)$$

The workspace of the intraocular shaft manipulator, generated by (6.52)-(6.62), is shown in Fig. 6.7. The tremor space, approximated by a $\varnothing 50 \mu\text{m}$ sphere is also plotted for comparison.

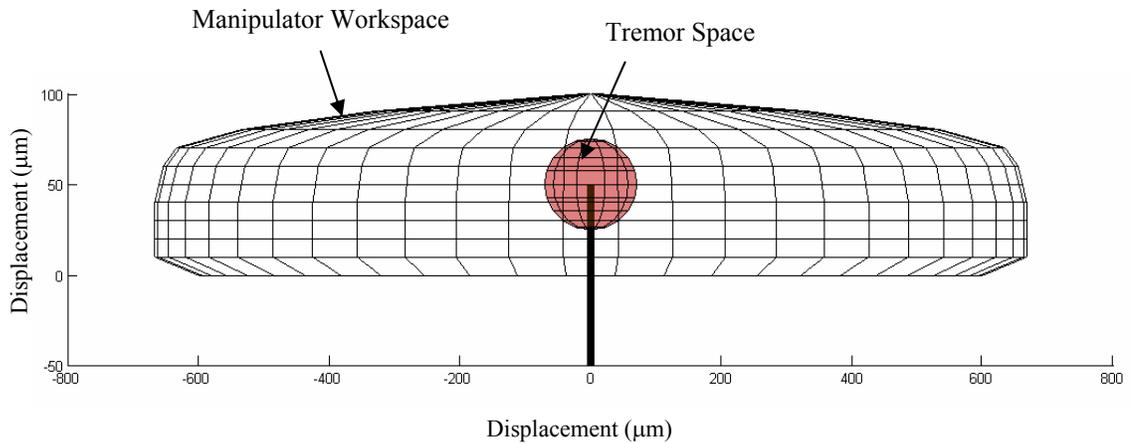


Fig. 6.7 Comparison of the intraocular shaft manipulator workspace versus the tremor space. Tremor space is approximated by a $\varnothing 50 \mu\text{m}$ sphere.

References

- [Gupta 99] P. K. Gupta, P. S. Jensen, E. de Juan, Jr., "Surgical Forces and Tactile Perception During Retinal Microsurgery," *Medical Image Computing and Computer-Assisted Interventions (MICCAI)*, Cambridge, U.K., Springer, Vol. 1679, *Lecture Notes in Computer Science*, pp. 1218-1225, 1999.
- [Polytec-PI] http://www.physikinstrumente.de/pdf/Picma_Datasheet.pdf
- [Lee 88] K.-M. Lee and D. K. Shah, "Kinematic analysis of a three-degrees-of-freedom in-parallel actuated manipulator," *IEEE Trans. Robotics & Automation*, Vol. 4, No. 3, pp.354-360, 1988.
- [Press 99] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, "Numerical Recipes in C," Cambridge University Press, Cambridge, U.K., 2nd Ed., Ch.9, 1999.

Chapter 7

Control of Manipulator System

7.1 Piezoelectric Actuator Hysteresis

A piezoelectric ceramic is an excellent choice as a micropositioning actuator because of its high output force, large bandwidth and fast response time. However, the existence of nonlinear multi-path hysteresis in piezoelectric material complicates the control of a piezoelectric actuator in high precision applications. The maximum hysteretic error is typically about 15% in static positioning applications. Still worse, the hysteresis is rate-dependent, increasing with the rate of control input, as shown in Fig. 7.1.

While the formation theory of hysteresis [Chen 80] and its complex looping behavior [Krasnosel'skii 89] have been well documented, literature on piezoelectric hysteresis rate dependence is scarce. Lacking a more profound understanding of the underlying physics, any attempt to account for the hysteresis rate dependence has to be phenomenological. Current research in hysteresis modeling and compensation can be broadly classified into three categories:

- (I) Electric charge control;
- (II) Closed-loop displacement control;
- (III) Linear control with feedforward inverse hysteresis model.

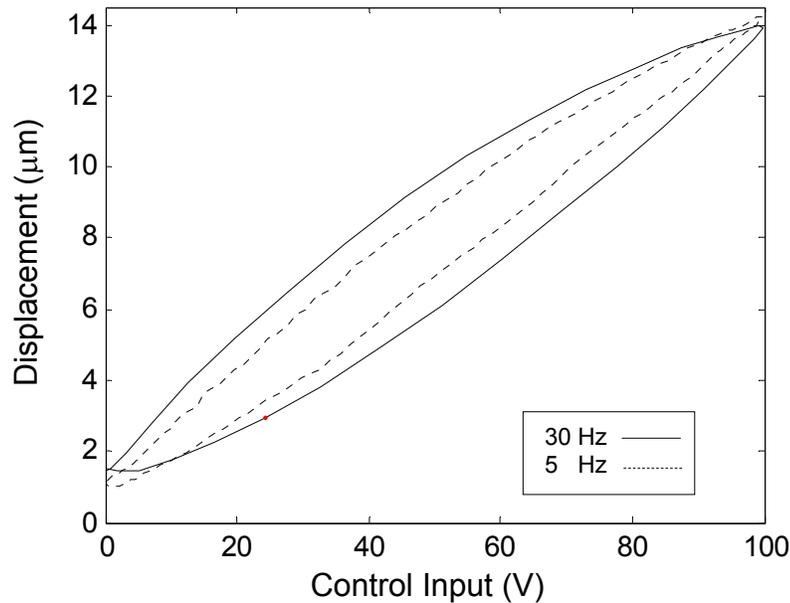


Fig. 7.1 Hysteresis of piezoelectric material is rate-dependent. The plots show the response of a piezoelectric actuator at two different driving frequencies.

The first category exploits the fact that the relationship between the deformation of a piezoceramic and the induced charge has significantly less hysteresis than that between deformation and applied voltage [Newcomb 82], [Furutani 98]. However, this approach requires specialized equipment to measure and amplify the induced charge and will inevitably reduce the responsiveness of the actuator. There has been little or no discussion on the effectiveness of this method in tracking more complex, nonstationary motion profiles, where the hysteretic rate dependence comes into play.

Most commercial systems (e.g. Polytec PI, Inc., Dynamic Structures and Materials, LLC., Melles Griot, Inc., Michigan Aerospace Corp. etc.) fall into the second category, normally using strain gauges as the feedback sensors. These systems can achieve sub-micron and even nano-level positioning precision but are generally more suitable for static positioning applications. When driven to track a 12.5 μm p-p sinusoid, the Polytec PI NanoCube™ P-611 three-axis micropositioner (Polytec PI, Inc., Germany) exhibits a system response that resembles that of a low-pass filter, i.e. diminishing magnitude gain with frequency increment as shown in Fig. 7.2. The effect of hysteresis remains evident

(Fig. 7.3) and the closed-loop controller manages rms tracking errors of about $40.0 \mu\text{m}$ (40% of p-p amplitude) at 10 Hz, without taking into account the phase lag.

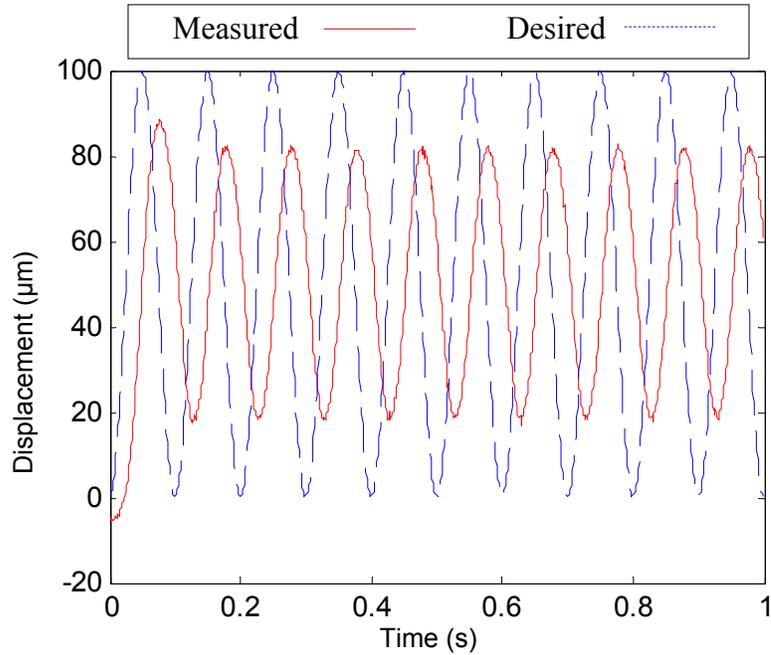


Fig. 7.2 The response of the Polytec PI NanoCube™ P-611 three-axis micropositioner (Polytec PI, Inc., Germany) with 10 Hz p-p sinusoidal input. The closed-loop controller exhibits a system response that resembles that of a low-pass filter, with attenuated amplitude and phase lag.

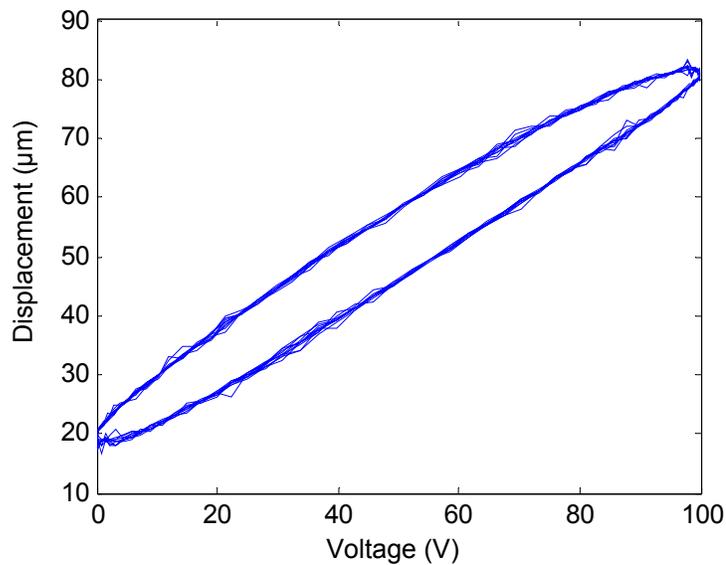


Fig. 7.3 Hysteresis is evident in the response of the Polytec PI NanoCube™ P-611 three-axis micropositioner (Polytec PI, Inc., Germany) with 10 Hz p-p sinusoidal input.

Among the proposed closed-loop schemes capable of tracking control, a few use different varieties of linear control schemes after linearizing the hysteretic nonlinearity [Chen 99], [Jan 00]; Tao et al. [Tao 95] uses adaptive control with an approximate model of the hysteresis; others propose using a neural network to learn the nonlinearity [Ku 00] or a combination of neural network with adaptive control [Hwang 03]. Simulation results of [Hwang 03] show an error of about 5% tracking a multiple frequencies stationary sinusoid with amplitude close to $70 \mu\text{m}$ p-p. None of these control schemes are designed to handle more dynamic tracking scenarios.

The main idea of category III is to obtain a mathematical model that closely describes the complex hysteretic behavior, and then to implement an inverse feedforward controller based on the inverse hysteresis model to linearize the actuator response (see Fig.7.4).

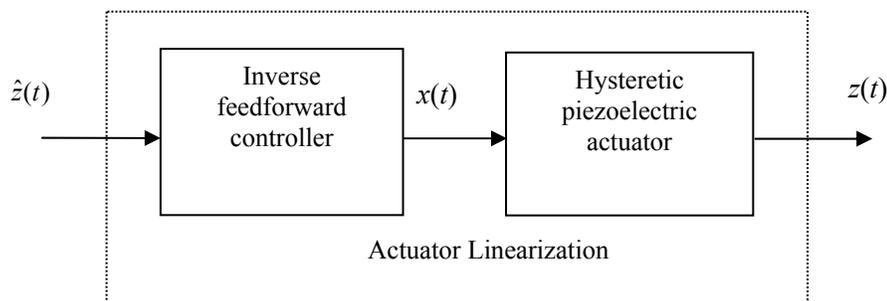


Fig. 7.4 Piezoelectric actuator linearization with inverse feedforward control strategy.

Among the proposed hysteresis models, e.g. the Maxwell's slip model [Goldfarb 97], [Choi 97], the Duhem model [Stepanenko 98], polynomials approximation [Chonan 96] etc., the Preisach model [Mayergoyz 91; Hughes 95; Hu 99] and its variations are by far the most well-known and widely used in both closed-loop [Ge 96; Majima 01; Cruz-Hernandez 01] and open-loop [Galinaitis] systems. Ge and Jouaneh [Ge 96] combine the feedforward model with a PID feedback loop and their implementation attains a maximum error of about 1.5 – 3.5% tracking $13 \mu\text{m}$ p-p stationary sinusoids from 0.1 – 20 Hz. However this method does not work for non-stationary sinusoid because of the intrinsic properties of the classical Preisach model. Galinaitis [Galinaitis] uses a Preisach type Kransnosel'skii-Pokrovskii (KP) operator to model and control piezoelectric actuators in open-loop and reports a maximum tracking error of 3.9% with a 0.05 Hz

stationary sinusoid. The tracking errors increase to about 6 – 9.4% when tracking sinusoids with reduced amplitudes.

Another important subclass of the Preisach model is the Prandtl-Ishlinskii (PI) model. The main advantages of this approach over the classical Preisach operator are that it is less complex and its inverse can be computed analytically, thus making it more attractive for real-time applications [Kuhlen 01]. These PI approaches reduce maximum hysteretic error to about 1 – 3% in open-loop control with quasi-static tracking [Krejci 01], [Kuhlen 02] and to about 1% tracking a non-stationary constant rate saw-tooth profile with closed-loop adaptive control [Kuhlen 99].

One basic assumption of the Preisach type model is that hysteresis is rate-independent. To date, little work has been done to explicitly model the rate dependence of hysteresis. Tan and Baras [Tan 02] extend the Preisach operator to model and control magnetostrictive actuators at > 5 Hz where hysteresis can no longer be assumed to be rate-independent. They report a maximum error of about 7.5% when tracking a nonstationary dynamic motion profile in closed-loop control. Smith et al. [Smith 01] report that hysteresis of piezoelectric materials is rate-dependent even at very low frequencies (< 1 Hz).

In this chapter, we present an extension to the PI operator to also model the rate-dependent hysteresis characteristic of a piezoelectric actuator. We will show experimentally that the slope of the hysteresis loading curve is dependent on the rate of the control input. We implement the rate-dependent PI hysteresis model with open-loop control and compare the experimental results with the rate-independent case. A discussion on the significance of the result and the limitations of our model is also presented. At the end of the chapter, the proposed controller is implemented to control the 3-DOF parallel intraocular shaft manipulator. Experimental tracking result of the manipulator system will also be presented.

7.2 Prandtl-Ishlinskii (PI) Hysteresis Model

This section describes the modeling of hysteresis using the modified PI operator proposed by Kuhnen et al. [Kuhlen 01; Kuhlen 99], with a slightly different treatment to

account for the one sided characteristic of many commercial piezoelectric actuators that are driven by non-negative control voltage. We also introduce modifications to Kuhnen's approach to improve the modeling accuracy, which will be discussed in greater details at due course.

7.2.1 Prandtl-Ishlinskii (PI) Operator

The elementary operator in the PI hysteresis model is a rate-independent backlash or linear-play operator. It is commonly used in the modeling of backlash between gears with one degree of freedom. A backlash operator is defined by

$$\begin{aligned} y(t) &= H_r[x, y_0](t) \\ &= \max\{x(t) - r, \min\{x(t) + r, y(t - T)\}\} \end{aligned} \quad (7.1)$$

where x is the control input, y is the actuator response, r is the control input threshold value or the magnitude of the backlash, and T is the sampling period. The initial consistency condition of (7.1) is given by

$$y(0) = \max\{x(0) - r, \min\{x(0) + r, y_0\}\} \quad (7.2)$$

with $y_0 \in \mathfrak{R}$, and is usually but not necessarily initialized to 0. Multiplying the backlash operator H_r by a weight value w_h , we have the generalized backlash operator,

$$y(t) = w_h H_r[x, y_0](t). \quad (7.3)$$

The weight w_h defines the gain of the backlash operator ($w_h = 1$ represents a 45° slope) and may be viewed as the gear ratio in gear mechanical play analogy (see Fig.7.5).

Complex hysteretic nonlinearity can be modeled by a linearly weighted superposition of many backlash operators with different threshold and weight values,

$$y(t) = \vec{w}_h^T \vec{H}_r[x, \vec{y}_0](t). \quad (7.4)$$

with weight vector $\vec{w}_h^T = [w_{h0} \dots w_{hn}]$ and $\vec{H}_r[x, \vec{y}_0](t) = [H_{r0}[x, y_{00}](t) \dots H_{rn}[x, y_{0n}](t)]^T$ with the threshold vector $\vec{r} = [r_0 \dots r_n]^T$ where $0 = r_0 < \dots < r_n$, and the initial state vector $\vec{y}_0 = [y_{00} \dots y_{0n}]^T$. The control input threshold values \vec{r} are usually, but not necessarily, chosen to be equal intervals. If the hysteretic actuator starts in its de-energized state, $\vec{y}_0 = \vec{0}_{n \times 1}$.

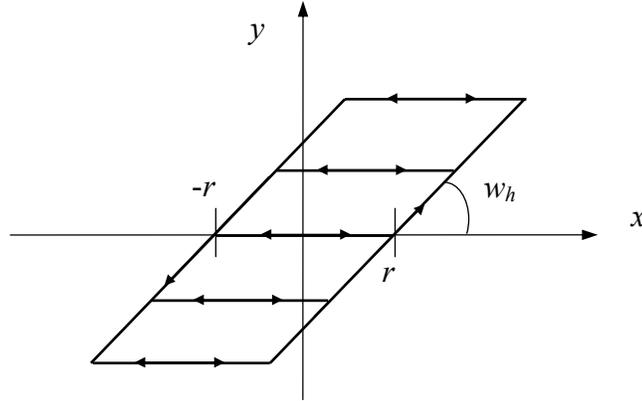


Fig. 7.5 The rate-independent generalized backlash operator is characterized by the threshold or backlash magnitude, r , and the weight or backlash gain, w_h .

Equation (7.4) is the PI hysteresis operator in its threshold discrete form. The hysteresis model formed by the PI operator is characterized by the initial loading curve (see Fig. 7.6). It is a special branch traversed by equation (7.4) when driven by a monotonically increasing control input with its state initialized to zero (i.e. $y(0) = 0$). The initial loading curve is defined by the weight values \bar{w}_h and threshold values \bar{r} ,

$$\varphi(r) = \sum_{j=0}^i w_{hj} (r - r_j), \quad r_i \leq r < r_{i+1}, \quad i = 0, \dots, n. \quad (7.5)$$

The slope of the piecewise linear curve at interval i is defined by W_{hi} , the sum of the weights up to i ,

$$W_{hi} = \frac{d}{dr} \varphi(r) = \sum_{j=0}^i w_{hj}. \quad (7.6)$$

The subsequent trajectory of the PI operator beyond the initial loading curve, with non negative control input is shown as the dotted loop in Fig. 7.6. The hysteresis loop formed by the PI operator does not return to zero with the control input. This behavior of the PI operator closely resembles the hysteresis of a piezoelectric actuator.

The backlash operators cause each of the piecewise linear segments to have a threshold width of $2r$ beyond the initial loading curve. As such, there is no need to define any backlash operators beyond the mid-point of the control input range, i.e. $r_n \leq \frac{1}{2} \max[2]$. This also implies that the backlash operators have descending importance from

the first to the last, since the first operator is always used and the subsequent operators are only used when the control inputs go beyond their respective threshold values, r_i 's. Moreover, observations from the piezoelectric hysteretic curves suggest that more drastic changes in the slope occur after the turning points, i.e. in the region of the first few backlash operators. To strike a balance between model accuracy and complexity, we propose to importance sampled the threshold intervals \bar{r} , i.e. to have finer intervals for the first few backlash operators and have increasing intervals for the subsequent ones. More details on the trade-offs of choosing different threshold intervals will be discussed in Section 7.5.

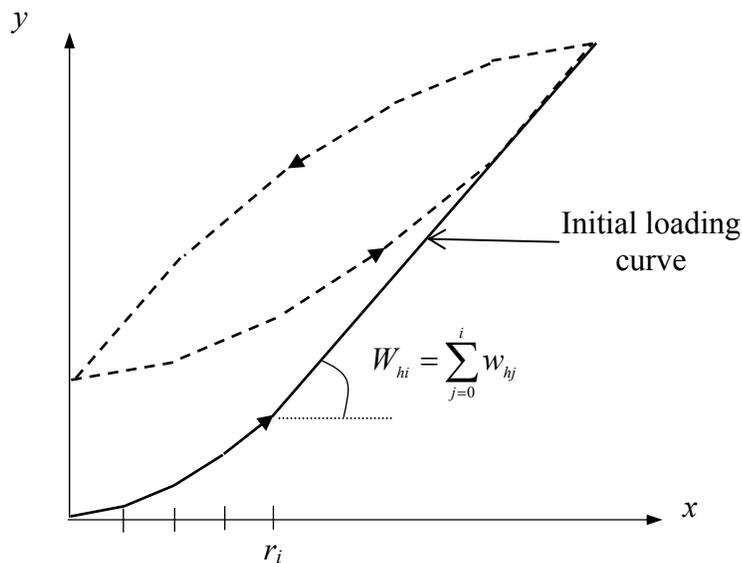


Fig. 7.6 The PI hysteresis model with $n = 4$. The hysteresis model is characterized by the initial loading curve. The piecewise linear curve is defined by the equally spaced threshold values \bar{r} and the sum of the weight values \bar{w}_h .

7.2.2 Modified Prandtl-Ishlinskii (PI) Operator

The PI operator inherited the symmetry property of the backlash operator about the center point of the loop formed by the operator. The fact that most real actuator hysteretic loops are not symmetric weakens the model accuracy of the PI operator. To overcome this overly restrictive property, a saturation operator is combined in series with the hysteresis operator. A saturation operator is a weighted linear superposition of linear-stop

or one-sided dead zone operators; a dead zone operator is a non-convex, non-symmetrical, and memory free nonlinear operator as shown in Fig. 7.7.

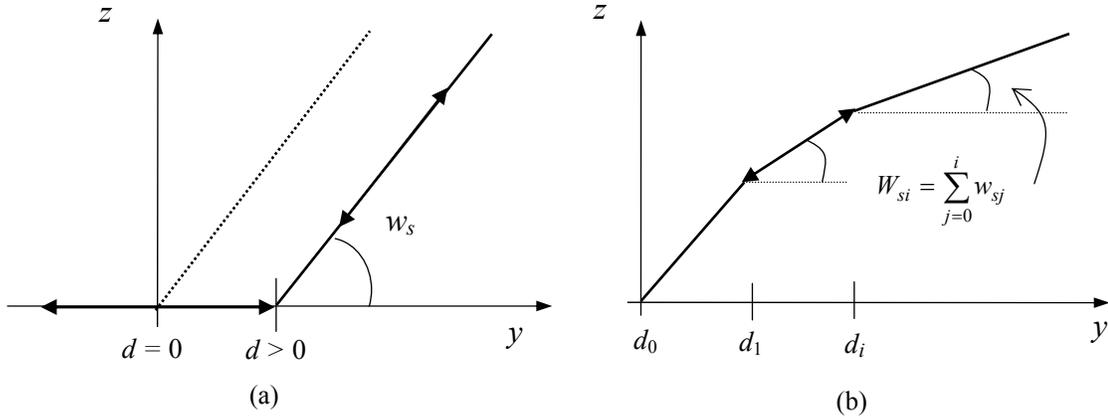


Fig. 7.7 (a) The one-sided dead-zone operator is characterized by the threshold, d , and the gain w_s .
 (b) The saturation operator with $m = 2$. The slope of the piecewise linear curve at interval i , W_{si} is defined by the sum of the weights up to i .

A one-sided dead-zone operator and a saturation operator are given by

$$S_d[y](t) = \begin{cases} \max\{y(t) - d, 0\}, & d > 0 \\ y(t), & d = 0 \end{cases} \quad (7.7)$$

$$z(t) = \vec{w}_s^T \vec{S}_d[y](t). \quad (7.8)$$

where y is the output of the hysteresis operator, z is the actuator response, $\vec{w}_s^T = [w_{s0} \dots w_{sm}]$ is the weight vector, $\vec{S}_d[y](t) = [S_{d0}[y](t) \dots S_{dm}[y](t)]^T$ with the threshold vector $\vec{d} = (d_0 \dots d_m)^T$ where $0 = d_0 < d_1 < \dots < d_m$. For convenience, equal intervals of \vec{d} are usually, but not necessarily, chosen between d_0 and d_m . Good selection of \vec{d} depends on the shape of the hysteresis loop and some trials and errors.

The modified PI operator is thus

$$z(t) = \Gamma[x](t) = \vec{w}_s^T \vec{S}_d[\vec{w}_h^T \vec{H}_r[x, \vec{y}_0]](t). \quad (7.9)$$

7.2.3 Parameter Identification

To find the hysteresis model parameters, we first have to measure experimentally the responses of the piezoelectric actuator subjected to periodic control inputs. A good set of

identification data is one that covers the entire operational range of the piezoelectric actuator and at the nominal operational frequency. Next we decide the order of the PI operator (i.e. n) and the saturation operator (i.e. m), and set the threshold values \vec{r} and \vec{d} as described in the previous section. The weight parameters \vec{w}_h and \vec{w}_s are found by performing a least squares fit of (7.9) to the measured actuator response, i.e. minimizing the error equation:

$$E[x, z](\vec{w}_h, \vec{w}_s, t) = \vec{w}_s^T \vec{S}_d[\vec{w}_h^T \vec{H}_r[x, \vec{y}_0]](t) - z(t) \quad (7.10)$$

Fig. 7.8 shows superposition of the identified modified PI hysteresis model on the measured piezoelectric actuator response, subjected to a sinusoidal control input.

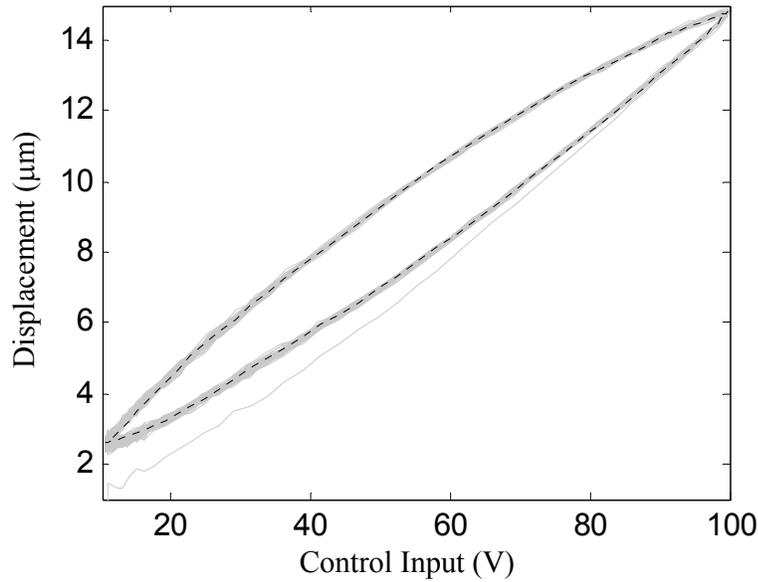


Fig. 7.8 The lighter solid lines are the measured piezoelectric actuator response, subjected to a 10 Hz, 12.5 μm p-p sinusoidal control input. The dark dotted line is the identified modified PI hysteresis model.

7.2.4 Inverse Modified Prandtl-Ishlinskii (PI) Operator

The key idea of an inverse feedforward controller is to cascade the inverse hysteresis operator, Γ^{-1} with the actual hysteresis which is represented by the hysteresis operator, Γ to obtain an identity mapping between the desired actuator output $x(t)$ and actuator response $z(t)$,

$$z(t) = \Gamma[\Gamma^{-1}[x]](t) = I[x](t) = x(t). \quad (7.11)$$

The operation of the inverse feedforward controller is depicted in Fig.7.9.

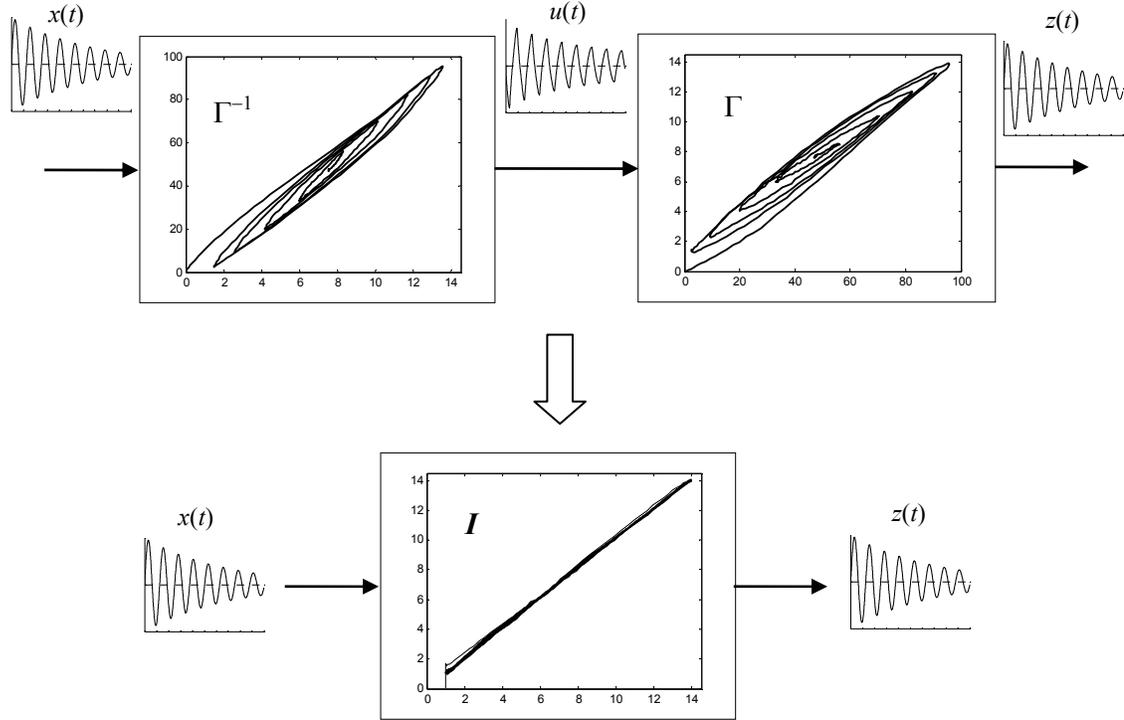


Fig. 7.9 An inverse feedforward controller. For a desired periodic actuator output $x(t)$, the inverse modified PI operator Γ^{-1} transforms it into a control input $u(t)$ that would produce a response $z(t)$ in the hysteretic system that closely resembles $x(t)$. This produces an equivalent control system with identity mapping between the desired output and the actual actuator response.

The inverse of a PI operator is also of the PI type. The inverse PI operator is given by

$$\Gamma^{-1}[z](t) = \bar{w}'_h{}^T \bar{H}_{r'} [\bar{w}'_s{}^T \bar{S}_{d'} [z], \bar{y}'_0](t) \quad (7.12)$$

where the inverse modified PI parameters can be found by

$$w'_{h0} = \frac{1}{w_{h0}}; \quad w'_{hi} = \frac{-w_{hi}}{(\sum_{j=0}^i w_{hj})(\sum_{j=0}^{i-1} w_{hj})}, \quad i = 1 \dots n;$$

$$r'_i = \sum_{j=0}^i w_{hj} (r_i - r_j), \quad y'_{0i} = \sum_{j=0}^i w_{hj} y_{0i} + \sum_{j=i+1}^n w_{hj} y_{0j}, \quad i = 0 \dots n; \quad (7.13)$$

$$w'_{s0} = \frac{1}{w_{s0}}; \quad w'_{si} = \frac{-w_{si}}{(\sum_{j=0}^i w_{sj})(\sum_{j=0}^{i-1} w_{sj})}, \quad i = 1 \dots m;$$

$$d'_i = \sum_{j=0}^i w_{sj} (d_i - d_j), \quad i = 0 \dots m; \quad (7.14)$$

The intuition of computing the inverse is to find the reflection of the resultant hysteresis looping curves about the 45° line as shown in Fig. 7.9.

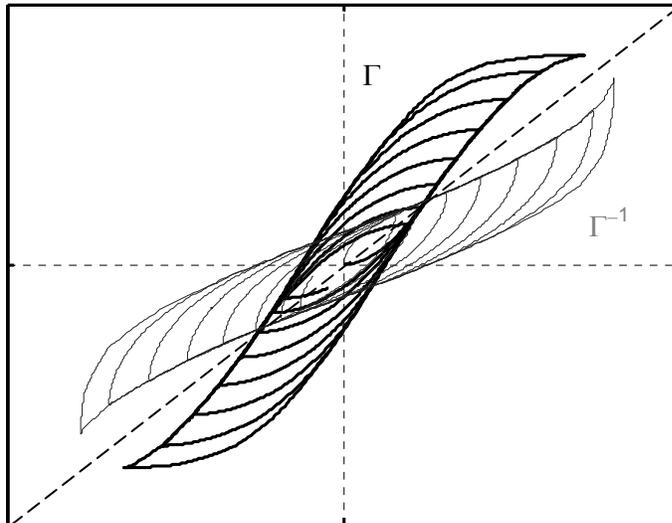


Fig. 7.10 The darker thick line is the modified PI hysteresis model, Γ . The inverse modified PI hysteresis model Γ^{-1} , represented by the lighter thin line, is the mirror image of the hysteresis model about the 45° line.

7.3 Rate-Dependent Prandtl-Ishlinskii (PI) Hysteresis Model

7.3.1 Rate-dependent Hysteresis Slope

We propose in this section an extension to the modified PI operator to also model the rate-dependent characteristics of the piezoelectric hysteresis.

One of the advantages of the PI hysteresis model is that it is purely phenomenological; there are no direct relationships between the modeling parameters and the physics of the hysteresis. Therefore we would model the rate-dependent hysteresis with reference only to the experimental observations. While the rate dependence of hysteresis is evident from Fig. 7.1, the sensitivity of actuator saturation to the actuation rate is not apparent. Hence we assume that saturation is not rate-dependent and hold the

saturation weights, \bar{w}_s as well as the threshold values, \bar{r} and \bar{d} , constant while attempting to construct a relationship between hysteresis and the rate of actuation $\dot{x}(t)$. We model the slope of the hysteresis curve (i.e. sum of the PI weights) at time t as the sum of the referenced hysteresis slope and a rate-dependent function,

$$W_{hi}(\dot{x}(t)) = \hat{W}_{hi} + f(\dot{x}(t)), \quad i = 1 \dots n; \quad (7.15)$$

where

$$\dot{x}(t) = \frac{x(t) - x(t-T)}{T}, \quad \dot{x}(0) = 0. \quad (7.16)$$

Equation (7.15) will be reduced to the referenced hysteresis slope, \hat{W}_{hi} , or to the rate-independent case if the rate-dependent term is zero.

7.3.2 Rate-dependent Model Identification

The response of a piezoelectric actuator subjected to periodic constant rate or saw-tooth control inputs are first measured. Measurements are made over a frequency band whose equivalent rate values cover the entire operational range of the actuation rates. For example, in an application tracking sinusoids of up to 12.5 μm p-p in the band of 1 to 19 Hz, the operational range of the actuation rate is from 0 to 746 $\mu\text{m/s}$, which corresponds to the rate of 12.5 μm p-p saw-tooth waveforms of up to about 60 Hz. The PI parameters identifications are then performed on each set of measured actuator responses.

The sum of the hysteresis weights W_{hi} , $i = 0 \dots n$, of each identification are plotted against the actuation rate $\dot{x}(t)$ as shown in Fig. 7.11.

We observed that the hysteresis slope of the piezoelectric actuator varies linearly with the actuation rate. Thus the rate-dependent hysteresis slope model would be:

$$W_{hi}(\dot{x}(t)) = \hat{W}_{hi} + c_i \dot{x}(t), \quad i = 0 \dots n \quad (7.17)$$

where c_i is the slope of the best fit line through the W_{hi} 's and the referenced slope \hat{W}_{hi} is the intercept of the best fit line with the vertical W_h axis or the slope at zero actuation. The individual rate-dependent hysteresis weight values can be calculated from

$$w_{hi}(\dot{x}(t)) = W_{hi}(\dot{x}(t)) - W_{h(i-1)}(\dot{x}(t)), \quad i = 1 \dots n;$$

$$w_{h0}(\dot{x}(t)) = W_{h0}(\dot{x}(t)). \quad (7.18)$$

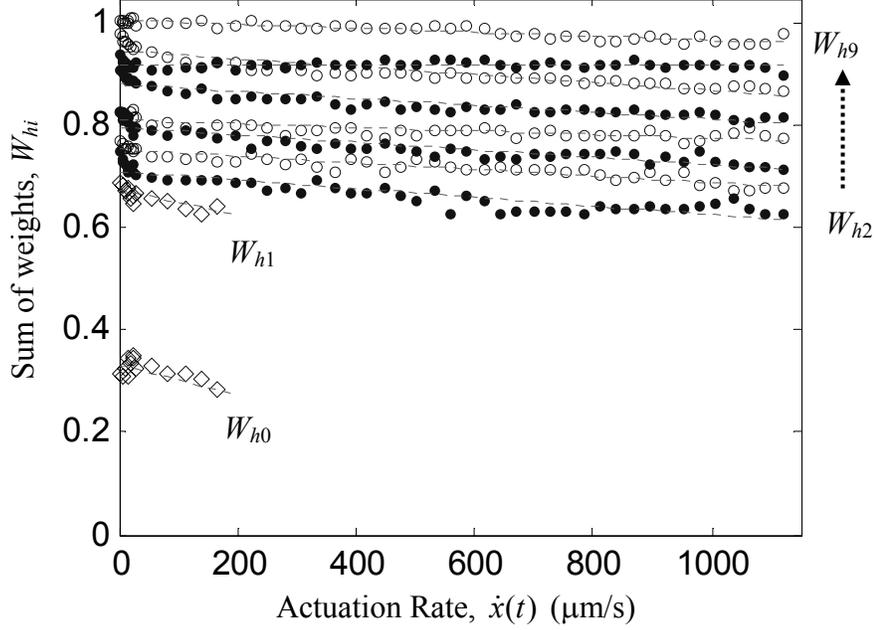


Fig. 7.11 Plot of the hysteresis slopes W_{hi} , $n = 9$, vs. actuation rate, $\dot{x}(t)$. Since the actuation rate is always slow at the turning points of a sinusoid, the first two sum of weights, W_{h0} and W_{h1} are modeled only up till $200 \mu\text{m/s}$.

7.3.3 Rate-dependent Modified Prandtl-Ishlinskii Operator

The rate-dependent modified PI operator is defined by

$$z(t) = \Gamma[x, \dot{x}](t) = \bar{w}_s^T \bar{S}_d \left[\bar{w}_h^T(\dot{x}) \bar{H}_r [x, \bar{y}_0] \right](t) \quad (7.19)$$

The inverse rate-dependent modified PI operator is also of the PI type:

$$\Gamma^{-1}[z](t) = \bar{w}_h^T(\dot{x}) \bar{H}_r \left[\bar{w}_s^T \bar{S}_d [z, \bar{y}'_0] \right](t). \quad (7.20)$$

The inverse rate-dependent parameters can be found by (7.13), replacing \bar{w}_h with the rate-dependent $\bar{w}_h(\dot{x})$.

$$w'_{h0}(\dot{x}(t)) = \frac{1}{w_{h0}(\dot{x}(t))}; \quad w'_{hi}(\dot{x}(t)) = \frac{-w_{hi}(\dot{x}(t))}{W_{hi}(\dot{x}(t))W_{h(i-1)}(\dot{x}(t))}, \quad i = 1 \dots n;$$

$$\begin{aligned}
r'_i &= \sum_{j=0}^i w_{hj}(\dot{x}(t))(r_i - r_j), \quad i = 0 \dots n; \\
y'_{0i} &= \sum_{j=0}^i w_{hj}(\dot{x}(t))y_{0i} + \sum_{j=i+1}^n w_{hj}(\dot{x}(t))y_{0j}, \quad i = 0 \dots n;
\end{aligned} \tag{7.21}$$

7.4 Motion Tracking Experiments – Piezoelectric Actuator

The proposed rate-dependent modified PI operators in modeling piezoelectric ceramic hysteretic nonlinearity, and its application in tracking control is evaluated on a piezoelectric actuator P-885.50 (Polytec PI, Inc.). Open-loop controllers with feedforward inverse rate-independent and rate-dependent modified PI hysteresis models are implemented. The displacement of the piezoelectric actuator in test is measured by an infrared interferometer (Philtec, Inc., Model D63) at 1 kHz sampling rate. The mean measurement noise of the interferometer is 0.07 μm rms.

The rate-independent model uses a PI operators of order 9 ($n = 9$, i.e. 10 backlash operators) and a saturation operator of order 3 ($m = 3$, i.e. 4 dead-zone operators); the PI thresholds \vec{r} selected are multiples of 5 from 0 to 45 and the saturation thresholds are $\vec{d} = [0 \ 63.3 \ 74.8 \ 87.3]^T$. The identification of the PI and saturation weights are based on measured response of the piezoelectric actuators subjected to a 10 Hz, 12.5 μm p-p sinusoid control input.

The rate-dependent model uses the same order of modified PI operator and saturation operator, i.e. $n = 9$ and $m = 3$. Importance sampled PI thresholds are used, with $\vec{r} = [0 \ 4 \ 8 \ 12 \ 16 \ 20 \ 25 \ 31 \ 38 \ 45]^T$. Identification of PI parameters are performed on the measured actuator response subjected to 12.5 μm p-p saw-tooth control input at intervals of 0.1 Hz in the band of 0.1 – 5.0 Hz, and at intervals of 1 Hz in the band of 5 – 40 Hz. Since we assume the actuator saturation is rate-independent, the same saturation thresholds and weights of the rate-independent model are used.

The first experiment compares the performance of the rate-independent and rate-dependent modified PI models based open-loop feedforward controllers in tracking a 10

Hz, 12.5 μm p-p stationary sinusoid. Next, the experiment is repeated to track 12.5 μm p-p stationary sinusoids at 1, 4, 7, 13, 16 and 19 Hz. The tracking rmse and maximum error of the controllers at each frequency is summarized in Table 7.1 and plotted in Fig. 7.12. Fig. 7.13(a) plots the hysteretic response of the piezoelectric actuator without any controller; Fig. 7.13 (b) and (c) show the tracking results of the rate-independent and rate-dependent inverse feedforward controllers.

Table 7.1 Measured Performance of the Rate-independent and Rate-dependent Inverse Feedforward Controllers in Tracking 12.5 μm p-p Stationary Sinusoids.

Freq. (Hz)	Without Model		Rate-independent		Rate-dependent	
	rmse (μm)	max ε (μm)	rmse (μm)	max ε (μm)	rmse (μm)	max ε (μm)
1	1.13	2.11	0.25	0.63	0.21	0.57
4	1.12	2.07	0.19	0.67	0.16	0.46
7	1.23	2.24	0.18	0.52	0.16	0.50
10	1.19	2.26	0.14	0.46	0.17	0.47
13	1.21	2.31	0.19	0.53	0.17	0.55
16	1.30	2.49	0.27	0.59	0.17	0.53
19	1.37	2.61	0.34	0.70	0.18	0.59
Mean	1.22	2.30	0.23	0.59	0.18	0.52
$\pm \sigma$	± 0.09	± 0.19	± 0.07	± 0.8	± 0.02	± 0.05

The rmse's and max errors are the mean results over a set of three 5-seconds (5000 data points) experiments.

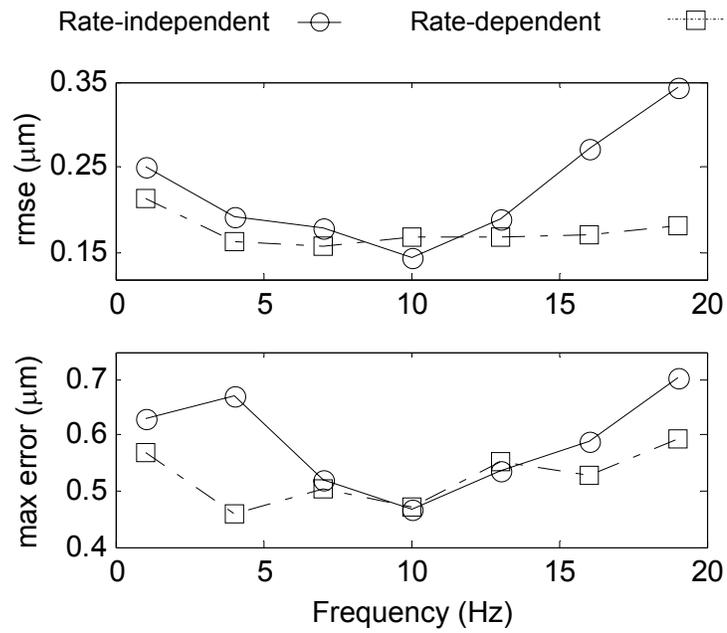
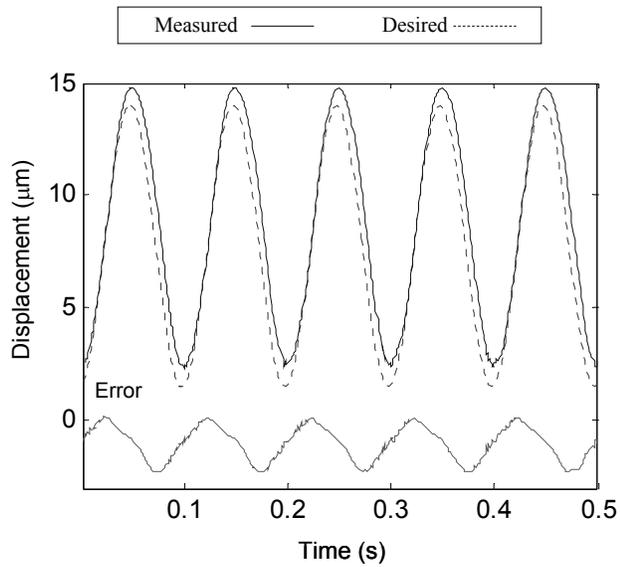


Fig. 7.12 Rmse and maximum errors of the rate-independent and rate-dependent controllers in tracking $12.5\mu\text{m}$ p-p stationary sinusoids at different frequencies. The rate-independent controller is based on the modified PI hysteresis model identified at the same 10Hz , $12.5\mu\text{m}$ p-p sinusoid.



(a) Without controller

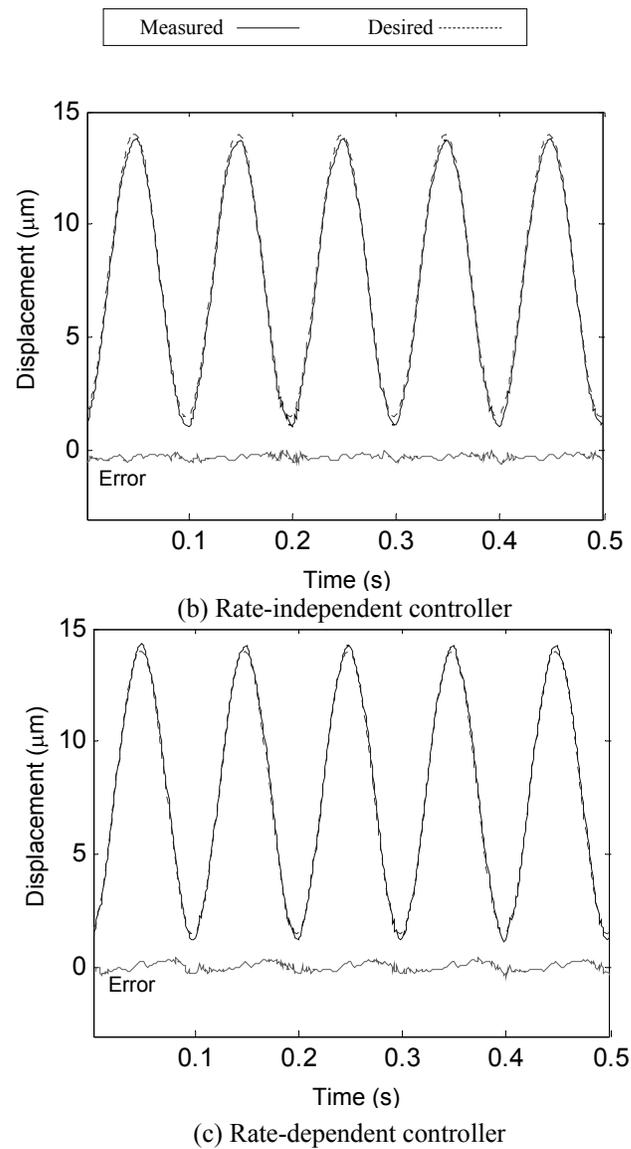


Fig. 7.13 Experimental open-loop tracking results of stationary $12.5 \mu\text{m}$ p-p sinusoids at 10 Hz. The rate-independent controller is based on the modified PI hysteresis model identified at the same 10Hz, $12.5 \mu\text{m}$ p-p sinusoid.

The second experiment compares the performance of the controllers in tracking a multi-frequency, non-stationary dynamic motion profile. The motion profile is made up of superimposed modulated 1, 10, and 19 Hz sinusoids with time-varying amplitudes.

The graphical results are plotted in Fig. 7.14, and the numerical results are summarized in Table 7.2.

Table 7.2 Measured Performance of the Rate-independent and Rate-dependent Inverse Feedforward Controllers in Tracking Multi-Frequency (1, 10, and 19 Hz) Non-Stationary Sinusoids.

	Without Model	Rate-independent	Rate-dependent
rmse $\pm \sigma$ (μm)	1.09 ± 0.02	0.18 ± 0.01	0.13 ± 0.01
$\frac{\text{rmse}}{\text{p-p amplitude}}$ (%)	8.7	1.5	1.1
max error $\pm \sigma$ (μm)	1.93 ± 0.03	0.68 ± 0.05	0.50 ± 0.08
$\frac{\text{max error}}{\text{p-p amplitude}}$ (%)	15.5	5.5	4.0

The rmse's and max errors are the mean results over a set of seven 5-seconds (5000 data points) experiments. The consistency of each set of experiments is indicated by the standard deviation σ .

Since physiological tremor is modulating and dynamic in nature, in order to cancel tremor or to track the negated tremor profile effectively, the piezoelectric actuator controller has to account for the rate-dependence of the hysteresis. The open-loop inverse feedforward controller with rate-dependent modified PI hysteresis model proposed in the previous sections is well-suited for this application

Fig. 7.15 shows the experimental results of a P-885.50 (Polytec PI, Inc.) piezoelectric actuator tracking real hand tremor recorded from ophthalmologic surgeons [Riviere 97], driven by an open-loop rate-dependent inverse feedforward controller. The tracking result of a rate-independent controller is also presented for comparison. The tracking errors are tabulated in Table 7.3.

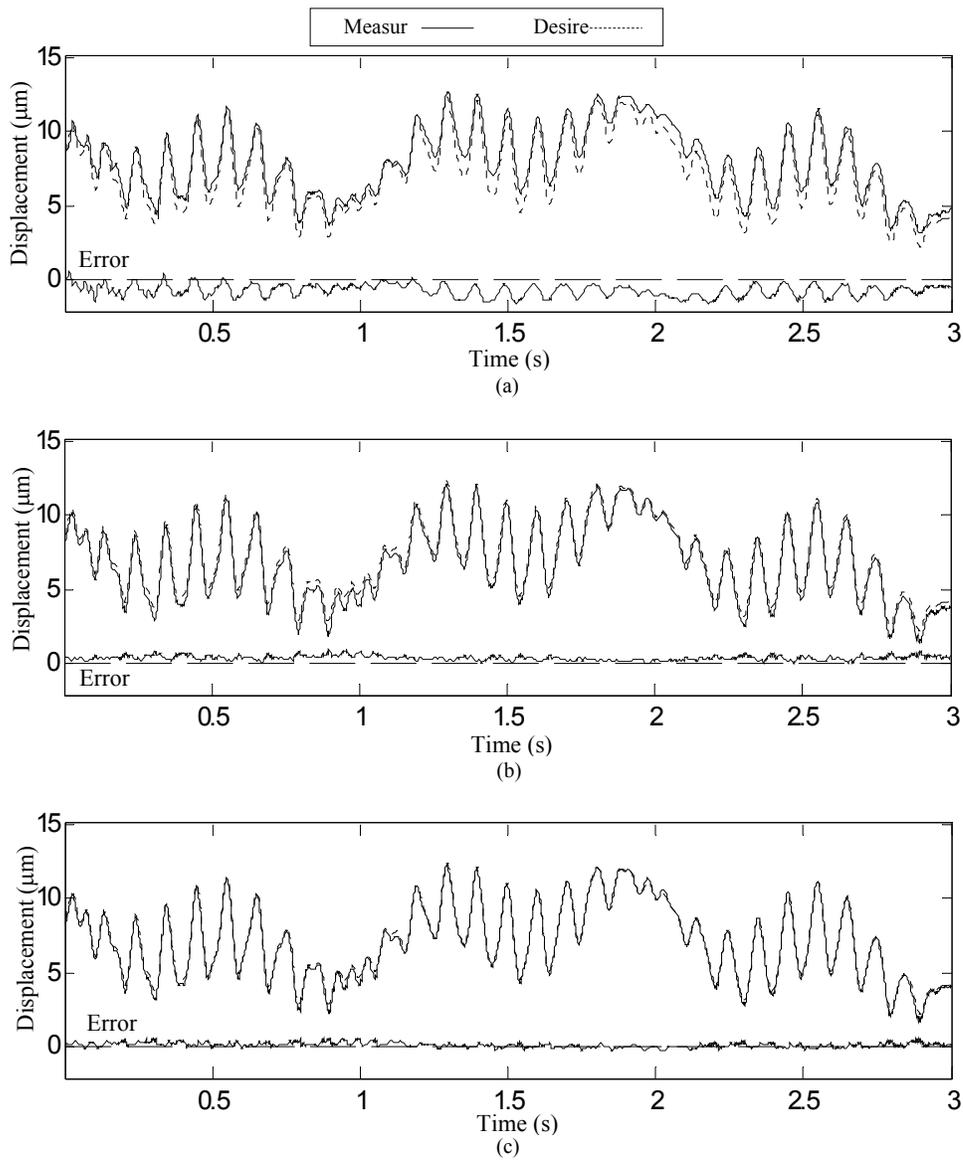
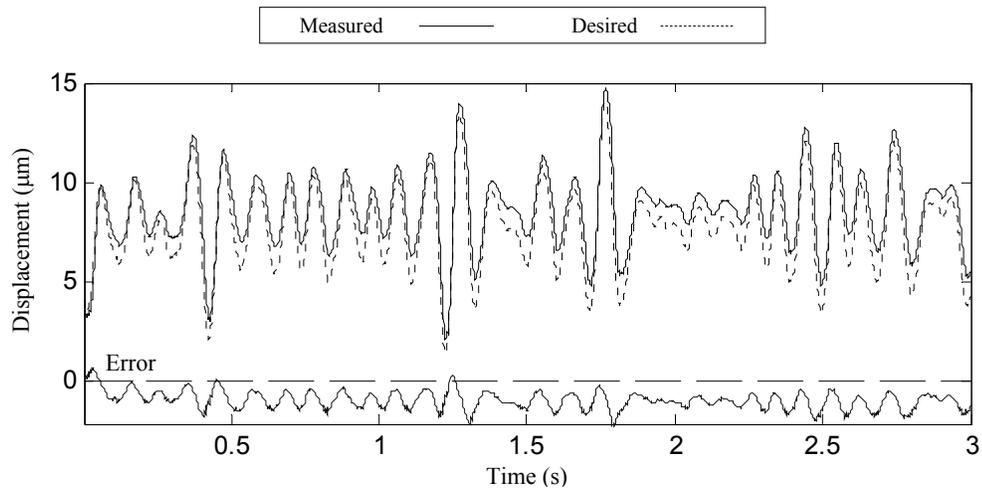
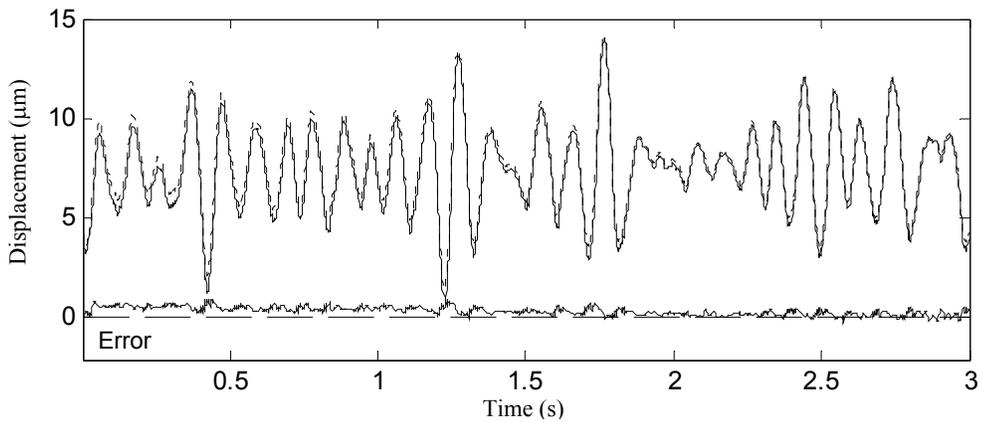


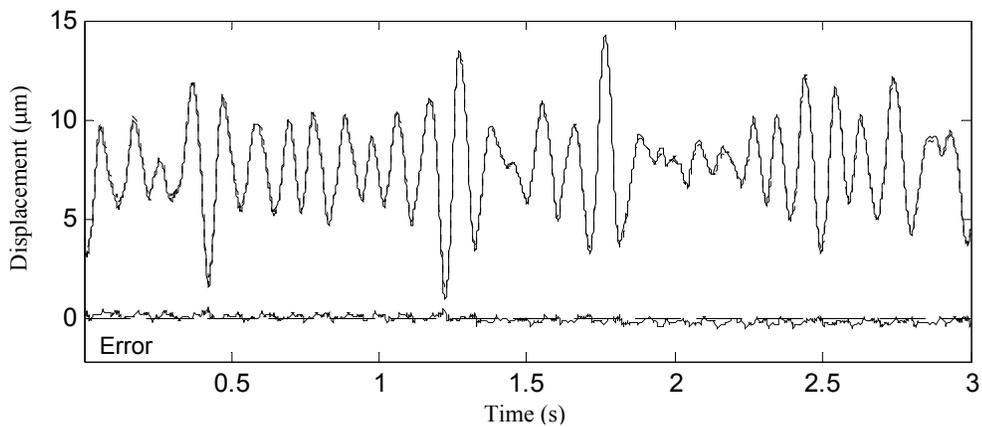
Fig. 7.14 Experimental open-loop tracking results of a multi-frequency, non-stationary dynamic motion profile. The motion profile is made up of superimposed modulated 1, 10, and 19 Hz sinusoids with time-varying amplitudes. The rate-independent controller is based on the modified PI hysteresis model identified at the same 10Hz, 12.5 μm p-p sinusoid. Transient error is observed for the rate-independent controller in the first 2 seconds. (a) Without compensation. (b) Rate-independent controller. (c) Rate-dependent controller.



(a) Without Controller



(b) Rate-independent Controller



(c) Rate-dependent Controller

Fig. 7.15 Experimental open-loop tracking results of a multi-frequency, non-stationary dynamic motion profile. The motion profile is made up of superimposed modulated 1, 10, and 19 Hz sinusoids with time-varying amplitudes. The rate-independent controller is based on the modified PI hysteresis model identified at the same 10Hz, 12.5 μm p-p sinusoid. Transient error is observed for the rate-independent controller in the first 2 seconds.

Table 7.3 Measured Performance of the Rate-independent and Rate-dependent Inverse Feedforward Controllers in Tracking Recorded Real Tremor of Ophthalmologic Surgeons.

	Without Model	Rate-independent	Rate-dependent
rmse $\pm \sigma$ (μm)	1.05 ± 0.05	0.34 ± 0.04	0.18 ± 0.02
$\frac{\text{rmse}}{\text{p-p amplitude}}$ (%)	8.3	2.7	1.4
max error $\pm \sigma$ (μm)	2.26 ± 0.06	0.95 ± 0.09	0.60 ± 0.06
$\frac{\text{max error}}{\text{p-p amplitude}}$ (%)	17.9	7.5	4.8

The rmse's and max errors are the mean results over a set of seven 5-seconds (5000 data points) experiments. The consistency of each set of experiments is indicated by the standard deviation σ .

7.5 Discussions

One limitation of all PI-type hysteresis models is that the inverse PI operator fails when the slope of the hysteresis loading curve or the sum of weights becomes zero or negative. Singularity occurs when the first PI weight w_{h0} is zero, the inverse weight becomes undefined (refer to (7.13) and (7.21)). When the slope is negative, the inverse hysteresis loading curve violates the fundamental assumption that it should be monotonically increasing, and since the one-to-one mapping relationship between the direct and the inverse model is lost, the PI operator breaks down. Singularities occur at higher frequency, where the hysteresis loop gets larger and is more rounded at the turning points. For a given piezoelectric actuator, the singular frequency of a PI hysteresis model depends on the choice of the thresholds \vec{r} . Choosing a larger first interval r_1 can raise the singularity frequency, but the trade off would be poorer modeling accuracy at the turning points. The singularity of our implementation for tracking $12.5 \mu\text{m}$ p-p sinusoids occurs at around 36 Hz.

Most of the tracking error of the rate-dependent controller is at the turning points of the motion profiles, where the actuation rate is slow. The inadequacy of the rate-

dependent modified PI hysteresis model at low rate is evident in Fig. 7.11, where it can be seen that the sum of the PI weights are no longer linear at rates lower than 20 $\mu\text{m/s}$. This is again manifested in Fig. 7.12, where the tracking rmse at 1 Hz is clearly the highest. This is again the consequence of the choice of the first PI threshold interval r_1 . A balance between tracking accuracy and tracking bandwidth should be made depending on the application.

Creep is not modeled in the proposed model because its effect is negligible for periodic excitation with frequency higher than 1 Hz. If quasi-static tracking is desired, since the rate-dependent model and its inverse are also of the PI type, the creep model proposed by Krejci et al. [Krejci 01] can be incorporated.

In the first experiment tracking 12.5 μm p-p stationary sinusoids, both the rate-independent and rate-dependent controllers significantly reduced the tracking error due to the hysteretic nonlinearity of the piezoelectric actuator. The rate-independent controller on the average reduces the tracking rmse and maximum error by 81.7% and 74.4% respectively in the band of 1 – 19 Hz. The best performance occurs at 10 Hz, in which its modified PI hysteresis model parameters are identified. The tracking accuracy deteriorates as the tracking frequency deviates from 10 Hz. The rate-dependent controller outperforms its rate-independent counterpart with tracking rmse and maximum error reduction of 85.6% and 77.2% respectively. The tracking accuracy remains consistent across the entire 1 – 19 Hz band. At 19 Hz, the tracking rmse is almost half of that of the rate-independent controller. Maximum tracking errors for both controllers occur in the transient phase at beginning of the test.

In the second experiment tracking a multi-frequency (1, 10, and 19 Hz) non-stationary motion profile, similar results are observed. Both controllers continue to perform well, with the rate-independent controller reducing the rmse and maximum error by 80% and 70%, and the rate-dependent controller doing noticeably better at 85% and 75% respectively. The rate-dependent controller registers 40% smaller tracking rmse as a result of the transient error, which is significant for the rate-independent controller in the first 2 seconds. One possible explanation for this behavior is that a piezoelectric ceramic has shape memory, transient error takes place until the actuator is driven into a steady-

state path in which the rate-independent modified PI hysteresis model is identified. The rate-dependent controller shows a smaller and faster transient because of the adaptive characteristic of the rate-dependent modified PI hysteresis model. Maximum tracking errors for both controllers again occur in the transient phase at beginning of the test. This could be the reason why the rate-dependent controller consistently performs only slightly better than the rate-independent controller in term of maximum error.

The open-loop rate-dependent controller continues to show superior tracking performance to the rate-independent controller. The tracking rmse of 1.4% of the maximum p-p tremor amplitude is very consistent with the previous experiments. This implies that the 3 DOF parallel manipulator driven by the P-885.50 piezoelectric actuators would be able to cancel a tremor profile (or to track the negated tremor profile) of 50 μm p-p amplitude with a rmse of about 0.7 μm . With this tracking accuracy, not only that it meets the application objective, the open-loop rate-dependent controller also outperforms other feedback controllers reviewed in Section 7.1.

A well-implemented open-loop controller with all the performance prerequisites is definitely more desirable than a feedback controller, especially from a system implementation standpoint. Beside the apparent cost reduction, the riddance of displacement feedback sensors (e.g. strain gauges) and their signal conditioning circuitry also eliminates many associated implementation issues – gauge mounting and alignment issues, heating problem, complication of mechanical designs, etc. In addition, an open-loop controller does not have instability problem and is almost always more responsive.

References

- [Charles 96] S. Charles, “Dexterity enhancement for surgery,” in *Computer Integrated Surgery: Technology and Clinical Applications*, R. H. Taylor, S. Lavallée, G. C. Burdea, and R. Mösges, Eds. Cambridge, Mass.: MIT Press, pp. 467-471, 1996.
- [Chen 80] P. Chen and S. Montgomery, “A macroscopic theory for the existence of the hysteresis and butterfly loops in ferroelectricity,” *Ferroelectrics*, vol.23, pp.199-207, 1980.
- [Chen 99] B. M. Chen, T. H. Lee, C.C. Hang, Y. Guo, and S. Weerasooriya, “An H_∞ Almost Disturbance Decoupling Robust Controller Design for a Piezoelectric

- Bimorph Actuator with Hysteresis”, *IEEE Tran. Automation and Control*, Vol. 7, No. 2, pp.160-174, Mar 1999.
- [Choi 97] G. S. Choi, H-S. Kim, and G.H. Choi, “A Study on Position Control of Piezoelectric Actuators”, *Proc. IEEE Int. Symp. Indust. Electr.*, pp.851-855, Jul 1997.
- [Chonan 96] S. Chonan, Z. Jiang, and T. Yamamoto, “Nonlinear hysteresis compensation of piezoelectric ceramic actuators”, *J. Intell. Mater. Syst. Struct.*, vol.7, no.2, pp.150-156, 1996.
- [Cruz-Hernandez 01] J.M. Cruz-Hernandez and V. Hayward, “Phase Control Approach to Hysteresis Reduction”, *IEEE Trans. Control & System Tech.*, vol.9, no.1, pp.17-26, Jan 2001.
- [Elble 90] R. J. Elble and W. C. Koller, *Tremor*. Baltimore: Johns Hopkins University Press, 1990, p. 1.
- [Furutani 98] K. Furutani, M. Urushibata, and N. Mohri, “Displacement control of piezoelectric element by feedback of induced charge”, *Nanotechnology*, 9, pp.93-98, 1998.
- [Galinaitis] W. S. Galinaitis, “Two Methods for Modeling Scalar Hysteresis and their use in Controlling Actuators with Hysteresis”, *Ph.D. Thesis*, Virginia Polytechnic Institute and State University, Blacksburg, VA.
- [Ge 96] P. Ge and M. Jouaneh, “Tracking Control of a Piezoceramic Actuator”, *IEEE Trans. Control and System Tech.*, Vol.4, No.3, pp.209-216, May 1996.
- [Goldfarb 97] M. Goldfarb and N. Celanovic, “Modeling Piezoelectric Stack Actuators for Control of Micromanipulation”, *IEEE Control System Magazine*, vol.17, issue 3, pp.69-79, June 1997.
- [Hu 99] H. Hu and R. Ben Mrad, “Preisach Modeling of a Hysteresis in a Piezoceramic Actuator,” *2nd Canadian Smart Structures and Materials Conference*, Sep. 13-14, St. Huber, Quebec, Canada, pp. 31-40, 1999.
- [Huges 95] D. Huges and J. T. Wen, “Preisach Modeling of Piezoceramic and Shape Memory Alloy Hysteresis,” *Proc. 4th IEEE Conf. on Control Applications*, pp. 1086-1091, 1995.
- [Hwang 03] C-L. Hwang and C. Jan, “A Reinforcement Discrete Neuro-Adaptive Control for Unknown Piezoelectric Actuator Systems with Dominant Hysteresis”, *IEEE Trans. Neural Net.*, vol.14, no.1, pp.66-78, Jan 2003.
- [Jan 00] C. Jan and C-L. Hwang, “Robust Control Design for a Piezoelectric Actuator System with Dominant Hysteresis”, *Proc. 26th Annual Conf. of the IEEE IECON 2000*, Vol.3 pp.1515-1520, 2000.
- [Krasnosel’skii 89] M.A. Krasnosel’skii and A.V. Pokrovskii, *Systems with Hysteresis*, New York: Springer-Verlag, 1989.
- [Krejci 01] P. Krejci and K. Kuhnen, “Inverse control of systems with hysteresis and creep”, *IEE Proc. Control Theory Appl.*, vol. 148, no.3, May 2001.
- [Ku 00] S-S. Ku, U. Pinoson, S. Cetinkunt, and S. Nakjima, “Design, Fabrication, and Real-time Neural Network of a Three-Degrees-of-Freedom Nanopositioner”, *IEEE / ASME Trans. Mechatronics*, vol. 5, no. 3, pp.273-280, Sep. 2000.
- [Kuhnen 01] K. Kuhnen and H. Janocha, “Inverse Feedforward Controller for Complex Hysteretic Nonlinearities in Smart-Material Systems,” *Control and Intelligent System*, Vol. 29, No. 3, pp. 74-83, 2001.
- [Kuhnen 02] K. Kuhnen and H. Janocha, “Complex hysteresis modeling of a broad class of hysteretic nonlinearities”, *Proc. of the 8th Int. Conf. on New Actuators*, Bremen, pp. 688-691, Jun 2002.
- [Kuhnen 99] K. Kuhnen and H. Janocha, “Adaptive Inverse Control of Piezoelectric Actuators with Hysteresis Operators”, *Proc. European Control Conf. ECC’99*, Karlsruhe, Germany, 1999.

- [Majima 01] S. Majima, K. Kodama, and T. Hasegawa, "Modeling of Shape Memory Alloy Actuator and Tracking Control System with the Model", *IEEE Trans. Cont. & System Tech.*, vol.9, no.1, pp.54-59, Jan 2001.
- [Mayergoyz 91] I. D. Mayergoyz, *Mathematical Models of Hysteresis*. New York: Springer-Verlag, 1991.
- [Newcomb 82] C. Newcomb and I. Filnn, "Improving linearity of piezoelectric ceramic actuators", *Electronics Letters*, Vol.18, No.11, pp.442-444, May 1982.
- [Riviere 03] C. N. Riviere, W.T. Ang, and P. K. Khosla, "Toward Intelligent Microsurgical Instrument", *IEEE Transactions on Robotics and Automation*, Vol. 19, No. 5, Oct. 2003, pp. 793-800.
- [Riviere 97] C. N. Riviere and P. K. Khosla, "Augmenting the human-machine interface: Improving manual accuracy," *Proc. IEEE Int. Conf. Robot. Autom.*, Albuquerque, N.M. Apr. 20-25, 1997.
- [Smith 01] R. C. Smith, Z. Ounaies, and R. Wieman, "A Model for Rate-dependent Hysteresis in Piezoceramic Materials Operating at Low Frequencies", *Technical Report*, NASA/CR-2001-211-62, NASA Langley Research Center, 2001.
- [Stepanenko 98] Y. Stepanenko and C-Y. Su, "Intelligent Control of Piezoelectric Actuators," *Proc. IEEE Conf. Decision & Control*, Florida, USA, pp. 4234-4239, Dec. 1998.
- [Tan 02] X. Tan and J. S. Baras, "Control of Hysteresis in Smart Actuators, Part I: Modeling, Parameter Identification, and Inverse Control", *Technical Research Report*, CDCSS TR 2002-8, Center for Dynamics and Control of Smart Structures, 2002.
- [Tao 95] G. Tao and P. V. Kokotovic, "Adaptive Control of Plants with Unknown Hystereses", *IEEE Tran. Auto. & Control*, vol.40, no.2, pp.200-212, Feb 1995.

Chapter 8

Micron Instrument

8.1 Introduction

This chapter discusses the integration and implementation issues of the hardware and software subsystems of Micron described in Chapter 2 to 7. Section 8.2 gives the overview and details of the system peripheral hardware and software architecture. Learning from the previous iteration of the intelligent instrument, a brief discussion on the current redesign effort that focuses on improving the grasping ergonomic of the instrument is also presented.

8.2 System Integration

8.2.1 Micron Instrument

A picture of the integrated Micron instrument is shown in Fig. 8.1. The instrument weighs less than 100 g and measures 180 mm in length, excluding the intraocular shaft. The diameter of the handle is $\varnothing 20$ mm with $\varnothing 17$ mm at the narrowest part of the contoured section to aid grasping. The back sensor suite that houses the three-axis magnetometer and two dual-axis accelerometer measures 26 mm \times 26 mm \times 36 mm. The entire handle body and the back sensor suite housing are made of black polycarbonate.

The intraocular shaft is a disposable surgical needle with standard plastic Luer female connector.

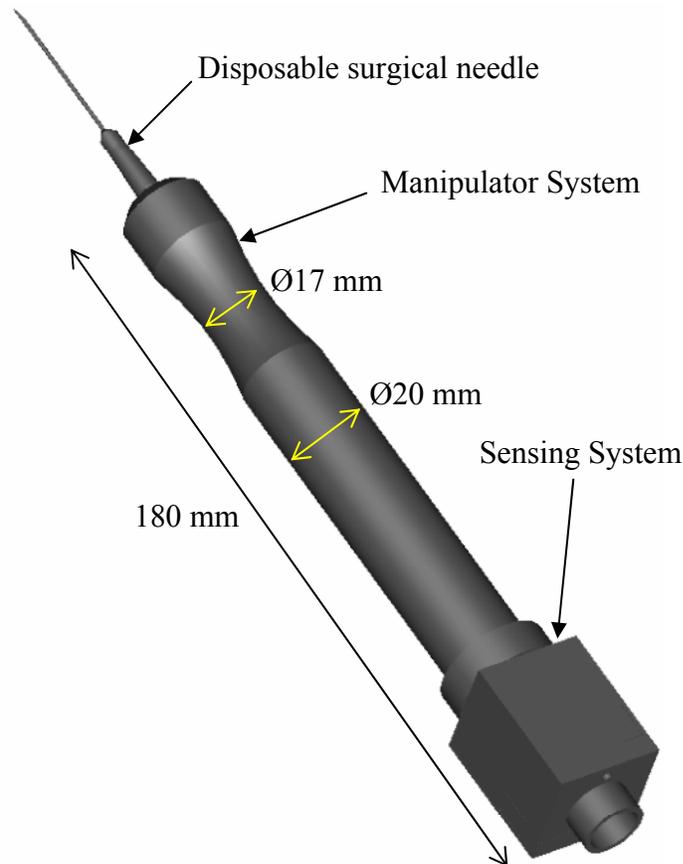


Fig. 8.1 The Micron instrument.

8.2.1.1 Grasping Ergonomic

One key redesign effort of the current Micron prototype is to improve the grasping ergonomic of the instrument. The first Micron prototype [Ang 00; Ang 01], as shown in Fig. 8.2, has the same length as the current prototype, a slightly larger diameter and weights almost twice the current design at 170 g. Surgeons' evaluation of the first prototype indicated that the size and weight of the instrument, despite larger and heavier than typical vitreoretinal microsurgical instruments (< 100 mm long, < Ø15 mm, and < 50 g), are within acceptable limits. However, the concentration of the weight at the back end of the handle creates a turning moment and stresses the finger grasping the handle

over time. An analysis of the weight distribution of a few commercial pens and pencils that feature ‘reduced grasping stress’ design has found that the center of gravity (CG) lies within the circle formed by the grasping index finger and thumb, as illustrated in Fig. 8.2. Therefore, for the current prototype, the manipulator is designed to be close to the front end of the instrument to counter balance the weight of the back sensor suite to bring the CG to lie within the grasping circle. Moreover, reducing the diameter of grasping section and the overall weight would also improve the grasping ergonomic of the instrument.

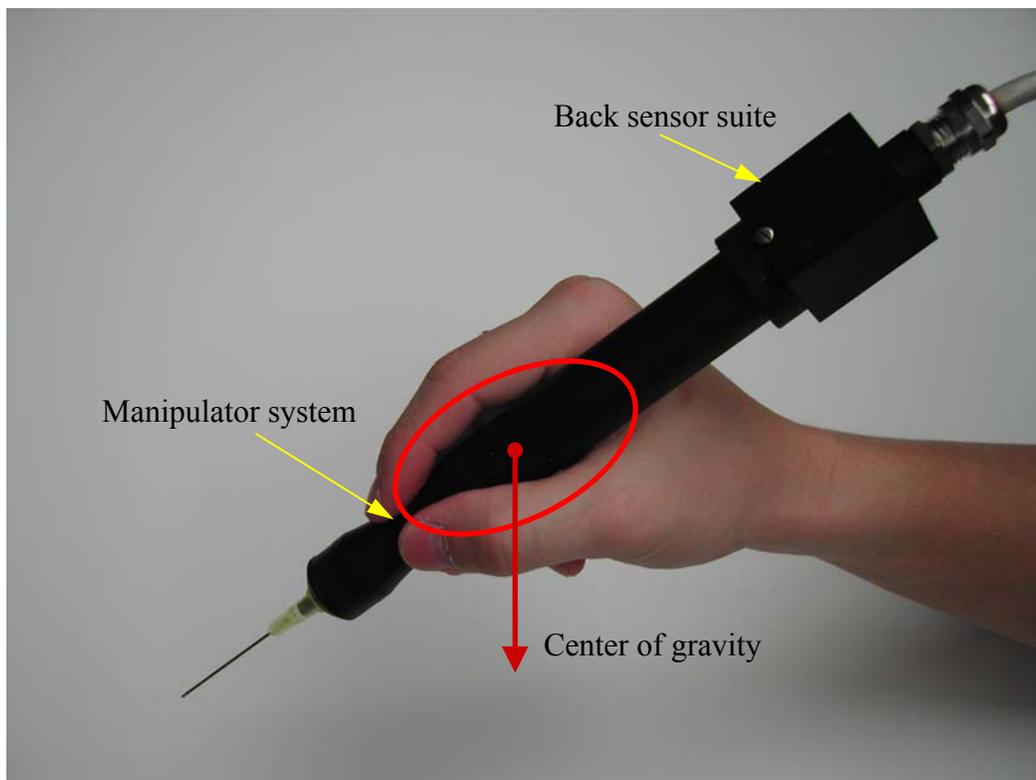


Fig. 8.2 The manipulator system is designed to shift the CG of the instrument to within the circle formed by the grasping index finger and the thumb.

8.2.2 System Hardware and Software

An overview of the complete system is depicted in Fig. 8.3. The on-board magnetometer-aided all-accelerometer IMU described in Chapter 2 senses the motion of

the instrument and feed the voltage output of sensors to the host computer via an analog-digital converter (ADC). The host computer is a Pentium 4 2.4 GHz desktop computer running Windows XP operating system. The ADC is a Microstar (Microstar Laboratory, Inc., WA) DAP 5216a data acquisition board, a PCI bus card with on-board AMD-K6 III+ 400 MHz processor and 32 MB RAM. It has 16 analog input channels and 2 analog output channels, both with 16-bit resolution. Additional 4 analog outputs are provided by an expansion board MSXB 022-01 which takes up another PCI slot of the computer.

The software systems, denoted by italic fonts in Fig. 8.2, are programmed in Visual C++ on the host computer and then compiled into DAPL command lines and upload to the memory of the DAP 5216a. DAPL is a real-time operating system for the DAP 5216a. The sensor fusion algorithm, presented in Chapter 4, estimates the body translational and angular displacement, based on the measurement model proposed in Chapter 3. The tip displacement is computed by the forward kinematics algorithm in Chapter 2, before it is fed into the erroneous motion filtering algorithms described in Chapter 5. It is to be noted that only the weighted-frequency Fourier linear combiner (WFLC) algorithm is implemented in the current system. The WFLC algorithm outputs the negated tremulous component of the tip displacement to the inverse kinematics algorithm given in Chapter 6, which in turn computes the joint variables or required displacement of the piezoelectric actuators. The open-loop feedforward controller with inverse rate-dependent hysteresis model proposed in Chapter 7 then controls the piezoelectric actuators to drive the parallel intraocular shaft manipulator to cancel the erroneous motion.

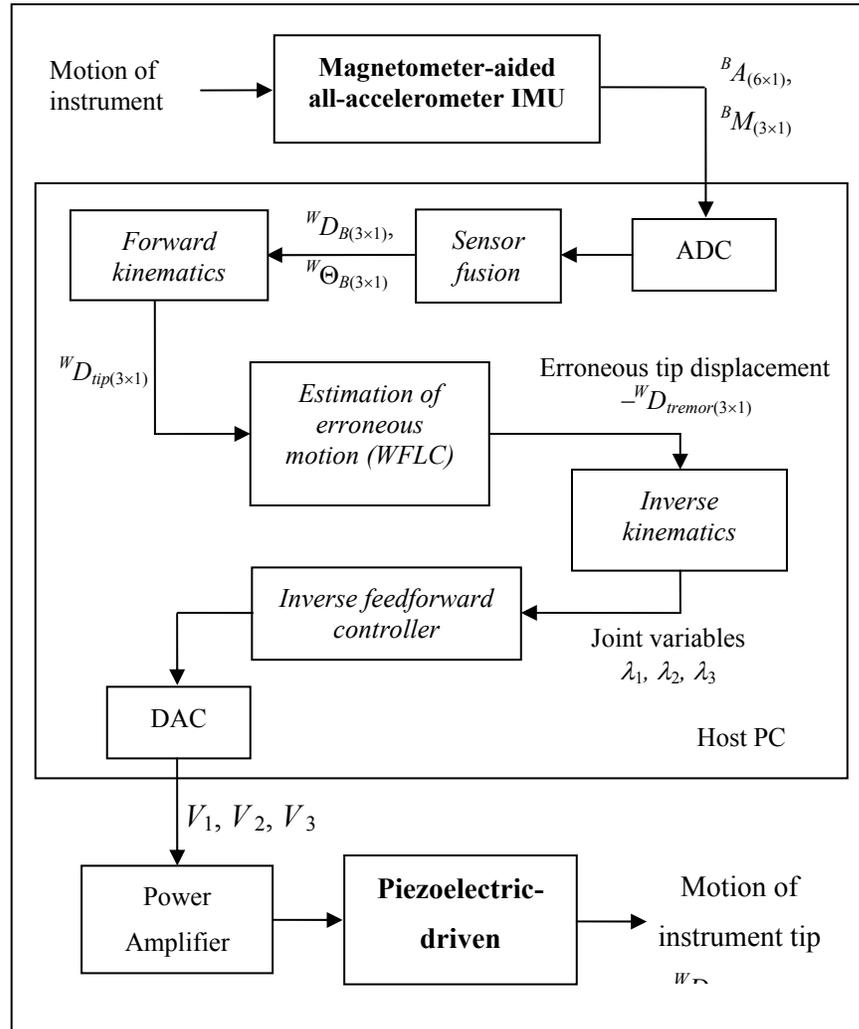


Fig. 8.3 Overview of Micron system. Micron subsystems are denoted by bold font, peripheral hardware and software systems are denoted by italic and regular font respectively.

References

- [Hotraphiyo 01] Hotraphiyo, L. F. and Riviere, C. N., "Precision measurement for microsurgical instrument evaluation", *Proceedings of the 23rd Annu. Conf. IEEE Eng. Med. Biol. Soc.*, Volume: 4, 2001, pp. 3454 -3457.

- [Ang 00] W. T. Ang, C. N. Riviere, and P. K. Khosla, "An Active Hand-held Instrument for Enhanced Microsurgical Accuracy", *Proc of the Third International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, October, 2000, pp. 878-886.
- [Ang 01] W. T. Ang, C. N. Riviere, and P. K. Khosla, "Design and implementation of active error canceling in hand-held microsurgical instrument", *Proc. IEEE/RSJ Intl. Conf. Intelligent Robots and Systems*, Vol. 2, Oct 2001, pp. 1106-1111.

Chapter 9

Conclusion

9.1 Conclusion

Chapter 1 discusses the problem of manual positioning inaccuracy in micromanipulations and the nature and characteristics of physiological tremor and other erroneous sources affecting human performance in delicate tasks such as microsurgery. It then presents the novel idea of performing active error compensation within a completely handheld instrument and compares it to other approaches. The chapter also defines the problem statements of our approach and previews the organization of the thesis with respect to these key technical challenges.

Chapter 2 presents the design of the motion sensing system. The detail design of the new magnetometer-aided all-accelerometer inertial measurement unit is described, and the key design choices are explained. Two algorithms to determine the orientation of the instrument are presented. The first is a differential sensing kinematics algorithm that derives angular motion information from the accelerometers, and the second is the TRIAD algorithm that uses the magnetic North and gravity vectors to estimate the orientation. The chapter ends with an error variance analysis to demonstrate the superiority of the proposed sensing system in the angular sensing resolution versus a conventional IMU equipped with rate gyroscopes.

Based on the sensing system design in Chapter 2, Chapter 3 put forward a comprehensive measurement model that models and compensates sensor errors. A physical model is proposed to account for the deterministic errors of the sensors, these include hysteretic behavior, non-linearity, cross-axis effects, temperature drift, misalignments, and environmental factors. Simple mathematical models suitable for real-time implementation are used, using parameters identified from experimental observations. A stochastic model based on Allan variance analysis is proposed to understand the noise content of the sensors.

Experiments involving static orientation sensing, tremor-like translational motion tracking, and real-time 6 DOF motion tracking of the instrument are described in Chapter 4. Real-time orientation tracking is implemented via an augmented state quaternion-based Kalman filtering. There are two sources of orientation information from the magnetometer-aided all-accelerometer IMU as described in Chapter 2. The angular displacement information from the differential kinematics algorithm has very high resolution but suffers from integration drift, on the other hand, the TRIAD algorithm gives noisy but non-drifting orientation estimates. The proposed Kalman filter fuses these two complementary sources to yield a non-drifting reduced noise orientation estimate.

Chapter 5 describes two algorithms for erroneous motion estimation. The weighted-frequency Fourier linear combiner (WFLC) is used to model and filter tremulous motion. The WFLC is a notch filter that uses a dynamic sinusoidal model to adaptively model the phase, amplitude, and frequency of incoming signal. The WFLC algorithm is an adaptation of Riviere's dissertation work, it is not the author's original work. In the absence of a more profound understanding of the structure of non-tremulous errors, a neural network technique is proposed to perform black-box modeling and compensation. Only WFLC is implemented with the Micron system currently.

A three DOF parallel mechanism is proposed to manipulate the intraocular shaft to cancel the erroneous motion in Chapter 6. Piezoelectric actuators are chosen to drive the mechanism because of their high actuation bandwidth. The inverse and forward kinematics and workspace analysis of the mechanism are also presented.

All favorable characteristics of the piezoelectric actuator notwithstanding, its actuation accuracy is hampered by the highly non-linear hysteresis. Chapter 7 proposes an open-loop feedforward controller with inverse hysteresis model. The chapter extends the Prandtl-Ishlinskii hysteresis model to incorporate the ability to model the rate-dependence behavior of piezoelectric actuators. Experimental motion tracking results are also shown.

Integrating all the subsystems of the Micron instrument, Chapter 8 describes the hardware and software components of the complete system in details.

9.2 Contributions

The major contributions of this dissertation work are listed as follow:

- 1) Design and implementation of a novel all-accelerometer inertial measurement unit that uses differential kinematics to obtain high resolution angular motion information.
- 2) Proposal and implementation of a physical model to account for the deterministic errors of the accelerometer.
- 3) Proposal and implementation of an augmented state quaternion-based Kalman filtering for real-time fusion of orientation sensing information.
- 4) Design and implementation of the 3 DOF parallel intraocular shaft manipulator.
- 5) Proposal and implementation of an open-loop feedforward controller based on the rate-dependent Prandtl-Ishlinskii hysteresis model for accurate tracking control using piezoelectric actuators.

9.3 Future Work

The integrated Micron system has not reached its full design and theoretical performance potential as demonstrated by separate testing of both its software and hardware subsystems. More engineering effort will be required to tweak the system parameters to make the subsystems to work more seamlessly together.

User acceptance evaluation of the instrument by eye surgeons has identified a few future design changes to make Micron more coherent with the operating room setup and surgical procedures. One of these is the overall length of the Micron instrument, which has to be reduced to less than 100 mm including the intraocular shaft. This is because the surgical microscopes typically used in eye microsurgery have lenses with 175 – 200 mm focal length, which means the clearance between the patient’s eye and the microscope is equal to the focal length. Another design deficiency highlighted is the umbilical cable at the backend of Micron is restricting the hand motion of the surgeon. Other potential improvements include redesigning the tool attachment interface to incorporate the actuation mechanism for more sophisticated surgical tools such as a forceps.

Although this dissertation work focuses on microsurgery, the concept of performing active error compensation within a completely handheld instrument has numerous applications in micromanipulation tasks that require accuracy enhancement. Micromanipulation tasks are those that involve small signal-to-noise (voluntary-to-erroneous motion) ratio, for example, cell manipulation in the biotech industry, gun-sights or handheld military tracking equipment, and handheld video photography etc.

The inertial sensing and piezoelectric control technology developed in this dissertation also have applications beyond active error compensation. The magnetometer-aided all-accelerometer IMU can serve as an alternative to the conventional three gyros three accelerometers IMU, especially in low-g motion sensing applications. The open-loop inverse feedforward controller with rate-dependent PI hysteresis model outperforms all currently commercially available piezoelectric driven micropositioner in tracking control applications. Potential industrial applications include high-speed high-precision manufacturing and driving the read/write heads in computer hard drives.