DISS. ETH NO. 17953

VISUAL TRACKING FOR NANOROBOTIC MANIPULATION AND 3D RECONSTRUCTION IN AN ELECTRON MICROSCOPE

A dissertation submitted to

ETH ZURICH

for the degree of

Doctor of Sciences

presented by

BRADLEY E. KRATOCHVIL

M.S. University of Minnesota
B.S. University of Nebraska

born on June 30, 1978

citizen of
United States of America
Republic of Italy

accepted on the recommendation of

Prof. Bradley J. Nelson, examiner
Dr. Lixin Dong, co-examiner
Prof. Sergej Fatikow, co-examiner

2008

To my father who taught me to work,
To my mother who taught me to read,
To my brothers who taught me to play,
To my wife who taught me to love

Preface

Graduate studies progress in different ways for different people. In my case the road wandered a bit, but I was fortunate to share the journey with a large number of capable and engaging people. Without them I would not be where I am today. First and foremost, I would like to thank my adviser Brad Nelson for allowing me to make the transition not only from one academic field to another, but from one continent to another. I would additionally like to thank my co-adviser Lixin Dong for his time and effort with my projects. His input and guidance helped me avoid getting lost in the world of nanotechnology. I would also like to make special note of two colleagues, Dominic Frutiger and Karl Vollmers, without whom I would not have been able to maintain my sanity through this process.

I would like to thank Jake Abbott for helping me unravel twists and Brigitte Geissmann for helping to keep me from unravelling when dealing with ETH paperwork. Ruedi Borer was instrumental in a number of projects and helped me build robotic systems which operated over eight orders of magnitude. He deserves as many orders of magnitude of appreciation. I would like to also not only thank the members of the Institute of Robotics and Intelligent Systems, but the students I worked with over the years. Through interactions with them all I was able to develop myself both professionally and personally.

Directly regarding my thesis, a special thanks goes to Gian-Marco Baschera for his help with the SEM imaging calibration. Without Felix Beyeler, Simon Muntwyler, and FemtoTools I would have had nothing to track. I would like to thank Li Zhang for providing the nanohelices and to Zeiss, especially Andreas Schertel, for time on the FIB and the data sets we gathered there. I would also like to express my appreciation to both Christos Bergeles and Daniela Schuler for help along the way to the 3D reconstruction system.

Last, but certainly not least, I would like to thank my wife Astrid for all her support. Without her help and faith, I would never have been able to accomplish so much in these past years.

Abstract

Robotic micromanipulation is changing the way microelectromechanical systems (MEMS) are designed and fabricated as well as how cellular and molecular biological structures are handled. Recent efforts in manipulating nanometer sized structures such as carbon nanotubes and nanocoils have shown great promise in fabricating nanoelectromechanical systems (NEMS). Although many tasks at this scale involve robotic manipulators, the control of these devices has been difficult to automate in any meaningful way. In order to enable researchers to perform more kinematically complex manipulations and to help free them from the burden of manually performing every task better manipulation strategies are needed.

Manipulations are often performed inside of scanning electron microscopes (SEM) or transmission electron microscope (TEM) which typically provide sensory feedback in the form of an monocular image. We will investigate the imaging properties of an SEM to better describe the tool's capabilities and limitations regarding visual servoing applications. To overcome one of the SEM's principle limitations, feedback rate, we propose using a rigid-body tracking system for guiding manipulation tasks. This system allows the operator to easily incorporate domain-specific knowledge into the tracking task and to reduce the scan area of the system to increase the imaging frame rate.

A variety of different methods exist for gathering three-dimensional information for micro- and nanoscale objects. Tilting of samples in a scanning electron microscope provides a nondestructive way of generating this data. Traditionally the reconstruction of this image data is performed by stereo photogrammetric methods which compare features from two or three frames. We propose the application of techniques from the structure-from-motion community as being efficient, high-precision alternatives to stereo methods which allows for automated utilization of a large number of sampled images. We propose the use of nanohelices to generate localized rotational motions. Using this method alleviates the demand of high-precision actuators, allows 360° rotations, and provides a useful tool for micro- and nanomanipulation.

Zusammenfassung

Neueste Anstrengungen in der Manipulation von Strukturen im Nanometerbereich, wie Kohlenstoff Nanoröhren und Nanospiralen, haben vielversprechende Möglichkeiten für die Fabrikation von Nanoelektromechanischen Systemen (NEMS) gezeigt. Während die Aufgaben in diesem Grössenbereich robotische Manipulatoren erfordern, stellt eine sinnvolle Automation dieser Werkzeug eine grosse Herausforderung dar. Um Forschern die Ausführung von kinematisch komplexen Manipulationen zu ermöglichen und sie von der Last vieler einzelner manueller Arbeitsschritte zu befreien werden bessere Strategien benötigt.

Manipulationen werden in der Regel in einem Rasterelektronenmikroskop (REM) durchgeführt, wobei ein monokulares Bild die einzige sensorische Rückmeldung des Systems darstellt. Wir werden die bildgebenden Eigenschaften eines REM untersuchen um die Möglichkeiten und Grenzen in Bezug auf Anwendungen einer optischen Steuerung besser beschreiben zu können. Um den grössten Schwachpunkt eines REM, die tiefe Bildrate, zu überbrücken, schlagen wir ein System zur Verfolgung von Festkörpern als Steuerung für Manipulationsaufgaben vor. Dieses System erlaubt es dem Bedienenden in einfacher Weise aufgabenspezifisches Vorwissen in die Objektverfolgung einfliessen zu lassen und dank einer Verkleinerung des Abtastbereichs die Bildrate zu erhöhen.

Es existieren verschiedene Methoden um dreidimensionale Informationen von Objekten im Mikro- und Nanobereich zu sammeln. Das Neigen von Proben im Rasterelektronenmikroskop ist eine Möglichkeit zur nichtdesktruktiven Generierung solcher Daten. Wir empfehlen die Anwendung von Techniken der 'Struktur-aus-Bewegung'-Gemeinde als effiziente und hochpräzise Alternativen zu Stereomethoden, was die automatisierte Verarbeitung grosser Bilddatensätze erlaubt. Weiter schlagen wir den Einsatz von Nanospiralen zur Generierung von örtlich begrenzten Rotationsbewegungen vor. Die Verwendung dieser Methode mildert die Notwendigkeit von hochpräzisen Aktuatoren, erlaubt 360° Rotationen und stellt ein nützliches Werkzeug für Mikro- und Nanomanipulationen dar.

Contents

List of Tables

List of Figures

1 Introduction

Robotic micromanipulation is changing the way microelectromechanical systems (MEMS) are designed and fabricated, as well as how cellular and molecular biological structures are handled. Recent efforts in manipulating nanometer- sized structures such as carbon nanotubes and nanocoils have shown great promise in characterizing nanoelectromechanical systems (NEMS). In order to enable researchers to perform more kinematically complex manipulations, and to help free them from the burden of manually performing every task, better manipulation tools are needed.

Nanomanipulation began in the early 1980's with the use of Scanning Tunneling Microscopes to manipulate individual atoms. Since then, new actuator and sensing technologies have enabled nanomanipulation at a variety of different scales [Weir *et al.*, 2005]. One commonality of most nanomanipulations to date is that they are performed using open-loop manipulators where a user must interpret the various forms of feedback to accomplish the desired task. This requires considerable amounts of time from experienced and well-trained users. These limitations considerably reduce the speed at which researchers can explore new phenomena and technologies at this scale. Thus far, most research in nanorobotics has focused on experimentation with little regard for developing better methods for the experimenter. Although the field of robotics is devoted to the intelligent control of machines, much of this work is yet to be utilized in nanorobotic systems. This thesis explores the use of robotic technologies to bridge the gap between macro- and nanoscale robotics and to help close the loop between the feedback system and the manipulator.

Manipulations are often performed in conjunction with devices such as optical microscopes, scanning electron microscopes (SEM), or transmission electron microscopes (TEM), which all provide sensory feedback in the form of a monocular image. The SEM is often favored for nanomanipulation due to its high resolution and high depth of field. Thus far, most nanomanipulations have focused on manual strategies for characterizing nanostructures [Weir *et al.*, 2005], but recent work has explored the use of template matching [Siev-

ers and Fatikow, 2005] and active contours [Fatikow *et al.*, 2006] for semi-automated visual servoing tasks.

One major challenge of performing visual servoing tasks inside an SEM is balancing the needs of image quality with real-time imaging. The sequential scanning used to create an SEM image necessitates lower frame rates than those available with optical cameras. SEM imaging quality is also highly dependent on the scale and material properties of the area being viewed. Often these parameters cannot be changed because they are directly related to the task being performed. Thus, methodologies must be developed to work with the limited frame rates and imaging restrictions of the SEM.

We will first examine the state-of-the-art tools for observation and manipulation in the nano domain. Of the observation strategies available to date, the Scanning Electron Microscope (SEM) provides the best compromise of speed, flexibility, and resolution for robotic manipulation tasks. Since the use of electron microscopes for computer vision tasks is still a relatively new application [Ribeiro and Shah, 2006], we will explore the imaging system of the scanning electron microscope in terms of projection models, image calibration, and imaging optimization. Central to the use of the SEM for real-time imaging is the trade-off between image quality and frame rate. To make more effective use of the available information, we propose using a rigid-body tracking scheme to provide additional information for the tracking task at hand. The addition can take the form of object geometry as well as constraints on how the object can move in its environment. To alleviate the frame rate challenges due to the serial sensing nature of the SEM, a selective region-of-interest method is proposed to enable selective scanning of the pertinent areas of the image during a manipulation task. This system automatically uses knowledge of the geometry of the tracked object to select the most efficient scanning regions given the frame rate constraints of the task at hand.

A variety of different methods exist for gathering three-dimensional information for micro- and nanoscale objects. Tilting of samples, in a scanning or transmission electron microscope, provides a nondestructive way of generating this data, but requires a sample to be eucentrically tilted. Scaling of rota-

tional and translational motions occur along significantly different paradigms. To scale a translational motion, i.e.from macroscale to microscale, the primary concern is in the precision of the actuator, while the exact orientation of the motion axis is of less consequence. Alternatively, rotational motions require consideration of the pose of the axis to be taken into account if the rotation is to occur at a desired location. This phenomenon has led to the limited use of rotational motions during micro- and nanomanipulation tasks. When it does occur, such as in a eucentric sample stage, it requires the use of high precision actuators as well as frequent calibration. One method for alleviating this problem is to align the rotation axes of the manipulator at a coincident point [Probst *et al.*, 2007]. Alternatively, the entire actuator can be scaled to a size commensurate with the task at hand. One possible solution to this problem is the use of a helical nanocoil to generate the rotational motion. We demonstrate the applicability of helical nanocoils to generate rotations and demonstrate how these motions can be tracked and utilized for gathering information from the sample. Using this method alleviates the demand for high precision actuators, allows full 360° rotations, and provides a useful tool for micro- and nanomanipulation.

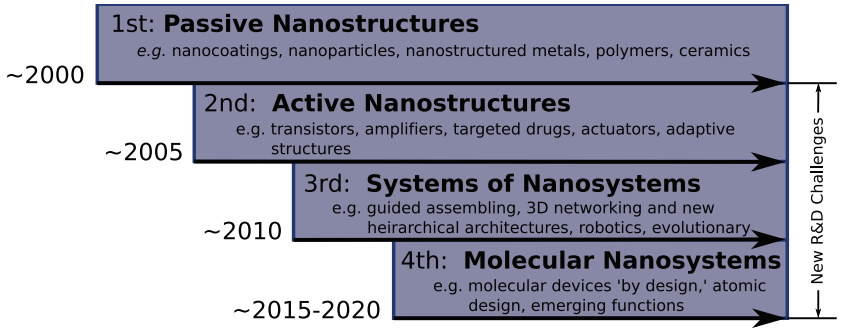


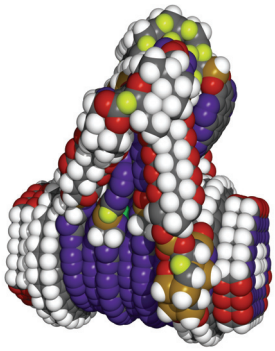
Figure 2.1: four generations of products: timeline for beginning of industrial prototyping and nanotechnology commercialization [Roco, 2005].

2 Related Work

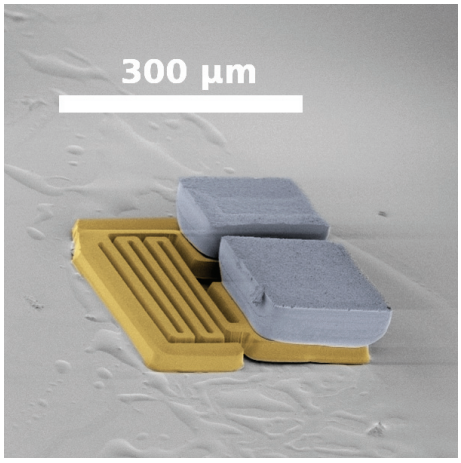
2.1 Nanotechnology

The first hint of Nanotechnology came with Feynman's famous speech entitled "There's plenty of room at the bottom" [1959]. In his talk, Feynman mentioned how the field of miniaturization was barely in its infancy and great progress would be made towards the world of the very, very small.

Nanotechnology consists of technological developments on the nanometer scale. In this domain, researchers work at atomic and molecular levels to develop new technologies. Technologies at this scale typically involve working with particles on the order of 0.01 to 100 nm. This diverse field which encompasses chemistry, physics, molecular biology, material science, computer science, and engineering is garnering an increasing amount of funding throughout the world. Many attribute the genesis of significant research in nanotechnology to the discovery of fullerines, or "Buckyballs", in 1985 by Nobel Laureate Richard Smalley [1997]. Since that time, research in the field has been expanding at an increasing rate. From 1997 to 2005, government funding for nanotechnology research areas has increased ninefold. After over 20 years of research, nanotechnology is rapidly moving from its infancy, where much of the work focused on building fundamental knowledge and infrastructure, to an application domain [Roco, 2005]. Estimates by the National Nanotechnology Initiative (NNI) show by the year 2015 world-



(a) a molecular machine as proposed by [Drexler, 1992]



(b) a microfabricated mobile microrobot [Vollmers *et al.*, 2008]

Figure 2.2: Bottom-up strategies, shown in (*a*), build from the molecular level, while top-down strategies, shown in (*b*), work to make existing microfabrication strategies smaller.

wide annual industrial production in nanotechnology is expected to exceed $1 trillion and employ around 2 million workers [Roco, 2004]. This output will encompass products such as electronic, magnetic, optoelectronic, biomedical, pharmaceutical, cosmetic, energy, catalytic, and materials applications [Weir *et al.*, 2005].

A variety of different strategies exist for developing these technologies (Figure 2.2). Dong [2003] delineates two major approaches for research in building micro- and nanoscale devices:

- *Bottom-up strategies* encompass approaches such as those suggested by Drexler [1986; 1992] where machines would be engineered at a molecular level. These technologies focus on building mechanisms by precisely controlling the organization of the atomic structure of a device. Much of the motivation for such devices comes from observation of similar structures in nature.

- *Top-down strategies* involve scaling or modifying existing technologies used in microfabrication. Technologies such as photolithography, thin

film processes, and etching currently enable the construction of devices with minimum feature sizes on the order of 1 µm, and research is actively targeting methods for reducing these to scales suitable for nanotechnology.

With efforts approaching from both ends, the meeting point between micro- and nanoscale technologies can become blurred. In the future, many of the new products in this domain will be a integration of both micro- and nanotechnologies [Hsu, 2004].

2.2 Nanoscale Sensing

The resolving power of a lensed system is physically limited by diffraction. An empirical value for this limit is given by the Rayleigh criteria. Typical optical microscopes are limited to approximately 200 to 300 nm resolution using tradition methods [Goldstein *et al.*, 2003]. Although newer methods in optical microscopy such as Near-Field Scanning Optical Microscopy (NSOM) use subwavelength light [Dürig *et al.*, 1986] to enable resolutions in the 20 nm range, alternative methods must be used for resolving nanoscale objects.

Nanoscale imaging techniques (Table 2.1) can be broadly divided into two methodologies based on their imaging modes. Scanning Probe Microscopy (SPM) consists of methods such as the Scanning Tunneling Microscope (STM) and the Atomic Force Microscope (AFM) where a physical probe is used for gathering the imaging data. Electron microscopy encompasses both the Transmission Electron Microscope (TEM) and Scanning Electron Microscope (SEM).

2.2.1 Scanning Tunneling Microscopy

With the invention of the Scanning Tunneling Microscope in 1982 by a group of researchers led by Binning and Rohrer at the IBM lab in Switzerland, researchers were able to observe phenomena with atomic precision [Binning *et al.*, 1982]. The device works by rastering an atomic scale probe tip across the surface in question. As the scanning takes place, electrons flow across

properties	items	AFM	STM	TEM	SEM
manipulation	interaction mode	contact, noncontact, intermittent-contact	noncontact	noncontact	noncontact
	sample holder	small	small	small	large
imaging	principle	interatomic force	tunnelling current	electron transmission	electron emission
	resolution	>0.1 nm	>0.01 nm	>0.05 nm	>1 nm
	materials	all	conductor or semiconductor	thin samples	conductor semiconductor
	dimension	3D	3D	2D	2D
real-time sensing		force/image	image	image	image

Table 2.1: methods for nanoscale imaging, adapted from [Sitti, 2001; Du *et al.*, 2006]

the gap between the probe and sample due to a small electrical difference. The tunneling current is then measured and kept constant by using the value as feedback in a closed-loop system which adjusts the probe height. These height values can then be recorded as the topography of the surface. Alternatively, the probe can be left at a constant height and the current value used to infer the height of the probe. In order for electron tunneling to occur, the probe tip must be held within a few nanometers of the surface and be electrically conductive [Cohen and Lightbody, 2004].

2.2.2 Atomic Force Microscopy

The development of the Atomic Force Microscope came after the development of the SPM and was developed in part to remove the conductivity requirement of the STM [Binnig *et al.*, 1986]. The AFM works on a similar principle as the STM in that it uses an atomically sharp probe to raster the surface of the sample. As opposed to using the electron tunneling current for distance measurement though, the AFM relies on the interatomic forces between the sample and the probe tip which is mounted on a cantilever beam. When the probe is close to the sample, these forces cause the cantilever to deflect. These deflections are measured by reflecting a laser off the cantilever and onto a phototedetector. Using Hooke's law, the deflections can

be related to the topography of the surface if the stiffness of the cantilever is known. Atomic Force Microscopes are capable of better than 1 nm resolution in lateral displacements and 0.01 nm in height measurements [Cohen and Lightbody, 2004].

2.2.3 Transmission Electron Microscopy

The first usable Transmission Electron Microscope was built by Prebus and Hillier in 1938 [Reimer, 1993]. The TEM accelerates a beam of electrons which then pass through the sample before being detected. This necessitates a high acceleration voltage, 50-500kV, as well as thin samples ($< 1\,\mu m$). The denser portions of the material allow fewer electrons to hit the detecting screen, which provides the contrast in the resultant image. The TEM has resolutions on the order of 0.1 nm. The high precision of this device requires an ultra-high-vacuum to prevent the electron scattering as well as avoiding contamination on the sample.

2.2.4 Scanning Electron Microscopy

The Scanning Electron Microscope (SEM) similarly relies on a beam of electrons for imaging. The SEM works by rastering a beam of electrons across the specimen and sampling the electrons that are emitted through elastic and inelastic scattering. The imaging method of a SEM allows for lower acceleration voltages (1 to 30 kV) and lower vacuum than the TEM. Current SEMs are capable of imaging resolutions in the order of 1 to 10 nm and imaging both conductive and non-conductive samples. A variety of different system configurations currently exist, Figure 2.3, that range from pure imaging systems to devices with an integrated Focused Ion Beam (FIB), which allows for selective milling or deposition of a sample. Smaller desktop units have also recently become available with magnifications on the range of 20 to 20,000×, providing a lower-cost alternative to the high performance systems. The SEM will be discussed in detail in section 3.

(a) Zeiss NVision 40 SEM with integrated FIB (b) FEI Phenom desktop SEM

Figure 2.3: examples of current SEM systems

2.2.5 Comparison

Scanning Probe Microscopy methods have the capability of measuring extremely small displacements in the topography of a surface. Although a high-vacuum is often employed to reduce contaminants, these measurements can take place in environmental conditions which greatly reduces the complexity of sample preparation and loading as well as allows for measurement on biological entities. Images gathered from SPMs directly contain the depth information of the sample, which is not available from other imaging systems. The key disadvantage of SPM systems for manipulation tasks is that the imaging and manipulating systems are shared. Thus, one can observe or manipulate, but not both. Due to the scanning nature of these systems, high resolution imaging and manipulating take considerable amounts of time to allow for the physical motion of the system.

Electron microscopes require a high or ultra-high vacuum environment which limits the samples to be imaged. Electron microscopy tools have the advantage for manipulations in that their sensors and actuators are separable. When performing manipulations, the vacuum chamber size becomes

one of the significant limiting factors when designing robots to be used. The comparatively large chamber size of the SEM lends itself to the insertion of nanorobotic manipulators. Although the scanning mechanism for the SEM is significantly faster than that of an AFM or SPM, it still requires more time than parallel processes such as optical microscopes.

2.3 Nanorobotics

As nanotechnology research shifts to designing more complex systems, robotic manipulatiors will play a key role in enabling scientists and engineers to explore and interact with the nano domain. Similar to nanotechnology, researchers from a multitude of different academic backgrounds take interest in nanorobotics. Topics such as self-assembly, nanorobotic-assembly, hybrid assembly are studied with the aim of building more complex nanoelectromechanical systems (NEMS). Generally speaking, the term nanorobotics refers to the study of robotics at the nanometer scale. Specifically, nanorobotics has evolved into the study of two main focus areas [Requicha, 1999; 2003; Dong and Nelson, 2007].

In one branch, scientists work to develop robots, sometimes referred to as *nanites*, with principle dimensions in the 1 nm to 1 µm range. Due to the difficulty in fabricating devices at this scale, much of the work in this area remains theoretical. For inspiration, researchers in this field often look to nature for understanding the mechanisms needed to develop the necessary technologies. This has led much of the work in building nanorobots to have close ties with biological and medical applications [Nelson, 2004; Freitas, 2005; Patel *et al.*, 2006]. Recent advances in the field of microrobotics show progress toward their nanorobotic counterparts. To avoid difficulties in storing energy, many of these devices harvest it from the environment in a variety of forms [Abbott *et al.*, 2007]. In particular, Donald *et al.* have demonstrated some of the first untethered mobile microrobots [2006]. These devices are based upon an electrostatic scratch drive and can move on specially prepared substrates. Systems also exist that harness other forms of energy such as heat [Sul *et al.*, 2006] and magnetics [Vollmers *et al.*, 2008;

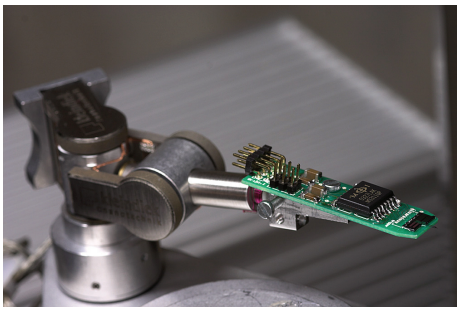
Frutiger *et al.*, 2008]. All of the previously noted devices have principle dimensions on the order of 10 to 100 µm.

The remainder of this thesis will be devoted to the discussion of the second branch of nanorobotics, which focuses on the use of larger robotic manipulators capable of handling nanoscale objects with nanometer precision. These nanorobotic manipulators are used to characterize phenomenon at this scale as well as investigate novel assembly techniques to aid in prototyping devices.

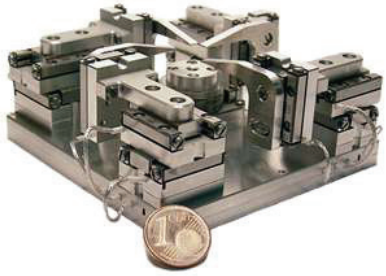
2.3.1 Micro- and Nanoscale Interactions

Compensating for the mechanics of micromanipulation is a principle challenge of robotic manipulation at the micro- and nano scales. When scaling from the macro world to the micro world, volume effects associated with topics such as weight, inertia, heat capacity, and body forces give way to surface forces associated with friction, heat transfer, and adhesive forces [Abbott *et al.*, 2007]. The dominant forces in the micro world are due to van der Waals forces, electrostatic forces, and surface tension as opposed to gravity and inertia. Often, these forces are significantly affected by environmental factors such as temperature or humidity [Zhou *et al.*, 2002]. In addition to the microscale forces, nanorobotic manipulations must deal with intermolecular and intramolecular physical interactions [Dong and Nelson, 2007]. These forces cause micro- and nanoparticles to interact with other objects in the environment in ways that appear counterintuitive when thinking in relation to macro-scale physics. Due to the change in importance from volume to surface effects, parts may appear to "defy gravity" as they are electrostatically attracted to an end effector. Phenomenon like this are the source of significant challenges when manipulating objects at these scales.

Although understanding of these forces is improving, the large number of factors influencing them often makes utilizing this knowledge difficult in terms of manipulation [Fearing, 1995; Böhringer *et al.*, 1999; Wautelet, 2001; Wautelet and Duvivier, 2007]. Often, systems are designed to minimize the influence of particular forces [Arai *et al.*, 1995]. As opposed to minimizing the effect microscale forces, strategies have been proposed that capitalize on



(a) Kleindiek MM3A micromanipulator with FemtoTools microgripper



(b) SmarAct 13D manipulator

Figure 2.4: commercially available micromanipulators

electrostatic force [Cohn, 1992] and surface tension [Bowden *et al.*, 1997; Hosokawa *et al.*, 1996] to construct self-assembling devices. Another approach to negating the effects of microscale forces is to ensure that they are dominated by friction forces and/or geometrical constraints [Popa and Stephanou, 2004].

2.3.2 Nanorobotic Manipulators

A variety of different approaches exist for nanomanipulations as shown in Figure 2.5. Some of the first work at manipulating nanoscale particles was through the efforts of [Eigler and Schweizer, 1990]. In this seminal work, they arranged 35 xenon atoms to form the IBM logo. Since then, other researchers have extended SPM manipulations to the AFM [Junno *et al.*, 1995; Hansen *et al.*, 1998]. One significant challenge of using an atomic force microscope for nanomanipulation is the fact that the sensor and end effector are combined. A first step towards overcoming this limitation was the combination of the AFM with a haptics device [Sitti, 1999; Guthold *et al.*, 2000]. In these systems, a user is conveyed real-time force information from the AFM. Virtual reality interfaces are used to display static virtual environments and the dynamic position of the tip. Although simulated results can be used for dynamic feedback [Vogl *et al.*, 2004], actual environmental feedback can only be acquired by stopping the manip-

ulation. Thermal drift, creep, hysteresis, and other nonlinearities in the actuation system cause spatial uncertainties during the manipulation [Requicha, 1999; Sitti, 2004]. These factors can cause significant problems during manipulation. Later systems sought to alleviate this problem through performing the SPM manipulations inside a SEM [Kikukawa *et al.*, 1993; Ermakov and Garfunkel, 1994; Fahlbusch *et al.*, 2002; Fukushima *et al.*, 2002; Joachimsthaler *et al.*, 2003].

Aside from using SPMs for manipulations, a number of unique nanomanipulation systems have been proposed in the literature. These systems often use piezoelectric actuators when performing nanomanipulations, which allow for precise positioning by changing the voltage potential across the piezoelectric element. Devices are typically assembled in a serial fashion to develop a multi-degree-of-freedom positioner. To enable larger workspaces for the manipulators, the actuators are often driven in a "stick-slip" mode [Zesch, 1997; Breguet, 1998; Bergander, 2003]. Using this method, actuators can have a range on the order of centimeters, but a resolution on the order of nanometers.

The pioneering work for nanomanipulation in an electron microscope was performed at the University of Tokyo [Hatamura and Morishita, 1990; Sato *et al.*, 1995] as early as 1990. This work involved the use of two separate nanorobots for pick and place particles smaller than 100 µm. Since then, other serial manipulation systems [Yu *et al.*, 1999; Fukuda *et al.*, 2003] have been built for the manipulation and characterization of carbon nanotubes.

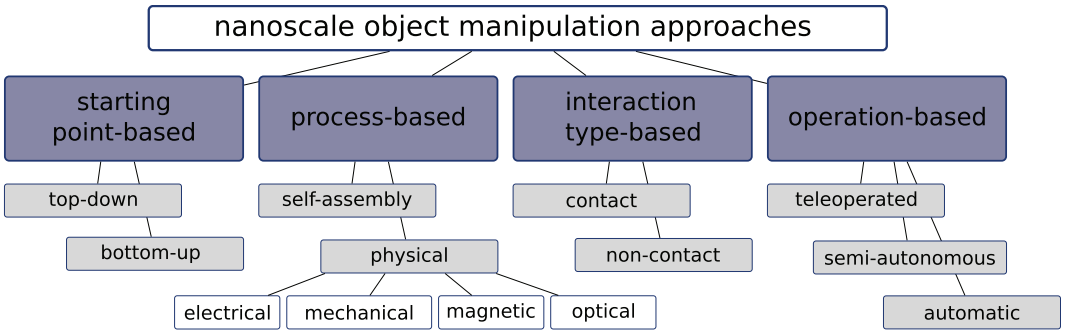
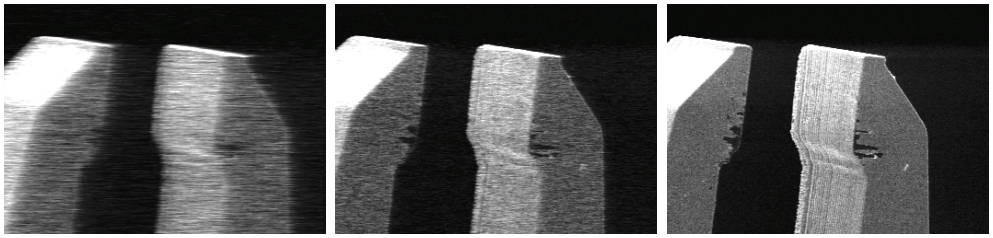


Figure 2.5: taxonomy of nanomanipulation methods [Sitti, 2001].

Serial robotic nanomanipulators typically have relatively large workspaces and the capability of rotational degrees-of-freedom. A number of mobile manipulation systems have also been proposed by [Fahlbusch *et al.*, 1999; Kortschack *et al.*, 2003]. The MINIMAN system is capable of moving throughout its environment with coarse motions as well as performing delicate manipulations with nanometer end effector resolutions. Recently, researchers have investigated the use of parallel manipulators for nanomanipulation [Dagalakis and Amatucci, 2001; Dwarakanath *et al.*, 2000; Moon and Kota, 2002; Culpepper and Anderson, 2004; Zhu and Cui, 2007]. The physical size of these manipulators can be relatively large, and some have looked at expanding the vacuum chamber to accommodate the manipulators [Weck *et al.*, 1997]. This can be prohibitively expensive in terms of money and time, which has prompted researchers to investigate using MEMS technologies to build entire manipulation systems "on-chip" [Gorman *et al.*, 2006].

In all of these systems, automation is the exception rather than the rule. Due to difficulties in building encoders with the required precision and the influence of environmental factors, the serial manipulators are driven in joint-space with the operator required to perform the inverse kinematics. While many of these manipulations are performed with real-time feedback (i.e. SEM imaging), only [Kasaya *et al.*, 1999; Sievers and Fatikow, 2005; Sievers *et al.*, 2006] have shown any significant degree of automation.



(a) 16 Hz (b) 4 Hz (c) 0.5 Hz

Figure 3.1: SEM images of a microgripper at different frame-rates.

3 Imaging System

The scanning electron microscope was proposed in 1938 by Manfred von Ardenne [Goldstein *et al.*, 2003]. Since its inception, it has enabled researchers to gain insights into realms not observable by light microscopy. In more recent times, it has garnered the interest of the robotics community as they investigate new ways to build micro- and nanorobotic devices. Although the system has been used for a number of years for computer applications, it has yet to be thoroughly described from a computer vision perspective. The SEM is subject to numerous sources of noise, which can influence computer vision tasks. Ribeiro and Shah [Ribeiro and Shah, 2006] summarize the following items as the principle challenges of dealing with electron microscope images:

- *Complex image formation geometry.* The images acquired are often gathered in a weak perspective or orthographic setting which makes gathering information along the optical axis difficult.

- *Extremely low signal-to-noise (SNR) ratio.* Analyses are often performed near the imaging limits of the machine. This typically necessitates reducing the image contrast, reducing the image resolution, and increasing the sampling time. The high levels of noise in the images are significantly above those used in optical systems.

- *Limited number of images per observation.* The destructive nature of the imaging process and potentially long imaging times can prevent the acquisition of a large number of images of a particular specimen. This

is a particular hindrance to algorithms that involve large training sets, which has traditionally been available to computer vision tasks.

- *Transparent appearance of specimens.* Depending on the type of materials being imaged and the methodologies used, an object with the same physical geometry can appear significantly different.

While the fourth item is entirely task specific, the first three items are pertinent to general use of these systems. The remainder of this section addresses the image formation geometry of the scanning electron microscope. In Section 4, we propose the use of a rigid-body tracking system to deal with the low SNR images and an adaptive region-of-interest (ROI) system for increasing the system's frame rate.

3.1 Operating Principles

It is often remarked that given the significant difference between how optical and scanning electron microscope images are generated, the latter are easily interpreted by even novice users. This is due in part to the fact that the image is mapped similarly to a traditional optical system. A typical scanning electron microscope image is generated by rastering a stream of electrons across the sample of interest. The beam generated at the electron gun generally has a spot size too large to produce a sharp image and thus a series of electron lenses must be used to focus the beam (Figure 3.2).

The beam is generated at the electron gun which accelerates the electrons to an energy in the range of $0.1 - 30\,\text{keV}$. The beam then passes through a series of condenser lenses, which reduces the spot size. Lowering the beam size increases the resolution of the system, but also increases the signal noise due to the decrease in energy being returned from the sample. The scan coils are then used to raster the beam across the image, as opposed to an optical image sensor, which acquires all the data in parallel. Typically, the sample time for each pixel ranges in the order of $50 - 250\,\text{ns}$, where multiple samples at individual pixel locations are used to increase the signal-to-noise ratio (Figure 3.1). Finally, the beam passes through an objective lens, which focuses on the observed object [Goldstein *et al.*, 2003].

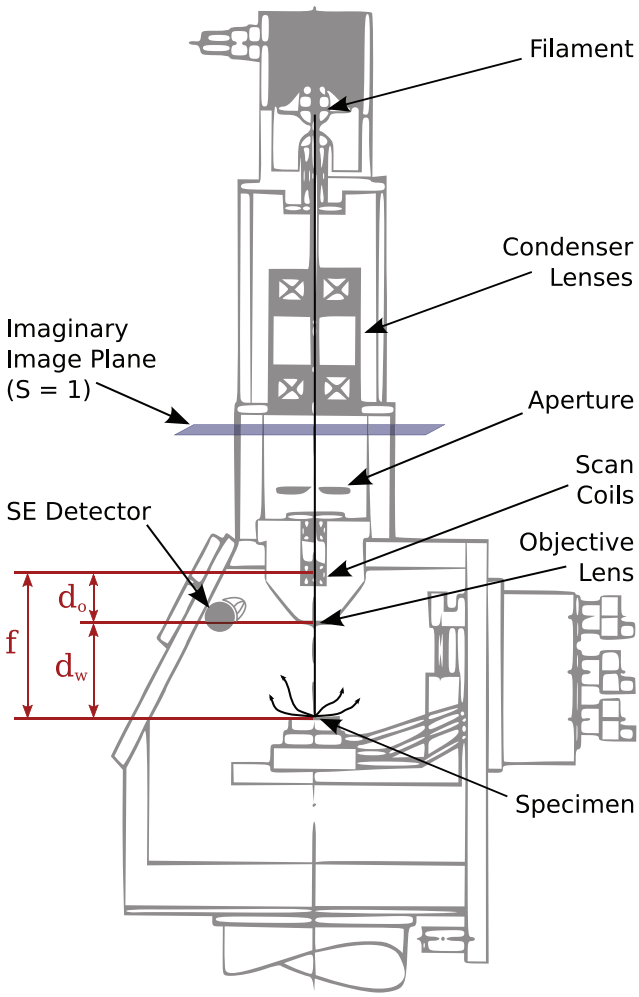
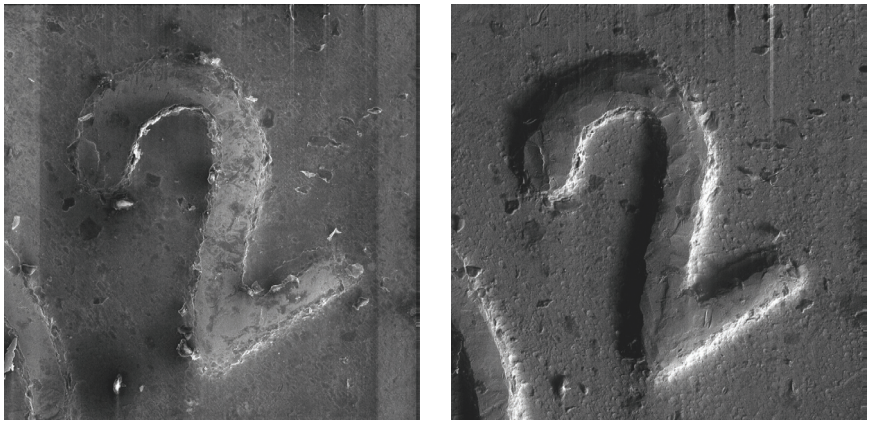


Figure 3.2: schematic drawing of electron column

When the electron beam scans across the material of interest, a number of signals are generated. Of primary concern to computer vision applications are the secondary and backscattered electrons. Secondary electrons (SE) are loosely bound outer-shell electrons that receive enough kinetic energy to be ejected by the inelastic scattering of the beam electrons. These are typically low-energy ($< 50\,\mathrm{eV}$) electrons and provide a diffuse image of the sample. Backscattered electrons (BSE) are high-energy electrons that are generated by elastic scattering and generate a more directional image. Although different detectors are used to sense each signal, the lenses that steer the electrons



(a) SE (b) BSE

Figure 3.3: Images captured with SE and BSE detectors. The same imaging model can be used for various detector types as long as the path of the electron to the sample is the same.

to the sample remain the same for each image type, which enables us to use the same model for different imaging modalities (Figure 3.3).

3.2 Sensor Calibration

In an optical system, a photon of light travels from the object being imaged through a lens system to the sensor which is the image plane. In the SEM, the electron travels from the source through a series of lenses to the specimen and is then sampled as the beam rasters across the image. Although there is no physical image plane in a SEM, a perspective camera model, similar to one used in optical systems, can be used to model the projection process as shown in Figure 3.5.

Although a variety of camera calibration algorithms exist in the literature [Tsai, 1987; Weng *et al.*, 1992; Zhang, 2000], most of them rely on specially engineered calibration targets to assist with the acquisition process. When calibrating an imaging system capable of magnifications over six orders of magnitude, it becomes difficult to find a precise calibration target suitable at a sufficient number of resolutions. Although targets have recently been designed that allow for calibration on multiple scales [Ritter *et*

al., 2004], we chose to utilize a method proposed by Cornille [Cornille, 2005; Cornille *et al.*, 2003] due to its independence from a structured calibrating environment.

In Cornille's thesis, he analyzes the microscope with the intention to reconstruct three-dimensional models of a sample of interest. Analogous to optical systems, the electromagnetic lenses in electron microscopes are also subject to irregularities that can lead to spatial and time-varying distortions. Cornille proposes a general method for assessing and compensating for these distortions. His work focuses on reconstruction of static objects and thus, setup and calibration time are not major considerations. During a typical manipulation task though, the microscope is used at a variety of magnifications and working distances.

Therefore, we extend the calibration scheme proposed by Cornille to be more suitable for visual servoing tasks. First in section 3.3.1, we propose a modification to the projection model which allows us to calibrate the SEM for any magnification and working distance. In this process, we also eliminate the need of a scaling factor as discussed in Cornille's work. This methodology is then used to calibrate two separate microscopes for later use with the rigid-body tracking systems, and the results discussed.

3.3 Projection Models

In optical microscopy, the perspective projection is often ignored in favor of the simpler parallel projection model. This is primarily due to the limited depth of field of an optical microscope. Using the pinhole projection model, a lens is only in focus at exactly the focal plane. At depths other than that, the image starts to become blurred. At small distances the defocused area is indistinguishable from a point. The depth of focus relates to the distance from the focal plane where the image becomes unsharp. The typical microscope lens with a viewing area on the order of hundreds of microns will have a depth-of-field of only a few microns. In this scenario, with motion along the optical axis the object will be out of focus before the image is significantly changed due to the projection. Microscopy systems often capitalize

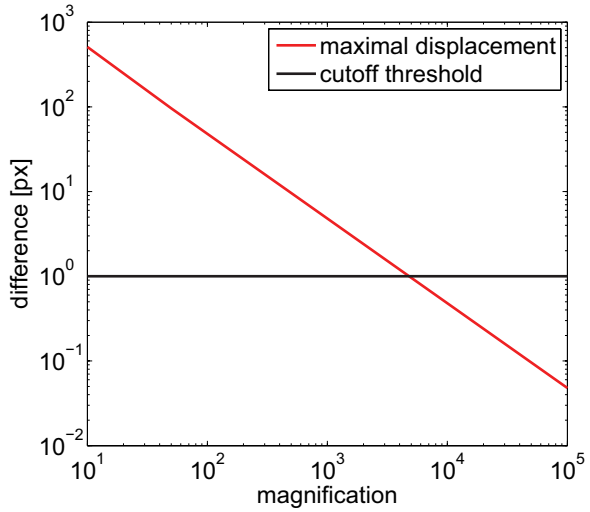


Figure 3.4: Limits of perspective projection: Over 4000× magnification movements that remain in focus cause less than one pixel motion in the image plane due to perspective projection.

on this fact for gathering additional three-dimensional information from the sample by taking images at a series different working distances to create a z-stack. This z-stack can be qualitatively evalutated or segmented and then quantitatively analyzed for additional information.

The larger working distances and smaller apertures of a scanning electron microscope allow for significantly larger displacements along the optical axis and thus at low magnifications, the perspective imaging model cannot be ignored. Depending on the imaging parameters such as aperture and working distance, the switching point between perspective and parallel projection can be calculated by determining the point at which the maximum change in position of the object along the optical axis creates less image plane motion than one pixel. Figure 3.4 shows one such estimation for a typical SEM configuration. In the following discussion, using no subpixel accuracy for the feature point detection, a valid calibration was possible up to a 1000× magnification. Although a perspective projection model may be used for magnifications beyond this, a parallel projection is also valid.

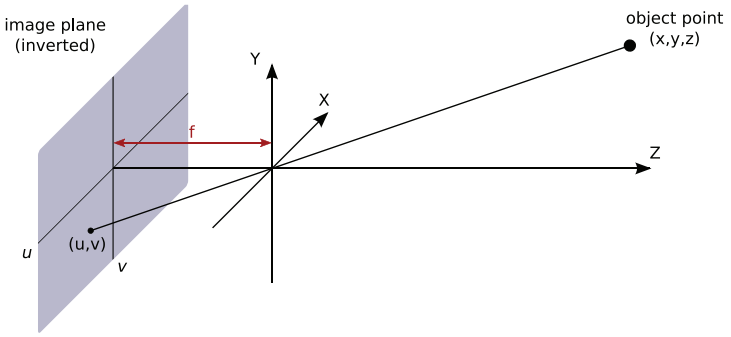


Figure 3.5: In a perspective projection, a ray of light is reflected from point $[x\ y\ z]^T$ and travels through the focal point after which it is detected at point $[u\ v\ f]^T$.

3.3.1 Perspective Projection

The imaging system of a SEM can be modeled by a pinhole camera model, which is a projection from 3D world points to 2D image points. This process is described using projective geometry [Faugeras, 1993]. A point in projective space, P^n, is defined by a $n+1$ coordinate vector $\mathbf{p} = \begin{bmatrix} p_1 & p_2 & \dots & p_{n+1} \end{bmatrix}$. The values of $\mathbf{p}$ denote the homogeneous coordinates of the point. Two points are equal in projective space if they lie on the same line through the focal point as shown in Figure 3.5. If the image point is located at $\mathbf{p} = \begin{bmatrix} u & v & f \end{bmatrix}$ then any point $\mathbf{x} = \begin{bmatrix} x & y & z \end{bmatrix}$ on the corresponding line is equal,

$$\frac{u}{f} = \frac{x}{z} \tag{3.1}$$

$$\frac{v}{f} = \frac{y}{z} \tag{3.2}$$

where f is the focal length of the system. Thus, a point in the camera's coordinate frame can be mapped to the SEM image plane as

$$\begin{bmatrix} U \\ V \\ S \end{bmatrix} = \underbrace{\begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}}_{K} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \tag{3.3}$$

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} \frac{U}{S} \\ \\ \frac{V}{S} \end{bmatrix} \tag{3.4}$$

Affine transformations are used to account for non-projective factors in the imaging process.

$$A = \begin{bmatrix} a_x & s & u_0 & 0 \\ 0 & a_y & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \tag{3.5}$$

The skew variable, s, determines the angle between the x and y axes. The variables u_0 and v_0 are coordinates of the principle point, which is where the optical axis intersects the image plane. The variables a_u and a_v traditionally relate to the physical pixel size in an optical system. Since there is no planar image sensor, the system cannot rely on the physical size of the sensing element for determination of f. Instead, these values are determined by the magnification factor as well as the image resolution of the SEM.

To account for this, we combine the matrices K and A to form the *intrinsic projection matrix* P.

$$P = KA = \begin{bmatrix} f_u & s & u_0 & 0 \\ 0 & f_v & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \tag{3.6}$$

To determine the values of f_u and f_v information must be known about the configuration of the microscope. Although the magnification and image size can be directly read from the SEM, the distance from the optical center to the specimen is unknown. In the case of an electron microscope, the optical center of the system is at the scan coils which deflect the beam (Figure 3.2). Although the beam must still pass through the objective lens, this only further focuses the beam and does not affect the image projection. The distance from the objective lens to the sample (d_w) is easily attainable for most SEMs, but the distance from the objective lens to the scan coils (d_o)

must be determined through calibration. The equation

$$f = \lambda M(d_w + d_o) \tag{3.7}$$

provides us with our focal length based on the magnification M, the working distance d_w, and the calibrated parameters scale λ and offset d_o. A series of multiple calibrations must performed to determine the parameters λ and d_o as discussed in Section 3.5.

In the event that the camera and world frames are not coincident, the homogeneous transformation matrix H of the *extrinsic parameters* is required to relate the camera's coordinate frame to the world coordinate frame. Using this relationship, we are now able to fully relate points in image space to vectors in world space [Faugeras, 1993].

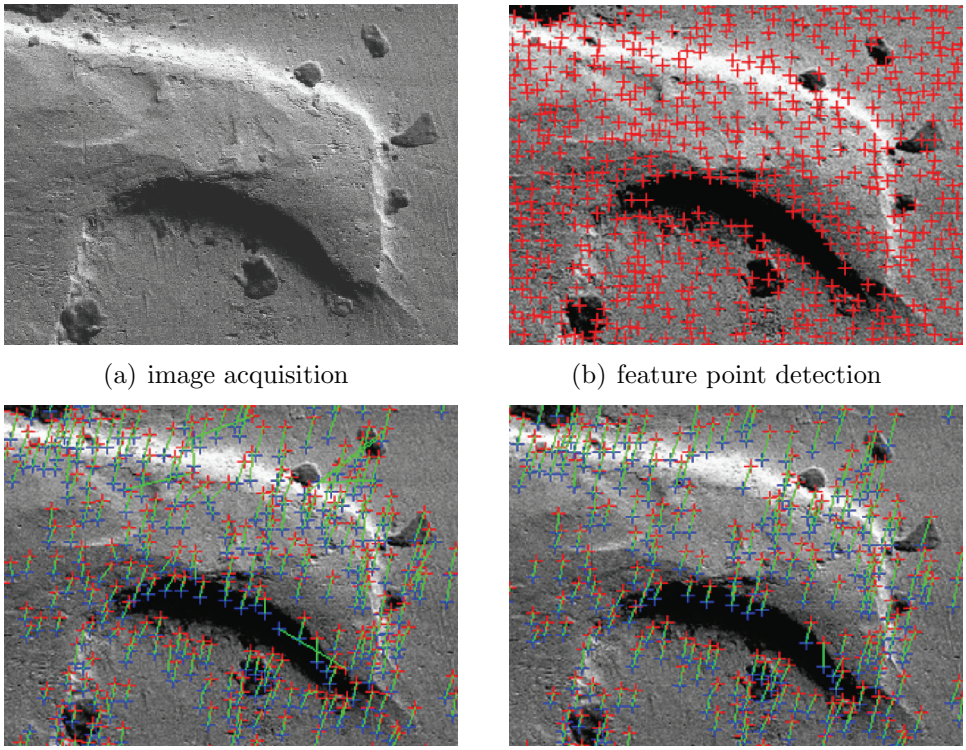
$$\begin{bmatrix} U \\ V \\ S \end{bmatrix} = PH\mathbf{x} \tag{3.8}$$

3.3.2 Parallel Projection

As the distance from the object to the image plane increases, the z distance between features becomes relatively small when compared with the viewing distance. When this occurs the image approaches an orthographic projection and the projection matrix for the SEM can be a simple scaling factor k and a skew s such that

$$\begin{bmatrix} U \\ V \\ 1 \end{bmatrix} = \begin{bmatrix} k & s & 0 & 0 \\ 0 & k & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \tag{3.9}$$

For magnifications greater than 1000× this assumption can be made because the effective size of the pixel becomes more than the shift in the image plane due to the perspective projection [Cornille *et al.*, 2003].



(a) image acquisition (b) feature point detection

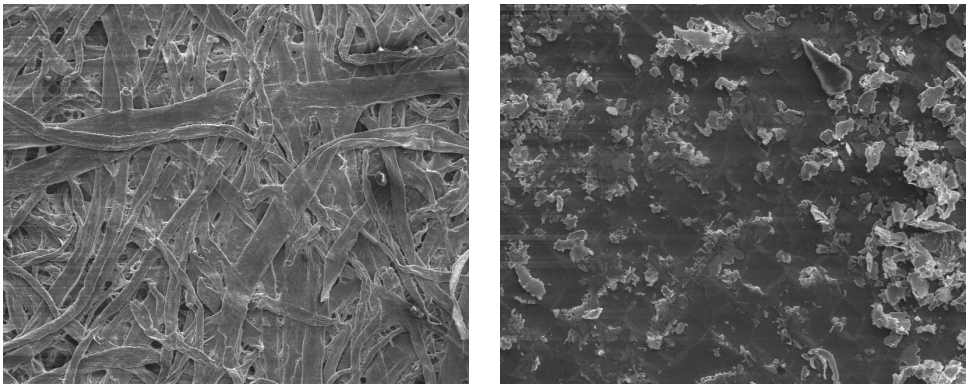
(c) feature point matching (d) robust filtering

Figure 3.6: calibration workflow

3.4 Calibration Procedure

The goal of the calibration process is to determine the intrinsic parameters of the SEM imaging model. Along with that, we will analyze the spatial and time varying distortions of the sensor. As previously noted, in Section 3, most camera calibration algorithms require known checkerboard patterns for calibration, which are difficult to manufacture with the required precisions at a multitude of scales. Also, non-parametric image distortions have been reported in the literature [Cornille *et al.*, 2003; Cornille, 2005; Sutton *et al.*, 2006], and in order to investigate these a large number of feature points are required. Due to these considerations, we elected to use a feature based method proposed by Cornille [2005].

Using this method, a series of tilted images are first acquired (Figure 3.6(a)). Features are located (Figure 3.6(b)) and matched between the multiple im-



(a) organic specimen at 100× magnification

(b) back of a wafer at 400×

Figure 3.7: These are segments of images of the specimens used for calibration. They are nearly planar and provide a large set of well distributed feature points.

ages (Figure 3.6(c)). Section 5.1.1 discusses the detection and selection methods more thoroughly. To prevent outliers from corrupting the optimization, a final filtering step is performed to remove any erronously matched features (Figure 3.6(d)). Finally, the entire group of matches is used to optimize the camera parameters. Since our goal is to fully calibrate the SEM, we perform this process at a variety of different magnifications and working distances. Using the updated focal length model of Section 3.3.1, we are able to then fit the remaining parameters for the fully calibrated system.

3.4.1 Image Acquisition

Although the calibration method does not require a manufactured target, it does require a large number of target points. These points need to be roughly planar and during motion it is desirable that few points become occluded. For low magnifications, an organic sample was chosen (Figure 3.7(a)) and for high magnifications, the back of a silicon wafer was chosen (Figure 3.7(b)). The images for calibration were captured by eucentrically tilting the sample as shown in Figure 3.8, which provide the critical movements for focal length calibration [Sturm, 2002].

3.4.2 Feature Point Detection and Matching

The Harris corner detector [Harris and Stephens, 1988] was selected for feature detection. This method uses the local auto-correlation to detect edges and corners in the image. These features must then be matched to those found in a second image. The small image displacements allow us to define a region of possible correspondences. These regions were then searched using a normalized cross correlation method [Stefano *et al.*, 2003], which is similar to a template matching operation near each of the possible correspondences. Finally, a Least Median of Squares [Rousseeuw and Leroy, 2003] estimator is used to remove invalid matches based on epipolar geometry constraints.

3.4.3 Sparse Bundle Adjustment

The set of feature matches serve as input to the bundle adjustment cycle [Triggs *et al.*, 2000]. Initially, estimates are determined for both the pose of the individual feature points as well as the pose of the target object. Based on these poses, the projection errors are determined for each of the different images. These reprojection errors are then minimized using the following:

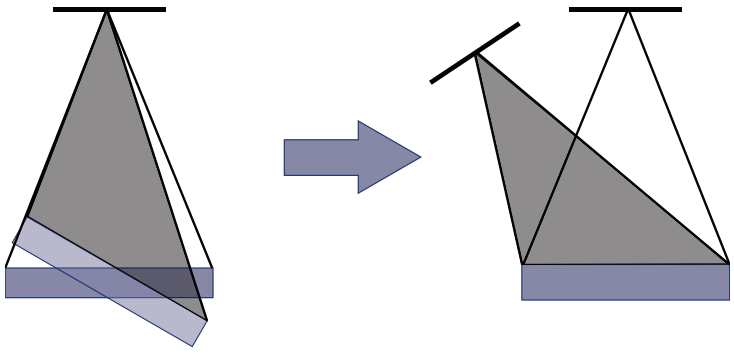


Figure 3.8: The tilting of the specimen corresponds to a rotation of the camera around the specimen.

$$\begin{bmatrix} U_i^j \\ V_i^j \\ S_i^j \end{bmatrix} = PH^j\mathbf{x_i} \tag{3.10}$$

$$\begin{bmatrix} u_i^j \\ v_i^j \end{bmatrix} = \begin{bmatrix} \frac{U_i^j}{S_i^j} \\ \\ \frac{V_i^j}{S_i^j} \end{bmatrix} \tag{3.11}$$

$$\min_{\mathbf{x_i},H^j,P} \sum_i \sum_j \left\| \begin{bmatrix} u_i^j \\ v_i^j \end{bmatrix} - \begin{bmatrix} \tilde{u}_i^j \\ \tilde{v}_i^j \end{bmatrix} \right\|_2 \tag{3.12}$$

Where u_i^j and v_i^j are the estimated pixel coordinates of the i^{th} point in the j^{th} image, and $\tilde{u}_i^j$ and $\tilde{v}_i^j$ are the observed positions. The fitting step is performed over all of the intrinsic camera parameters, the pose estimates of each image, and the pose of each feature in relation to the particular frame. This non-linear optimization process can be solved by readily available software packages such as Matlab.

Figure 3.9 shows how the adjustment of the camera parameters and 3D coordinates minimize the reprojection errors:

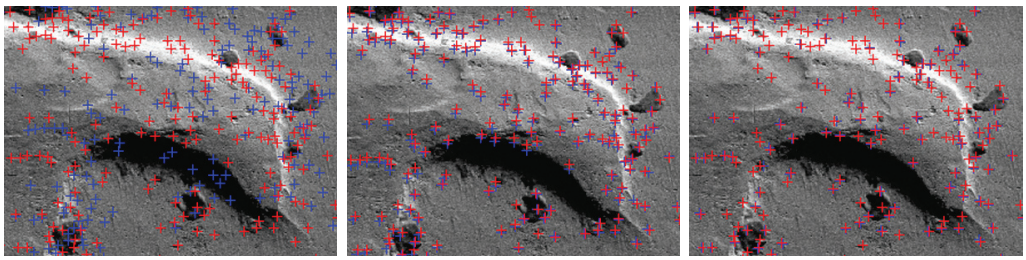


Figure 3.9: The bundle adjustment adjusts camera parameters and 3D coordinates to reduce the reprojection errors. The reprojected 3D points indicated by blue crosses converge to the feature points fixed in the image indicated by red crosses.

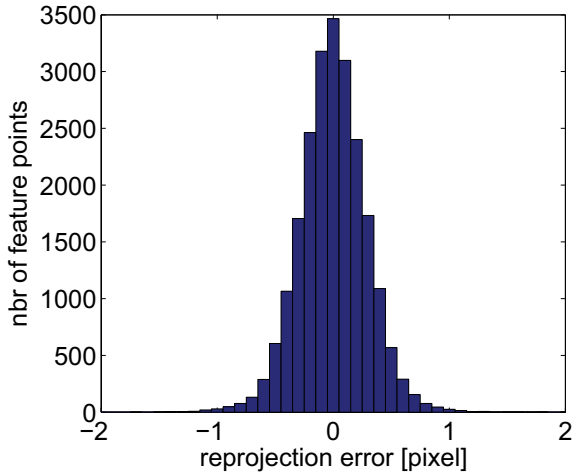


Figure 3.10: The reprojection errors after a calibration run with a 100× magnification and $d_W = 15$.

3.5 Experimental Results

The calibration algorithm was evaluated on a Zeiss DSM 962 and a Zeiss ULTRA 55. The manufacturer states that DSM 962 has approximately 10 nm resolution, while the Ultra 55 has approximately 1 nm resolution.

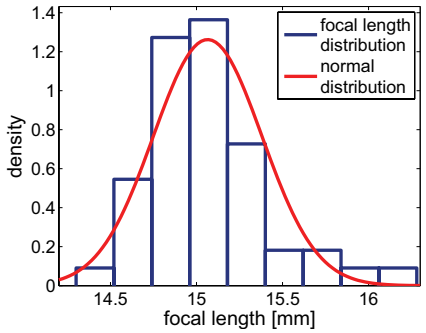
Before analyzing the performance of the calibration algorithm as a whole, it is beneficial to analyze the performance of the feature detection portion. Figure 3.10 shows the results of the reprojection errors after calibration. Since our feature detection algorithm did not have sub-pixel accuracy, errors are to be expected ± 0.5 pixels. The standard deviation of the reprojection error was 0.41 pixels.

Due to the diminishing effects of perspective projection with increased magnification, Figure 3.11 shows that the standard deviation of the focal length estimation increases with increased magnification. In all cases, this deviation was below 5% of the target value.

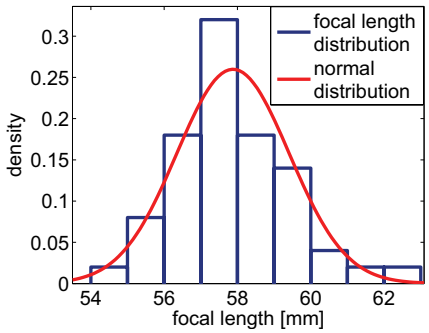
The accuracy of the focal length estimation is also influenced by the resolution of the images acquired. Table 3.1 shows the influence of resolution on the calibration procedure. Although there is a significant decrease in standard deviation between the 500×500 image and the 1000×1000 image, there is not a significant benefit to image sizes higher than this. Thus, for

Table 3.1: The standard deviation of focal length estimation out of constant input images at 100× and $d_W = 15\,\text{mm}$ using different resolutions.

Resolution:	500 × 500	1000 × 1000	1500 × 1500
Standard deviation [mm]**:**	0.103	0.051	0.053
Focal length [mm]**:**	18.6	18.6	18.6
Relative error [%]**:**	0.55	0.27	0.28

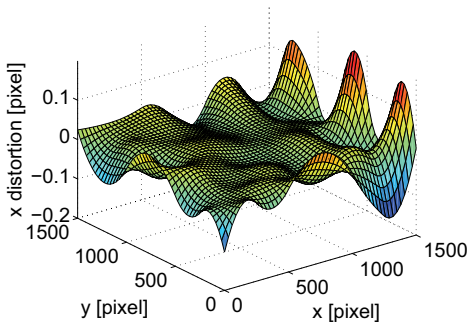


(a) focal length estimation at 100×

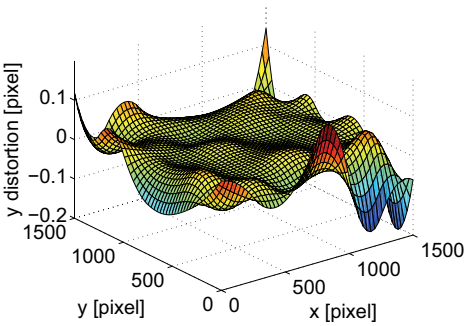


(b) focal length estimation at 400×

Figure 3.11: Distribution of focal length estimation at $d_W = 15$. At 100× the focal length distribution has a standard deviation of 0.3 mm. At 400× it is already 1.6 mm.



(a) distortion along x axis



(b) distortion along y axis

Figure 3.12: Along both axes, the distortion of the image stays between -0.2 and 0.2 pixel and is therefore neglectable.

the remainder of the discussion 1000 × 1000 images are used.

Sutten *et al.* [2006] examine these phenomenon for metrology applications and find that, although the spatial distortion is relatively small across the

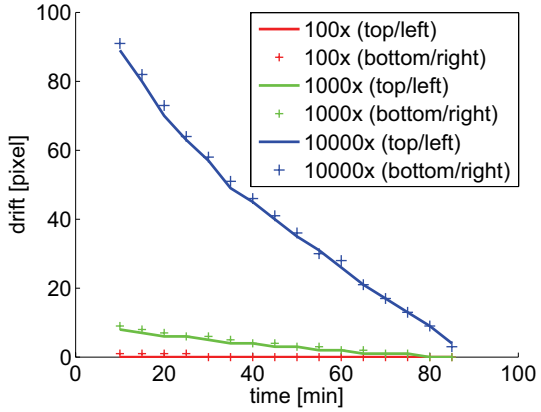


Figure 3.13: At $10,000\times$ magnification, the drift is $\sim$1 pix/min and needs to be corrected. At lower magnifications such as $100\times$, the drift is $\sim$0.01 pix/min and can be neglected.

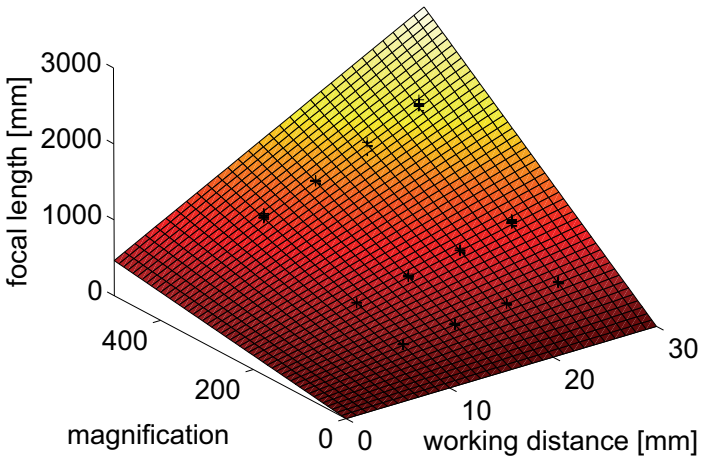


Figure 3.14: the calibrated focal length for a variety of magnifications and working distances

image (approximately 0.5 pixel maximum), the drift ranges up to 20 pixels over 2 h at the magnifications used.

Figure 3.12 is taken from the DSM 962 and demonstrates that the distortions of the image are much smaller than those found by Cornille, and for the case of visual servoing can safely be disregarded. The phenomenon of drift across the image was noted on both SEMs, but only tracked on the

DSM 962. In this case, shown in Figure 3.13, the drift was characterized at a variety of different magnifications and found to affect the entire viewing area homogeneously at a rate of approximately 10 nm/min. Thus, the need for drift compensation is highly dependent on the task being performed.

In the case of visual servoing for robotic manipulation, it is desirable to rapidly know the calibration parameters at a variety of magnification and focal distances to allow for variable zoom and decreased experimental setup time, which makes calibration for each experiment impractical. To determine the intrinsic parameters of the system at any system configuration, the parameters λ and d_o must be determined. To find these parameters, a series of twelve calibration experiments were performed with each of the SEMs. The output from the DSM 962 is shown in Figure 3.14. Using this method, the imaging parameters for the SEM can be reliably determined for any set of parameters used during a manipulation task.

3.6 Discussion

This section examined the imaging characteristics of a scanning electron microscope and particularly how they pertain to visual servoing applications. Independent of the electron detector used, the system can be modeled with a pinhole camera. With this knowledge, we were able to fully calibrate the imaging system for magnifications below 1000× as a perspective camera. This scenario is most useful for microassembly tasks that would like to benefit from the high-resolution and large depth-of-field that a SEM imaging system has to offer. For higher magnifications, an orthographic projection model should be used, and the drift and spatial distortion of the sensor are of more interest. We found that the spatial distortions of the two SEMs characterized were not as significant as that reported by Cornille [2003], allowing future work to omit this calibration phase unless extreme high fidelity is required. The drift of the system analyzed was linear, uniform across the image, and independent of magnification. This indicates that while for low-magnification applications it can safely be ignored, it may need to be taken into consideration for high-magnification.

4 Rigid-body Tracking for Nanorobotic Manipulation

Electron microscopes provide some of the most challenging imaging conditions for real-time computer vision applications. As noted in Section 3, due to measurement constraints, images are often low-contrast, noisy, and have slow refresh rates. For this reason, the majority of computer vision applications presented thus far work with offline images for tasks such as 3D reconstruction.

Some of the first real-time tracking applications in the SEM were performed by Kasaya *et al.* [1999]. They used Hough transforms to detect 30 µm spheres to enable automatic pick-and-place operations. With this system, they were able to localize the center of the spheres to approximately 3 µm and the end effector to 6 µm.

Sievers and Fatikow [2005] proposed using a cross-correlation based method for tracking a tungsten probe tip. To improve imaging frame-rate, they reduced the area of interest to a 128×128 window around the target to enable 10 Hz tracking. One challenge with this system is that template images must be stored for each magnification and rotated versions of each template must also be stored for angular detection. Another difficulty with correlation based methods, is that they tend to fail if the object is occluded or deformed.

Later work by Sievers *et al.* [2006], used active contours [Kass *et al.*, 1988] to track objects in the SEM. These methods use energy-minimizing splines to track a shape or region in the image. The position of the nodes in the contour is determined by balancing internal forces which help guide the shape of the contour, with image forces, which attract the nodes to features in the image. Contour based methods handle the changing shape of a tracked object by themselves deforming. Due to the dynamic nature of SEM image, gradient based features were selected *in lieu* of intensity based ones. Since these methods are purely image-based, they are restricted to using information available in the image. Thus, they are limited to in-plane rotations and translations. Many of the probes and manipulators being tracked are rigid-bodies, and occlusions can potentially be misinterpreted as deformations by contour based

methods. Objects of interest may have multiple polygonal surfaces, which cannot be utilized for tracking by deformable contour methods.

A number of successful tracking methods have been developed for tracking rigid objects in complex scenes [Lowe, 1991; Rehg and Kanade, 1995; Wunsch and Hirzinger, 1997; Yesin and Nelson, 2005]. These methods benefit from being robust from noise because they use a large number of image points to perform the estimation task. The addition of model geometry allows the methods to operate in environments where the objects are largely occluded. Tracking applications differ from pose estimation algorithms in that they rely on previously known pose information to restrict the scope of the search. Current tracking algorithms can roughly be partitioned into feature-based, model-based methods, and hybrid methods. Feature-based methods generally encompass methods such as active contours [Kass *et al.*, 1988], affine covariant region detectors [Mikolajczyk *et al.*, 2005], or techniques for detecting geometric primitives [Illingworth and Kittler, 1988]. Model-based methods capitalize on pre-existing knowledge of the tracking problem such as the geometrical structure of the target [Drummond and Cipolla, 2002] or its interaction with the environment [Greminger and Nelson, 2004]. Hybrid methods incorporate aspects of both [Pressigout and Marchand, 2007] with the desire to be robust to factors which would cause problems to methods employing only a single modality.

Although electron microscopes provide a host of new challenges to imaging systems, they also allow for some simplifying assumptions to be valid in real-world applications. In particular, manipulation tasks in a SEM generally have structured environments without a large number of occlusions or distractors. Many of the objects present in a manipulation task are either previously modeled or can be easily modeled due to the microfabricated nature of their construction. These facts have led us to choose rigid-body methods based on [Drummond and Cipolla, 2002; Yesin and Nelson, 2005] to help provide more precise motion feedback for manipulation and measurement tasks. The model-based approach allows us to exploit knowledge of the task at hand to increase tracking precision and frame rate.

Bregler and Malik [1998] first proposed a tracking method that used

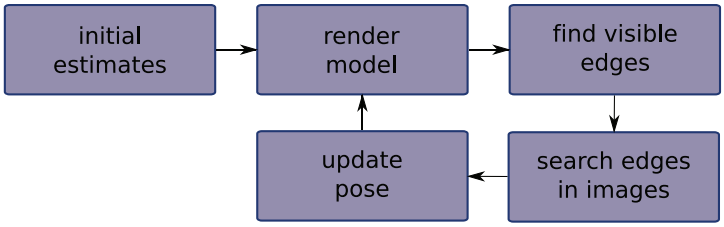


Figure 4.1: flowchart of tracking system operation

twists [Ball, 1876] to simplify the pose estimation framework. Later, Drummond and Cipolla [2002] extended and formalized the framework using Lie algebra. In this work, we pose the work of Drummond and Cipolla as a least-squares optimization, which enables us to use standard constrained least-squares optimization techniques for solving the fitting problem. The least-squares formulation allows us to use linear-regression techniques to determine at runtime the optimal regions-of-interest for tracking any target object. We then demonstrate how these strategies improve tracking performance at the nano-scale by providing the system with additional information and higher quality images.

4.1 Rigid-body Motion

The model pose with respect to the camera is stored in the matrix $H \in SE(3)$ such that a model point $\mathbf{x}$ is projected into image space by

$$\begin{bmatrix} U \\ V \\ S \end{bmatrix} = PH\mathbf{x}. \tag{4.1}$$

The Special Euclidean group (3) is commonly know in the robotics literature as homogeneous transformations. The Lie algebra of SE(3), denoted $se(3)$,

is identified by a 4×4 skew symmetric matrix of the form:

$$\begin{bmatrix} 0 & -\omega_3 & \omega_2 & v_1 \\ \omega_3 & 0 & -\omega_1 & v_2 \\ -\omega_2 & \omega_1 & 0 & v_3 \\ 0 & 0 & 0 & 0 \end{bmatrix} = \hat{\xi} \tag{4.2}$$

The mapping from $se(3)$ to $SE(3)$ is performed by the exponential formula $H = e^{\hat{\xi}}$ and a closed-form solution exists through the Rodriguez formula. We refer to the matrix $\hat{\xi}$ as a *twist*. Similar to Murray [Murray *et al.*, 1994], we define the $\vee$ (vee) operator to extract the six-dimensional *twist coordinates* which parametrize a twist,

$$\begin{bmatrix} 0 & -\omega_3 & \omega_2 & v_1 \\ \omega_3 & 0 & -\omega_1 & v_2 \\ -\omega_2 & \omega_1 & 0 & v_3 \\ 0 & 0 & 0 & 0 \end{bmatrix}^{\vee} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ \omega_1 \\ \omega_2 \\ \omega_3 \end{bmatrix} = \xi \tag{4.3}$$

The motion between consecutive frames can be represented by right multiplication of H with a motion matrix M. The generators of the group $SE(3)$ can be defined as the translations in the x, y, and z directions, as well as rotations about their respective axes expressed in the current model frame. The *twist coordinates* for these generators are:

$$g_i = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \tag{4.4}$$

These generators form a basis for the vector space of derivatives of $SE(3)$ at

the identity and use the exponential map to obtain group elements such that

$$M = e^{\sum_{i=1}^{6} \alpha_i \hat{g}_i} \tag{4.5}$$

The values of α are scalar quantities such that $\alpha = [\alpha_1 \dots \alpha_6]^T$. Assuming that the motion between frames is small, a linear approximation can be made for M using Taylor's theorem.

$$M \approx I + \sum_{i=1}^{6} \alpha_i \hat{g}_i \tag{4.6}$$

Since the motion is approximately a linear sum of each of the generators, the partial derivatives (illustrated in Figure 4.2) for each point $\mathbf{x}$ with respect to each generating motion can be computed as:

$$\begin{bmatrix} U_i' \\ V_i' \\ S_i' \end{bmatrix} \approx PH\hat{g}_i\mathbf{x} \tag{4.7}$$

$$L_i = \begin{bmatrix} \frac{\partial u_i}{\partial \alpha_i} \\ \\ \frac{\partial v_i}{\partial \alpha_i} \end{bmatrix} = \begin{bmatrix} \frac{U_i'}{S} - \frac{US_i'}{S^2} \\ \\ \frac{V_i'}{S} - \frac{VS_i'}{S^2} \end{bmatrix} \tag{4.8}$$

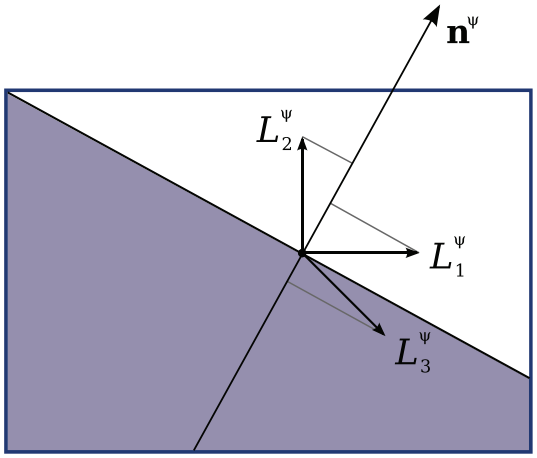


Figure 4.2: translational generating motions for $SE(3)$

4.2 Edge Tracking

The edge tracking algorithm leverages the aperture problem, which states that only motion normal to an edge can be detected. Thus, the edge-tracking phase can be simplified to searching for intensity gradients along a one-dimensional path normal to a point on the edge. The first stage is to determine which edges are currently visible in the view. The model is provided for the system in the form of a boundary representation model that can be drawn in commercially available CAD packages. These models consist of triangular meshes that cover the surface of the model. An initial preprocessing step can be used to classify these edges into either "invisible," "silhouette," or "sharp." Invisible edges occur when a flat surface is converted into a mesh of triangles and cannot be viewed. Sharp edges are generated by two faces of the original model coming into contact. Silhouette edges are ones that are generated along the face of a curved surface, and are only capable of being viewed when they make up part of the silhouette of the image. The use of triangular meshes allows for arbitrary shapes to be modeled using this method.

The system initially renders the model using estimates for the object pose H. These estimates can be supplied by the user or another image localization package. Using a method such as z-buffering [Isenberg *et al.*, 2003] or binary

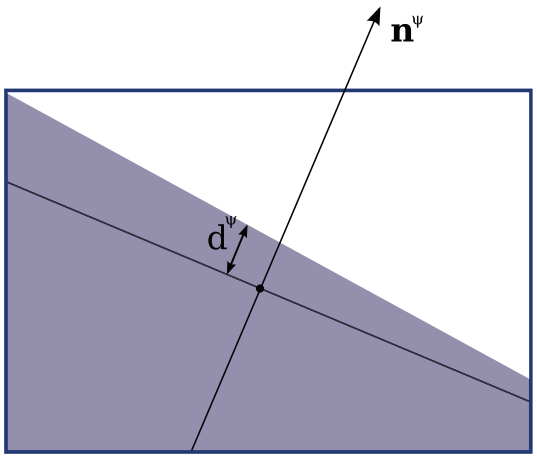


Figure 4.3: error measurement along normal direction

space partition (BSP) trees, a user can determine which edges of the polygon mesh are visible. These model edges are then divided into m small segments for tracking. Once the midpoints are determined and the unit-normal $\mathbf{n}^{\psi}$ constructed, where ψ denotes the ψ^{th} segment being analyzed, a line of l points is searched for the edge location (Figure 4.3). A variety of detection methods such as the Canny or Laplacian operators can be used for edge detection. For speed purposes, we used a gradient kernel operator. This is performed by convolving the search line with a one-dimensional kernel $\begin{bmatrix} -1 & 0 & 1 \end{bmatrix}$ and analyzing the resulting values.

Different strategies exist for selecting the edge from the resultant data. Yesin [Yesin, 2003] suggests selecting the mean location of all points that surpass a common threshold. In this case, the algorithm's performance can be severely influenced by neighboring edges. Although methods such as image space potentials exist for rectifying this problem, they are processor intensive. Instead we propose using either the maximum edge along the search path or the closest point to the midpoint that supasses a prescribed threshold. In all cases, the distance from the midpoint to the detected edge is then recorded as d_{ψ}. Other useful information can also be recorded in this step for use with heuristics to improving tracking. More information on this topic is provided in Section 4.3.1.

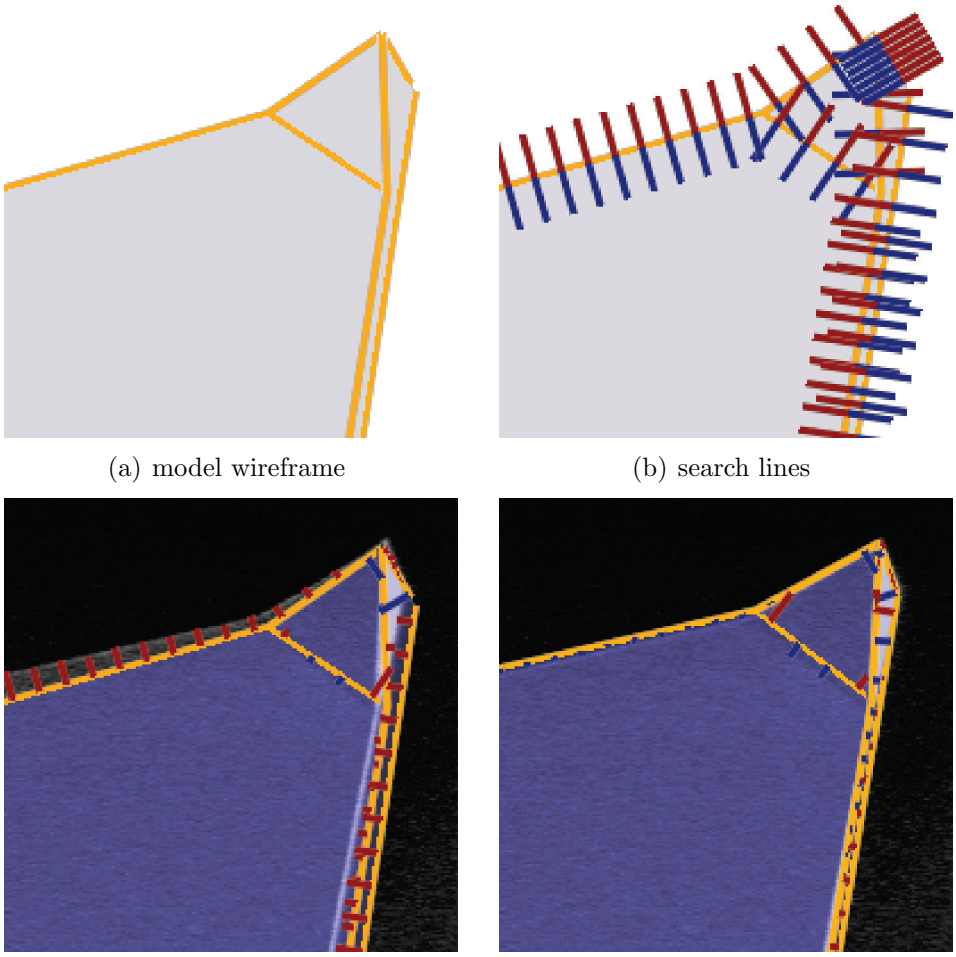
4.3 Pose Estimation

The final step is pose estimation. With knowledge of both the edge normal and the partial derivatives with respect to generating motions, we are able to build an $m \times 6$ image Jacobian that describes the normal component of observed motion with respect to the i^{th} generating motion as

$$J_{\psi i} = \frac{\partial d_{\psi}}{\partial \alpha_i} = \mathbf{n}^{\psi} \cdot L_i^{\psi} \tag{4.9}$$

such that

$$J\alpha = \mathbf{d} \tag{4.10}$$



(a) model wireframe (b) search lines

(c) edges detected (d) new pose

Figure 4.4: The images display one iteration of the pose estimation sequence. In a) the hidden edges are removed b) the normal lines are searched c) the edges are located and d) a new pose is estimated.

where $\mathbf{d}$ is the $m \times 1$ vector of d_ψ values. Formulating the problem in this fashion allows us to use a least-squares approach to find the updated pose. If J has full column rank, then $J^T J$ is invertable and the following equation can be used to solve for α:

$$\alpha = \left(J^T J\right)^{-1} J^T \mathbf{d} \tag{4.11}$$

In the case that J does not have full column rank, methods such as those based on singular value decomposition can be utilized to find the best estimate of α. Once the new update of α is available, the pose of the tracked object is updated before repeating the tracking cycle.

$$H_{t+1} = H_t M \tag{4.12}$$

The flow of the algorithm in illustrated in Figures 4.1 and 4.4.

4.3.1 Row Weighting

The large number of samples helps reduce the influence of outliers in the least-squares estimation, but additional steps can be taken to improve the quality of the fitting. If extra information is known about particular samples, this data can be used to weight the influence of a particular measurement.

$$\begin{aligned} WJ\alpha &= W\mathbf{d} \\ W &= \text{diag}(w_1, w_2, ..., w_m) \end{aligned} \tag{4.13}$$

This scaling transforms the problem into an iterative least-squares problem [Golub and Loan, 1996]. The iterations can either occur at each new image frame, or due to the relatively slow refresh rate of the SEM can occur multiple times on a single image. Drummond and Cipolla [Drummond and Cipolla, 2002] suggest a variety of different methods for choosing these weighting functions which attempt to solve different domain-specific problems. These criteria and others can be multiplicatively combined to provide an estimate for w_ψ:

- **False Matches** Since least-squares assumes a Gaussian distribution, outliers in the edge detection step can cause a large change in the estimated pose. To alleviate this problem the weighting function can be adjusted to:

$$w_\psi = \frac{1}{c + |d_\psi|} \tag{4.14}$$

A good value for the constant c is one standard deviation of the error data [Drummond and Cipolla, 2002].

- **Multiple Edges** If multiple edges are on the search line, the tracker is capable of choosing the wrong one. This problem can be reduced by making the weight inversely proportional to the number of edges detected on the search line.

- **Edge Quality**

  The edge strength can be cited as a helpful metric because the weaker edges are harder to track. A set of two thresholds can be used for generating the weights. Below a certain threshold no edge is detected. Between the two thresholds the weight varies linearly, and above a threshold the weight is not adjusted.

4.4 Constrained Tracking

The tracking performance can be improved if domain-specific information can be factored into the pose estimation process. One of the simplest ways to incorporate this extra domain knowledge is by reformulating the problem to perform a least-squares minimization with equality or inequality constraints:

$$\begin{aligned} J\alpha &= \mathbf{d} \\ C\alpha + \beta &\leq \mathbf{0} \\ C_{eq}\alpha &= \mathbf{0} \end{aligned} \tag{4.15}$$

A variety of efficient software libraries such as LAPACK, Matlab, and levmar [Lourakis, 2004] exist for solving these types of problems. To better illustrate how these constraints are built we will briefly discuss a few examples.

4.4.1 Planar Constraint Example

The simplest case is to put a constraint on tracking a single object. For example, this can be due to the object resting on a ground plane. In this

case, we would like to restrict translations along the z axis and not allow rotations around the x and y axes of the object being tracked.

$$c^{1,2,3} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}^T, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}^T, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}^T \tag{4.16}$$

We can write the constraint equations as

$$\begin{bmatrix} c^1 \\ c^2 \\ c^3 \end{bmatrix} \alpha = \begin{bmatrix} \mathbf{0_{3\times 1}} \end{bmatrix} \tag{4.17}$$

and solve the constrained least-squares optimization problem.

4.4.2 Optical Axis Constraint Example

Due to the projection mode of the SEM, the image data contains minimal depth information. The absence of this information can cause instabilities in the fitting operation. As shown in Section 4.7, it is beneficial to restrict motion along the optical axis with constraints in the camera frame such as:

$$c^1 = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix} \tag{4.18}$$

Since the pose is estimated in the body frame, the constraint must be transformed from the camera frame. The adjoint operator provides a convenient method for transforming a twist from one coordinate frame to another [Murray *et al.*, 1994]. Given $M \in SE(3)$, the adjoint transform is a 6×6 matrix which transforms twists from one coordinate frame to another.

$$M = \begin{bmatrix} R & \mathbf{t} \\ \mathbf{0_{1\times 3}} & 1 \end{bmatrix} \tag{4.19}$$

$$Ad(M) = \begin{bmatrix} R & \hat{\mathbf{t}}R \\ \mathbf{0_{3\times 3}} & R \end{bmatrix} \tag{4.20}$$

The adjoint operator is invertible, and is given by:

$$Ad^{-1}(M) = \begin{bmatrix} R^T & -R^T\hat{\mathbf{t}} \\ \mathbf{0_{3\times 3}} & R^T \end{bmatrix} \tag{4.21}$$

The adjoint operator can thus be used to express the constraints in the camera frame as:

$$c^1 \mathrm{Ad}(H)\alpha = 0 \tag{4.22}$$

4.4.3 Multiple Object Constraint Example

Although multiple objects can be tracked independently, often one also wishes to track articulated objects. We can first build a sparse matrix that includes the Jacobians of both objects to be tracked.

$$\begin{bmatrix} J_1 & \mathbf{0_{m\times 6}} \\ \mathbf{0_{m\times 6}} & J_2 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} = \begin{bmatrix} \mathbf{d_1} \\ \mathbf{d_2} \end{bmatrix} \tag{4.23}$$

In the case of multiple objects each of their points are defined in their local frames. If we decide to choose the constraints in the local frame of α_1, then we must transform the motions of α_2 to the frame of α_1 for constraint checking. Since the transformation from the coordinate frame of α_2 to α_1 is described by the homogeneous matrix $H_1^{-1}H_2$, we can state:

$$c_1^i\alpha_1 + c_2^i \mathrm{Ad}(H_1^{-1}H_2)\alpha_2 = 0 \tag{4.24}$$

and thus

$$\begin{bmatrix} c_1^1 & c_2^1\mathrm{Ad}(H_1^{-1}H_2) \\ \vdots & \vdots \\ c_1^\varsigma & c_2^\varsigma\mathrm{Ad}(H_1^{-1}H_2) \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} = \begin{bmatrix} \mathbf{0_{\varsigma\times 1}} \end{bmatrix} \tag{4.25}$$

The vectors $c^{1..\varsigma}$ are then defined in the frame of α_1. Since we would like the combined motion of α_1 and α_2 along the constrained axes to be zero, the constraints can be described as:

$$c_2^i = -c_1^i \tag{4.26}$$

In the case of tracking the arms of a microgripper, we would like to restrict the motion between the gripper arms to planar motions and ensure that no rotations are allowed between the bodies. Thus, planar constraints similar to equation 4.16 can be used for c_1.

4.4.4 End Effector Tracking Example

In the micro- and nano-domains, we often have the ability to move with higher resolution than we can accurately measure. It is also difficult to calibrate manipulators at this scale which hinders the manipulator's ability to accurately move to positions in the environment. One possible use of visual feedback is to close the loop and help increase the accuracy of positioning motions. Although the feedback from the actuators is less accurate than we would like, information from the kinematics of the robot can be used to help constrain the pose tracking of the end effector during motion.

The body Jacobian of a manipulator maps joint velocities to end effector velocity with respect to the tool frame, and is defined as:

$$J_{st}^b(\theta) = \left[\begin{array}{cccc} \xi_1^\dagger & \xi_2^\dagger & \dots & \xi_n^\dagger \end{array} \right] \tag{4.27}$$

$$\xi_i^\dagger = Ad^{-1}(e^{\hat{\xi_i}\theta_i} \dots e^{\hat{\xi_n}\theta_n H})\xi_i \tag{4.28}$$

Where ξ_i is the twist coordinate representation related to the i^th joint, $H \in SE(3)$ which relates the tool frame to the station frame, and θ is the current joint position. If we had perfect knowledge of the manipulator kinematics, we could use the constraint:

$$\begin{aligned} K &= J_{st}^b(\theta)(\Delta\theta) \\ K^T\alpha &= K^T K \end{aligned} \tag{4.29}$$

Even though it is unrealistic to assume perfect knowledge of the manipulator kinematics, the robot kinematics can still be leveraged to provide the direction of motion.

Many micro and nanomanipulators are restricted to purely translational motions. In these and other kinematically simple configurations, the manipulator Jacobians can be analyzed and constraints that restrict motion along particular axes can be created. In the case of manipulators with rotational axes, we can still constrain the direction of motion to that of the manipulator.

$$\begin{aligned} c_i &= \begin{cases} -1 & \text{if } K_i \geq 0, \\ \\ 1 & \text{if } K_i < 0 \end{cases} \\ \text{diag}(c)\alpha &\leq 0 \end{aligned} \tag{4.30}$$

These are just a few examples of how domain-specific knowledge can be used to provide extra information to the pose estimation process.

4.5 Tracking in $SE(2)$

The previous examples were provided for tracking objects with three-dimensional coordinates. Two-dimensional tracking is useful for objects that will not rotate out of plane during a manipulation as well as objects that are not precisely modeled prior to the experiment. In the latter case, the rigid body model can automatically be created from an image of the object taken at the beginning of an experiment. The previously discussed algorithms can be analogously applied to tracking planar objects in $SE(2)$. An element in $se(2)$ is represented as

$$\begin{bmatrix} 0 & -\omega_1 & v_1 \\ \omega_1 & 0 & v_2 \\ 0 & 0 & 0 \end{bmatrix}^{\vee} = \begin{bmatrix} v_1 \\ v_2 \\ \omega_1 \end{bmatrix} \tag{4.31}$$

Given $M \in SE(2)$, the adjoint transform is the 3×3 matrix

$$Ad(M) = \begin{bmatrix} R & \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \mathbf{t} \\ \mathbf{0_{1\times 2}} & 1 \end{bmatrix} \tag{4.32}$$

The adjoint operator is invertible, and is given by:

$$Ad^{-1}(M) = \begin{bmatrix} R^T & -R^T \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \mathbf{t} \\ \mathbf{0_{1\times 2}} & 1 \end{bmatrix} \tag{4.33}$$

In this case we use parallel projection, and our image projection becomes:

$$\begin{bmatrix} U \\ V \\ 1 \end{bmatrix} = \begin{bmatrix} k & s & 0 \\ 0 & k & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \tag{4.34}$$

and the generating motions are:

$$g_i = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \tag{4.35}$$

Since there is no dependence on Z, the partial derivatives with respect to the generating motions are simply:

$$L_i = \begin{bmatrix} U_i' \\ V_i' \end{bmatrix} \tag{4.36}$$

All other operations can be applied in the same fashion as outlined above. Selection between tracking in $SE(2)$ or $SE(3)$ depends on the projection model selected as well as the desire to track out-of-plane motions. Often manipulators used at the micro- and nanoscales are constructed entirely of prismatic joints, making the $SE(2)$ tracking method an attractive alternative to three-dimensional tracking.

pixel time	image size 100 × 100	200 × 200	500 × 500	1000 × 1000
250 ns	400 Hz	100 Hz	16 Hz	4 Hz
500 ns	200 Hz	50 Hz	8 Hz	2 Hz
1000 ns	100 Hz	25 Hz	4 Hz	1 Hz
2000 ns	50 Hz	12.50 Hz	2 Hz	0.50 Hz
4000 ns	25 Hz	6.25 Hz	1 Hz	0.25 Hz

Table 4.1: the frame-rate versus image size and sample time of the DISS-5 imaging system

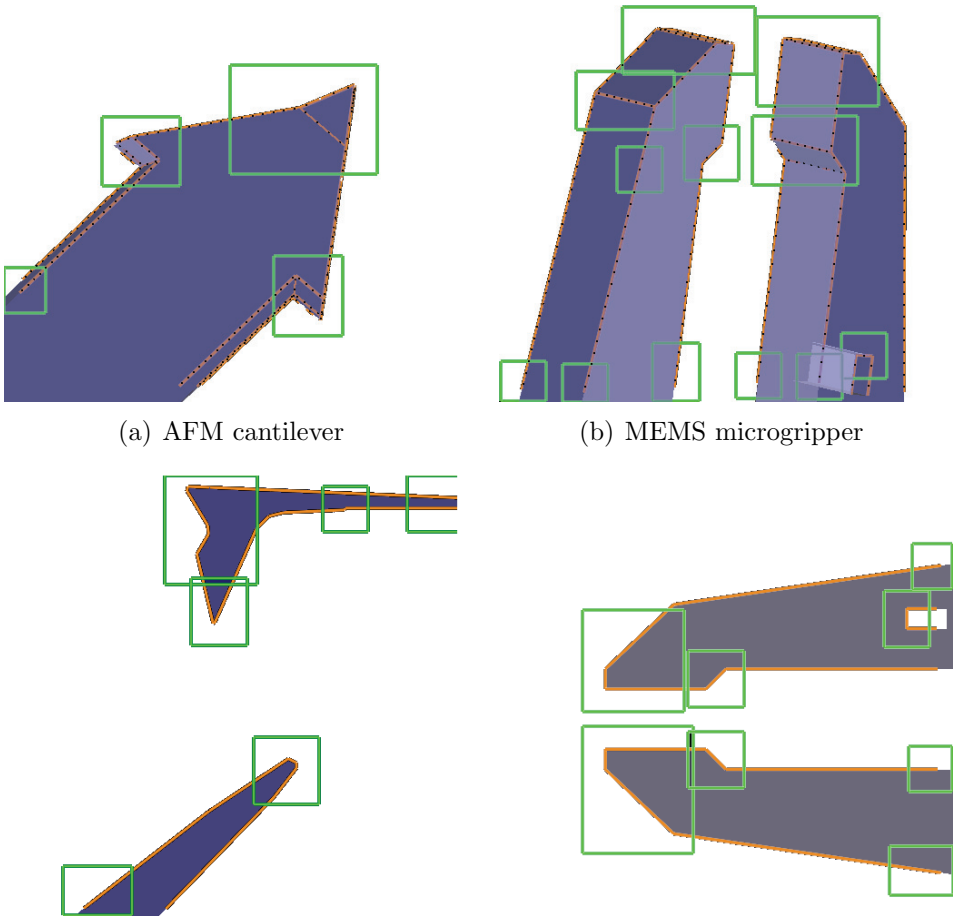
4.6 Region-of-Interest Tracking

As shown in Table 4.1, the image size has a significant influence on the achievable frame rates of any tracking system used in the microscope. It becomes desirable then to reduce the number of pixels sampled while retaining the information content of the image. The image generation process of the SEM lends itself to selectively scanning different regions-of-interest. These regions can be statically chosen at the beginning of an experiment or preferably dynamically adjusted at runtime to fit the task at hand.

In the case where J has full column rank, the least-squares solution to the pose estimation can be solved using equation 4.11. In this case, the vector of fitted values $\hat{\mathbf{d}}$ corresponding to the observed values $\mathbf{d}$ is

$$\hat{\mathbf{d}} = J\hat{\alpha} = J\left(J^T J\right)^{-1} J^T \mathbf{d} \tag{4.37}$$

The $m \times m$ matrix $\mathbf{H} = J(J^T J)^{-1} J^T$ is often called the hat matrix, and maps the vector of observed values into a vector of fitted values [Montgomery, 2005]. The hat matrix can be used to identify influential observations based on their effect on the fitting operation. The elements $\mathbf{H}_{ij}$ can be interpreted as the amount of leverage exerted by d_j on $\hat{d}_i$. By inspecting the values of $\mathbf{H}$, we can determine points that are more influential due to their spatial position. We focus on the diagonal elements $\mathbf{H}_{ii}$, which have an average value of n/m, where $n = 6$ for $SE(3)$. Heuristically, we can select any point that is greater than $\lambda n/m$, where λ is a user chosen threshold, as a point of high



(a) AFM cantilever (b) MEMS microgripper

(c) AFM cantilever and probe (d) MEMS microgripper

Figure 4.5: automatically selected regions-of-interest with a threshold $\lambda = 1.2$

importance and use it as the center of a region-of-interest. Alternatively, a target frame rate can be set and λ dynamically chosen such that a given frame rate is achieved. Since multiple points in a region of the tracked object may meet this criteria, the selected points are clustered to reduce the amount of area to be scanned multiple times. The results of the automatic region-of-interest selection are shown in Figure 4.5 for multiple tracking objects. This automatic region selection is then added into the imaging loop. After each pose estimation, a new set of regions is dynamically calculated and

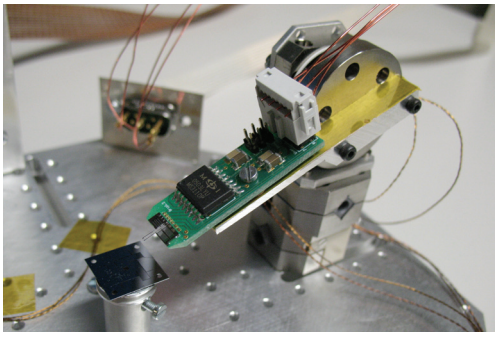


Figure 4.6: Attocube manipulator with FemtoTools FTG-30 gripper

subsequently sampled from the SEM.

4.7 Experimental Tracking Performance

The following experiments were performed with a Zeiss DSM 962 scanning electron microscope. The manufacturer states that this microscope has approximately 10 nm resolution. The image acquisition is handled by a DISS-5 scan controller by Point Electronic GmbH. For more information please refer to Appendix A. In the subsequent examples, an image size of 500×500 was selected to allow for a large viewing area and the possibility for sampling at a real-time rate (from 4 to 16 Hz). Unless otherwise mentioned, the data presented is for images acquired with a pixel sampling time of 500 ns, or equivalently 4 Hz.

A MEMS microgripper by FemtoTools GmbH (shown in Figure 4.6) and an AFM cantilever have been selected as the tracking target and will be used as the baseline object for our tracking evaluation. These devices allows for tracking experiments at a variety of magnification factors. The microgripper also provides an opportunity to apply the constraints discussed in Section 4.4. The tracking targets were attached to a serial link micromanipulator based on Attocube Systems' single degree-of-freedom stages. These devices are based on piezoelectric stick-slip actuators and can be configured in a variety of ways depending on the task at hand. The stages are used in open-loop mode and provide approximately 50 nm and 0.001° minimum step sizes depending on the actuation mode. In the current tracking tasks, the devices have been

configured in a PPR configuration.

4.7.1 $SE(3)$

The algorithm's precision in $SE(3)$ was determined by tracking a static target over a series of 100 frames and observing the variance of the resulting poses. In the following experiments the search length was set to 16 pixels, the multiple image and edge strength saliency metrics enabled, and the measurements with largest 5% of displacement were classified as noise and removed from the fitting process.

The uncertainty of the pose is determined by both the point matching errors as well as the directional-dependent sensitivity, or resolvability [Nelson and Khosla, 1996], of the system. As expected, Figure 4.7 clearly shows that the tracker's precision is significantly decreased along the optical axis of the system.

The influence of using constraints on precision was determined by tracking one gripper arm in the same sequence of images both with and without

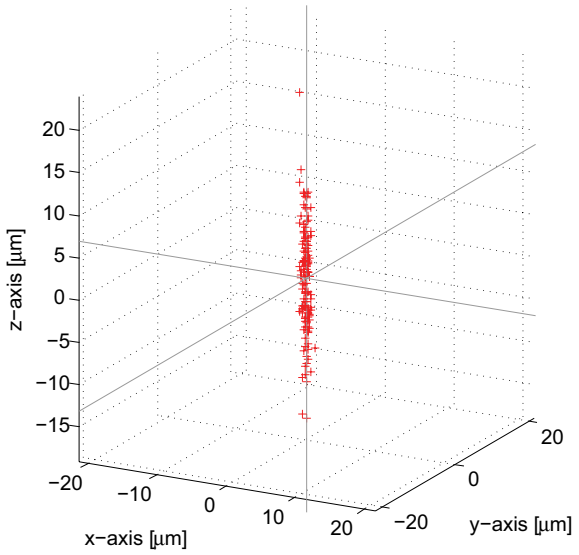


Figure 4.7: The position estimates from 100 iterations of a static microgripper at magnification 500× and pixel sample time 500 ns. The error along the z axis coincides with the optical axis of the SEM and is the most prone to noise.

	σ_x	σ_y	σ_z
unconstrained	0.102 µm	0.279 µm	40.686 µm
multiple object	0.094 µm	0.081 µm	28.105 µm
optical axis	0.076 µm	0.047 µm	0.002 µm
multiple object + optical axis	0.035 µm	0.037 µm	0.000 µm
	σ_ϕ	σ_β	σ_γ
unconstrained	0.002 rad	0.009 rad	0.001 rad
multiple object	0.002 rad	0.001 rad	0.001 rad
optical axis	0.004 rad	0.001 rad	0.001 rad
multiple object + optical axis	0.004 rad	0.001 rad	0.001 rad

Table 4.2: the standard deviation of a single gripper arm in constrained and unconstrained tracking modes with a pixel size of 0.229 µm

	pixel size	σ_x	σ_y	σ_z
200×	1.143 µm	0.112 µm	0.475 µm	0.004 µm
500×	0.547 µm	0.103 µm	0.056 µm	0.003 µm
1000×	0.229 µm	0.058 µm	0.034 µm	0.002 µm
	pixel size	σ_ϕ	σ_β	σ_γ
200×	1.143 µm	0.008 rad	0.025 rad	0.000 rad
500×	0.547 µm	0.010 rad	0.023 rad	0.001 rad
1000×	0.229 µm	0.004 rad	0.003 rad	0.001 rad

Table 4.3: the standard deviation of the tracked gripper pose at different magnifications.

constraints. The constraints between the gripper arms were defined as described in Section 4.4.3. When running the constrained tracker, the distance between points was increased so that the total number of points being tracked matched that of a single arm at ~ 480 points. The results are shown in Table 4.2 and show a significant improvement in x- and y-axis precision when the constraints for z-axis motion in the camera frame are applied. An additional improvement is shown when both gripper arms are used in the tracking process. This is most likely due to the increased separation of feature points caused by the addition of the second gripper. These additional points cause the entire estimation problem to be mathematically better conditioned than when the points are physically close. Due to the improved performance of

	σ_x	σ_y	σ_z
∼120 points	0.075 µm	0.063 µm	0.001 µm
∼240 points	0.055 µm	0.043 µm	0.001 µm
∼480 points	0.035 µm	0.037 µm	0.000 µm
∼960 points	0.036 µm	0.024 µm	0.000 µm

Table 4.4: the standard deviation of the Cartesian coordinates versus number of sampled points with a pixel size of 0.229 µm

algorithm using constraints, the remainder of the experiments utilize both the multiple object and optical axis constraints.

The tracking precision at a number of different magnifications was analyzed and the results shown in Table 4.3. In every case, the standard deviation of the x and y positions were below the effective pixel size of the image. At all magnifications, the rotational variance around the x, y, and z axes (σ_ϕ, σ_β, and σ_γ, respectively) remained similar. This is due to the fact that translational errors scale based on magnification, while rotational ones do not. It should also be noted that when switching between 200× and 500× magnification a large decrease in the standard deviation along the y-direction. This highlights the influence of the object geometery on the tracking precision of the system. At 200× magnification, the geometry of the gripper was such that there were very few edges visible orthogonal, or nearly so, to the y-axis. This means that the tracking problem was mathematically ill-posed along this axis. With higher magnification, more edges were visible in this orientation and thus tracking performance improved.

The effect of the number of feature points on the pose estimation was analyzed as shown in Table 4.4. The performance improves with the number of, but plateaus after a sufficient number is reached to dampen image noise. The rotational standard deviations are not shown, but all varied similarly between 0.001 rad and 0.003 rad. Table 4.5 demonstrates that sampling rate has a significant influence on the tracking precision. Similar to the number of feature points though, this influence tapirs off and provides a useful metric for evaluating the minimum sampling time for tracking.

	250×		500×	
sample rate	σ_x	σ_y	σ_x	σ_y
16 Hz	0.950 µm	8.429 µm	0.682 µm	0.481 µm
8 Hz	0.180 µm	0.215 µm	0.243 µm	0.315 µm
4 Hz	0.076 µm	0.107 µm	0.093 µm	0.097 µm
2 Hz	0.106 µm	0.109 µm	0.056 µm	0.088 µm
1 Hz	0.051 µm	0.120 µm	0.021 µm	0.063 µm

Table 4.5: the standard deviation of the tracked gripper pose at different sampling rates

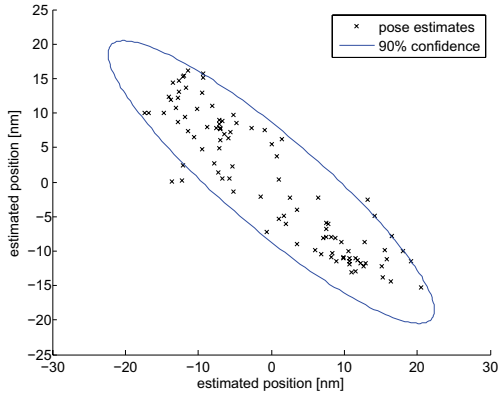


Figure 4.8: results of tracking a static AFM cantilever for 100 frames

4.7.2 $SE(2)$

The algorithm's precision was determined by tracking a static target over a series of 100 frames and observing the variance of the resulting poses. In the following experiments the search length was set to 10 pixels, the multiple image and edge strength saliency metrics enabled, and the measurements with largest 2% of displacement were classified as noise and removed from the fitting process. The system was allowed to converge on a pose for 20 turns before data was stored. Figure 4.8 shows the subsequent results of tracking 100 different frames which resulted in standard deviations of 10.4 nm and 9.56 nm in the x and y directions, respectively. The pixel size of the images used was 222 nm, which indicates that the tracking precision was better than $1/20^{\text{th}}$ of a pixel.

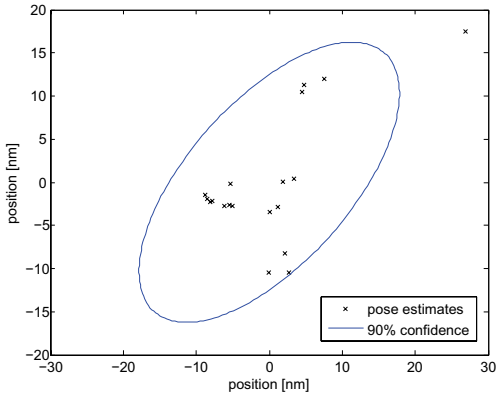


Figure 4.9: results of tracking a single image with different starting poses

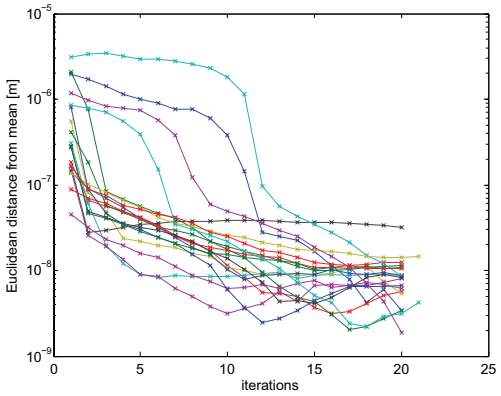
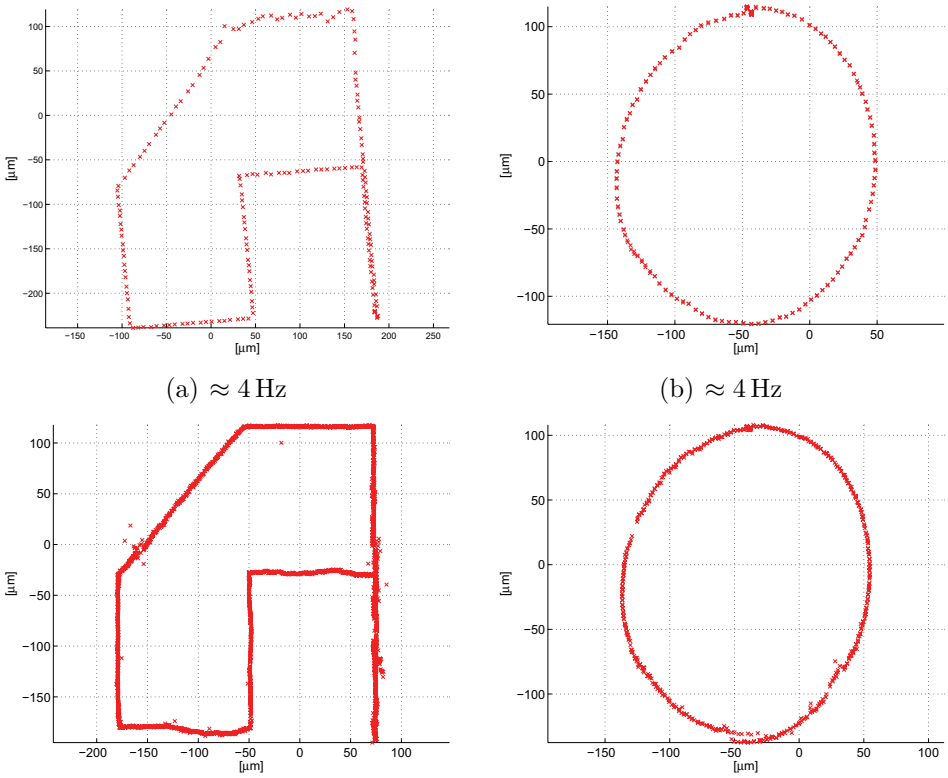


Figure 4.10: convergence of different starting poses over multiple iterations

Although the least-squares estimation step is unbiased, the fact that each new pose is estimated relative to the previous one suggests that there could be some bias in the estimation process. To investigate this, a single image from the previous precision experiment was tracked with different starting poses. These positions were taken from a uniform distribution on the intervals $\pm 3\,\mu\text{m}$ and $\pm 0.05\,\text{rad}$. Figure 4.9 shows the results of 20 different initial poses after 20 iterations. The resulting standard deviations of 8.3 nm and 7.6 nm are similar in magnitude to the precision of the system as previously mentioned, indicating that the tracking algorithm is not largely biased by the initial estimates.

It then becomes interesting to analyze the number of iterations until the system reaches its precision limit. Figure 4.10 shows the convergence of the



(a) ≈ 4 Hz (b) ≈ 4 Hz

(c) ≈ 20 Hz (d) ≈ 20 Hz

Figure 4.11: These are typical results of the trajectory tracking experiments. Trajectories (a) and (b) were tracked using a full 500×500 image, while (c) and (d) were tracked using automatically generated regions-of-interest. Note the higher density due to the increased frame-rate of the ROI poses.

20 different starting poses on their final estimates. One noticeable feature of this chart is that once the pose estimate approaches a neighborhood of approximately equivalent to 2 pixels in image space from the final pose estimation, the algorithm only requires 2-3 iterations to approach the precision limit. This is partially a result of the linearization step in equation 4.6 as well as the fact that the closer the estimate is to the actual pose, the larger number of edges that can be tracked. The initial poses in the experiment shown in figure 4.10 were intentionally chosen to be a large distance from the true value. When tracking objects that only move a small number of pixels between subsequent frames, the algorithm can rapidly converge on the

target value with as few as one or two iterations. Since the image refresh rate of the SEM for similar imaging operations is on the order of 10 Hz and the current tracking implementation performs in excess of 100 Hz with ∼170 tracked points, as many as 10 iterations can be performed on each new SEM image which can aid in tracking faster moving objects.

4.7.3 Region-of-Interest

To better understand the algorithms performance in both full-frame and region-of-interest tracking scenarios, two test trajectories were generated. The rectangle trajectory was intentionally designed with a variety of straight motions segments in different orientations to qualitatively evaluate the algorithm's performance. Although the actuators are driven in open-loop mode, the straight lines are generated by a single actuator, which allows us better to seperate the possible actuator noise from that of the tracking algorithm. The circular pattern was chosen because it involves translational motions in all directions and should thus have limited bias based on the geometry of the tracked object. The following experiments were performed with the SE(2) tracker a search length of ±12 pixels along the edge normal and with the tracking constraints between the gripper arms applied.

Figure 4.11 shows sample trajectories tracked with the algorithm at 200× magnification. In all four cases shown, the algorithm successfully tracked the target for the entire run. Qualitatively, it can be seen in the rectangular path, that there is more noise in the pose estimates along the y axis than along that of the x axis. This is primarily due to the orientation of the gripper during the experiment. The gripper geometry is such that there are significantly more feature points along the y axis than the x, which creates for a more precise measurement. While Figures 4.11(a) and 4.11(b) were tracked using the full sensor frame at ≈ 4 Hz, Figures 4.11(c) and 4.11(d) were tracked using the region-of-interest strategy. Aside from significantly denser poses, the tracking precision of the two methods does not appear to differ significantly.

A series of tests were then run on the algorithm using both the full-frame

velocity		full-frame (≈ 4 Hz)	ROI (≈ 20 Hz)
138 µm/s	122 pix/s	0%	60%
135 µm/s	120 pix/s	0%	80%
84 µm/s	74 pix/s	0%	100%
36 µm/s	32 pix/s	100%	100%
28 µm/s	25 pix/s	100%	100%

Table 4.6: the success rate of tracking the circle trajectory at different velocities out of 5 trials each

velocity		full-frame (≈ 4 Hz)	ROI (≈ 20 Hz)
138 µm/s	122 pix/s	0%	40%
135 µm/s	120 pix/s	0%	80%
84 µm/s	74 pix/s	0%	100%
36 µm/s	32 pix/s	100%	100%
28 µm/s	25 pix/s	100%	100%

Table 4.7: the success rate of tracking the rectangle trajectory at different velocities out of 5 trials each

and region-of-interest strategies. In both cases, the gripper was tracked over the trajectory at a series of different velocities. The results are shown in Tables 4.6 and 4.7. As anticipated these results demonstrate that the region-of-interest strategy performs significantly better at high-speed tracking tasks than its full-frame equivalent due to the increased frame rate. During the course of these tracking tasks, the scan regions are dynamically generated at each frame. These regions were selected to provide an anticipated frame rate of 25 Hz based on the sample time at each pixel and the area of the scanned regions. Figure 4.12 demonstrates that while the desired frame rate was not always reached, it did remain constant for the duration of the experiment. This difference is most likely caused by communication overhead to the controller and general processing overhead in the system, and can be easily taken into consideration when choosing the target frame rate.

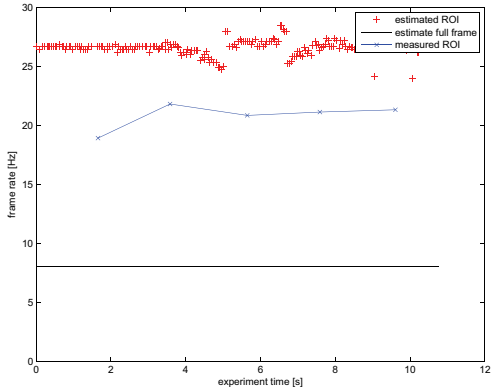


Figure 4.12: estimated and measured frame-rates over a tracking trajectory

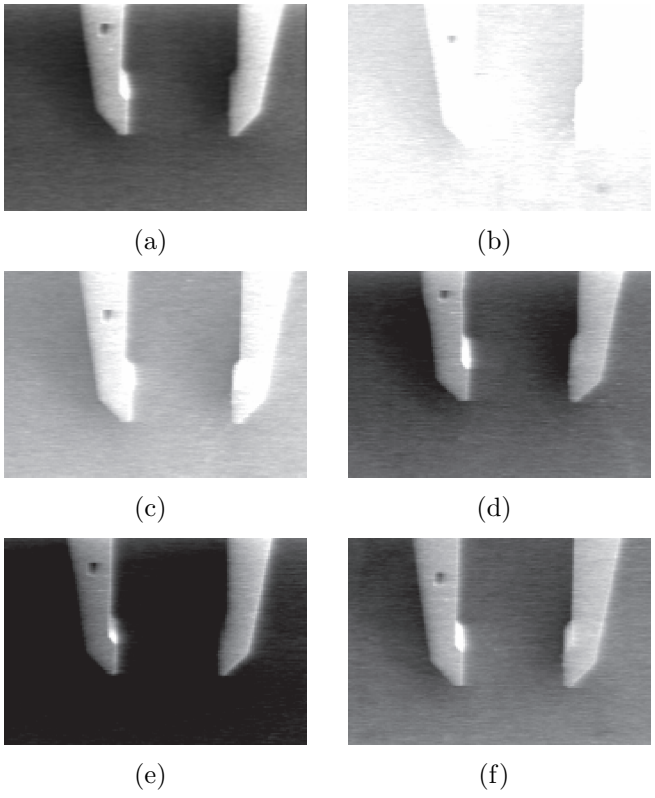
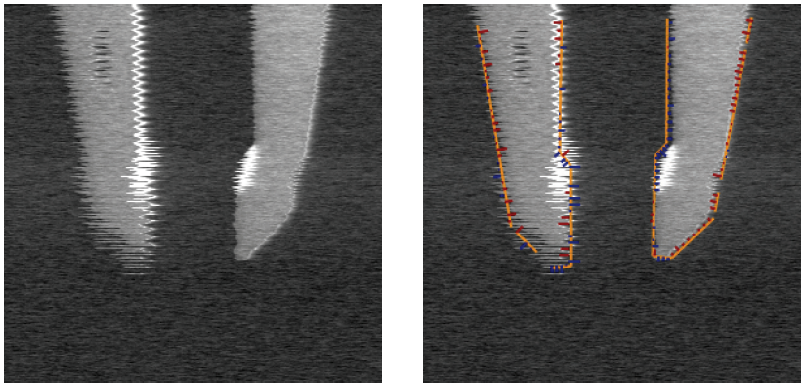


Figure 4.13: These are highlights of noisy images excerpted from a rectangular tracking image series during region-of-interest operation. The system successfully dealt with all of these variations and tracked to the end of the experiment.



(a) high frequency motion (b) tracked pose

Figure 4.14: image distortion caused by high frequency vibration of the gripper due to its low stiffness and the motion of the stick-slip actuator

4.8 Imaging Noise

As mentioned in Section 3, a variety of influences contributes to image noise when working with a scanning electron microscope. Much of the speckle noise seen on an image is due to the nature of detecting the emitted electrons and can be modeled by a Gaussian [Prasad and Joy, 2003]. This noise can often be directly mitigated through a reduction of the sampling rate of the image at the expense of frame rate. Additional noise in the imaging pipeline is caused by charging effects as a manipulator interacts with the environment. This can take the shape of large intensity and contrast variations in the image. Figure 4.13 demonstrates a number of different effects of this charging taken from a single experimental run. Additionally, many of the actuators used for manipulation are based on stick-slip actuation principles, which can cause vibrations in the end effectors during motion. Certain gripper arms have been shown to vibrate when the piezo actuation in the positioning stage excites a system resonance as shown in Figure 4.14. Although the system successfully tracked during these abberations, they pose a serious challenge to the robustness of any visual servoing system and effort should be given to reducing these effects. Even in the best cases, image noise and distortions cannot be entirely eliminated and future systems should consider

hybrid feature and model-based methods [Pressigout and Marchand, 2007] to provide added system stability.

4.9 Example Tasks

4.9.1 Vision-based Force Sensing

Ultra-small force sensing has attracted research interest for the application in chemical sensing [Battiston *et al.*, 2001], bio-sensing [Merkel *et al.*, 1999] and property characterization of nanomaterials and devices. To measure ultra-small forces, a direct method is to detect the deformation of an elastic element such as an AFM cantilever or a spring with known stiffness. The resolution of the measurement is restricted by the stiffness of of the elastic element and the precision of the deformation measurement. The most flexible commercially available AFM cantilever has a spring constant on the order of 10^{-2}N/m Though in laboratories AFM cantilevers have reached a spring constant on the order of 10^{-6}N/m [Stowe *et al.*, 1997], the necessitation of ultra-thin silicon cantilevers (60 nm thick) strongly limits their practical applications. Continuous improvements are still being made in probe geometry, material, and the sensing mechanisms.

On the other hand, three-dimensional helical structures with nanofea-

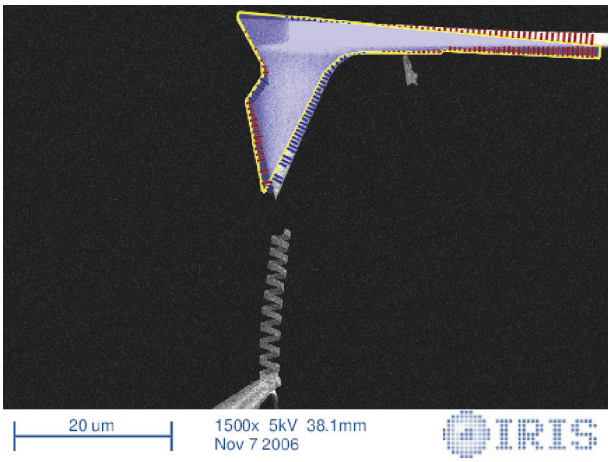


Figure 4.15: SEM image of an AFM cantilever about to be tracked with model overlay

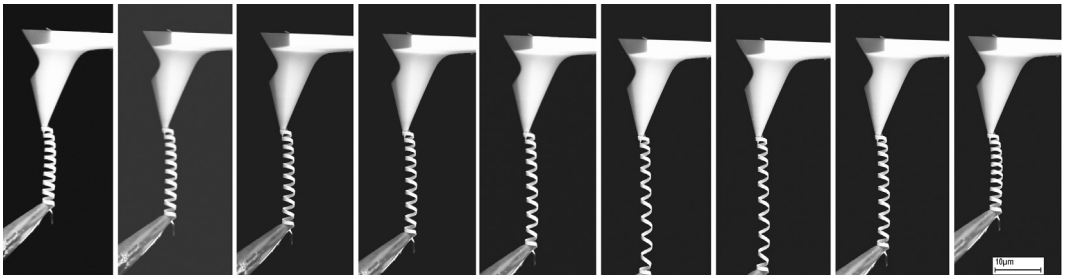


Figure 4.16: pulling the coil for stiffness characterization

tures, such as helical carbon nanotubes [Amelinckx *et al.*, 1994], coiled zinc oxide nanobelts [Gao *et al.*, 2005] and rolled-up semiconductive nanostructures [Prinz *et al.*, 1999], have attracted intensive research interest [Cho, 2006] because of their potential applications in nanoelectromechanical systems (NEMS) as springs, electromagnets, inductors, resonators, sensors, and actuators [Bell *et al.*, 2006; Zhang *et al.*, 2006]. Shrinking device size to these dimensions presents many fascinating opportunities such as measuring mass in femto-gram (fg) ranges, sensing forces at pico-Newton (pN) scales, and inducing giga-hertz (GHz) motion, among other new possibilities waiting to be discovered. As ultra-small force sensing elements for applications in chemical sensing, bio-sensing and property characterization of nanomaterials and devices, these helical nanobelts are expected to have ultra-flexibility and a wide linear range. Though the intrinsic piezoelectricity or piezoresistivity is potentially a mechanism for their deformation readout, it is strongly material dependent.

Vision-based force measurement [Greminger and Nelson, 2004] has the advantage that it can be used with existing elastic elements. Using this methodology, the deflection measurements are performed in conjunction with devices such as optical microscopes, scanning electron microscopes [Dong *et al.*, 2004], or transmission electron microscope [Kis *et al.*, 2006] which all provide sensory feedback in the form of a monocular image making visual methods an attractive data acquisition method.

Recently, Zhang *et al.* have demonstrated that small-pitched nanohelices can be fabricated by anomalous scrolling of strained SiGe/Si bilayer based

heterostructures due to the relaxation of the stress at the edges of narrow stripes [Zhang *et al.*, 2006]. When the edge effect dominates the coiling process of these stripe-like patterned thin films, the fabrication of small-pitched SiGe/Si and Cr coated SiGe/Si nanohelices with pitch angles less than 10° can be achieved in a highly controllable fashion. The reduction in helical nanobelt pitch also increases their flexibility. Here we describe an experimental investigation of the mechanical properties of SiGe/Si/Cr nanohelices using nanorobotic manipulation in 3D free space.

Figure 4.16 shows a typical experiment on an individual $Si_{0.6}Ge_{0.4}Si/Cr$ nanospring with layer thickness of 11/8/21 nm. The nanospring was first picked up using the manipulator using a sticky tungsten probe (Picoprobe, $T-4-10-1\,mm$) as an end effector. The probe was prepared by dipping the tip into a double-sided SEM silver conductive tape. The free end of the spring was then clamped onto an AFM cantilever (Mikromasch, CSC38/Ti-Pt, nominal stiffness 0.03 N/m) using electron-beam-induced deposition (EBID). The spring was loaded under tension and released for several cycles.

Applying the rigid-body tracking routine to the images from these experiments, the deformation of the spring and the AFM cantilever has been obtained. The load was then calculated according to the deflections of the

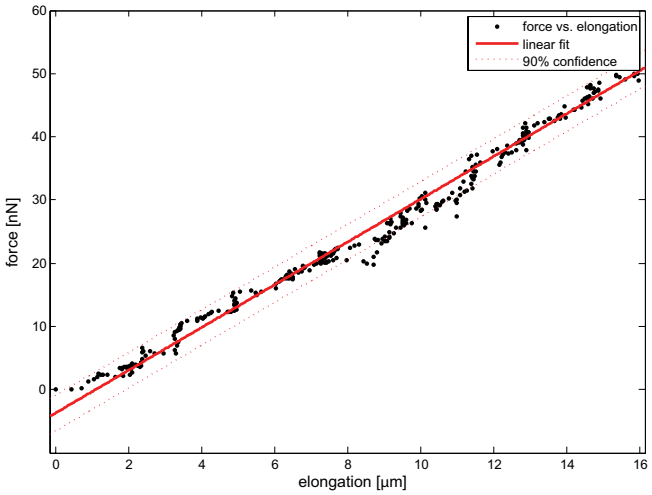
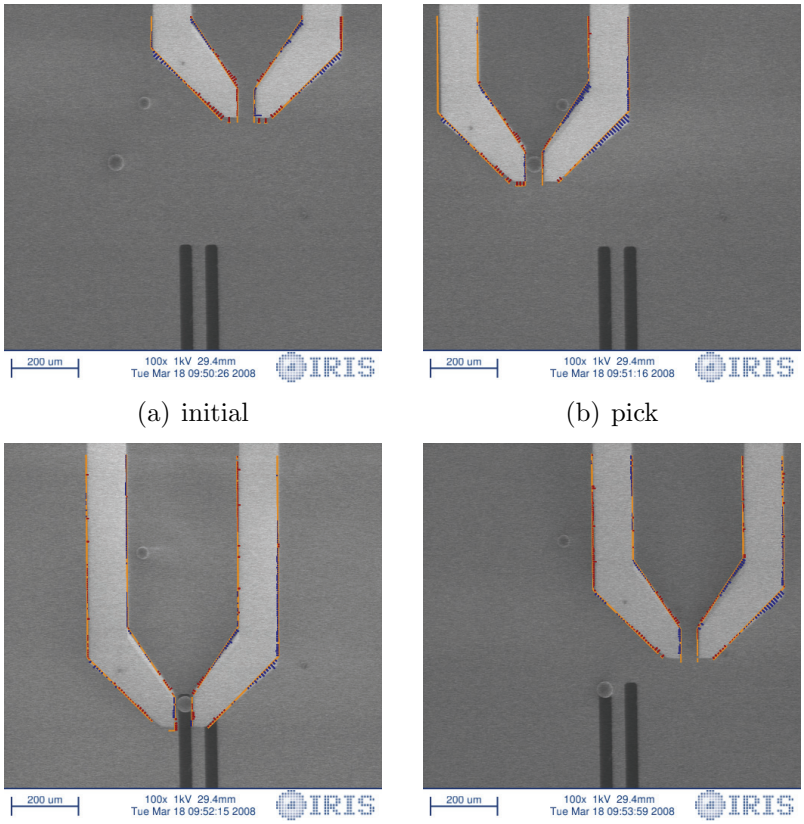


Figure 4.17: automated acquisition of elongation load curve of N series coil



(a) initial (b) pick

(c) placement (d) release

Figure 4.18: tracking during a manual pick and place of a 30 µm glass sphere

AFM cantilever (calibrated stiffness: 0.038 N/m). Figure 4.17 shows the results along with the 90% confidence interval of the fit parameters acquired from a single experiment where a load was applied and released. The data was acquired from images at 512×384 resolution.

Using it as a visual-based force sensor, this nanospring can provide a resolution of 3 pN/nm if an FESEM is adopted as a displacement measurement tool and assuming an imaging resolution of approximately 1 nm. With smaller stripe widths, more turns, higher resolution microscopy, or other readout techniques, nanosprings can potentially provide femto-Newton (fN) to atto-Newtom (aN) resolution. Aside from the wide linear region and ultra-high flexibility, the anomalous rolled-up SiGe/Si/Cr helical nanobelts

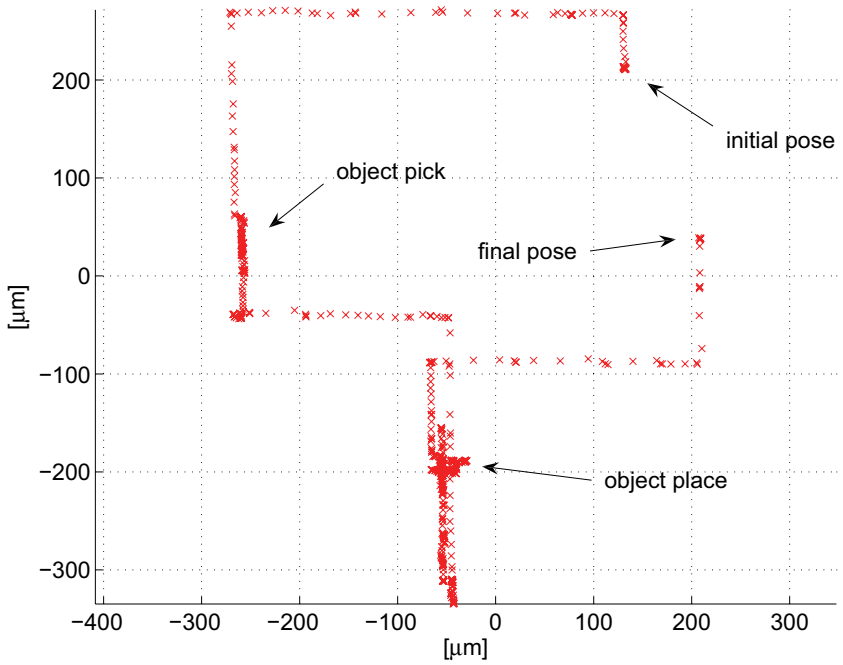


Figure 4.19: tracked poses for duration of pick-and-place task

are promising for batch fabrication and practical application as elastic elements in ultra-sensitive, large-range force/mass sensors because the geometrical parameters such as diameter, chirality, helicity angle, and pitch can be well controlled.

4.9.2 Tracking for Micro-object Handling

The SEM has a number of benefits which make it particularly compelling for microassembly tasks. The magnification capabilities of an electron microscope are significantly higher than those available to their optical counterparts alleviating the precision constraints related to imaging resolution. Microassembly systems must also often deal with limited depth-of-field for tracking, which is not generally a challenge in SEM systems. The major detriments to using SEMs for microassembly systems has been their relatively high cost and requirement of operation in a vacuum. The recent introduction of low-cost SEM imaging systems such as the FEI Phenom, has reduced the cost of systems suitable for microassembly to within that of a

high-quality optical microscope. Principle environmental factors that influence typical microassembly tasks are due to temperature and humidity [Zhou *et al.*, 2002]. These factors are reduced or removed in the vacuum chamber of the SEM, which can also help simplify the task at hand, thus reducing the burden of having to operate in a vacuum. A pick and place scenario was selected as an application for evaluation of the tracking system. Pick and place operations are tasks familiar to the domain of microassembly, and generally consist of transferring parts from an initial carrier to a final assembled location.

Figure 4.18 shows an image sequence where the tracker was applied to a pick-and-place manipulation task. The manipulation task was performed on a micromachined silicon substrate to provide the placement target. The objects manipulated were 30 µm glass spheres, and this particular experimental run took approximately 220 s. In addition to image distortions due to interactions with the environment, this task has the additional challenge of there being other objects in the environment that could distract the tracker from its task. Figure 4.19 displays the resultant poses from the entire manipulation sequence, and qualitatively demonstrates its suitability for such work.

4.10 Discussion

We have developed a rigid-body tracking system for observing objects in a scanning electon microscope and demonstrated its applicability. This system can be configured for two or three-dimensional objects with up to six degrees-of-freedom. To increase the stability and precision of the method, the system allows for constraints to be rapidly put on the tracking problem in an intuitive fashion. This system can be applied in real-time to full-frame tracking problems, and has been demonstrated to perform with sub-pixel precision. On microscopes that allow a user more intimate access to the scan controller, a selective region-of-interest strategy has been implemented to increase the sampling rate of the acquisition process. This method dynamically changes the scan regions based on the task at hand, and has been demonstrated to

provide 3-5 times the frame rate as a full frame acquisition process. Due to the imaging process of the SEM, a variety of image distortions can occur during a SEM observation task due to effects such as sample charging or are resultant of the low refresh rate of the sensor. This system has demonstrated its ability to function correctly in the presence of these distortions. Subsequent efforts for tracking and manipulation in an electron microscope must take these factors into consideration when designing both the system and the task to be achieved. This tool has been evaluated as a data acquisition system for vision-based force sensing and as the first step toward an automated visual servoing system. These tools provide useful steps toward creation of automated or semi-automated nanohandling systems.

5 3D Electron Microscopy

Since the early days of electron microscopy, the acquisition of three- dimensional information through multiple images has been an interest to a variety of researchers. These methods generally rely on observing the sample from different angles and using the relative motion of features to recreate the object. The different viewing angles can be accomplished through either specially designed detection systems [Suganuma, 1985; Jähnisch and Fatikow, 2007] or through the rotation of the sample stage [Ribeiro and Shah, 2006]. In both cases, systems are needed for interpreting the measured results to provide the three-dimensional reconstruction. This information has been used to measure a few feature points, extract line profiles, or for complete three-dimensional reconstruction [Raspanti *et al.*, 2005], as well as being used in conjunction with AFM data for multi-modal sensing [Seeger, 2004]. Before discussing our contribution to the effort, we will briefly discuss the state-of-the-art for image-based three-dimensional data acquisition.

When compared to electron microscopes, optical microscopes have a significantly reduced depth-of-field. Optical microscopists have capitalized on this fact and often rely on this narrow depth-of-field to determine the precise depth of each image acquired. A series of images are then gathered into a z-stack and a three-dimensional viewing volume is generated [Niederoöst *et al.*, 2003]. These image stacks must then be segmented if more quantitative data is desired.

Recent advances in Focused-Ion-Beam (FIB) technology have generated a new field of three-dimensional microscopy. These systems enable a microscopist to physically cut away sections of the sample. This method is similarly used to generate a z-stack of an image by sequentially cutting away a particular sample and observing the cut segment with an electron microscope [Bansal *et al.*, 2006; Groeber *et al.*, 2006].

Transmission electron microscopes are often used for generating tomographic reconstructions of samples. In these systems, the samples are moved through a series of different tilt angles. These images are then aligned and projected into a three-dimensional Fourier space. This space is then

re-sampled and projected back into two-dimensional images to once-again generate a z-stack of images [Lučić *et al.*, 2005; Fernandez *et al.*, 2006; Ribeiro and Shah, 2006]. The nature of this process requires precise image alignment as well as an equal sampling over all the possible viewing angles. The physical limitations of observing from high incidence angles limits the effective tilting of a sample to approximately $\pm 70^\circ$ for a single tilt axis, leading to the missing conic in Fourier space and effectively a limit to the resolution of the reprojection. To improve this limitation, dual-tilt tomographic systems have been created that allow for a second tilt of the image along an axis perpendicular to the first, which reduces the size of the missing conic. This second axis comes at a cost of either expensive hardware and electronics or of physically removing and rotating the sample 90° before re-imaging. In the latter case, it can be time consuming and difficult to locate a precise feature again.

Critical to this acquisition process is precision of the rotation and the alignment of the sample. For eucentric rotation, the system must be able to rotate around an arbitrary location on the sample. This is typically provided for by a kinematically calibrated macro-scale system which attempts to provide high precision motions. In spite of high-performance actuation stages, errors necessarily occur and generally tomographic systems must be re-aligned based on the observed images. In past times, this was manually done by a technician and was very time consuming. More modern systems allow the use of visual tracking systems such as cross-correlation and feature-based methods [Ribeiro and Shah, 2006]. One significant challenge for the automation of these systems is the low signal-to-noise ratio of many of the images as well as the propensity of certain samples to degrade or deform under the electron beam. This requires systems not only to track the motions, but to track deformations as well.

The remainder of this thesis is dedicated to addressing some of the challenges listed above. In particular, we investigate feature-based methods for use with generating three-dimensional data from a series of SEM images. As opposed to using purely stereogrammatic methods which generally focus on either a pair or triple of images, we focus on large series of images. A larger

number of viewing perspectives improves the system resolvability which improves the overall precision [Nelson and Khosla, 1996]. Due to the increased difficulty in detecting features at higher incidence angles, this necessitates methods which allow for features in images to be occluded and new ones added to the feature matrix. Due to the desire for high-precision measurements methods will be analyzed which allow for accurate estimation of the image acquisition position as well as the features of interest. We also examine data acquisition with a novel rotational actuator which enables localized rotations at the micro- and nanoscales. This actuator has the potential to alleviate the challenge of designing high-precision rotation stages and possible applications in dual-tilt tomographic systems.

5.1 Structure-from-Motion

Structure from motion is the generic term given to a class of computer vision problems that generate three-dimensional information from a series of two-dimensional images that are taken in the presence of motion. In macroscale robotic applications, this often takes the form of a mobile observer, while in the case of electron microscopy the sample necessarily must move due to the bulk of the electron column. Although this may appear to be a semantic delineation, there are real constraints on the suitable motions for three-dimensional reconstruction [Sturm, 2002].

Central to this problem is the notion of matrix factorization. In this case, measured data such as feature points are recorded over multiple image frames. These features can be added to a $2m \times n$ measurement matrix M where the $2m$ rows consist of the recorded features locations for each of the m frames and the n columns denote the values for each individual feature. This matrix can be factored into the desired information of the structure $B \in \mathbb{R}^{2m \times r}$ and motion $A \in \mathbb{R}^{n \times r}$. If all the points are observable at each frame, then the problem can be written as:

$$M = AB^T \tag{5.1}$$

In real situations, the matrix M is not known and its noise-corrupted version

is represented by $\tilde{M}$. The goal then is to find estimates A and B such that

$$\min ||\tilde{M} - AB^T||_F^2 \tag{5.2}$$

where $||\cdot||_F$ denotes the Frobenius norm [Buchanan and Fitzgibbon, 2005].

Tomasi and Kanade noted that with all features present under orthography, the measurement can be centered making it a rank three problem which can be optimally solved up to an affine transformation using singular value decomposition (SVD) [Tomasi and Kanade, 1992].

$$\begin{aligned} \tilde{M} &= U\Sigma V^T \\ \hat{A} &= U\Sigma^{1/2} \\ \hat{B} &= \Sigma^{1/2} V \end{aligned} \tag{5.3}$$

Although this provides us with an estimate for the structure and the motion, the true values are still unknown up to an affine transformation because of the possibility of $Q_{3\times 3}$ such that

$$\tilde{M} = \hat{A}(QQ^T)\hat{B}^T \tag{5.4}$$

One further step is now required called a Euclidean upgrade. In this we can observe that in the final motion matrix there must be a rotation matrix, meaning that for each frame the rows must be magnitude 1 and orthogonal to each other. This problem is classically solved by methods such as those proposed by [Quan, 1996], or newer methods such as [Guilbert *et al.*, 2006], leaving us with the final estimates for structure and motion as:

$$\begin{aligned} A &= \hat{A}Q \\ B &= Q^T\hat{B} \end{aligned} \tag{5.5}$$

These methods work well for small sets of images where features are visible in all images, but in larger data sets this is often rarely the case. A more

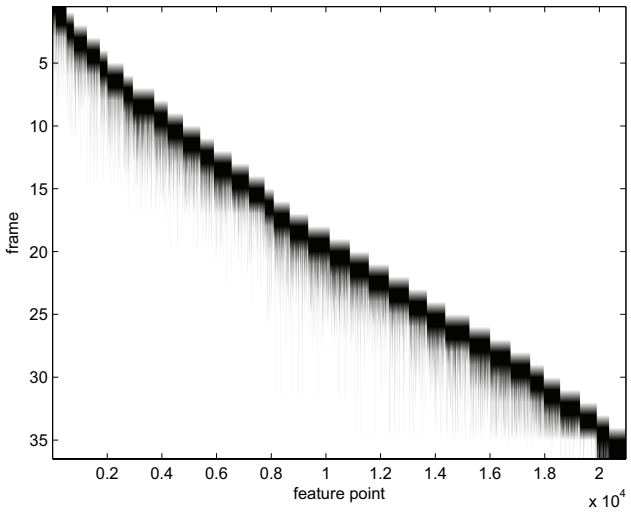


Figure 5.1: visibility matrix of the pollen image series

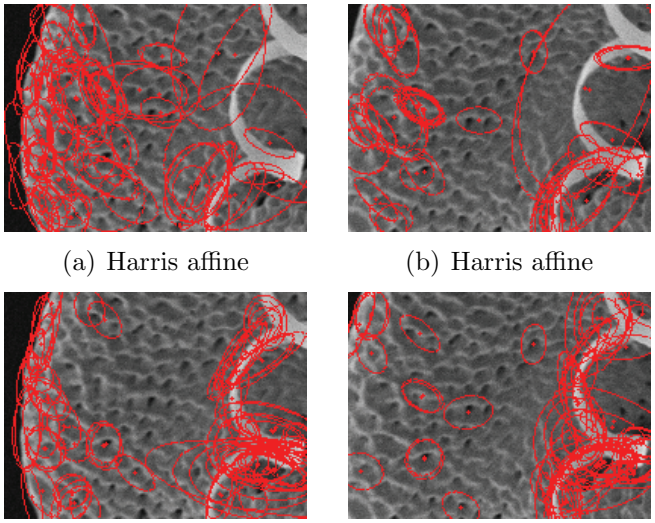
appropriate formulation of equation 5.2 is then

$$\min ||W \odot (\tilde{M} - AB^T)||_F^2 \tag{5.6}$$

where $\odot$ is the Hadamard product[1]. The matrix W is generally a sparse matrix with a roughly block diagonal layout. Figure 5.1 demonstrates one such matrix where the rows represent frames and the columns represent individual features. As the image series progresses, new features are added and old features are lost due to occlusions.

With equation 5.6 firmly in mind, we can now delineate the structure-from-motion problem into two subproblems which are often treated separately in the literature. In feature matching, the goal is to match corresponding points between multiple images to generate the measurement matrix $\tilde{M}$, while that of three-dimensional reconstruction consists of methods for interpreting matrix $\tilde{M}$ to determine Euclidean structure and motion.

[1]the entrywise product such that $R = P \odot Q \Leftrightarrow r_{ij} = p_{ij}q_{ij}$



(a) Harris affine (b) Harris affine

(c) Hessian affine (d) Hessian affine

Figure 5.2: example regions detected by different detectors

5.1.1 Affine Feature Tracking

While an exhaustive discussion of feature tracking algorithms is beyond the scope of this thesis, we would be remiss not to mention in brief the challenges and potential solutions to this topic. For a more thorough survey of different methods, we refer the reader to [Schmid *et al.*, 2000; Mikolajczyk *et al.*, 2005; Mikolajczyk and Schmid, 2005]. To focus the discussion, we define a region as a subset of pixels from a particular image. A feature can then be defined as a metric representation of that region that describes it in an invariant way.

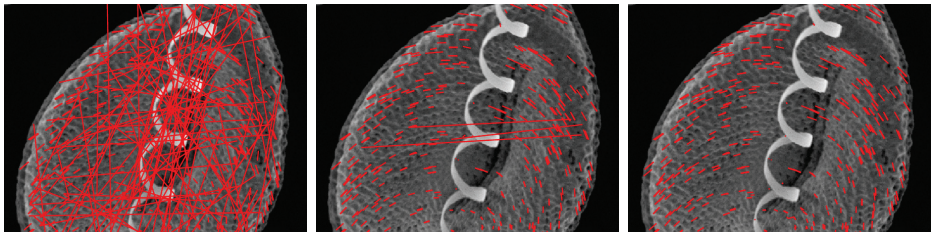
A typical feature matching operation consists of first identifying the regions-of-interest in the images to be analyzed. An image can contain anywhere from a few features up to thousands, depending on the task at hand. Each of these features is then analyzed and feature descriptor vectors are created which represent the regions. The descriptors can be tailored to meet a variety of different operating objectives, but in general the metrics they use to describe the regions are chosen to be invariant to viewpoint changes. After these descriptors are independently calculated for the pertinent images, they are matched to determine correspondences. Working within this framework,

the principle questions then become: "How do we select these regions?" and "How do we determine features that allow subsequent matching?"

A variety of different strategies exist in the literature for selecting regions. One of the detectors favored among the literature is the Harris detector [Harris and Stephens, 1988]. This technique computes the local image derivatives, which are then averaged using a Gaussian window. The eigenvalues of this matrix are then analyzed to determine if the signal change in the neighborhood is significant. This method has been used for a number of years and has been shown to provide dense, reliable interest points. The Hessian matrix can be similarly used for detection, but has been shown to vary with the scale of the feature. To improve the performance of both detectors, a scale selection method can be implemented which calculates the size of a region independently for each point [Lindeberg, 1998]. An affine region is then selected around the feature point by looking at the eigenvalues of the second moment matrix [Lindeberg and Garding, 1994] or the dominant gradient orientation [Lowe, 1999]. The detected regions can generally be represented by an ellipse and some examples are shown in Figure 5.2.

Mikolajczyk *et al.* do a comprehensive review of the state-of-the-art for feature detectors [2005], and a thorough evaluation of their performance in a variety of scenarios. As is often the case with computer vision methods, there is no single methodology that consistently outperforms the others in all cases. They did find that the Hessian affine detectors performed quite well in many of the cases, and that both the Hessian affine and Harris affine methods provide more regions than the other detectors. The latter result is of paramount import to this work due to the desire of finding dense mappings of features. The performance of all region detectors evaluated declines at roughly the same rate as the change of viewpoint angle increases. They also found that the various methodologies often found disjoint sets of features due to their affinity for different types of features. This implies that multiple detectors can be used with minimal overlap and potentially denser features, which we will examine in the experimental section.

Once the regions are selected, a variety of options exist for calculating the feature descriptors and matches. These include descriptors such as normal-



(a) unfiltered (b) RANSAC (c) heuristic

Figure 5.3: the different stages of matched feature point filtering

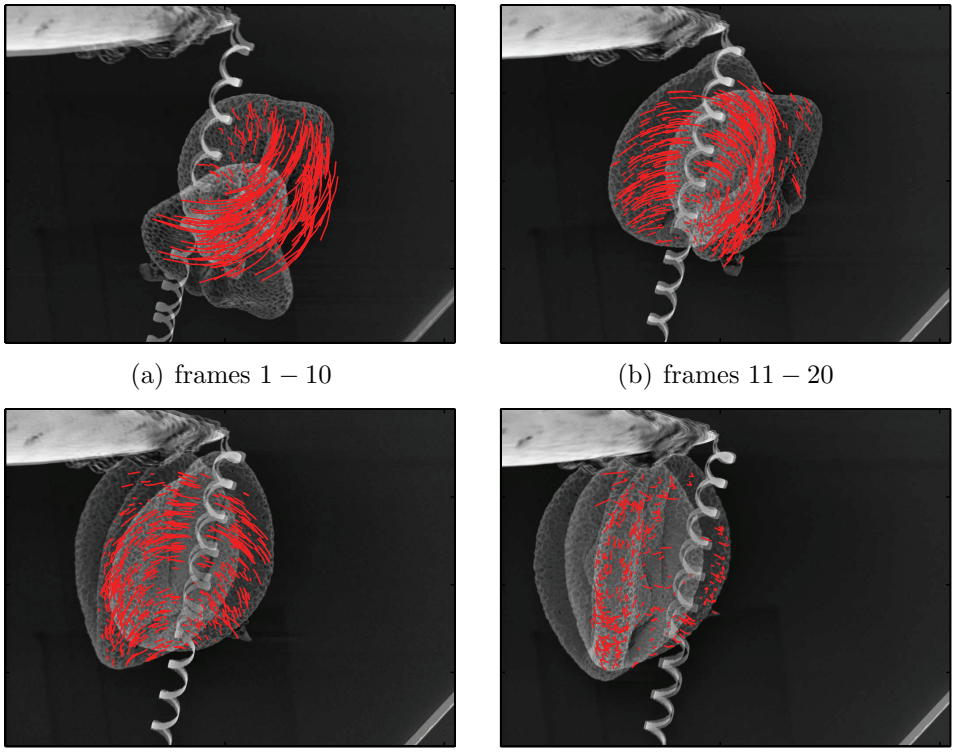
ized cross-correlation [Stefano *et al.*, 2003], moment invariants [Gool *et al.*, 1996], steerable filters [Freeman and Adelson, 1991], shape context [Belongie *et al.*, 2002], and SIFT [Lowe, 2004]. Due to their superb performance in a variety of conditions, we elected to use the SIFT feature descriptors for subsequent operations [Mikolajczyk and Schmid, 2005].

The calculation of the Scale Invariant Feature Transform (SIFT) descriptor involves first computing the gradient magnitude and orientation for each image sample point in the region. The samples are then partitioned into a $n \times n$ subregion, and histograms are calculated for each of the regions. Lowe experimentally determined that using a subregion where $n = 4$ and the histograms are binned into 8 orientation bins works well, yielding a 128 element feature vector. To improve robustness to image intensity changes, the vectors can be thresholded and normalized to unit length. This method has been additionally extended to yield the Gradient Location-Orientation Histogram (GLOH) by [Mikolajczyk and Schmid, 2005]. This method changes the region grid into a log-polar grid with 3 bins in the radial direction and 8 in the angular one. These additional bins creates a 272 bin histogram which necessitates the use of principle component analysis (PCA) to reduce the descriptor size to 128. Instead of using the gradient histograms, the gradient of the region can be directly used as discussed by [Ke and Sukthankar, 2004]. In this method, PCA is applied directly to the gradient of the sample patch to generate the descriptors. In all of the above SIFT cases, the Euclidean distance is used to calculate the metric between different features.

Even with state-of-the-art feature matching methods errors will occur

in the process, which can cause significant problems in the reconstruction phase. Heuristic methods have been shown to work well in filtering these errors at low computational cost, and have been employed at different stages during the process. Following the initial matching task, the feature points are filtered using the RANSAC method for estimating the fundamental matrix between two views [Fischler and Bolles, 1981; Torr and Murray, 1997]. The fundamental matrix, F, is a 3×3 matrix which relates the corresponding points between two images. If m_1 and m_2 represent homogeneous coordinates in two separate images, the epipolar constraint on this system dictates that $m_2^T F m_1 = 0$. The matrix F is rank two, and can be estimated given at least seven correspondences [Faugeras, 1993]. With large sets of matches, the redundancy of information can be exploited to reduce the false matches. The RANSAC method accomplishes this by randomly selecting a subset of the data to estimate a fundamental matrix. The remainder of the points are compared to determine if they are consistent with this solution. If a sufficient number of inliers are found, the entire set is used to compute a new model estimate. If the error of the inlying points relative to the model is sufficient, the points are selected. Otherwise, the system starts with a new random selection and repeats the process. Figure 5.3(b) demonstrates the results of this filtering step.

Although the fundamental matrix filter does an excellent job of removing false matches, some additional steps can help further refine the data. If matching with no assumptions as to the motion of the camera between frames, the feature matching algorithms must work across the entire image. In some cases, false matches occur which are consistent with the estimated fundamental matrix of the image pair. These cases often manifest themselves with a Euclidean distance in image space significantly larger than their correct counterparts. To remove these points, we have implemented a simple distance filter. Experimentally, a distance threshold of 2-3 times the standard deviation of matched distances worked well for this task (Figure 5.3(c)). The output from the entire feature detection process (Figure 5.4) is then aggregated into the feature matrix $\tilde{M}$ and passed on to the 3D reconstruction stage.



(a) frames $1 - 10$ (b) frames $11 - 20$

(c) frames $21 - 30$ (d) frames $31 - 37$

Figure 5.4: example of matched feature point trajectories between multiple frames

5.1.2 3D Reconstruction

As previously mentioned, if there is no missing data structure-from-motion problems can be efficiently and optimally solved using methods such as those proposed by Tomasi and Kanada [Tomasi and Kanade, 1992] for both orthographic and perspective images. Unfortunately, the real-world tends to have other factors such as occlusions, errors in feature detections, and limited fields-of-view that make the unaltered application of these methods untenable. Thus, much of current stereometric and structure-from-motion research focuses on methods for robustly identifying and handling occluded or missing features [Brown *et al.*, 2003; Buchanan and Fitzgibbon, 2005].

The methods for matrix factorization can be roughly divided into three categories: hierarchical, iterative, and batch. Hierarchical methods divide the

solution into a series of overlapping subproblems [Morita and Kanade, 1997; Pollefeys *et al.*, 2004]. Iterative methods focus on directly minimizing the factorization error using methods such as non-linear optimization [Buchanan and Fitzgibbon, 2005] or alternation [Hartley and Schaffalitzky, 2003; Vidal and Hartley, 2004]. These methods rely on initial approximations, but good results have been shown when initializing the algorithms with random solutions. Batch methods are usually considered complementary to iterative methods in that they can provide good starting positions for iterative methods at low computational costs [Tardif *et al.*, 2007]. They work by minimizing and approximation to the reprojection error to simplify the later iterative optimization [Guilbert *et al.*, 2006; Martinec and Pajdla, 2005].

Overall, hierarchical methods can be viewed as a superset of iterative and batch methods in that either can be used as part of the structure generation phase. Hierarchical methods tend to be used in large problems where either the amount of data is intractable or must be dealt with online. Due to the finite nature of data sampling for both SEM and TEM applications, we chose a strategy that builds upon using batch methods for initialization of an iterative method. This methodology provides a robust solution which is also computationally efficient.

The method chosen for batch factorization was that of using Camera Basis Constraints (CBC) proposed by [Tardif *et al.*, 2007]. This method was chosen over other existing methods due to its efficiency and excellent performance characteristics when compared to other state-of-the-art tools. This method proceeds as follows:

1. Determine sub-matrices without missing data.
2. Use the sub-matrices to compute a constraint on the joint projection matrix.
3. Combine the constraints to estimate the joint projection matrix.
4. Estimate the second factor by minimizing the reprojection error.

To accomplish this, we must first find a sub-block of $\tilde{M}$ without missing data. This selection is performed by left multiplying $\tilde{M}$ by a row-amputated

block diagonal matrix Π with 2×2 identity blocks. The feature subset is likewise extracted though multiplying to the right by Γ, which is an identity matrix amputated of some of its columns.

$$\hat{M}_{2\hat{m}\times\hat{n}} = \Pi\tilde{M}\Gamma \tag{5.7}$$

We can then separate the rotational and translational components of $\hat{M}$ similar to Tomasi and Kanade [Tomasi and Kanade, 1992].

$$\hat{M} = \Pi M\Gamma = \Pi RS\Gamma + \Pi\mathbf{t}\mathbf{1}^T\Gamma = \hat{R}\hat{S} + \hat{\mathbf{t}}\mathbf{1}^T \tag{5.8}$$

where $\hat{R}$ is the approximate rotation between the frames, $\hat{S}$ is the mean centered structure, and $\hat{t}$ is the translation between the frames. We now define $\boldsymbol{\mu}_n$ as:

$$\boldsymbol{\mu}_n = \frac{1}{n}\mathbf{1}_{n\times 1} \tag{5.9}$$

as the matrix which computes the column means of a $m \times n$ matrix by right-multiplication. We can now develop a centered measurement matrix as:

$$\bar{M} = \hat{M} - \hat{M}\boldsymbol{\mu}_n\mathbf{1}_{1\times\hat{n}}^T \tag{5.10}$$

The use of basis constraints for the structure and motion reconstruction consists of aligning the bases of the projection matrices of partial reconstructions performed on measurement sub-matrices. This is equivalent to solving a number of partial affine reconstructions and then aligning the basis constraints. Once this is done, the structure can be computed by minimizing the reprojection error.

Compute the singular value decomposition of the sub-matrix $\bar{M}$ as $\bar{M} = U\Sigma V^T$. The three leading columns of U form a basis of $\hat{R}$ meaning that there is a 3×3 invertable matrix, Z such that:

$$\hat{R} = \bar{U}Z \tag{5.11}$$

which leads to the CBC:

$$\Pi R = \bar{U}Z \tag{5.12}$$

To align the bases together, we must minimize

$$\sum_{k=1}^{l} ||\bar{U}_k Z_k - \Pi_k R||^2 = \sum_{k=1}^{l} \left\| (\Pi_k - \bar{U}_k) \begin{bmatrix} R \\ Z_k \end{bmatrix} \right\|^2 \tag{5.13}$$

This can be rewritten as a least-squares estimation as:

$$\left\| \underbrace{\begin{bmatrix} \Pi_1 & -\bar{U}_1 & & \mathbf{0} \\ \vdots & & \ddots & \\ \Pi_l & \mathbf{0} & & -\bar{U}_l \end{bmatrix}}_{D} \underbrace{\begin{bmatrix} R \\ Z_1 \\ \vdots \\ Z_l \end{bmatrix}}_{X} \right\|^2 \tag{5.14}$$

This equation can be efficiently minimized because of its sparseness and the fact that we do not need to calculate the aligning matrices Z_k. To ensure a non-trivial solution, we need to fix the gauge of the regression problem in the form of constraints on the solution. For this, the alignment of two affine cameras can be done by constraining the two first frames to

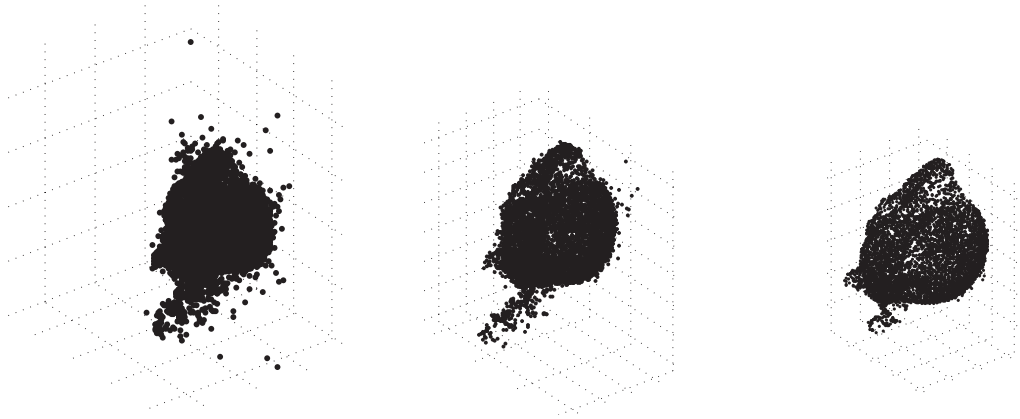
$$\begin{bmatrix} R_a & t_a \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \tag{5.15}$$

$$\begin{bmatrix} R_b & t_b \end{bmatrix} = \begin{bmatrix} 0 & 1 & 1 & 0 \\ * & * & * & * \end{bmatrix} \tag{5.16}$$

when solving the least-squares problem.

We now have the estimated joint projection matrix for the image sequence. Due to efficiency concerns, we now calculate only the joint translation vector $\mathbf{t}$ and use this to triangulate each feature. This is accomplished by first combining the computed basis $\bar{U}$ with the translation of the partial reconstructions:

$$\bar{\mathbf{t}} = \hat{M} \boldsymbol{\mu}_{\hat{n}} \tag{5.17}$$



(a) (b) (c)

Figure 5.5: the progression of the structure estimation: a) before robustly removing features b) after robustly removing features c) after bundle-adjustment

The joint translation vector can now be estimated by:

$$\left\| D \begin{pmatrix} \mathbf{t} \\ \mathbf{v}_1 \\ \vdots \\ \mathbf{v}_l \end{pmatrix} - \begin{pmatrix} \bar{\mathbf{t}}_1 \\ \vdots \\ \bar{\mathbf{t}}_l \end{pmatrix} \right\|^2 \tag{5.18}$$

where $\mathbf{v}$ is the equation when multiplying to the right subtracts the row mean from each entry in a matrix

$$\mathbf{v} = \mathbf{I} - \boldsymbol{\mu}_\mathbf{m}\mathbf{1} \tag{5.19}$$

With both the joint projection matrix and translation vector, it becomes a simple matter of triangulation to compute the remaining structure matrix S. Before this triangulation step takes place, the matrices R and S must be upgraded to Euclidean matrices as discussed in Section 5.1. Since the points are often tracked for a number of frames, it becomes useful to robustly estimate their position during the structure estimation phase. In this step,

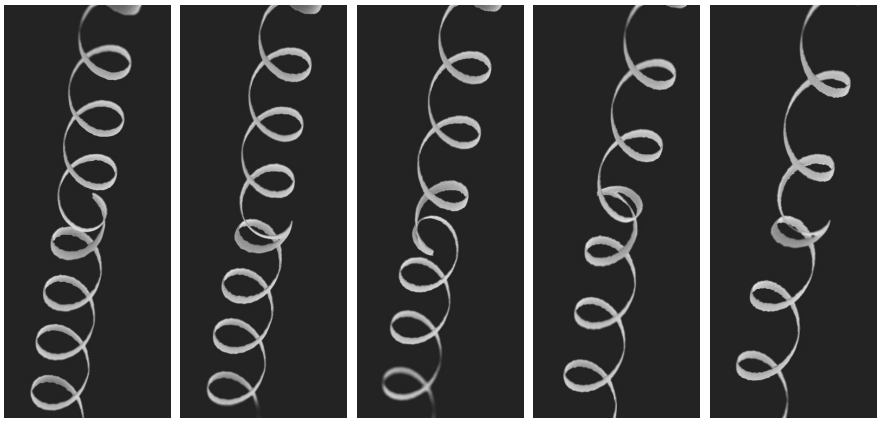
we use RANSAC and the reprojection error to reject any feature points which may be inaccurately matched. The results are shown in Figure 5.5.

Now that initial estimates for both structure and motion have been found, non-linear iterative methods will be employed to find the final estimates. In this step, we use sparse bundle adjustment, from Equation 5.6, based on the Levenberg-Marquardt method [Lourakis and Argyros, 2004].

5.2 Dual-Chirality Nanohelicies as Goniometers

Fundamental to the task of generating three-dimensional information from a sample in an electron microscope is the ability to view the sample from a variety of different angles. To accomplish this, a sample is rotated about an axis within the field of view of the image. Previously, the inability to generate local rotations at suitable scales for tomography has caused the rotational actuator to be succeeded by translational actuators that move the area of interest to the estimated rotation axis. Due to factors such as calibration, drift, and backlash the system does not perform a true eucentric rotation and data capture must be performed in a closed-loop fashion.

To avoid these problems, a dual-chirality nanohelix [Dong *et al.*, 2008] can be used to generate localized rotations, thus simplifying the requirements of



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

Figure 5.6: rotation series of a helical nanocoil

the sample stage. These structures rely on the phenomenon that if a spring is constructed with both a clockwise and counterclockwise pitch, the point at which the pitch direction changes will rotate as the ends of the spring are displaced (Figure 5.6). This rotation has been demonstrated to vary linearly with the displacement and the capability to deliver a relatively large linear motion to rotation conversion factor of 171.3°/µm. We investigate their efficacy as rotation stages for SEM imaging as a first step toward their possible use in TEM systems. The use of such localized goniometers, allows for a simpler sampling paradigm, an increased specimen rotation range, and reduces the requirements for the automated translation stage and the data acquisition.

Principle among the benefits of this method is the simplification of the rotation stage. Current systems use DC motors, stepper motors, and piezoelectric stick-slip mechanisms for generating the eucentric tilt. With the incorporation of a localized tilt-mechanism, the high-precision part of the stage can be limited to an xyz translation axis that needs a range on the order of a few micron. This would allow for a simpler element, such as a piezoelectric stack, to be used in this capacity.

One additional challenge with current TEM systems is that at high angles of rotation, the sample holder occludes the area of interest, which effectively limits the tilt of the sample. This limitation reduces the resolution along the z-axis of the reconstructed image stack. The geometry of dual-chirality nanohelices allows for higher rotation angles, and effectively limits the rotation to restrictions imposed by the sample.

Currently available TEM tomographic methods allow for tilts limited to approximately ±70° due to restrictions of the microscope, sample, and holder. This leads to a 'missing wedge' in the Fourier space, which leads to a blurring along the optical axis after reconstruction. By rotating the sample ±90° around the optical axis and taking a second data set, the missing data can be reduced.

The second data set can be acquired by manually removing the sample and rotating it or by holders that allow the specimen to be rotated inside the microscope. In both cases, the target area must be relocated before

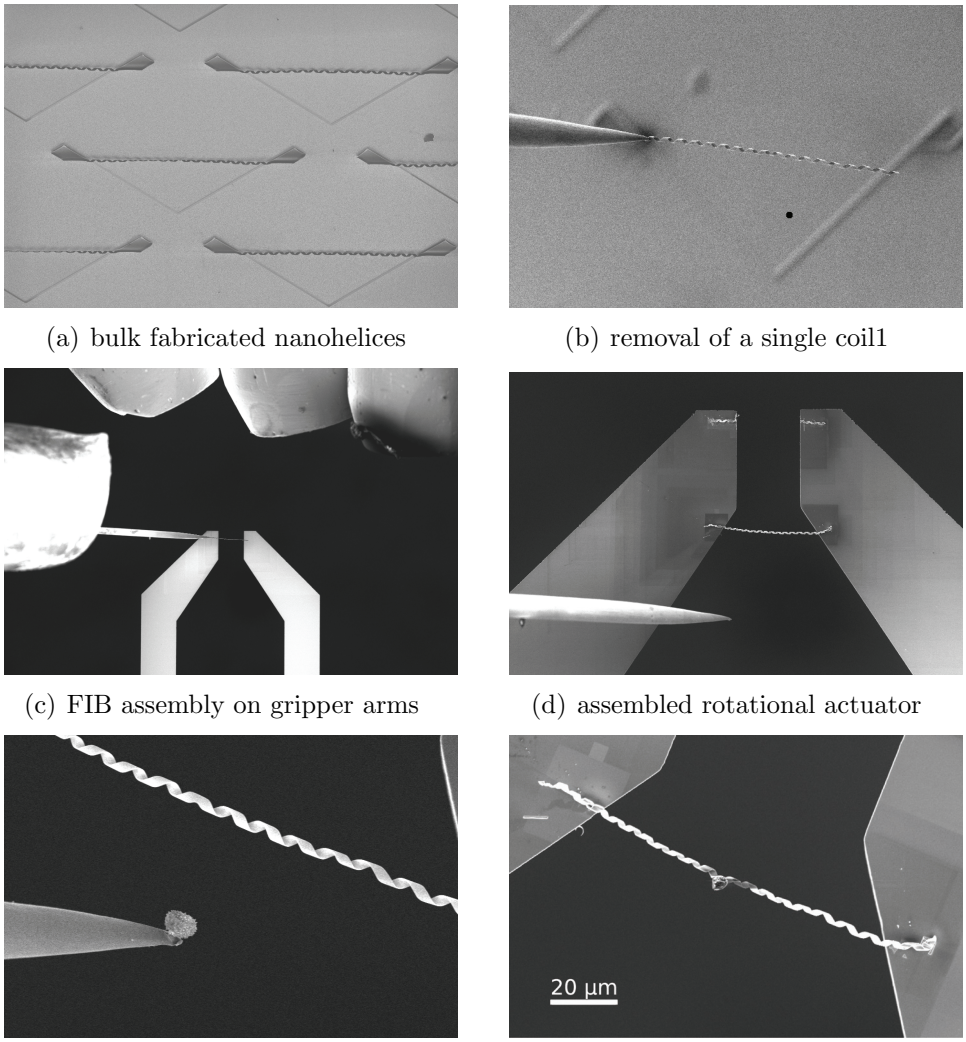
the scanning can resume. The use of a localized tilt-stage can enable the use of dual-tilt tomography without having to physically adjust the sample by using it in conjunction with an automated tomographic stage. In this method, a tilt-series would be acquired through the traditional automated sample stage. A second tilt-series can be then generated with the nanohelix actuation system orthogonal to the first data set.

Many older TEM systems are not equipped with automated sample stages. In these systems, a trained user must manually perform all the data acquisition steps, and the results can vary according to user experience. The use of nanohelicess for rotational actuators gives the ability to retrofit these older systems and make them suitable for tomographic tasks. Since the lifespan of a TEM is often in excess of twenty years, this could improve tomographic reconstruction on a large number of systems.

5.2.1 Fabrication Through Hybrid Bulk Micromachining and Robotic Assembly

Although eventual adoption of this technology for general use will require the monolithic bulk fabrication of the rotation and actuation components of the motion stage, it is sufficient for proof of concept to assemble these devices in a serial fashion. The nanogoniometer prototype was fabricated using a hybrid process consisting of bulk microfabrication and nanorobotic assembly. In this process, an as-fabricated nanohelix is assembled between the arms of an electrostatic gripper.

A focused ion beam in a scanning electron microscope was used to affix the two parts together by depositing a thin layer at the intersection of the two devices. After building this device, a sample is then affixed to the nanohelix center using a similar process. In Figure 5.7 we demonstrate one application of such a process. The data was acquired in a Carl Zeiss NVision 40 CrossBeam workstation. This device is equipped with a Gemini SEM Column and a SIINT 100 mm zeta FIB column. A gas injection system was used to supply carbon to the chamber for deposition with the FIB. SiGe/Si nanohelices [Zhang *et al.*, 2006] were used for the rotational actuator and an



(a) bulk fabricated nanohelices

(b) removal of a single coil1

(c) FIB assembly on gripper arms

(d) assembled rotational actuator

(e) a single mold spore

(f) attached to coil

Figure 5.7: sample preparation

electrostatic microgripper, specifically a FemtoTools FTG-G100, was used as the linear actuator [Beyeler *et al.*, 2007].

The first step in the process was selection and removal of a bulk-fabricated nanohelix. This was accomplished by first cutting one side with the FIB away from the substrate. The second side was then attached to a tungsten probe attached to a Kleindiek MM3A micromanipulator and subsequently cut from the the substrate (Figures 5.7(a) and 5.7(b)). The device was then

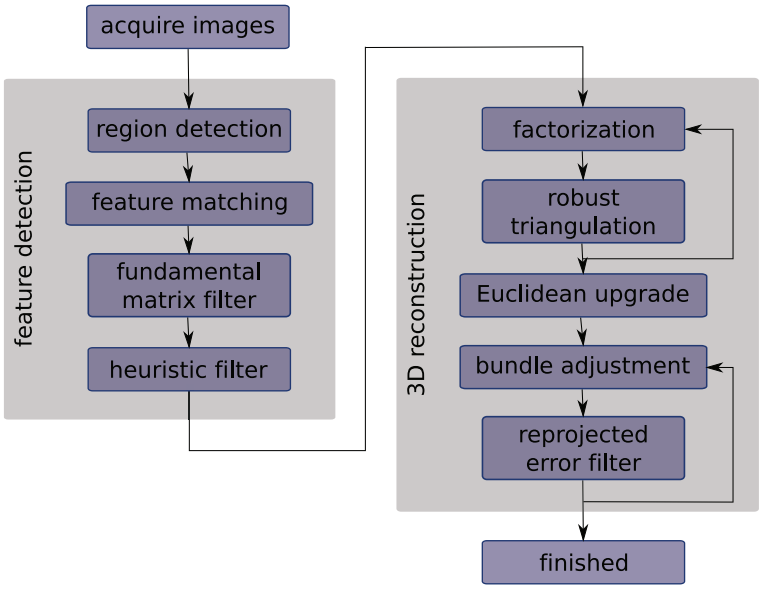


Figure 5.8: complete reconstruction algorithm

transported to the linear actuator and gas injection needles were inserted into the workspace (Figure 5.7(c)). The free end of the nanohelix was then attached to the arm of the microgripper with a small, approximately 1 µm, carbon deposition. This task was then repeated with the other side and the coil was detached from the tungsten probe (Figure 5.7(d)). Finally a sample was picked up using electron-beam induced deposition and attached to the coil for rotation (Figures 5.7(e) and 5.7(f)).

5.3 Experimental Results

The final reconstruction algorithm is summarized in Figure 5.8. Unlike the previously noted methods, we have elected to use an iterative filtering method for detecting additional outliers during three-dimensional reconstruction. After both the triangulation step and the bundle adjustment step, we perform filtering based on the reprojected errors of the points. Because of the least-squares nature of the estimation, outliers can potentially have a large influence on the resultant estimation. To minimize this effect, we first filter with a larger threshold to remove the distant outliers. Then, the factor-

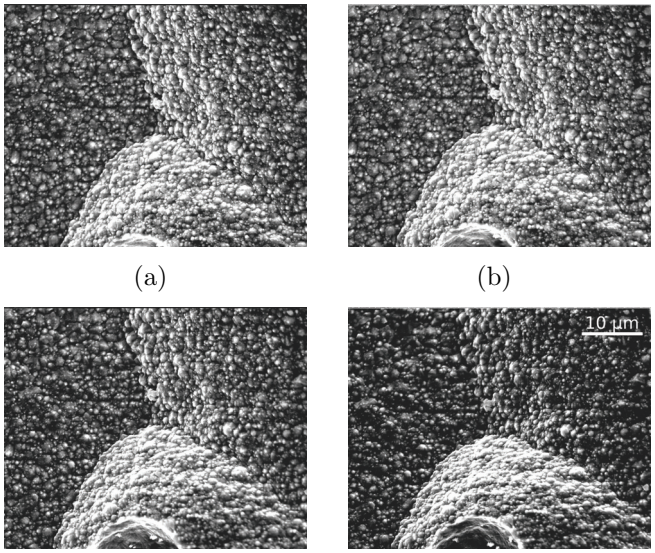
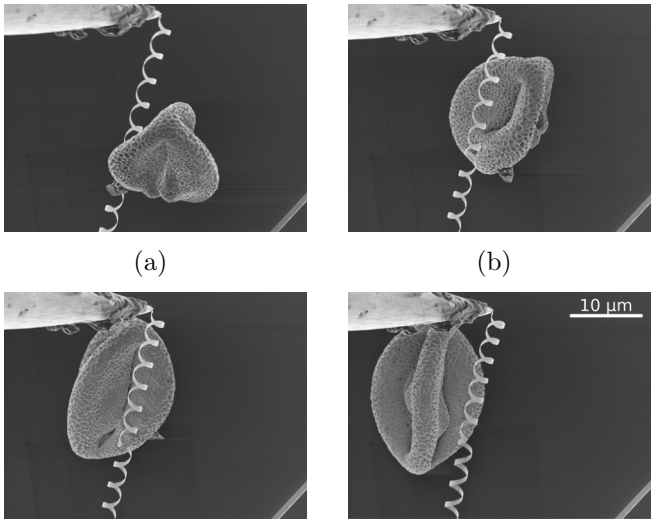


Figure 5.9: tilt-series of electroplated nickel

ization, or bundle adjustment, is calculated again and subsequently filtered with a smaller threshold. Experimentally, two or three iterations provides good results in a reasonable amount of time because of the efficiency of the factorization and the ability to use the previous results as a starting point for bundle adjustment. Due to the priority of flexibilty over processing time for the method evaluation, the software was developed in Matlab with certain processor intensive steps such as feature detection or bundle adjustment implemented in C [Lourakis and Argyros, 2004]. In its current form, the system is capable of running entirely automatically on a set of acquired images and has been evaluated at processing series up to 100 frames.

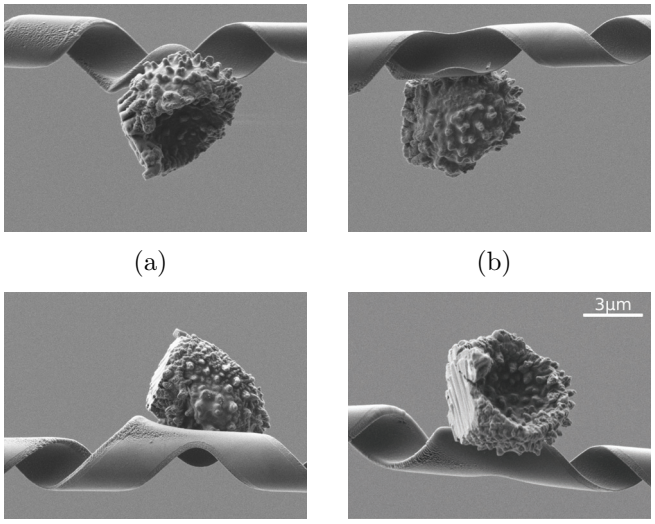
We will discuss the experimental results of this method on three different data sets. The first was acquired on a Zeiss DSM 962 by tilting the sample stage, and is gathered in a fashion similar to traditional methods for multi-view SEM reconstruction (Figure 5.9). The sample is the electroplated nickel surface which forms one of the soft magnetic bodies of the microrobot shown in Figure 2.2(b). We will also examine two data sets using the nanohelices as rotational actuators. The series shown in Figure 5.10 was acquired on



(a) (b)

(c) (d)

Figure 5.10: 180° rotation series of a single pollen grain

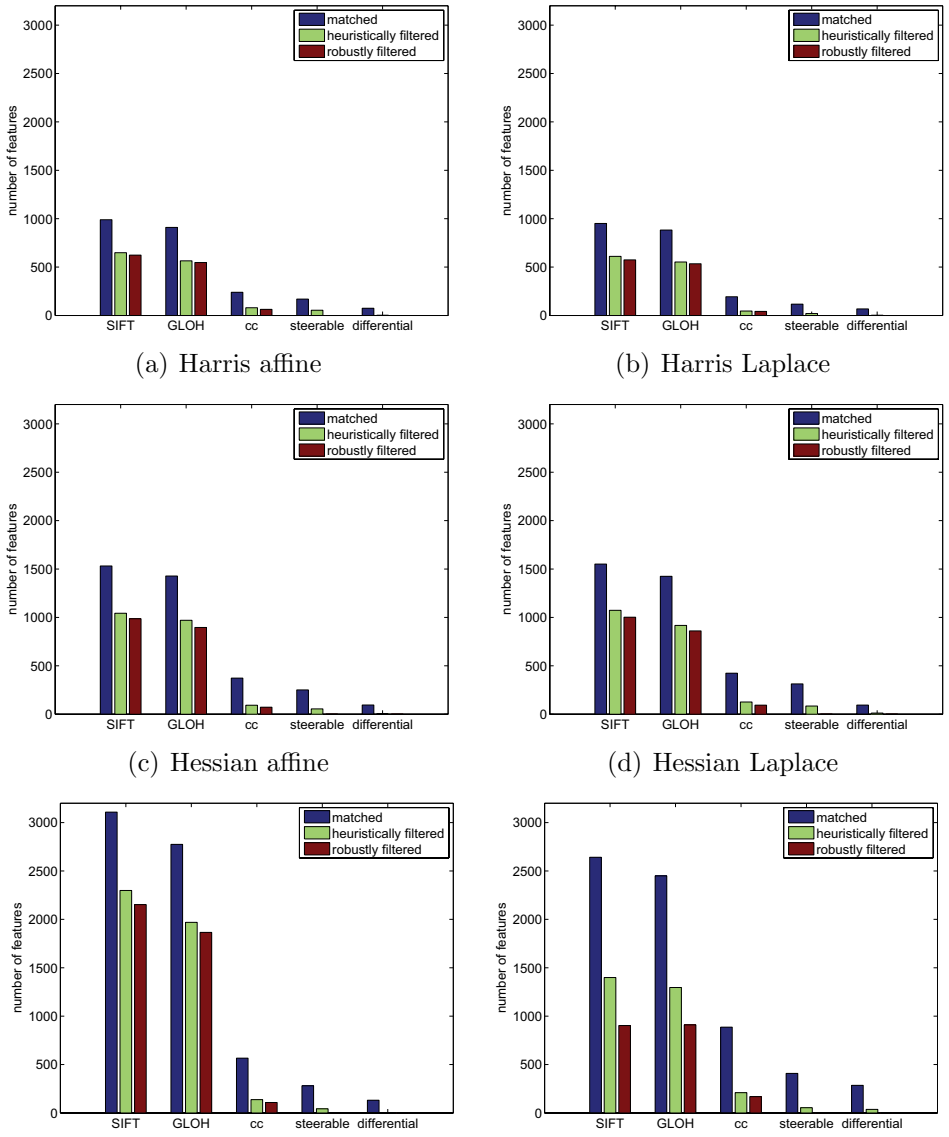


(a) (b)

(c) (d)

Figure 5.11: 360° rotation series of a single mold spore

a Zeiss Ultra 55 and the nanohelix was manipulated with a MM3A robotic manipulator. In this series, an individual pollen grain, ≈15µm in diameter was used for the sample specimen. The second, shown in Figure 5.11, was



(a) Harris affine

(b) Harris Laplace

(c) Hessian affine

(d) Hessian Laplace

(e) Harris Hessian Laplace

(f) edge Laplace

Figure 5.12: average number of features matched between ten frames of the pollen series using different descriptors and detectors

performed in a Zeiss NVision system as described in Section 5.2.1. The specimen was a single mold spore.

Before attempting the entire reconstruction process, we analyzed a number of different region descriptor methods using an excerpt from the pollen

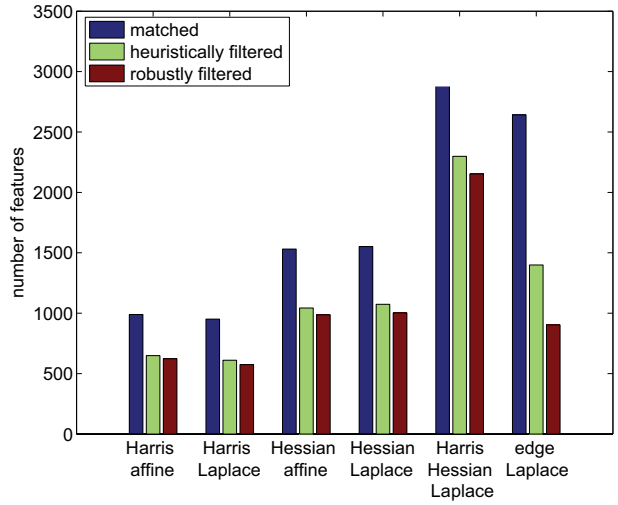


Figure 5.13: average number of features matched between ten frames of the pollen series using different region detection methods with SIFT descriptors

	Harris Laplacian	Hessian affine	Hessian Laplacian	Harris Hessian
Harris affine	69%	1%	0%	37%
Harris Laplacian		0%	1%	39%
Hessian affine			58%	15%
Hessian Laplacian				16%

Table 5.1: overlap between detector methods

image series. In agreement with Mikolajczyk *et al.* [2005], we found the SIFT-based methods to perform better than the other detectors in terms of both the number of features matched and the number of features kept after filtering (Figure 5.12). While the GLOH descriptor performed quite well, contrary to the Mikolajczyk and Schmid [2005] we did not find any performance benefits on the data series tested. In addition to the descriptor methods, we analyzed the influence of the region detectors for SIFT. Figure 5.13 shows the performance of the different region detection methods on the average matched points between frames. While both the Harris Hessian Laplace and edge Laplace detection methods found large numbers of features, many of the ones detected using the edge Laplacian methods were removed in the

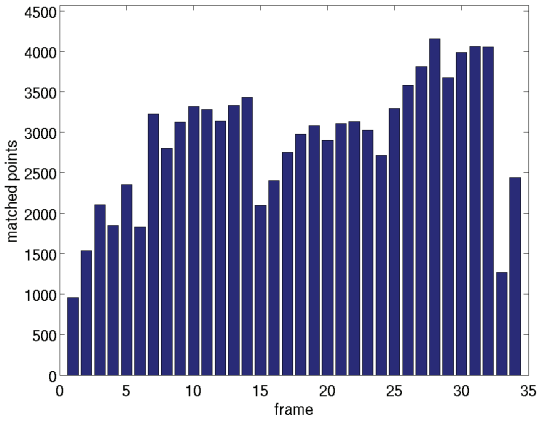
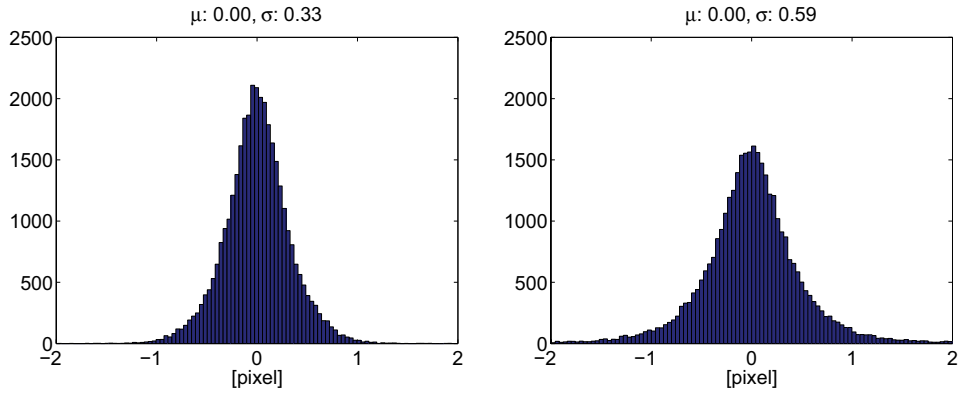


Figure 5.14: the number of matched points found in each frame of the pollen sequence



(a) x axis (b) y axis

Figure 5.15: pollen 3D reconstructed points reprojection error histogram

matching and filtering operations. Based on the previously discussed results, we chose the Harris Hessian Laplace region detector in conjunction with the SIFT detector for subsequent feature matching. If a denser feature matching is desired, it may be of interest to capitalize on multiple region detection methods when determining the matches. Table 5.1 demonstrates the measured overlap of features using these different methods. This figure indicates that the different region detectors do find disjoint sets of regions and that if necessary, additional features can be utilized through their combination.

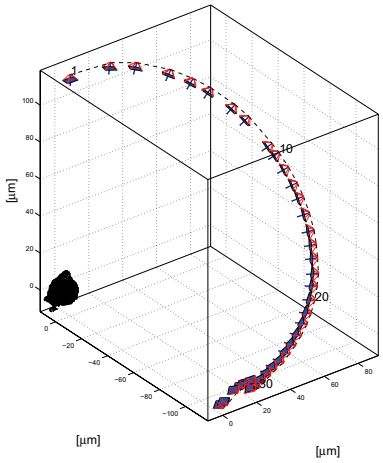


Figure 5.16: reconstructed camera trajectory of the pollen image series

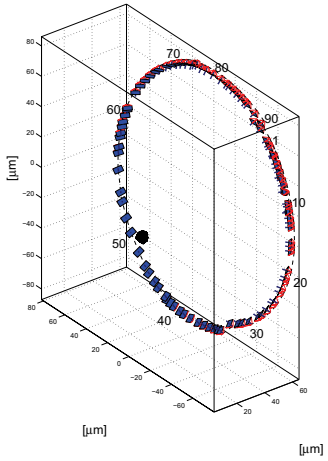
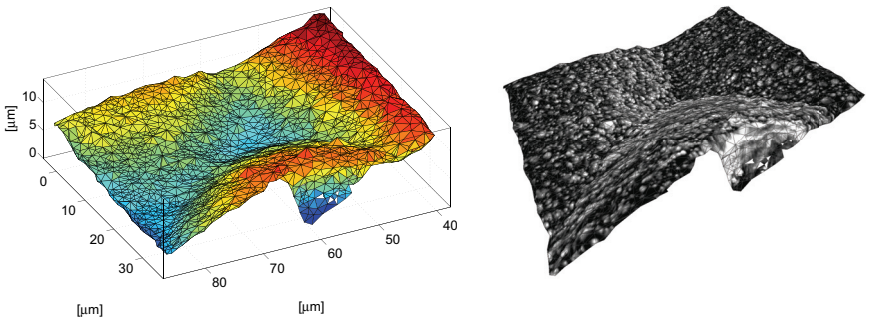


Figure 5.17: reconstructed camera trajectory of the spore image series

Figure 5.14 shows the number of matched points between frames over the entire pollen series of images. These features were then provided to the three-dimensional reconstruction software for further filtering and estimation. This process begins with CBC factorization and an iterative filtering process based on the reprojection error of the estimated points and poses. To generate the following reconstructions, the reprojection error filter threshold was graduated from eight pixels to one pixel over the course of four iterations. The estimate is subsequently further refined using bundle adjustment. Figure 5.15 represents typical reprojection errors of the estimated points and



(a) (b)

Figure 5.18: a) estimated structure b) with texture mapping

poses along the x- and y-axes after the bundle adjustment step. In this case, the data is provided for the pollen image series and one pixel equates to 40.5 nm.

Figures 5.16 and 5.17 demonstrate the reconstruction results of the pollen and mold spore data sets. Due to the high magnification of the SEM in both cases, an orthographic projection model was used. Thus, the distance from the object to the camera frames show in the figures has been arbitrarily set to allow both the resultant point cloud and the estimated camera positions to be observed. The point clouds for the pollen and spore structures contain 21,153 and 36,656 points respectively. In the case of the samples mounted to the helical nanocoils, a manual step was used to segment the coil from the sample after the reconstruction process finished successfully.

After reconstructing the structure and motion, different approaches can be used for interpreting the data. One of the most useful processing steps is the creation of surface meshes from the measured points. These methods use nonuniform rational B-splines (NURBS) or Delaunay triangulations to represent geometrical shapes in compact forms. Convex surfaces can be rapidly reconstructed with tools such as the Quickhull algorithm [Barber *et al.*, 1996], which can readily be found in software packages such as Matlab. More complex structures, such as those with concave segments, require consequently more complex techniques. Commercial software packages exists for this task

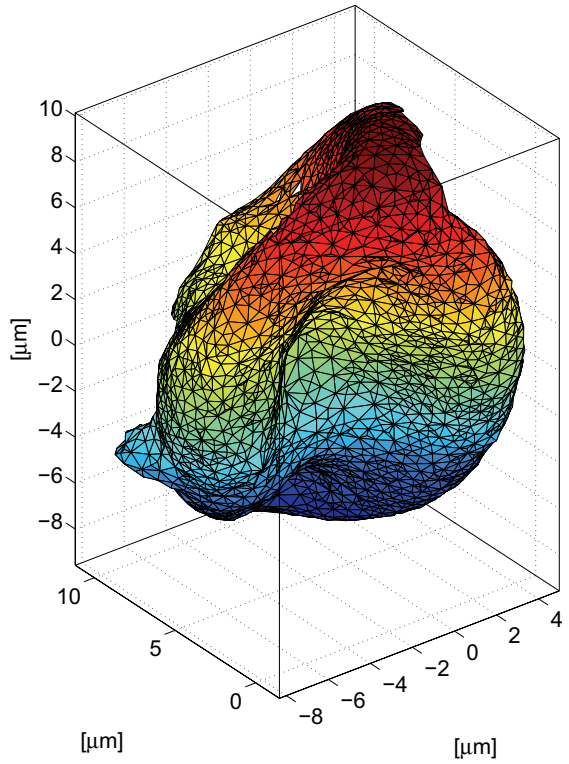


Figure 5.19: meshed pollen surface

such as RapidForm or Rhino 3D. We chose an open source alternative called MeshLab by Cignono *et al.* [2008]. This program provides facilities for both meshing a point cloud, using the ball-pivoting algorithm [Bernardini *et al.*, 1999], and editing the subsequent surface. Results of this processing step on the tilt-series and pollen data sets are shown in Figures 5.18 and 5.19 respectively. The mold spore data set was not meshed because its structure was too complex for the meshing tool.

To improve the meshing process in the future, dense stereo techniques can be used to provide even more features to simplify the meshing process [Sunyoto *et al.*, 2004]. These techniques generally first register the images with the epipolar lines along the horizontal image axis to simplify the feature detection computations [Brown, 1992]. Once the search space for features has been limited, denser feature detection algorithms can be used to find signficantly more feature points for the subsequent reconstruction.

5.4 Discussion

We have described and demonstrated an image-based method for reconstructing micro- and nanoscale objects in three-dimensions. Different from previously published reconstruction systems for scanning electron microscopy which rely primarily on techniques from stereogrammetry, this system is capable of efficiently using a large series of acquired images to generate the resultant structure. This has the benefits of enabling a larger rotation range and higher precision than methods based on only image pairs. The use of multiple images also accommodates the determination of the rotation angles between images, thus removing the burden of precise knowledge of the motion between frames. The methods described here are robust to imaging and feature detection noise, and have been used in a number of different imaging conditions.

Additionally, we have demonstrated the application of a novel localized rotational actuator. The use of nanohelices demonstrates the first microscale actuator capable of more than 360° rotation. By utilizing a rotational actuator with principle dimensions at the microscale, the image acquisition process is simplified from that of traditional systems. This system realizes the first step toward an array of different devices which have the possibility of being used in applications such as tomography or for creation of true three-dimensional atomic force microscopy.

6 Summary and Contributions

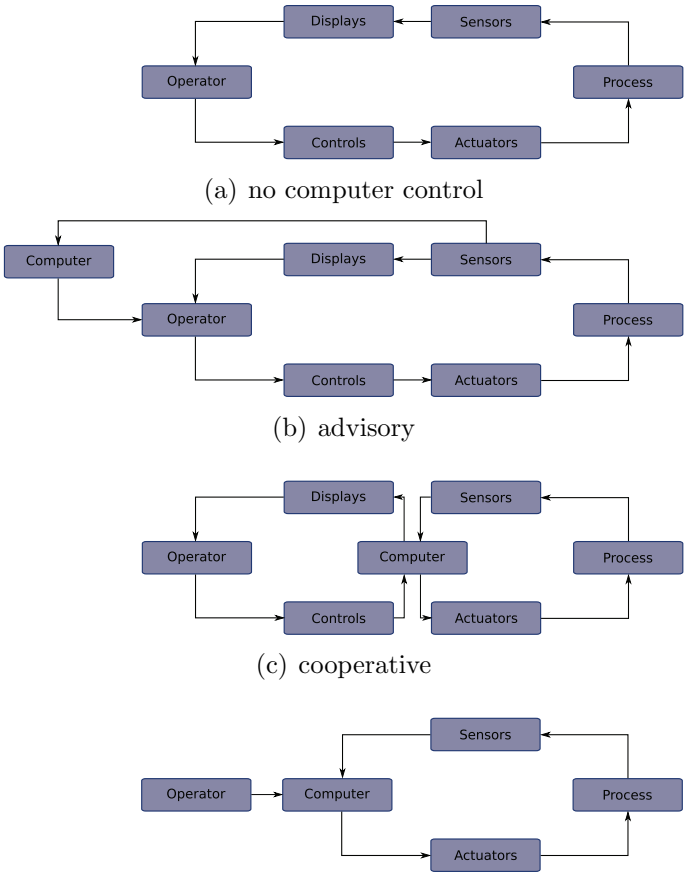
6.1 Discussion

Better tools are needed to improve interactions with the world of the very, very small. Robotics can help play a role in releasing the creativity of researchers as they explore this domain. Robotics is often designated as a field that searches for better ways to incorporate intelligence into control of motion. This intelligence is embodied in a number of different forms making robotics a diverse and challenging field. Up to now, much of the effort in nanorobotics has been given to developing the basic tools with the scales and precisions capable of working effectively at these scales, and little work has focused on adding intelligence to these systems.

With the advent of commercially available nanomanipulators, this will start to change as robotic technology becomes more ubiquitous when exploring nano-scale phenomenon. This increase in user base will in part provide new challenges and problems to explore through the use of robotics, and in part will demand more from the robotics community to improve the efficacy of the researchers performing the manipulations. The robotics field can contribute in multiple ways to the improvement of nanomanipulation by adding differing degrees of intelligence to these manipulators.

An initial phase of this added intelligence will come in the form of improved feedback to users during manipulation (Figure 6.1(b)). This can take various forms, such as better kinematic representations of the system, augmented reality systems, or automatically acquired manipulation metrics. The principle function of this information will be to increase the efficiency of a user performing manipulation. A simple example of this would be having a vision system that would automatically perform real-time measurements on the images as the user manipulates an object, or perhaps a three-dimensional representation of the position of the various manipulators in the workspace.

Technology will subsequently evolve such that a computer becomes more intimately involved in the control loop (Figure 6.1(c)). A simple example of this would be the evolution from joint space control to task space control.



(a) no computer control

(b) advisory

(c) cooperative

(d) complete control

Figure 6.1: the evolution of computers in control processes

Previously most robots were open-loop systems, which precludes the use of task space control. As more closed-loop systems become involved, whether it be due to improved actuator feedback or a visual tracking system, the possibilities for task-space control will increase. At this phase, haptic control interfaces have the ability to start to be of use to the user. Nanomanipulators often have significantly different actuation mechanisms for relatively large or small scale motions. Concepts such as virtual fixtures or virtual mechanisms have great possibilities to improve the speed and accuracy of nanorobotic manipulations as well as reduce the training time for new users. Simple forms of automation such as performing pre-recorded motions or au-

tomatically aligning the manipulator will start to be visible in this phase.

As more advanced systems become available, computers have the potential to take control of the manipulation task from the user (Figure 6.1(d)). At this stage, a user would request a high level task to be performed and the computer would manage the actual manipulation. This is the most difficult stage of robotic interaction at any scale. In macro-scale robotics, the task is largely handled by high-precision actuators performing easily repeatable, deterministic operations. Due to various interactions of nano-scale forces and the precision required from actuators, this type of manipulation is not possible without various forms of feedback.

Some may argue that the ultimate role of robotics in nanotechnology is the automation of assembly and manipulation tasks into a nanofactory analogous to their macro counterparts, but this is not the author's opinion on the matter. Due to the scales involved, any serial manipulation task at the nanoscale is necessarily relegated to a small amount of material when compared to the larger world. This does not mean that robotics has no future in nanotechnology, but rather that it should be targeted to problems where it has the most potential. Robotics provides the ability to rapidly prototype structures that may not be possible through currently existing bulk fabrication methods. The exploration of these and other structures can be greatly facilitated by robotic methods, and thus pave the way for future bulk or hybrid fabrication methods. Robotics also has the capability to help automate sample preparation, such as handling TEM lamella, or increase the yield of other handling tasks in research. Additionally, computer vision techniques have the ability to automate data acquisition when performing structure characterization. Where nano-scale computer vision algorithms often have to make simplifications on the environment they will interact in, nano-scale interactions have the advantage of a highly structured environment. This environment, albeit often with a low SNR and other imaging challenges, provides a domain where computer vision technologies can be practically applied in real-world conditions. In short, while the future of fully-automated nanomanipulation systems may be small, they show great potential as tools for assisting nanotechnology researchers for years to come.

6.2 Contributions

The focus of this thesis is the analysis and application of a scanning electron microscope as a visual sensor for nanorobotic manipulations. This has taken the form of three different areas of investigation: sensor calibration, visual tracking, and three-dimensional reconstruction. In brief, the contributions of this thesis can be summarized as:

- We have demonstrated the applicability of a perspective projection model for low-magnifications, $< 1000\times$, and a methodology for calibrating this model on any SEM without a high-precision calibration target.

- We have extended the calibration procedure to construct a parametrized perspective model for the SEM which is valid over a range of magnifications and working distances without the need to calibrate in each situation.

- We have developed and demonstrated the first rigid-body tracking system suitable for use in an SEM under a variety of conditions. As opposed to previously proposed SEM tracking systems which allow for only image-based tracking methodologies, the model-based paradigm enables additional information, such as three-dimensional structure or kinematic constraints, to be utilized. This additional information can be used to improve the system performance or enable previously untrackable motions, such as out-of-plane rotations, to be tracked. This methodology has shown an ability to work with noisy images and subpixel resolution at a number of magnifications, and the ability to track moving targets at real-time ($\sim 4 - 16$ fps) frame rates.

- We have developed the first multi-region-of-interest tracking system for a SEM. This method capitalizes on the structure of the observation at hand and can be used to automatically select regions which highly influence the tracking tasks. These regions can then be selectively scanned with the electron beam to provide a reduced scan area and thus higher frame rates, $3 - 5$ times that of full-frame imaging.

- We have developed and demonstrated a nondestructive three-dimensional data acquisition system for scanning electron microscopy based on structure-from-motion techniques. As opposed to the purely stereo methods discussed in the literature, this automated system is inherently capable of dealing with large degrees of rotation where points become occluded as well as automatically accounting for errors in the sample pose.

- We have demonstrated the first use of a localized rotational actuator for three-dimensional data acquisition. This actuator is capable of high-precision rotations in excess of 360° without the need of bulky macro-scale actuation stages, and provides the first steps toward new multiview nanoscale sensing methodologies.

6.3 Future Work

As contrast to optical imaging systems, which typically have low sensor-noise and parallel acquisition, the noisy and serial nature of SEM imaging systems provides many yet unobserved challenges and possibilities to traditional computer vision systems. Recently, the introduction of low cost scanning electron microscopes has introduced devices which approach the price of optical microscopes to the market. This, coupled with the high resolution and large depth-of-field, makes the SEM an attractive tool for microassembly tasks as well as nanorobotic applications. The previously discussed tracking system has demonstrated its ability to deal with some of these aspects, but additional steps can be taken to increase the system's performance reliability.

The current system implementation is unable to detect a loss of tracking, which is detrimental to visual servoing applications. Different monitoring systems can be added to the system such as monitoring velocity of the pose or detection of a rapid change in the number of tracked points. One common challenge to many tracking systems is an inability to localize the target from a random pose, i.e. once tracking has failed. While this is a mild limitation for tasks of augmenting user interactions with the system, it poses a serious challenge to fully automated systems. Future work towards automation will

combine this real-time method with slower more deliberate methods for both initial pose estimation and object reacquisition. One such possible method would be the combination of the model-based tracker with an image-based correlation tracker. While correlation systems are quite good at dealing with noisy images, they have major limitations in speed and their inability to easily track rotational or deformable motions. These limitations could be alleviated through combination with a model-based system by using the tracked model to selectively sample new correlation tracking regions at each frame for use with the subsequent one. These newly sampled regions could then easily be used to augment the current pose estimation, or in the case of a tracking loss, provide a new starting pose for the iterative methods.

Aside from region-of-interest strategies, little has been done to capitalize on the image formation process in the SEM for computer vision tasks. The serial, scanning nature of the sensor would naturally lend itself to methods such as image pyramiding for computer vision tasks. Traditionally, pyramiding is performed as a post processing step to reduce the complexity of an image for increased speed or robustness. The SEM provides a unique possibility to perform this image pyramiding with the sensor during the image acquisition as opposed to after-the-fact. This and other unique aspects of the SEM imaging process have yet to be fully exploited by vision researchers.

The technologies discussed for three-dimensional data acquisition show a great deal of promise in a variety of application fields. In addition to SEM data acquisition, the robust feature-based pose estimation system shows great potential for automated alignment of TEM tomographic images. Current alignment systems typically rely either on human tracked fiducial markers or time consuming automatic methods. The feature detection and factorization approach has the potential to provide increased data points and accuracy over manual methods as well as more efficient computations than current non-linear methods.

The use of nanohelices provides a new tool to the nanomanipulation repertoire. Although fabrication of these rotational actuators currently requires a serial assembly step, future work will look into the bulk fabrication of both the coils and the linear actuators. This technology can enable arrays of ro-

tational motion stages with potential uses in TEM tomography, or for the creation of three-dimensional AFM systems. In either case, a sample holder will need to be fabricated between the two coil segments which will facilitate more rapidly loading or exchanging samples. These sample holders can then be functionalized to enable the batch acquisition of biological or other fluidic samples to alleviate the micromanipulation step.

References

[Abbott *et al.*, 2007] J. J. Abbott, Z. Nagy, F. Beyeler, and B. J. Nelson. Robotics in the small part I: Microrobotics. *IEEE Robotics and Automation Magazine*, 14(2):92–103, 2007.

[Amelinckx *et al.*, 1994] S. Amelinckx, X. B. Zhang, D. Bernaerts, X. F. Zhang, V. Ivanov, and J. B. Nagy. A formation mechanism for catalytically grown helix-shaped graphite nanotubes. *Science*, 265:635–639, Jul 1994.

[Arai *et al.*, 1995] F. Arai, D. Ando, T. Fukuda, Y. Nododa, and T. Oota. Micro manipulation based on micro physics - strategy based on attractive force reduction and stress measurement. In *IEEE/RJS International Conference on Intelligent Robots and Systems*, volume 2, pages 263–241, 1995.

[Ball, 1876] R.S. Ball. The theory of screws: A study in the dynamics of a rigid body. *Mathematische Annalen*, 9(4):541–553, Dec 1876.

[Bansal *et al.*, 2006] R. K. Bansal, A. Kubis, R. Hull, and J. M. Fitz-Gerald. High-resolution three-dimensional reconstruction: A combined scanning electron microscope and focused ion-beam approach. *Journal of Vacuum Science and Technology B: Microelectronics and Nanometer Structures*, 24(2):554–561, Apr 2006.

[Barber *et al.*, 1996] C. B. Barber, D. P. Dobkin, and H. Huhdanpaa. The quickhull algorithm for convex hulls. *ACM Transactions on Mathematical Software*, 22(4):469–483, 1996.

[Battiston *et al.*, 2001] F. M. Battiston, J. P. Ramseyer, H. P. Lang, M. K. Baller, C. Gerber, J. K. Gimzewski, E. Meyer, and H. J. Guntherodt. A chemical sensor based on a microfabricated cantilever array with simultaneous resonance-frequency and bending readout. *Sensors and Actuators B-Chemical*, 77:122–131, 2001.

[Bell *et al.*, 2006] D. J. Bell, L. X. Dong, B. J. Nelson, M. Golling, L. Zhang, and D. Grützmacher. Fabrication and characterization of three-dimensional InGaAs/GaAs nanosprings. *Nano Letters*, 6:725–729, 2006.

[Belongie *et al.*, 2002] S. Belongie, J. Malik, and J. Puzicha. Shape matching and object recognition using shape contexts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(4):509–522, 2002.

[Bergander, 2003] Arvid Bergander. *Control, Wear Testing and Integration of Stick-Slip Micropositioning*. PhD thesis, Ecole Polytechnique Federale de Lausanne, 2003.

[Bernardini *et al.*, 1999] F. Bernardini, J. Mittleman, H. Rushmeier, C. Silva, and G. Taubin. The ball-pivoting algorithm for surface reconstruction. *IEEE Transactions on Visualization and Computer Graphics*, 05(4):349–359, 1999.

[Beyeler *et al.*, 2007] F. Beyeler, A. P. Neild, S. Oberti, D. J. Bell, Y. Sun, J. Dual, and B. J. Nelson. Monolithically fabricated micro-gripper with integrated force sensor for manipulating micro-objects and biological cells aligned in an ultrasonic field. *Journal of Microelectromechanical Systems*, 16(1):7–15, Feb 2007.

[Binnig *et al.*, 1986] G. Binnig, C.F. Quate, and C. Gerber. Atomic force microscope. *Physical Review Letters*, 56(9):930–933, Mar 1986.

[Binning *et al.*, 1982] G. Binning, H. Rohrer, C. Gerber, and E. Weibel. Surface studies by scanning tunneling microscopy. *Physical Review Letters*, 49(1):57–61, 1982.

[Böhringer *et al.*, 1999] K. F. Böhringer, R. S. Fearing, and K. Y. Goldberg. *The Handbook of Industrial Robotics*, chapter Microassembly. Wiley & Sons, 1999.

[Bowden *et al.*, 1997] N. Bowden, A. Terfort, J. Carbeck, and G. M. Whitesides. Self-assembly of mesoscale objects into ordered two-dimensional arrays. *Science*, 276(5310):233–235, 1997.

[Bregler and Malik, 1998] C. Bregler and J. Malik. Tracking people with twists and exponential maps. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 8–15, Jun 1998.

[Breguet, 1998] J.-M. Breguet. *Actionneurs "Stick and Slip" Pour Micro-Manipulateurs*. PhD thesis, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland, 1998.

[Brown *et al.*, 2003] M. Z. Brown, D. Burschka, and G. D. Hager. Advances in computational stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(8):993–1008, 2003.

[Brown, 1992] L. G. Brown. A survey of image registration techniques. *ACM Computing Surveys*, 24(4):325–376, 1992.

[Buchanan and Fitzgibbon, 2005] A. M. Buchanan and A. W. Fitzgibbon. Damped newton algorithms for matrix factorization with missing data. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2, pages 316–322, Jun 2005.

[Cho, 2006] A. Cho. News focus: Nanotechnology: Pretty as you please, curling films turn themselves into nanodevices. *Science*, 313:164–165, 2006.

[Cignoni *et al.*, 2008] P. Cignoni, M. Callieri, M. Corsini, M. Dellepiane F., Ganovelli, and G. Ranzuglia. Meshlab: an open-source mesh processing tool. In *Eurographics Italian Chapter Conference*, 2008.

[Cohen and Lightbody, 2004] S. H. Cohen and M. L. Lightbody, editors. *Atomic Force Microscopy/scanning Tunneling Microscopy 3*. Kluwer Academic/Plenum Publications, 2004.

[Cohn, 1992] M. B. Cohn. Self-assembly of microfabricated devices. United States Patent 5 355 577, Sep 1992.

[Cornille *et al.*, 2003] N. Cornille, D. Garcia, M. A. Sutton, S. McNeill, and J. J. Orteu. Automated 3D reconstruction using a scanning electron microscope. In *SEM Conference on Experimental and Applied Mechanics*, Jun 2003.

[Cornille, 2005] N. Cornille. *Accurate 3D Shape and Displacement Measurement using a Scanning Electron Microscope*. PhD thesis, Institut National des Science Appliquées, Toulouse, France, 2005.

[Culpepper and Anderson, 2004] M. L. Culpepper and G. Anderson. Design of a low-cost nano-manipulator which utilizes a monolithic, spatial compliant mechanism. *Precision Engineering*, 28(4):469–482, Oct 2004.

[Dagalakis and Amatucci, 2001] N. G. Dagalakis and E. Amatucci. Kinematic modeling of a 6 degree of freedom tri-stage micro-positioner. In *American Society for Precision Engineering*, pages 200–203, Nov 2001.

[Donald *et al.*, 2006] B. R. Donald, C. G. Levey, C. D. McGray, I. Paprotny, and D. Rus. An untethered, electrostatic, globally controllable MEMS micro-robot. *Journal of Microelectromechanical Systems*, 15(1):1–15, Feb 2006.

[Dong and Nelson, 2007] L. X. Dong and B. J. Nelson. Robotics in the small part II: Nanorobotics. *IEEE Robotics and Automation Magazine*, 14(3), 2007.

[Dong *et al.*, 2004] L. X. Dong, F. Arai, and T. Fukuda. Destructive constructions of nanostructures with carbon nanotubes through nanorobotic manipulation. *IEEE/ASME Transactions on Mechatronics*, 9:350–357, Jun 2004.

[Dong *et al.*, 2008] L. X. Dong, L. Zhang, B. E. Kratochvil, K. Shou, and B. J. Nelson. Dual-chirality helical nanobelts: A novel linear-to-rotary motion converter. In *IEEE International Conference on Micro Electro Mechanical Systems*, number 21, Jan 2008.

[Dong, 2003] L. X. Dong. *Robotic Manipulations of Carbon Nanotubes*. PhD thesis, Nagoya University, 2003.

[Drexler, 1986] K. E. Drexler. *Engines of creation*. Doubleday/Anchor Press, 1986.

[Drexler, 1992] K. E. Drexler. *Nanosystems: molecular machinery, manufacturing, and computation*. John Wiley and Sons, Inc., 1992.

[Drummond and Cipolla, 2002] T. Drummond and R. Cipolla. Real-time visual tracking of complex structures. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7):932–946, Jul 2002.

[Du *et al.*, 2006] E. Du, H. Cui, and Z. Zhu. Review of nanomanipulators for nanomanufacturing. *International Journal of Nanomanufacturing*, 1(1):83–104, 2006.

[Dürig *et al.*, 1986] U. Dürig, D. W. Pohl, and F. Rohner. Near-field optical-scanning microscopy. *Journal of Applied Physics*, 86(10):3318–3327, 1986.

[Dwarakanath *et al.*, 2000] T.A. Dwarakanath, C.D. Crane III, J. Duffy, and C. Tyler. In-parallel passive compliant coupler for robot force control. In *ASME Mechanisms Conference*, Sep 2000.

[Eigler and Schweizer, 1990] D. M. Eigler and E. K. Schweizer. Positioning single atoms with a scanning tunnelling microscope. *Nature*, 344:524–526, Apr 1990.

[Ermakov and Garfunkel, 1994] A. V. Ermakov and E. L. Garfunkel. A novel AFM/STM/SEM system. *Review of Scientific Instruments*, 65(9):2853–2854, Sep 1994.

[Fahlbusch *et al.*, 1999] S. Fahlbusch, S. Fatikow, J. Seyfried, and A. Buerkle. Flexible microrobotic system MINIMAN: design, actuation principleand control. In *Proceedings IEEE/ASME International Conference on Advanced Intelligent Mechatronics*, pages 156–161, 1999.

[Fahlbusch *et al.*, 2002] S. Fahlbusch, A. Shirinov, and S. Fatikow. AFM-based micro force sensor and haptic interface for a nanohandling robot. *IEEE/RSJ International Conference on Intelligent Robots and System*, 2:1772–1777, 2002.

[Fatikow *et al.*, 2006] S. Fatikow, T. Wich, H. Hülsen, T. Sievers, and M. Jähnisch. Microrobot system for automatic nanohandling inside a scanning electron microscope. In *IEEE International Conference on Robotics and Automation*, pages 1402–1407, May 2006.

[Faugeras, 1993] O. Faugeras. *Three-Dimensional Computer Vision*. The MIT Press, 1993.

[Fearing, 1995] R. S. Fearing. Survey of sticking effects for micro parts handling. In *IEEE/RJS International Conference on Intelligent Robots and Systems*, volume 2, pages 212–217, Aug 1995.

[Fernandez *et al.*, 2006] J. J. Fernandez, C. O. S. Sorzano, R. Marabini, and J. M. Carazo. Image processing and 3-D reconstruction in electron microscopy. *IEEE Signal Processing Magazine*, 23(3):84–94, May 2006.

[Feynman, 1959] R. P. Feynman. There's plenty of room at the bottom. Talk given at the annual meeting of the American Physical Society, Dec 1959.

[Fischler and Bolles, 1981] M. A. Fischler and R. C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981.

[Freeman and Adelson, 1991] W.T. Freeman and E.H. Adelson. The design and use of steerable filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(9):891–906, 1991.

[Freitas, 2005] J. R. A. Freitas. Nanotechnology, nanomedicine and nanosurgery. *International Journal of Surgery*, 3(4):243–246, 2005.

[Frutiger *et al.*, 2008] D. R. Frutiger, K. Vollmers, B. E. Kratochvil, and B. J. Nelson. Small, fast, and under control: wireless resonant magnetic micro-agents. In *International Symposium on Experimental Robotics 2008*, Jul 2008.

[Fukuda *et al.*, 2003] T. Fukuda, F. Arai, and L. X. Dong. Assembly of nanodevices with carbon nanotubes through nanorobotic manipulations. *Proceedings of the IEEE*, 91(11):1803–1818, Nov 2003.

[Fukushima *et al.*, 2002] K. Fukushima, S. Kawai, D. Saya, and H. Kawakatsu. Measurement of mechanical properties of three-dimensional nanometric objects by an atomic force microscope incorporated in a scanning electron microscope. *Review of Scientific Instruments*, 73(7):2647–2650, 2002.

[Gao *et al.*, 2005] P. X. Gao, Y. Ding, W. J. Mai, W. L. Hughes, C. S. Lao, and Z. L. Wang. Conversion of zinc oxide nanobelts into superlattice-structured nanohelices. *Science*, 309:1700–1704, Sep 2005.

[Goldstein *et al.*, 2003] J. Goldstein, D. Newbury, D. Joy, C. Lyman, P. Echlin, E. Lifshin, L. Sawyer, and J. Michael. *Scanning Electron Microscopy and X-Ray Microanalysis*. Springer, third edition, 2003.

[Golub and Loan, 1996] G. H. Golub and C. F. Van Loan. *Matrix Computations*. The Johns Hopkins University Press, third edition, 1996.

[Gool *et al.*, 1996] L. J. Van Gool, T. Moons, and D. Ungureanu. Affine/ photometric invariants for planar intensity patterns. In *European Conference on Computer Vision*, volume 1, pages 642–651, London, UK, 1996. Springer-Verlag.

[Gorman *et al.*, 2006] J. Gorman, Y-S. Kim, A. E. Vladar, and N. G. Dagalakis. Design of an on-chip micro-scale nanoassembly system. In *International Symposium on Nanomanufacturing*, 2006.

[Greminger and Nelson, 2004] M. A. Greminger and B. J. Nelson. Vision-based force measurement. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26:290–298, Mar 2004.

[Groeber *et al.*, 2006] M. A. Groeber, B. K. Haley, M. D. Uchic, D. M. Dimiduk, and S. Ghosh. 3D reconstruction and characterization of poly-

crystalline microstructures using a FIBSEM system. *Materials Characterization*, 57:259–273, Jan 2006.

[Guilbert *et al.*, 2006] N. Guilbert, A. Bartoli, and A. Heyden. Affine approximation for direct batch recovery of euclidian structure and motion from sparse data. *International Journal of Computer Vision*, 69(3):317–333, 2006.

[Guthold *et al.*, 2000] M. Guthold, M. R. Falvo, W. G. Matthews, S. Paulson, S. Washburn, D. A. Erie, R. Superfine, Jr. F. P. Brooks, and R. M. Taylor. Controlled manipulation of molecular samples with the nanomanipulator. *IEEE/ASME Transactions on Mechatronics*, 5(2):189–198, Jun 2000.

[Hansen *et al.*, 1998] L. T. Hansen, A. Kühle A, A. H. Sørensen, J. Bohr, and P. E. Lindelof. A technique for positioning nanoparticles using an atomic force microscope. *Nanotechnology*, 9(4), 1998.

[Harris and Stephens, 1988] C. Harris and M. Stephens. A combined corner and edge detector. In *Alvey Vision Conference*, 1988.

[Hartley and Schaffalitzky, 2003] R. Hartley and F. Schaffalitzky. Powerfactorization: 3D reconstruction with missing or uncertain data. In *Australia-Japan Advanced Workshop on Computer Vision*, 2003.

[Hatamura and Morishita, 1990] Y. Hatamura and H. Morishita. Direct coupling system between nanometer world and human world. In *IEEE Micro Electro Mechanical Systems*, pages 203–208, 1990.

[Hosokawa *et al.*, 1996] K. Hosokawa, I. Shimoyama, and H. Miura. Two-dimensional micro-self-assembly using the surface tension of water. In *Sensors and Actuators*, volume 57, pages 117–125, 1996.

[Hsu, 2004] T. R. Hsu. On the integration of micro and nano scale engineering. In *Modern Engineering and Technology Seminars*, volume 20, Nov 2004.

[Illingworth and Kittler, 1988] J. Illingworth and J. Kittler. A survey of the hough transform. *Computer Vision, Graphics, and Image Processing*, 44(1):86–116, 1988.

[Isenberg *et al.*, 2003] T. Isenberg, B. Freudenberg, N. Halper, and S. Schlechtweg. A developer's guide to silhouette algorithms for polygonal models. *IEEE Computer Graphics and Applications*, 23(4):28–37, Jul 2003.

[Jähnisch and Fatikow, 2007] M. Jähnisch and S. Fatikow. 3-D vision feedback for nanohandling monitoring in a scanning electron microscope. *International Journal of Optomechatronics*, 1:4–26, 2007.

[Joachimsthaler *et al.*, 2003] I. Joachimsthaler, R. Heiderhoff, and L. J. Balk. A universal scanning-probe-microscope-based hybrid system. *Measurement Science and Technology*, 14(1):87–96, 2003.

[Junno *et al.*, 1995] T. Junno, K. Deppert, L. Montelius, and L. Samuelson. Controlled manipulation of nanoparticles with an atomic force microscope. *Applied Physics Letters*, 66(26):3627–3629, 1995.

[Kasaya *et al.*, 1999] T. Kasaya, H. Miyazaki, S. Saito, and T. Sato. Micro object handling under SEM by vision-based automatic control. In *IEEE International Conference on Robotics and Automation*, volume 3, pages 2189–2196, 1999.

[Kass *et al.*, 1988] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *International Journal of Computer Vision*, 1(4):321–331, Nov 1988.

[Ke and Sukthankar, 2004] Y. Ke and R. Sukthankar. PCA-SIFT: A more distinctive representation for local image descriptors. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2:506–513, 2004.

[Kikukawa *et al.*, 1993] A. Kikukawa, S. Hosaka, Y. Honda, and H. Koyanagi. Magnetic force microscope combined with a scanning electron mi-

croscope. *Journal of Vacuum Science and Technology*, 11(6):3092–3098, 1993.

[Kis *et al.*, 2006] A. Kis, K. Jensen, S. Aloni, W. Mickelson, and A. Zettl. Interlayer forces and ultralow sliding friction in multiwalled carbon nanotubes. *Physical Review Letters*, 97, Jul 2006.

[Kortschack *et al.*, 2003] A. Kortschack, O.C. Hanbler, C. Rass, and S. Fatikow. Driving principles of mobile microrobots for micro- and nanohandling. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, volume 2, pages 1895–1900, 2003.

[Kratochvil *et al.*, 2008] B. E. Kratochvil, L. Dong, and B. J. Nelson. Real-time rigid-body visual tracking in a scanning electron microscope. *The International Journal of Robotics Research*, 2008.

[Lindeberg and Garding, 1994] T. Lindeberg and J. Garding. Shape-adapted smoothing in estimation of 3-D depth cues from affine distortions of local 2-D brightness structure. In *European Conference on Computer Vision*, pages 389–400, 1994.

[Lindeberg, 1998] T. Lindeberg. Feature detection with automatic scale selection. *International Journal of Computer Vision*, 30(2):79–116, Nov 1998.

[Lourakis and Argyros, 2004] M.I.A. Lourakis and A.A. Argyros. The design and implementation of a generic sparse bundle adjustment software package based on the levenberg-marquardt algorithm. Technical Report 340, Institute of Computer Science - FORTH, Heraklion, Crete, Greece, aug 2004.

[Lourakis, 2004] M.I.A. Lourakis. levmar: Levenberg-marquardt nonlinear least squares algorithms in C/C++, Jul 2004.

[Lowe, 1991] D. G. Lowe. Fitting parameterized three-dimensional models to images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(5):441–450, May 1991.

[Lowe, 1999] D. G. Lowe. Object recognition from local scale-invariant features. *International Conference on Computer Vision*, 2:1150–1157, 1999.

[Lowe, 2004] D G Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, Nov 2004.

[Lučić *et al.*, 2005] V. Lučić, F. Förster, and W. Baumeister. Structural studies by electron tomography: From cells to molecules. *Annual Review of Biochemistry*, 74:833–865, 2005.

[Martinec and Pajdla, 2005] D. Martinec and T. Pajdla. 3D reconstruction by fitting low-rank matrices with missing data. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1:198–205, 2005.

[Merkel *et al.*, 1999] R. Merkel, P. Nassoy, A. Leung, K. Ritchie, , and E. Evans. Energy landscapes of receptor-ligand bonds explored with dynamic force spectroscopy. *Nature*, 397:50–53, 1999.

[Mikolajczyk and Schmid, 2005] K. Mikolajczyk and C. Schmid. A performance evaluation of local descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(10):1615–1630, Oct 2005.

[Mikolajczyk *et al.*, 2005] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. Van Gool. A comparison of affine region detectors. *International Journal of Computer Vision*, 65(1/2):43–72, 2005.

[Montgomery, 2005] D. C. Montgomery. *Design and Analysis of Experiments*. John Wiley and Sons, Inc., Hoboken, NJ, 6th edition, 2005.

[Moon and Kota, 2002] Y-M. Moon and S. Kota. Design of compliant parallel kinematic machine. In *ASME Biennial Mechanisms and Robotics Conference*, pages 1–7, Sep 2002.

[Morita and Kanade, 1997] T. Morita and T. Kanade. A sequential factorization method for recovering shape and motion from image streams. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(8):858–867, Aug 1997.

[Murray *et al.*, 1994] R. M. Murray, Z. Li, and S. S. Sastry. *A Mathematical Introduction to Robotic Manipulation*. CRC Press, 1994.

[Nelson and Khosla, 1996] B.J. Nelson and P. K. Khosla. Force and vision resolvability for assimilating disparate sensory feedback. *IEEE Transactions on Robotics and Automation*, 12(5):714–731, Oct 1996.

[Nelson, 2004] B. J. Nelson. Bio micro robotics: Robotics for exploring life. *Bulletin Magazine*, Feb 2004.

[Niederošt *et al.*, 2003] M. Niederošt, J. Niederošt, and J. Scucka. Automatic 3d reconstruction and visualization of microscopic objects from a monoscopic multifocus image sequence. *International Archives of the Photogrammetery, Remote Sensing and Spatial Information Sciences*, 2003.

[Patel *et al.*, 2006] G.M. Patel, G.C. Patel, R.B. Patel, J.K. Patel, and M. Patel. Nanorobot: A versatile tool in nanomedicine. *Journal of Drug Targeting*, 14(2):63–67, Feb 2006.

[Pollefeys *et al.*, 2004] M. Pollefeys, L. V. Gool, M. Vergauwen, F. Verbiest, K. Cornelis, J. Tops, and R. Koch. Visual modeling with a hand-held camera. *International Journal of Computer Vision*, 59(3):1573–1405, Sep 2004.

[Popa and Stephanou, 2004] D. O. Popa and H. E. Stephanou. Micro and meso scale robotic assembly. *SME Journal of Manufacturing Systems*, 2004.

[Prasad and Joy, 2003] M. S. Prasad and D. C. Joy. Is SEM noise gaussian. *Microscopy Society of America*, pages 982–983, 2003.

[Pressigout and Marchand, 2007] M. Pressigout and E. Marchand. Real-time hybrid tracking using edge and texture information. *International Journal of Robotics Research*, 26(7):689–713, 2007.

[Prinz *et al.*, 1999] V. Y. Prinz, V. A. Seleznev, and A. K. Gutakovsky. Self-formed InGaAs/GaAs nanotubes: Concept, fabrication, properties. In

D. Gershoni, editor, *International Conference on the Physics of Semiconductors*, Singapore, 1999. World Scientific.

[Probst *et al.*, 2007] M. Probst, R. Borer, and B. J. Nelson. A microassembly system for manufacturing hybrid mems. In *IFToMM World Congress*, Jun 2007.

[Quan, 1996] L. Quan. Self-calibration of an affine camera from multiple views. *International Journal of Computer Vision*, 19(1):93–105, 1996.

[Raspanti *et al.*, 2005] M. Raspanti, E. Binaghi, I. Gallo, and A. Manelli. A vision-based, 3D reconstruction technique for scanning electron microscopy: Direct comparison with atomic force microscopy. *Microscopy Research and Technique*, 67(1):1–7, 2005.

[Rehg and Kanade, 1995] J. M. Rehg and T. Kanade. Model-based tracking of self-occluding articulated objects. In *International Conference on Computer Vision*, pages 612 – 617, 1995.

[Reimer, 1993] L. Reimer. *Transmission Electron Microscopy: Physics of Image Formation and Microanalysis*. Science, 1993.

[Requicha, 1999] A. A. G. Requicha. *Handbook of Industrial Robotics*, chapter Nanorobotics. Wiley and Sons, 1999.

[Requicha, 2003] A. A. G. Requicha. Nanorobots, NEMS, and nanoassembly. In *Proceedings of the IEEE*, volume 91 of *11*, pages 1922–1933, 2003.

[Ribeiro and Shah, 2006] E Ribeiro and M. Shah. Computer vision for nanoscale imaging. *Machine Vision and Applications*, 17(3):147–162, Jun 2006.

[Ritter *et al.*, 2004] M. Ritter, M. Hemmleb, O. Sinram, J. Albertz, and H. Hohenberg. A versatile 3D calibration object for various micro-range measurement methods. In *International Society for Photogrammetry and Remote Sensing Congress*, volume V of *WG 1*, 2004.

[Roco, 2004] M.C. Roco. The future of the national nanotechnology initiative. Technical report, National Academy of Sciences, 2004.

[Roco, 2005] M. C. Roco. International perspective on government nanotechnology funding in 2005. *Journal of Nanoparticle Research*, 7(6):707–712, 2005.

[Rousseeuw and Leroy, 2003] P. J. Rousseeuw and A. M. Leroy. *Robust Regression and Outlier Detection*, chapter Algorithms, pages 197–215. John Wiley and Sons, Inc., 2003.

[Sato *et al.*, 1995] T. Sato, T. Kameya, H. Miyazaki, and Y. Hatamura. Hand-eye system in nano manipulation world. In *International Conference on Robotics and Automation*, volume 1, pages 59–66, 1995.

[Schmid *et al.*, 2000] C. Schmid, R. Mohr, and C. Bauckhage. Evaluation of interest point detectors. *International Journal of Computer Vision*, 37:151–172, Jun 2000.

[Seeger, 2004] A. Seeger. *Surface Reconstruction from AFM and SEM Images*. PhD thesis, University of North Carolina at Chapel Hill, 2004.

[Sievers and Fatikow, 2005] T. Sievers and S. Fatikow. Visual servoing of a mobile microrobot inside a scanning electron microscope. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 1350–1354, Aug 2005.

[Sievers *et al.*, 2006] T. Sievers, M. Jähnisch, C. Schrader, and S. Fatikow. Vision feedback in an automatic nanohandling station inside an SEM. In Yoshitada Katagiri, editor, *Proceedings of SPIE*, volume 6376 of *Optomechatronic Micro/Nano Devices and Components*, Oct 2006.

[Sitti, 1999] M. Sitti. *Teleoperated 2-D Micro/Nanomanipulation Using Atomic Force Microscope*. PhD thesis, University of Tokyo, Tokyo, Jun 1999.

[Sitti, 2001] M. Sitti. Survey of nanomanipulation systems. In *IEEE Conference on Nanotechnology*, pages 75–80, 2001.

[Sitti, 2004] M. Sitti. Micro- and nano-scale robotics. In *American Control Conference*, volume 1, pages 1–8, 2004.

[Smalley, 1997] R. E Smalley. Discovering the fullerenes. *Review Modern Physics*, 69(3):723–730, Jul 1997.

[Stefano *et al.*, 2003] L. Stefano, S. Mattoccia, and F. Tombari. An algorithm for efficient and exhaustive template matching. In *Image Analysis and Recognition*, volume 3211, pages 408–415. Springer Verlag, 2003.

[Stowe *et al.*, 1997] T. D. Stowe, K. Yasumura, T. W. Kenny, D. Botkin, K. Wago, and D. Rugar. Attonewton force detection using ultrathin silicon cantilevers. *Applied Physics Letters*, 71:288–290, 1997.

[Sturm, 2002] P. F. Sturm. Critical motion sequences for the self-calibration of cameras and stereo systems with variable focal length. *Image and Vision Computing*, 20(5-6):415–426, Feb 2002.

[Suganuma, 1985] T. Suganuma. Measurement of surface topography using SEM with two secondary electron detectors. *Journal of Electron Microscopy (Tokyo)*, 34(4):328–337, 1985.

[Sul *et al.*, 2006] O. J. Sul, M. R. Falvo, II R. M. Taylor, S. Washburn, and R. Superfine. Thermally actuated untethered impact-driven locomotive microdevices. *Applied Physics Letters*, 89, 2006.

[Sunyoto *et al.*, 2004] H. Sunyoto, W. van der Mark, and D. M. Gavrila. A comparative study of fast dense stereo vision algorithms. *IEEE Intelligent Vehicles Symposium*, pages 319–0324, Jul 2004.

[Sutton *et al.*, 2006] M. Sutton, N. Li, D. Garcia, N. Cornille, J. J. Orteu, S. R. McNeill, H. W. Schreier, and X. Li. Metrology in a scanning electron microscope: theoretical developments and experimental validation. *Measurement Science and Technology*, 17(10):2613–2622, 2006.

[Tardif *et al.*, 2007] J.-P. Tardif, A. Bartoli, M. Trudeau, N. Guilbert, and S. Roy. Algorithms for batch matrix factorization with application to

structure-from-motion. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, Jun 2007.

[Tomasi and Kanade, 1992] C. Tomasi and T. Kanade. Shape and motion from image streams under orthography: a factorization method. *International Journal of Computer Vision*, 9(2):137–154, Nov 1992.

[Torr and Murray, 1997] P. H. S. Torr and D. W. Murray. The development and comparison of robust methodsfor estimating the fundamental matrix. *International Journal of Computer Vision*, 24(3):271–300, 1997.

[Triggs *et al.*, 2000] B. Triggs, P. F. McLauchlan, R. I. Hartley, and A. W. Fitzgibbon. Bundle adjustment - a modern synthesis. In *Vision Algorithms: Theory and Practice*, volume 1883, pages 298–372. Springer Verlag, Feb 2000.

[Tsai, 1987] R. Tsai. A versatile camera calibration technique for high-accuracy 3d machine vision metrology using off-the-shelf tv cameras and lenses. *IEEE Journal of Robotics and Automation*, 3(4):323–344, Aug 1987.

[Vidal and Hartley, 2004] R. Vidal and R. Hartley. Motion segmentation with missing data using PowerFactorization and GPCA. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2 of *2*, pages 310–316, 2004.

[Vogl *et al.*, 2004] W. Vogl, M. Sitti, and M. F. Zah. Nanomanipulation with 3D visual and force feedback using atomic force microscopes. In *IEEE Conference on Nanotechnology*, pages 349–351, 2004.

[Vollmers *et al.*, 2008] K. Vollmers, D. Frutiger, B. E. Kratochvil, and B. J. Nelson. Wireless resonant magnetic actuation for untethered microrobots. *Applied Physics Letters*, 92(14), Apr 2008.

[Wautelet and Duvivier, 2007] M. Wautelet and D. Duvivier. The characteristic dimensions of the nanoworld. *European Journal of Physics*, 28:953–959, 2007.

[Wautelet, 2001] M. Wautelet. Scaling laws in the macro-, micro-and nanoworlds. *European Journal of Physics*, 22:601–611, 2001.

[Weck *et al.*, 1997] M. Weck, J. Huemmler, and B. Petersen. Assembly of hybrid microsystems in a large-chamber scanning electron microscope by use of mechanical grippers. In *Proceedings of SPIE*, volume 3223 of *Micromachining and Microfabrication Process Technology III*, pages 223–229, Sep 1997.

[Weir *et al.*, 2005] N. A. Weir, D. P. Sierra, and J. F. Jones. A review of research in the field of nanorobotics. Technical Report SAND2005-6808, Sandia National Laboratories, Oct 2005.

[Weng *et al.*, 1992] J. Weng, P. Cohen, and M. Herniou. Camera calibration with distortion models and accuracy evaluation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(10):965–980, Oct 1992.

[Wunsch and Hirzinger, 1997] P. Wunsch and G. Hirzinger. Real-time visual tracking of 3D objects with dynamic handling of occlusion. In *IEEE International Conference on Robotics and Automation*, volume 4, pages 2868–2873, Apr 1997.

[Yesin and Nelson, 2005] K. B. Yesin and B. J. Nelson. A CAD model based tracking system for visually guided microassembly. *Robotica*, 23:409–418, Jun 2005.

[Yesin, 2003] Kemal Berk Yesin. *CAD Model Based Tracking for Visually Guided 3D Microassembly*. PhD thesis, University of Minnesota, 2003.

[Yu *et al.*, 1999] M. F. Yu, M. J. Dyer, G. D. Skidmore, H. W. Rohrs, X. K. Lu, K. D. Ausman, J. R. Ehr, and R. S. Ruoff. Three-dimensional manipulation of carbon nanotubes under a scanning electron microscope. *Nanotechnology*, 10(3):244–252, 1999.

[Zesch, 1997] W. Zesch. *Multi-Degree-of-Freedom Micropositioning Using Stepping Principles*. PhD thesis, ETH Zürich, 1997.

[Zhang *et al.*, 2006] L. Zhang, E. Ruh, D. Grützmacher, L. X. Dong, D. J. Bell, B. J. Nelson, and C. Schönenberger. Anomalous coiling of SiGe/Si and SiGe/Si/Cr helical nanobelts. *Nano Letters*, 6:1311–1317, 2006.

[Zhang, 2000] Z. Zhang. A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(11):1330–1334, Nov 2000.

[Zhou *et al.*, 2002] Q. Zhou, C. Corral, P. J. Esteban, A. Aurelian, and N. K. Heikki. Environmental influences on microassembly. In *IEEE/RJS International Conference on Intelligent Robots and Systems*, pages 1760–1765, Oct 2002.

[Zhu and Cui, 2007] Z. Zhu and H. Cui. Six-DOF nanomanipulation tools for nanoscale device prototyping and fabrication. *Solid State Phenomena*, 121–123:823–827, 2007.

Figure A.1: the DISS 5 scan controller

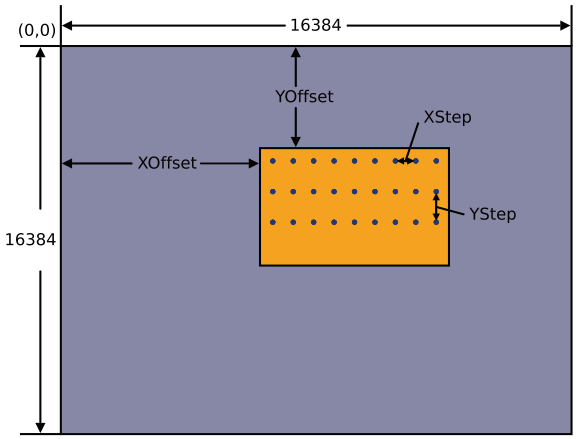


Figure A.2: layout of DISS 5 scan area

A Software

A.1 libdiss5

The scan controller used for the real-time tracking experiments with the Zeiss DSM 962 was the DISS 5 system from Point Electronics GmbH (Figure A.1). This was an after market add-on which allowed the DSM 962 to be computer controlled. Due to the nonconventional nature of the ROI scanning system, a new interface for the scan controller needed to be implemented named libdiss5. This software package provides low-level access to the scan system in the form of a C++ library on a Linux system.

The DISS 5 accepts commands over USB 2.0 which define a scan region as shown in Figure A.2. At a particular magnification, the scan controller is capable of 2^{14} pixels in both the horizontal and vertical directions. If they were scanned at native resolution, this would lead to a 256 MP image. Typically, the system is scanned by selecting an offset and a step size through

the image. Thus, a digital magnification can be accomplished by simply reducing the step size. This feature could readily be exploited for tracking systems wishing to employ a pyramiding or multi-resolution tracking scheme in hardware.

To achieve the multiple region-of-interest scanning modality, a controller determines the regions wished to be scanned. These can be manually chosen, defined by a tracking algorithm, or through a combination of methods. These regions are then queued and sequentially scanned by the controller. The current implementation then regards a group of scanned regions as part of a single virtual image which can be interchanged freely with a full-frame image in subsequent imaging computations.

A.2 Daedalus

The front-end for the libdiss5 library is manifested through the Daedalus GUI package written in QT 4.3 by Trolltech. This software is written in an inherently decoupled fashion to allow for rapidly exchanging and sharing

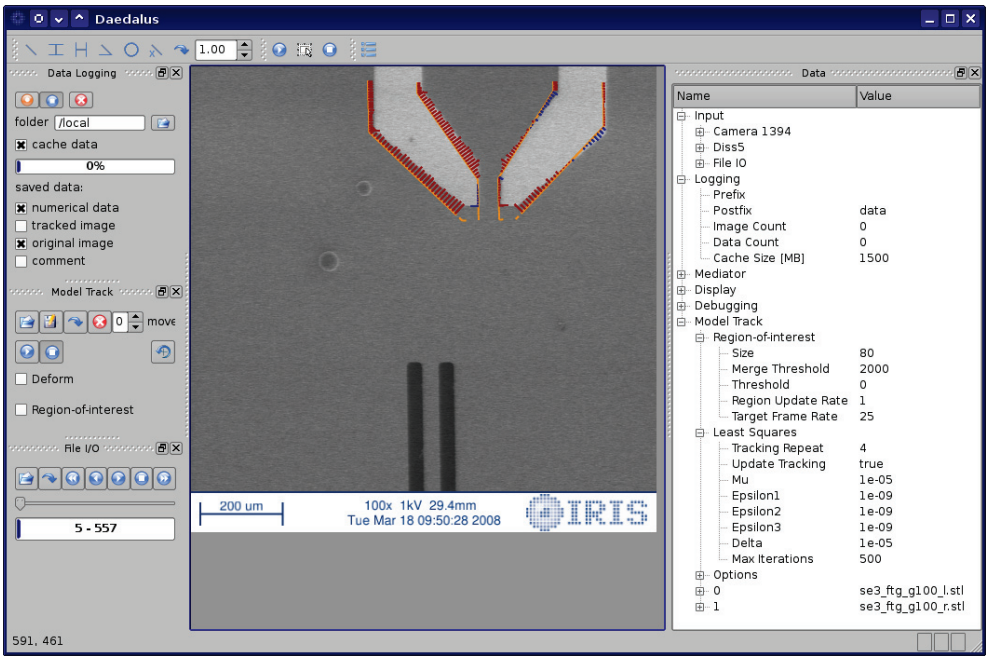


Figure A.3: Daedalus user interface for controlling the SEM

modules between different projects at the Institute of Robotics and Intelligent Systems.

Daedalus defines a number of class interfaces for different tasks commonly performed in the system, such as acquiring images, logging data, tracking images, visual servoing, or editing system parameters. These interfaces are then implemented by a series of plugins which are entirely separate from the main system. This loose-coupling allows for plugins to subscribe or publish data pipelines for transmission through the system. Each plugin can then be selectively used or omitted depending on the task at hand. This plugin modality has allowed the system to be used for a diverse range of applications from controlling the SEM [Kratochvil *et al.*, 2008] to controlling mobile micro-robots [Vollmers *et al.*, 2008; Frutiger *et al.*, 2008]. The common interface allows data processing tools to be easily shared and creates a common structure for subsequent robotics efforts.

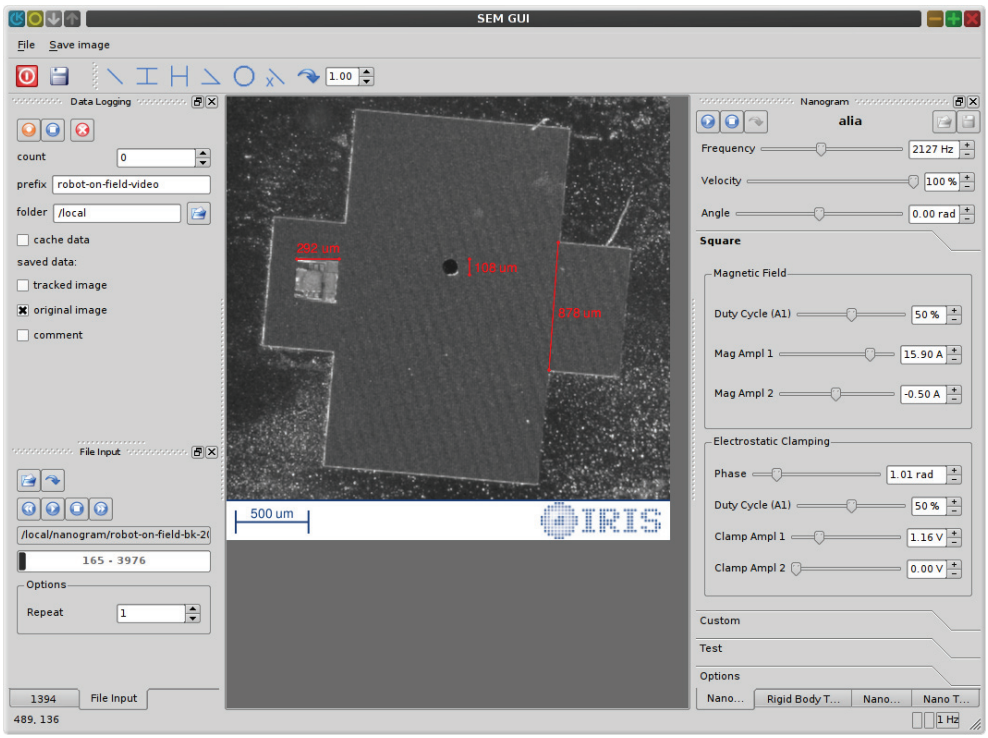
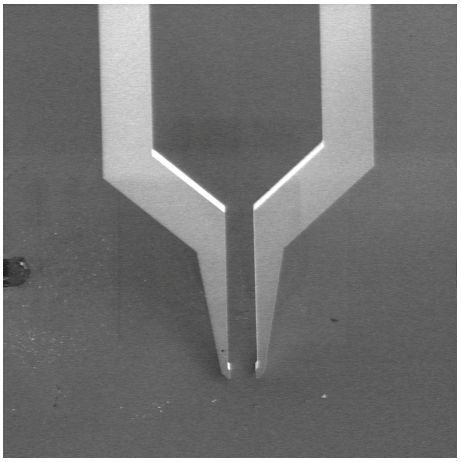


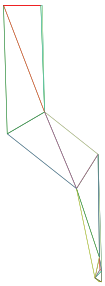
Figure A.4: Daedalus user interface for controlling mobile microrobots



(a) initial image

(b) manual segmentation

(c) edge detection



(d) polygon conversion

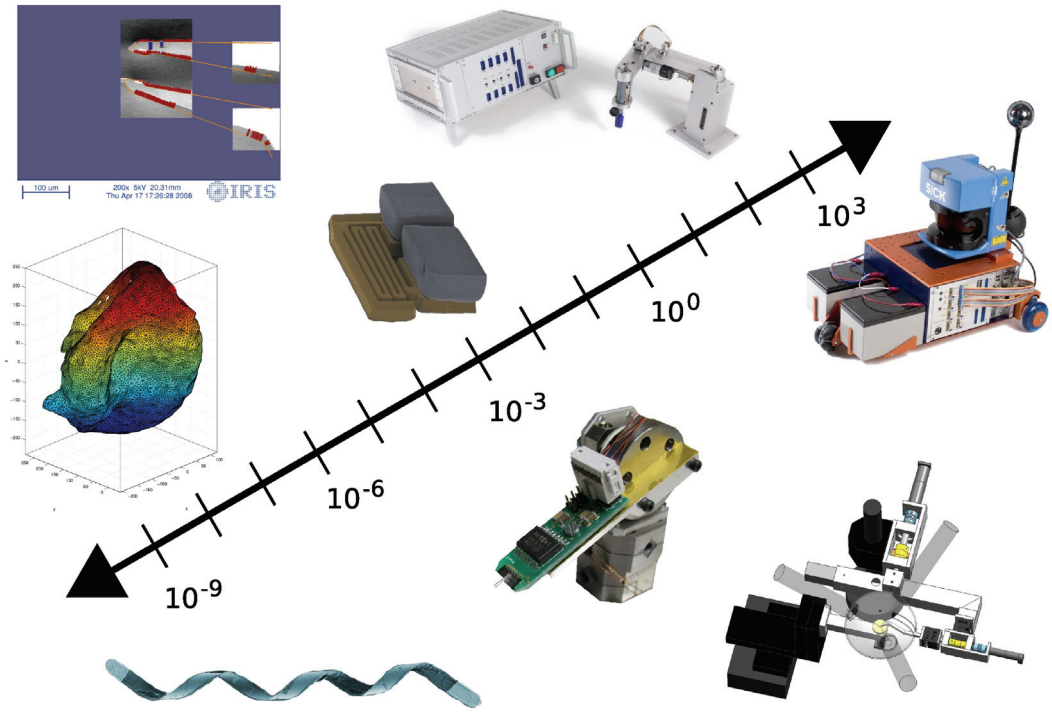
Figure B.1: semi-automated model generation for $SE(2)$

B Model Generation

The tracking system was designed to enable easy addition of new rigid-body models for tracking. To that extent, the main method for importing models is through the use of is a file format native to the stereolithography CAD software created by 3D Systems called STL. This format maintains a list of polygons that create the object of interest and is supported by various CAD software packages. When geometries of targets are know *a priori*, they can rapidly be imported through this format.

When tracking previously unknown objects, a new STL model can be rapidly generated from the image. Figure B.1 demonstrates the process of

generating these models. The original image is first manually segmented using a program such as Photoshop to provide a binary image. This step could also be automated, but at this stage manual segmentation was deemed most robust. Matlab is then used to search the subsequent binary image for edges which are linked together, and then converted to a polygon representation. This process is relatively error free and can be rapidly performed, in less than 5 minutes, during an experimentation session to guarantee a valid model for the target of interest. Future systems may implement algorithms such as active contours [Kass *et al.*, 1988] which would allow a user to automatically grow a region before a tracking experiment.



Epilogue

Over the course of my doctoral degree, I have had the privilege of working on robotic systems that either have principle dimensions or workspaces that span over eight orders of magnitude. This work has allowed me to expand my experience to encompass such diverse fields as software engineering, electronics design, CAD drafting, machining, electron microscopy, kinematics, control, haptics, microfabrication, and engineering education. This has been to a large part due to the diverse research areas of the Institute of Robotics and Intelligent System's Multi-Scale Robotics Lab, and the flexibility offered to me by my adviser.

One of the primary lessons learned from this work is the need for different strategies at different scales. As devices and tasks approach micro- and nanoscales, macroscale solutions are no longer feasable and radically different approaches must be considered. As we ever increasingly search for minaturization, microrobotics sits at a crossroads between different research fields

such as physics, biology, and chemistry. Robotics as a discipline has the capability to provide the tools and framework for dialog between the different disciplines in the quest to intelligently control the world of the "very, very small."

Confronted with this confluence of fields, a microrobotics engineer must walk the line of being a jack of all trades with the risk of mastering none. As my work progressed it became apparent that although my research necessarily dealt with a wide array of different technologies, my passion and focus lies with software and system's engineering. From this standpoint, concepts such as simplicity, modularity, and extenability apply at all scales of robotics systems, and have provided me a basis for confronting challenges at any order of magnitude.

Bradley E. Kratochvil

Institute of Robotics and Intelligent Systems
ETH Zurich
Tannenstrasse 3
CH-8092 Zürich
Switzerland

Tel: +41 (43) 632 69 21
Fax: +41 (43) 632 10 78
brad@kratochvil.name
http://brad.kratochvil.name/

Citizen of the United States of America
Citizen of the Republic of Italy

Education

Ph.D. Mechanical Enginnering, ETH - Zurich, 2008

M.S. Computer and Information Sciences, University of Minnesota - Minneapolis, 2003

B.S. Computer Engineering, University of Nebraska - Lincoln, 2001

Professional Experience

Research Assistant — 2003 - current
Institute of Robotics and Intelligent Systems — Zürich, Switzerland
Development of computer vision and control methods for micro- and nanorobotic systems.

Research Assistant — 2001–2003
Center for Distributed Robotics — Minneapolis, MN
Development lead on MegaScout and COTS Scout platforms.

Student Consultant — 2000–2001
Microsoft — Lincoln, MN
Promote and teach Microsoft technologies at the University of Nebraska.

Intern - Software Test Engineer — 2000
Microsoft — Redmond, WA
Design and implementation of GUI-based lab configuration utility.

Intern - Electronics Technician — 1998-2000
J. A. Woollam Co. Inc. — Lincoln, NE
Support and manufacture of numerous electronic platforms.

Research Assistant — 1997
Institute for Human Computer Interaction — Lincoln, NE
Designed a multimedia program to help teach C++.

Teaching Experience

Teaching Assistant, Advanced Robotics and Mechatronic Systems, ETH 2003-2008

Course Design and Teaching Assistant, Introduction to Robotics and Mechatronics, ETH 2003-2006

Teaching Assistant, Introduction to Robotics and Mechatronics, UMN 2003-2006

Teaching Assistant, Senior Design Project, UNL 2003-2006

Bradley E. Kratochvil

Activities

Student Representative, Curriculum Committee, UNL, 2000–2001

Vice Chairman, Association for Computing Machinery, UNL, 2000–2001

Chairman, Association for Computing Machinery, UNL, 1998–2000

Public Relations Chiar, Association for Computing Machinery, UNL, 1997–1998

Alumni Chair, Acacia Fraternity, UNL, 1998–1999

Recruitment Chair, Acacia Fraternity, UNL, 1998

Acacia Leadership Academy, 1997

LeaderShape Academy, 1997

Academic Publications

B. E. Kratochvil, L. X. Dong, L. Zhang, J. J. Abbott, and B. J. Nelson, "Nanohelices as motion converters," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, *to appear* 2009.

B. E. Kratochvil, L. Dong, and B. J. Nelson, "Real-time rigid-body visual tracking in a scanning electron microscope," *The International Journal of Robotics Research*, *to appear* 2009.

D. R. Frutiger, K. Vollmers, B. E. Kratochvil, and B. J. Nelson, "Small, fast, and under control: wireless resonant magnetic micro-agents," in *International Symposium on Experimental Robotics*, Jul 2008.

D. R. Frutiger, B. E. Kratochvil, K. Vollmers, and B. J. Nelson, "Magmites - wireless resonant magnetic microrobots," in *IEEE International Conference on Robotics and Automation*, 2008.

K. Vollmers, D. Frutiger, B. E. Kratochvil, and B. J. Nelson, "Wireless resonant magnetic actuation for untethered microrobots," *Applied Physics Letters*, vol. 92, no. 14, Apr 2008.

L. X. Dong, L. Zhang, B. E. Kratochvil, K. Shou, and B. J. Nelson, "Dual-chirality helical nanobelts: A novel linear-to-rotary motion converter," in *IEEE International Conference on Micro Electro Mechanical Systems*, no. 21, Jan 2008.

B. E. Kratochvil, L. Dong, L. Zhang, and B. J. Nelson, "Automatic nanorobotic characterization of anomalously rolled-up sige/si helical nanobelts through vision-based force measurement," in *IEEE Conference on Automation Science and Eneering*, Sep 2007.

B. E. Kratochvil, L. Dong, and B. J. Nelson, "Real-time rigid body visual tracking for micro/nanorobotic manipulation," in *IEEE International Conference on Nanotechnology*, Aug 2007.

M. Probst, K. Vollmers, B. E. Kratochvil, and B. J. Nelson, "Design of an advanced microassembly system for the automated assembly of bio-microrobots," in *International Workshop on Microfactories*, Oct 2006.

B. E. Kratochvil, K. B. Yesin, V. Hess, and B. J. Nelson, "Design of a visually guided 6DOF micromanipulator system for 3D assembly of hybrid MEMS," in *International Workshop on Microfactories*, Oct 2004.

B. E. Kratochvil, I. Burt, A. Drenner, D. Goerke, B. Jackson, C. McMillen, C. Olson, N. Papanikolopoulos, A. Pfeifer, S. A. Stoeter, K. Stubbs, and D. Waletzko, "Heterogeneous implemen-

tation of an adaptive robotic sensing team," in *IEEE International Conference on Robotics and Automation*, Apr 2003.

B. E. Kratochvil, I. Burt, T. Dahlin, A. Drenner, M. Gini, D. F. Hougen, C. McMillen, B. J. Nelson, N. Papanikolopoulos, P. E. Rybski, S. A. Stoeter, K. Stubbs, D. Waletzko, and K. B. Yesin, "A new generation of scout robots with multiple mobility modes," in *IEEE International Conference on Robotics and Automation*, May 2002.

A. Drenner, I. Burt, B. E. Kratochvil, B. J. Nelson, N. Papanikolopoulos, and K. B. Yesin, "Communication and mobility enhancements to the scout robot," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, Oct 2002.

A. Drenner, I. Burt, B. Chapeau, T. Dahlin, B. E. Kratochvil, C. McMillen, B. J. Nelson, N. Papanikolopoulos, P. E. Rybski, K. Stubbs, D. Waletzko, and K. B. Yesin, *Design of the UMN Multi-Robot System*, ser. Multi-Robot Systems: From Swarms to Intelligent Automata. Kluwer, 2002.