ANNÉE 2016

UNIVERSITÉ DE RENNES 1

UNIVERSITE BRETAGNE LOIRE

THÈSE / UNIVERSITÉ DE RENNES 1
sous le sceau de l'Université Bretagne Loire

pour le grade de

DOCTEUR DE L'UNIVERSITÉ DE RENNES 1

Mention : Informatique

École doctorale Matisse

présentée par

Le Cui

préparée à l'unité de recherche IRISA – UMR6074
Institut de Recherche en Informatique et Système Aléatoires
Université de Rennes 1

Robust micro/nano-positioning by visual servoing

Thèse soutenue à Rennes
le 26 janvier 2016

devant le jury composé de :

Stéphane RÉGNIER
Professeur à l'Université Pierre et Marie Curie / *Président du jury*

Antoine FERREIRA
Professeur à l'INSA Centre Val de Loire / *Rapporteur*

Jacques GANGLOFF
Professeur à l'Université de Strasbourg / *Rapporteur*

Nadine LE-FORT PIAT
Professeur à l'ENSMM / *Examinatrice*

Sounkalo DEMBÉLÉ
Maître de conférence à l'Université de Franche-Comté / *Examinateur*

Éric MARCHAND
Professeur à l'Université de Rennes 1 / *Directeur de thèse*

Acknowledgments

To my jury

I would like to express my appreciation to all the thesis committee members for having accepted to be in my jury. First of all, my special thanks to Prof. Stéphane Régnier, for having honored me by being the president of the jury, and also for his supervision and encouragement, when I performed the experiments at UPMC. I sincerely thank Prof. Antoine Ferreira and Prof. Jacques Gangloff for having accepted to read and review the manuscript of my work, with their proficiency and expertise. I would also like to especially thank Prof. Nadine Piat and Assoc. Prof. Sounkalo Dembélé, for their supervision and helpful advice to my work at FEMTO-ST.

I want to deeply thank Prof. Éric Marchand, my thesis advisor, for giving me this precious opportunity to work under his supervision. During all the three years of my Ph.D. study, his patience, his indispensable advice, his constant guidance, his scientific rigor and his confidence in me, all these have enabled me to overcome the challenges in my research. I also appreciate his effort on solving the particular administrative issues that I met in the first few months of my Ph.D. study, that have encouraged me to accomplish the goal of the research and to finish my thesis.

To my colleagues in the project

The work in this thesis has been realized in the context of ANR Nanorobust project, with the collaboration of ISIR-UPMC and FEMTO-ST institute. I am grateful to Prof. Philippe Lutz, for his direction and organization of the Nanorobust project. I would like to thank Assoc. Prof. Sinan Haliyo and Assoc. Prof. Mokrane Boudaoud, for their warm and continuous support and patience in setting up the experimental system at ISIR. I also thank Dr. Naresh Marturi and Dr. Brahim Tamadazte, for their necessary support and their knowledge on microscopes in my work at AS2M department, FEMTO-ST. My gratitude is also extended to Jean Abrahamians, Camille Dianoux and Soukeyna Bouchebout for their useful assistance in experimental setups.

To Lagadic and Irisa people

I would like to express my sincere thanks to Prof. François Chaumette, for his leadership of this dynamic and innovative research team and especially for his effort to help me to overcome the troublesome issues in this three years. My warmest thanks also go to Céline Ammoniaux, for her understanding of my difficulties and her essential assistance in the first few months of my Ph.D. study that encouraged me to finish my thesis and Hélène de la Ruée for her useful assistance and kind help in my work and my defense. I am grateful to Fabien Spindler, for his patience and professionalism, which helped me to solve technical problems. I appreciate Alexandre for his jokes and enthusiasm, Marie for her daily "bisou" and also Paolo and Vincent for their nice presentations on their works that broadened my outlook on robotics.

I am grateful to Frédéric Renouard and Stéphanie Gauvain, for their kindly help when I was in the office at EIT ICT labs.

The research team is dynamic, but its spirit remains always the same. I would like to thank François Pasteau for his useful experience and Super Mario games, Clément for his kindness, Vishnu for his enthusiasm and Pedro for his calm and kindness. It is a great pleasure for me to work with them in the same office. I thank Aurelien and Souriya for their useful help in software and coding. I thank Laurent, Antoine, Rafiq, Bertrand and Manikandan for their kindness and useful suggestions that helped me in these three years. I thank Riccardo for his hard-working that is a good example to me. I thank Aly, Suman, Giovanni, Nicolas, Pierre, Lucas, Lesley, Quentin, Jason, Fabrizio, Thomas and Noël, for all their nice company in these years. I am very grateful to all the team members who created a friendly atmosphere and organized some excellent activities which are quite unforgettable, in the working time and after working.

To my family and my friends

I wish to thank my parents and all my family members, for their love, their understanding, and their encouragement. This thesis cannot be finished without their support. My thanks also go to my cousin for his useful advice in the writing part. Finally, I would like to thank all my friends in Rennes, in France or elsewhere, whom I have not seen for years and who have always shown their useful support and help when I need.

Contents

Introduction

Microtechnology and nanotechnology have received much attention over the last couple of decades. They are expected to play an important role in electronics, medical, information technology, etc. Nowadays, it becomes possible to handle and assembly devices at micro- or nano-scale. In order to satisfy these increasing demands, developing reliable, efficient and robust micro- and nano-manipulation tasks is then necessary. Many works have been performed in these related fields, such as measurements of nano-scale objects, analysis of materials, as well as micro/nano-positioning and assembly. However, due to the difficulties to observe and to control tiny objects, the automation of micro/nano-manipulation is always a bottleneck in micro/nano-robotics.

Vision is one of the most indispensable ways to observe the world. Vision-based control is an efficient solution for control problems in robotics. However, the object at micro/nano-scale cannot be observed by an ordinary camera or human eyes. The micro/nano-vision is usually performed by a microscope (e.g., Scanning Electron Microscope (SEM)) where the size of the sample image is amplified for our observation. Due to particular conditions and image formation models in a microscope, the vision-based control under a microscope should be studied particularly. This work aims to analyze these problems and to propose solutions using vision-based control approach to perform the micro/nano-positioning tasks.

This work has been conducted in the context of the ANR NANOROBUST project from the French National Research Agency (Agence Nationale de la Recherche). The project is entitled "Multi-physics characterization and robotic manipulation of nano-objects in SEM". Four French laboratories participate in this project: FEMTO-ST (Besançon), IRISA (Rennes), ISIR (Paris), and LPN (Marcoussis). This project concerns two research themes: (1) manipulations of tiny objects by a control approach in order to put them on a base for transporting them to the measurement system; (2) analysis of the structural properties of these objects under a SEM without damaging or contaminating the objects.

This thesis concerns the vision-based control in SEM task in the project. The motivation of this thesis is to realize robust micro/nano-positioning tasks in a SEM. One of the challenges related to these tasks is that the SEM produces images differently from an optical microscope. In this case, the SEM imaging process has to be studied first, especially on SEM calibration process considering the distortions. In fact, at high

magnifications, the geometric projection model of SEM is different from that of an optical camera. Instead of using a perspective projection model, the parallel projection models should be considered at high magnifications. It should be noticed that it is difficult to observe the motion along the depth direction through the SEM image. In order to perform a 6-DoF positioning task in a SEM, the control of the motion along the depth direction should then be adequately studied. Generally, there are two possible approaches to perform a robust micro/nano-positioning task in a SEM. The first focuses on the pose estimation and using the image registration techniques to minimize the error between the projections of the object CAD model on the images for a given pose and the observations in the images. By estimating the pose of the objects, the positioning task can be achieved using a classic control method. It is obvious that different projection models should be considered for the pose estimation and control. The second solution, the direct visual servoing approach, does not need any a priori condition of the object. Unlike local geometrical features, such as the position of a point, a line, or the 3D pose, the image appearance information (e.g., intensity, gradient, etc.) provides novel solutions to the positioning problem. By developing a visual servoing approach based on this direct information, vision-based control in SEM can be achieved. In this thesis, the envisaged goal is to propose a reliable and robust solution for micro/nano-positioning task by visual servoing using the image appearance information.

Contribution of this thesis

With the above objectives, the contributions of this thesis are stated below. Based on the study on SEM image formation geometry and the sensor projection model, we propose to use a non-linear optimization process to perform the SEM calibration, which is fundamental on robotic vision studies in a SEM. In this study, we show that the depth information cannot be recovered from the variation of the features position on the SEM images. This work has been published in [C4] and [J1].

A visual servoing framework for automated micro-positioning has been proposed. We prove that the image photometric information can be used as a visual feature. In order to solve the previously mentioned particular problem along the depth direction and to perform a 6-DoF automated visual servoing task in a SEM, we propose to use the defocus information as a visual feature for visual servoing along the depth direction. A hybrid visual servoing scheme has been proposed for 6-DoF micro/nano-positioning using image appearance information. This method has been validated in a SEM. This work has been partially published in [C3] and [C2].

Using the image gradient as a sharpness function, a closed-loop control framework has been proposed for SEM autofocusing. In this work, the control law is designed to maximize the image gradient to achieve the optimal focus configuration. This work has been published in [C1].

Finally, considering the defocus information, we propose a template-based visual

tracking approach to estimate the 3D pose of a micro-object in a SEM. This method is robust to the defocus blur caused by the motion along the depth direction since the defocus level is modeled in the visual tracking framework.

Organization of this thesis

This thesis is organized as follows. Chapter 1 presents the background on SEM imaging. Among the numerous microscopes, SEM plays an important role in the work of this thesis. In this chapter, the SEM structure, SEM image formation, and some other relative issues are introduced.

Chapter 2 is dedicated to a fundamental support for the vision in SEM: calibration. The SEM projection models are discussed. A non-linear optimization process for SEM calibration as well as experimental results is introduced.

Chapter 3 to Chapter 5 present the major contribution in this thesis, the automation of 6-DoF micro/nano-positioning by visual servoing. As the most important tool for robot motion control, the basics of visual servoing, as well as the overview of its application in micro/nano-scale are reviewed in Chapter 3.

In Chapter 4, we focus on the main challenge in our work: visual servoing along the depth direction. The selection of visual features and the visual servoing scheme for the robot motion along the depth direction is presented.

Based on the study in Chapter 4, a hybrid visual servoing scheme for 6-DoF micro/nano-positioning is proposed in Chapter 5. The experimental validations using both optical camera and SEM are also illustrated.

Chapter 6 presents a closed-loop control scheme for SEM autofocusing and the experimental validations.

Chapter 7 addresses a visual tracking and 3D pose estimation process in a SEM. The performance of the proposed method is shown by experimental results.

Finally, we conclude the proposed approaches. Future perspectives are suggested.

Thesis relative publications

Academic Journals

[J1] L. Cui, E. Marchand. Scanning electron microscope calibration using a multi-image non-linear minimization process. *Int. Journal of Optomechatronics*, 9(2):151-169, May 2015.

International Conferences

[C1] L. Cui, N. Marturi, E. Marchand, S. Dembélé, N. Piat. Closed-Loop Autofocus Scheme for Scanning Electron Microscope. In *Int. Symp. of Optomechatronics Technology, ISOT 2015*, Neuchatel, Switzerland, October 2015.

[C2] L. Cui, E. Marchand, S. Haliyo, S. Régnier. Hybrid Automatic Visual Servoing Scheme using Defocus Information for 6-DoF Micropositioning. In *IEEE Int. Conf. on Robotics and Automation, ICRA'15*, pp. 6025-6030, Seattle, WA, May 2015.

[C3] L. Cui, E. Marchand, S. Haliyo, S. Régnier. 6-DoF automatic micropositioning using photometric information. In *IEEE/ASME Int Conf. on Advanced Intelligent Mechatronics, AIM'14*, pp. 918-923, Besançon, July 2014.

[C4] L. Cui, E. Marchand. Calibration of Scanning Electron Microscope using a multi-images non-linear minimization process. In *IEEE Int. Conf. on Robotics and Automation, ICRA'14*, pp. 5191-5196, Hong Kong, China, June 2014.

Chapter 1

Background on SEM imaging

MIRCROSCOPY is one of the most important techniques in observing the objects in micro/nano-scale. Different from other optical microscopes, a SEM produces images by scanning the sample surface with a focused beam of high-energy electrons. In this chapter, various backgrounds regarding the SEM imaging are presented. Starting from the various microscopy techniques, we detail the principle of the SEM structure and SEM image formation. At last, some important factors about the SEM image quality are addressed.

1.1 Observing the micro- and nano-world

In recent years, the rapid development in micro- and nano-technologies leads to significant research interests on the micro and nanorobotics tasks such as handling and assembly of objects on microscale or nanoscale. Visual information is one of the most important sense in MEMS. In order to visualize the manipulation tasks in micro/nano-scale, it is necessary to use a specific microscopic vision system to provide high-quality images. A microscope is an essential instrument to observe the micro/nano-scale objects. These microscopic systems can be categorized into various types based on the used imaging principle, such as optical microscopes, electron microscopes and scanning probe microscopes. They are briefly introduced below.

The optical microscope is also called light microscope. It uses visible light and a system of lenses to magnify images of small objects. The noted first microscope components were probably invented by Hans Lippershey and Zacharias Janssen in 1590s [Van Helden et al., 2010]. Another possible inventor of the microscope was Galileo Galilei. He developed a compound microscope with a convex and a concave lens in 1609. Later in the 17th century, the microscope was used for research in Italy, the Netherlands and England, including the work of Robert Hook (see Figure 1.1) and Antonie van Leeuwenhoek. It played an important role in the biologic and medicine research. Typical magnification of an optical microscope is up to 1250x with a theoretical resolution limit of around 0.250 micrometers or 250 nanometers.

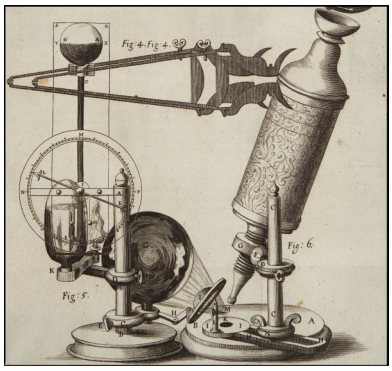


Figure 1.1: Robert Hooke's early microscope, from *Micrographia*, (London, 1665)

In 1878, Ernst Abbé proved that the resolution of the optical microscope is limited by the wavelength of the light [Freundlich, 1963]. Ernst Ruska and Max Knoll developed the first electron microscope (transmission electron microscope, TEM) in 1931. In a TEM, a beam of electrons is transmitted through an ultra-thin specimen, interacting with the specimen as it passes through. An image is formed from the interaction of the electrons transmitted through the specimen. In 1935, Max Knoll produced a photo with a 50-mm object-field-width showing channeling contrast by the use of an electron beam scanner. A scanning electron microscope (SEM) with high magnifications

by scanning the specimen with a demagnified and finely focused electron beam was invented by Manfred von Ardenne. A SEM produces images by probing the specimen with a focused electron beam that is scanned across a rectangular area of the specimen (raster scanning).

The first scanning probe microscope was the scanning tunneling microscope (STM), which was developed by Gerd Binnig and Heinrich Rohrer in 1981 [Binnig and Rohrer, 1983]. The STM is based on the concept of quantum tunneling. In an STM, a stylus analyzes the surface structure of the sample by scanning the surface from a specified distance. The STM can provide images with a resolution to 0.1 nm and can be used to obtain three-dimensional (3D) images of a sample (see Figure 1.2). In 1986, Gerd Binnig, Calvin Quate, and Christoph Gerber invented the atomic force microscope (AFM). The original AFM comprised a diamond shard attached to a gold foil strip. The sample surface is in direct contact with the diamond tip, and the interaction mechanism is provided by the interatomic van der Waals forces. With an AFM, it is possible to measure the roughness of a sample surface at a high resolution, to distinguish a sample based on its mechanical properties and to perform a microfabrication of a sample.

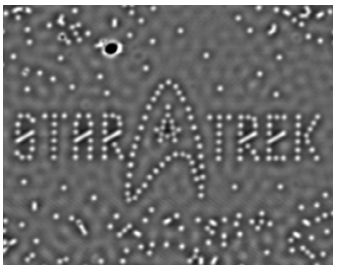


Figure 1.2: Manipulation and arrangement of atoms from IBM under a STM [IBM Research, 2009]

Recently, the fluorescence microscope, a type of optical microscope that uses fluorescence and phosphorescence to study properties of organic or inorganic substances, draws attention in the development of superresolution analysis of fluorescently labeled samples.

1.2 Scanning electron microscope

A scanning electron microscope (SEM) uses a focused beam of high-energy electrons to generate a variety of signals at the surface of samples. The signals that derive from electron-sample interactions reveal information about the sample including external morphology (texture), chemical composition, crystalline structure and orientation of materials making up the sample. In most applications, data is collected over a selected area of the surface of the sample, and a two-dimensional (2D) image is generated. Using conventional SEM techniques, areas ranging from approximately 1 cm to 1 µm

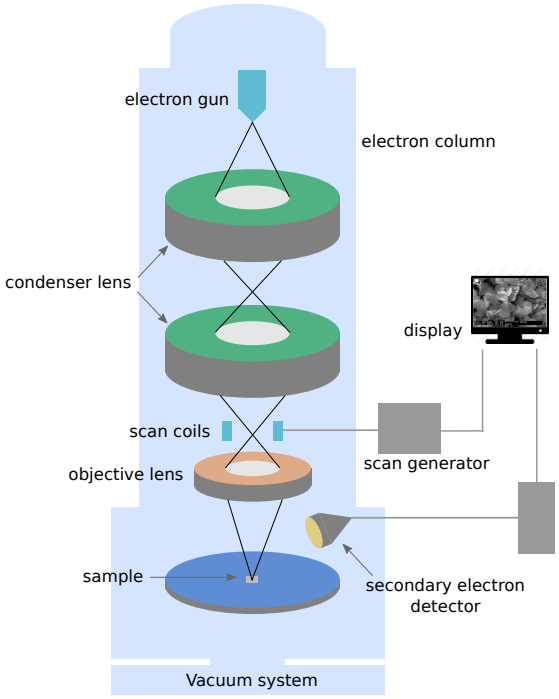


Figure 1.3: Conventional SEM architecture

in width can be imaged. In a typical SEM, the magnification can range from 10× to approximately 500,000×, and spatial resolution can attain better than 1 nm.

Compared with an optical microscope, SEM shows advantages in several areas. The first is the resolution and magnification. Resolution can be defined as the least distance between two close points, at which they are recognized as two separate entities. The best resolution possible in an optical microscope is about 200 nm whereas a typical SEM has a resolution of better than 1 nm.

Another advantage is the depth of field. The depth of field is the height of a sample that appears in-focus in an image. In a SEM, the depth of field can be more than 300 times than that of an optical microscope. This means that great topographical detail can be obtained. For many users, the 3D appearance of the sample image is the most valuable feature of the SEM. This is because such images, even at low magnifications, can provide much more information about a sample than is available using the optical microscope.

Last, the SEM provides not only the morphology information of the sample like an optical microscope but also the analysis of sample composition including chemical composition, as well as crystallographic, magnetic and electrical characteristics.

A conventional SEM architecture with the major components in the SEM electron column is shown in Figure 1.3. The essential components of a SEM include:

- Electron gun
- Electromagnetic lenses

- Scan coils
- Electron detectors for different signals
- Sample stage
- Display / data output devices
- Infrastructure requirements:
  - Power supply
  - Vacuum system
  - Cooling system
  - Vibration-free floor
  - Room free of ambient magnetic and electric fields

The details on SEM components has been presented in [Egerton, 2005, Goldstein et al., 2003]. The main components are described below.

1. Electron gun

An electron gun produces electrons by thermionic heating. The electrons are then accelerated to a voltage between 1-40 kV and condensed into a narrow beam which is used for imaging and analysis. There are three commonly used types of electrons sources: Tungsten (W) filament, solid state crystal (CeB_6 or LaB_6) and field emission gun (FEG). Tungsten filament consists of an inverted V-shaped wire of tungsten, about 100 µm long, which is heated resistively to produce electrons. This is the most basic type of electron source. On the other hand, cerium hexaboride (CeB_6) or lanthanum hexaboride (LaB_6) based electron gun is a thermionic emission gun. It is the most common high-brightness source. This solid state crystal source offers about 5-10 times the brightness and a much longer lifetime than tungsten. The FEG is a wire of tungsten with a very sharp tip, less than 100 nm, which uses field electron emission to produce the electron beam. The small tip radius improves emission and focusing ability.

2. Electromagnetic lenses

There are two sets of electromagnetic lens present in the electron column: the condenser lenses and the objective lenses. Condenser lenses focus the electron beam as it moves from the source down the column. Objectives lenses are under the aperture. It focuses the incoming beam on the sample surface. The narrower the beam the smaller the spot it will have when contacting the surface. This is always called spot size or probe diameter.

3. Scan coils

After the beam is focused, scan coils are used to deflect the beam in the X and Y axes so that it scans in a raster fashion over the surface of the sample.

4. Electron detector

SEMs always have at least one detector and most have additional detectors. The specific capabilities of a particular instrument are critically dependent on which detectors it accommodates. The detector collects the electrons coming from the sample surface. Two types of electrons are typically used for imaging: secondary electrons (SE) and backscattered electrons (BSE).

Secondary electron detector: secondary electrons are low energy electrons produced when electrons are ejected from the k-orbitals of the sample atoms by the beam. The most popular detector in SEMs is the Everhart-Thornley detector. It consists of a Faraday cage which accelerates the electrons towards a scintillator.

Backscattered electron detector: Backscattered electrons are higher energy electrons that are elastically backscattered by the atoms of the sample. Atoms with higher atomic numbers backscatter more efficiently and, therefore, this detector can give compositional information about the sample. These detectors can either be scintillators or semiconductors.

5. Vacuum chamber and sample stage

In general terms, samples are mounted into a vacuum chamber and placed on a positioning stage. The positioning stage consists of translation (along x, y, z axes), tilt (around x, y axes) and rotation (around z axis) movements.

A vacuum chamber is one of the mandatory conditions for an electron beam. It is a rigid enclosure from which air and other gases are removed by a vacuum pump. It is employed to avoid collisions between electrons and the extraneous gas molecules and to project the filament from oxidation. A typical pressure in a vacuum chamber in a SEM is about 10^{-4} to 10^{-6} Torr.

1.3 SEM image formation

The physics of the SEM image formation has been presented in [Reimer, 1998]. Different from an optical microscope where observation of the sample and formation of its image occurs simultaneously, the SEM constructs images progressively by scanning the surface of the sample using the electron beam generated from the electron gun. The electrons are redirected by the anode in the electron column. The electromagnetic lenses and apertures control the beam diameter and focus the beam on to the surface of the sample. Generally, an area of the sample surface is focused and scanned by the electron beam in both X and Y directions with a variable scan speed. The direction is controlled by the scan coils by changing the current passing through as a function of time according to a raster pattern. The emitted signals caused by the interaction of the electron beam on the surface of the sample are then detected by different electron detectors. The received

signals are amplified, processed and finally transferred to a monitor that displays the SEM image. This process is shown in Figure 1.4.

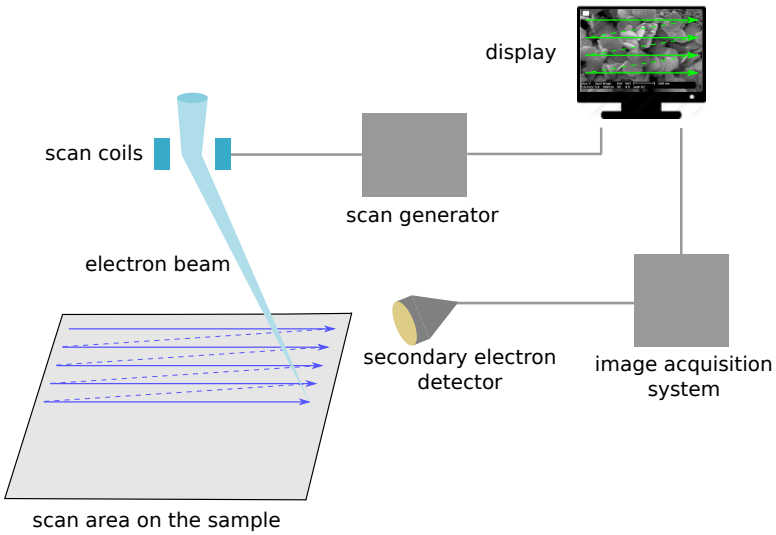


Figure 1.4: Image formation process in a SEM

There are some parameters playing an important role during the SEM image acquisition.

1. Magnification

The magnification M between the sample space and the display space is given by the ratio of the lengths of the scans:

$$M = \frac{L_{display}}{L_{sample}} \tag{1.1}$$

where L denotes the length of the scans. The numerical value of the magnification reported on the alphanumeric display typically refers to the final image format recorded on the SEM photographic system. Since the display length is fixed, increase or decrease in magnification is achieved by respectively reducing or increasing the length of the scan on the sample. It depends only on the excitation of the scan coils but not on the excitation of the objective lens, which determines the focus of the beam.

2. Scan speed

In an analog scanning system, the beam is moved continuously with a rapid scan along the X axis (line scan) supplemented by a stepwise slow scan along the Y axis at predefined lines. In a digital scanning system, only discrete beam locations are allowed. The beam is positioned at a particular location and remains there for a fixed time, called dwell time τ, and then it is moved to the next point. The

scan speed of a SEM can be measured as an amount of pixels (or images) that is scanned in units of time (e.g., second (s), millisecond (ms), microsecond (µs) or nanosecond (ns)). Generally, a SEM provides a wide range of available scan speeds. An image of 1024 × 768 pixels can be acquired in a hundred milliseconds by a fast scan speed or in more than a dozen seconds by a slow scan speed. It should be noticed that the scan speed plays an important role in the manipulation tasks. With a slow scan speed, the SEM produces the image in good quality but it costs a long time. A good manipulation task should consider both the time consumption and the image quality and find a balance on the scan speed.

3. Working Distance (WD)

In a SEM, the working distance is defined as the distance between the lower pole piece of the objective lens and the plane at which the electron beam is focused. It can be considered as a "focal length" of the SEM. It should be noticed that, using the definition from an optical microscope, the WD can also be expressed literally as the distance between the lower pole piece of the objective lens and the sample plane in some literature [Goldstein et al., 2003, Hafner, 2007]. It is equivalent to the former definition when the sample is focused (i.e. the general condition). As the former definition plays a vital role in SEM focusing, the former definition is used in this thesis in order to avoid any confusion.

4. Depth of Field

The depth of field that can be obtained is one of the most striking aspects of SEM images. The depth of field is the range of distances in object space for which object points are imaged with acceptable sharpness with a fixed position of the image plane. It corresponds to a range from the focused point to the top side and the bottom side in which the image is accepted to be in-focus. Out of this distance, the image is considered to be blurred or defocused.

The depth of field D can be expressed as [Brisset et al., 2012]:

$$D \approx \frac{2r_{pixel}}{\alpha M} \tag{1.2}$$

where r_{pixel} is the pixel size on the image, M is the magnification. α is the beam converge angle, the half-angle of the cone of electrons converging onto the sample. It can be approximated using the working distance W and the aperture radius R_a:

$$\alpha \approx \frac{R_a}{W} \tag{1.3}$$

The illustration of the depth of field D for a small aperture and a large aperture is shown in Figure 1.5.

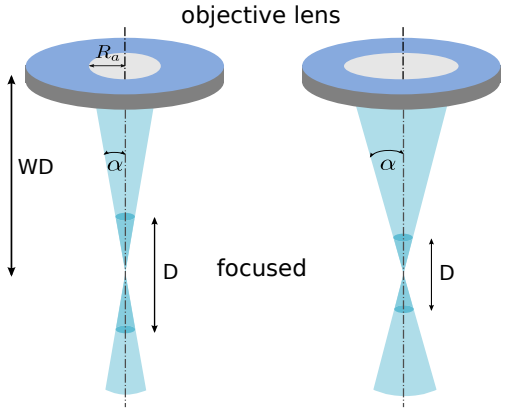


Figure 1.5: Depth of field for small aperture (left) and large aperture (right)

1.4 SEM image quality issues

Since SEM image formation is different from the image formation of an optical microscope, there are some particular issues about the SEM image quality. The major issues are discussed as follows.

1.4.1 Noise

As for most imaging devices, noise is a major issue in SEM. The noise appears at multiple stages in a SEM, each contributing its own noise component to the final SEM image. Generally, in the SEM image acquisition, the signal is affected during the beam production, interaction of the electrons on the sample surface and also by the presence of instabilities in the electron column [Reimer, 1998]. The major sources to be considered are noise in the primary beam, secondary emission noise, and noise in the final detection system [Sim et al., 2004]. It is difficult to model a single noise source into an image formation process.

In [Mulapudi and Joy, 2003], the noise on the final image from a thermionic gun SEM can be considered to follow Gaussian distributions. For the purpose of creating artificial SEM images, [Cizmar et al., 2008] has considered that the final image noise is an addition of a Poisson distribution representing primary emission and a Gaussian distribution representing the other types of noise in the SEM.

In signal processing, the signal-to-noise ratio (SNR) is a widely used indicator to measure the noise level of a signal. It is defined as the ratio of signal power to the noise power, often expressed in decibels. A general expression is

$$SNR = \frac{\sigma^2_{signal}}{\sigma^2_{noise}}. \tag{1.4}$$

A statistic model of SEM image SNR has been proposed in [Timischl et al., 2012]. The noise is modeled by Poisson-based statistics. A noise variance estimation approach has

been proposed in [Sim et al., 2013] using the image noise cross-correlation estimation model.

In order to reduce the noise (i.e. improve the SNR), a lot of methods have been proposed during the image acquisition as well as the image processing. In [Hafner, 2007], adequate probe current is essential to produce images with the necessary contrast and signal to noise ratio. This probe current I_p at the sample can be written as [Reimer, 1998]

$$I_p = \frac{j_p \pi d_p^2}{4}, \tag{1.5}$$

where j_p is the probe current density and d_p is the spot size (i.e. the diameter of the final beam at the surface of the sample). Considering the probe current density is proportional to the axial gun brightness, the SNR can be improved by increasing the axial gun brightness and increasing the spot size. Another way to improve the SNR is to employ a slow scan speed (i.e. a large dwell time). Indeed, it is necessary to use some combination of high beam current and a slow scan speed in order to detect objects of all size and low contrast in SEM.

After the acquisition of the image, one simple and commonly used method to reduce the noise is frame averaging. Many SEMs provide a hardware or software unit to average the frames before displays the SEM image on the monitor. The frame averaging can be expressed using:

$$f(x, y) = \frac{1}{N} \sum_{i=1}^{N} f_i(x, y). \tag{1.6}$$

The most widely discussed approach to reduce the noise in digital image processing is image filtering, including linear filtering and nonlinear filtering. A good quality image is acquired by passing the original image through a predefined filter in space, frequency or other transform domain to reduce the noise and keep the original information. The details about these methods can be found in [Gonzalez and Woods, 2008, Pratt, 2013].

1.4.2 Distortion

In general, the image in a SEM can be affected by two types of distortions: spatial distortion and time-dependent distortion (drift). There are many factors that cause the spatial distortion in the final SEM image. One possible distortion is introduced by the scanning system. In order to attain a low magnification, the scan area is relatively enlarged. The angle between the electron beam and the optical axis becomes important in these magnifications [Brisset et al., 2012]. In this case, the observed distortion increases with the distance d_{scan} between the optical axis and the beam focused area:

$$d_{scan} = W \tan \theta \tag{1.7}$$

where W is the working distance and θ is the angle between the electron beam and the optical axis. This relation is shown in Figure 1.6. When the magnification is reduced,

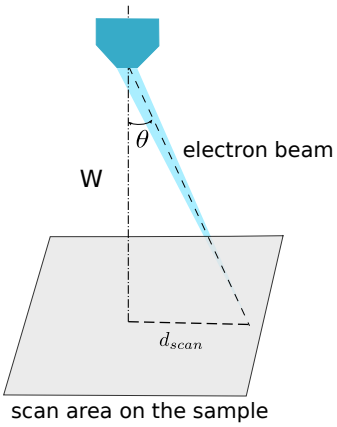


Figure 1.6: Distortion introduced by scanning system

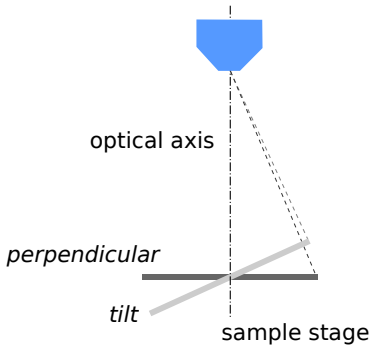


Figure 1.7: Distortion introduced by nonalignment of sample stage

the variation on d_{scan} could be no longer a linear progression. This non-linearity could produce an error in the final observed image.

Moreover, the distortion can also occur when the sample stage is not perpendicular with the optical axis. In the ideal case, the normal direction of the sample plane should be aligned with the optical axis in order to have a uniform electron beam on the sample surface. However, lack of this alignment could lead to the difference on the distance between the objective lens and the scan area (see Figure 1.7). This difference causes the different magnifications on the different scan areas. It should be noticed that this distortion can be particularly visible on the samples in the presence of the regular and orthogonal structure.

The final image can also be affected and deformed by other phenomena during the scanning process, such as the problem of synchronization of the scan on X and Y, the non-linearity on the scanning and the hysteresis phenomena on the scan speed [Goldstein et al., 2003].

1.4.3 Drift

The drift is mainly due to the presence of nonlinearities and instabilities when the electron beam scans a sample surface by [Maune, 1976, Mizuno et al., 1997]. It is always observed with consecutive scans although all the SEM parameters are unchanged. The drift can be characterized as the evolution of all the pixels during consecutive scans.

According to previous research [Cornille, 2005, Sutton et al., 2007], the drift between pixels or between lines of an image is negligible; and the drift between the two images can be considered as integrated. In order to remove or correct the drift, two types of compensation methods have been proposed. The first approach uses a reference image to estimate and correct the drift on-line [Snella, 2010, Cizmar et al., 2011, Malti et al., 2012b]. The alternative method is based on an empirical model [Cornille, 2005], which is used to compensate the drift. In [Sutton et al., 2007], the drift on each pixel is determined using a velocity and is fitted by B-splines with respect to time. Recently, the image drift has been compensated using an image-registration-based method [Marturi et al., 2013b]. In this method, the correction on the distorted image is performed by computing the homography, using the keypoint correspondences between the images.

1.5 Conclusion

In this chapter, we have presented fundamental background knowledge on SEM imaging. The structure and the components of a SEM are presented, including the electron gun, lenses, electron detectors, etc. As an important issue in our work on visual servoing in a SEM, the SEM image formation process and some other factors are detailed. With this basic knowledge, in the next chapter we will deal with the calibration method for a SEM.

Chapter 2

SEM Calibration

SEM is an electron microscope where a focused beam of electrons is used to scan the surface of a specimen. This is an essential instrument to display, measure and manipulate the micro and nano-structure with a micrometers or nanometers accuracy. When the task requires the computation of metric information from the acquired 2D images, the calibration of the SEM is an important issue to be considered. In this chapter, an overview on the calibration of an optical sensor and a SEM is first stated. As the basic knowledge, the camera geometrical imaging model and the projection models are presented. We propose to use a non-linear optimization process for SEM calibration. The SEM calibration for the intrinsic parameters is achieved by an iterative non-linear optimization algorithm which minimize the registration error between the current estimated position of the pattern and its observed position. The experimental results from two different SEMs proved the efficiency of the proposed approach. This work has been partially published in IEEE Int. Conf. on Robotics and Automation, ICRA 2014 [C4] and in International Journal of Optomechatronics [J1].

2.1 SEM Calibration overview

Calibration of an optical sensor has been widely investigated over the last decades. The goal of the calibration process is to determine the set of parameters which defines the relationship between the 3D coordinates of an object point on the observed specimen and its projection in the image plane (such parameters include, in an optical system, the focal length, the dimension of pixel, the location of principle points on the image plane are named intrinsic parameters). This issue is usually considered as a registration problem. Some authors use linear techniques (e.g., [Faugeras and Toscani, 1987]), where the least squares method is employed to estimate the intrinsic parameters and the pose (i.e., the position and the orientation of the calibration pattern frame in the sensor frame). Other techniques use non-linear optimization methods [Brown, 1971]. It consists in minimizing the error between the observation and the forward-projection of the model. In [Tsai, 1987] and [Wei and Ma, 1994], a linear estimation of some parameters is considered and the others are estimated iteratively. Alternatively, another technique [Ma et al., 2004], called self-calibration, does not use any calibration pattern. The parameters are estimated by moving a camera in a static scene, where constraints are provided by the scene rigidity in this approach.

Since the structure of a scanning electron microscope is very different from the structure of an optical microscope, it became apparent that novel image analysis, geometrical projection models and calibration processes would be necessary in order to extract accurate information from the SEM images. [Postek et al., 1993] has demonstrated that the accurate SEM calibration, as well as error analysis, was one of the major problems when considering such sensor. In earlier studies, the photogrammetric analysis of the SEM has been considered by some authors [Boyde, 1973, Ghosh, 1975]. Several photogrammetric related calibration methods [Boyde, 1970, Wergin, 1985, Minnich et al., 1999] have been proposed for the 3D imagery and reconstructions in SEM.

The projection models relate to a 3D point on a specimen in the observed space to its projection in the 2D image. The perspective projection, where objects are projected towards a point (the center of projection), is used in classical camera models. The parallel projection (typically orthographic projection) corresponds to a perspective projection with an infinite focal length. The projection rays and the image plane is perpendicular in parallel projection model. It is noticed that this projection model is similar to the model used for telecentric lenses [Li and Tian, 2013, Chen et al., 2014]. In [Chen et al., 2014], a telecentric stereo micro-vision system is calibrated by solving a problem of sign ambiguity induced by the planar-object-based calibration technique. Previous studies on SEM consider that at low magnifications, the perspective projection model can be applied because the observed area and the electron beam sweep angle are both large. At higher magnifications, the center of projection is usually considered at infinity so the parallel projection model is assumed. However, the practical limit

between the choice of the perspective projection and parallel projection model is not clear. Some experiments [Cornille et al., 2003, Sinram et al., 2002] show that parallel projection is assumed at a magnification of 1000× and higher. [Howell, 1978] has concluded that the use of the parallel projection depends on the desired accuracy for the calculation of the position of a point on the specimen.

As mentioned in Section 1.4, another important issue in calibration is the distortion. It contains the spatial distortion (static distortion) and the time-dependent drift (temporally-varying distortion). The drift is mainly due to the presence of nonlinearities and instabilities in the raster scan of a specimen surface by the electron beam [Maune, 1976, Mizuno et al., 1997]. This drift can be calibrated and be compensated as shown in [Cornille, 2005, Sutton et al., 2006, Malti et al., 2012b]. However, few authors have investigated the spatial distortion for an accurate calibration of SEM. One reason might be the complexity of modeling of distortions at a high magnification, where the common model of distortion is weakened. Several articles [Lacey et al., 1996, Sinram et al., 2002] ignore distortion and consider only a pure projection model. A few authors [Ghosh, 1975, Hemmleb and Albertz, 2000] consider the spatial distortion with parametric models. Spatial distortion including radial distortions and spiral distortions are introduced in their geometric model. [Schreier et al., 2004] has proposed to use a priori distortion estimation technique in combination with bundle-adjustment [Brown, 1976, Triggs et al., 2000] for an accurate calibration of SEM. In [Cornille et al., 2003], the distortion removal function is determined before the calibration stage. In this method, good guesses are required in the measurement to ensure the accuracy.

Furthermore, [El Ghazali, 1984] has proposed the so-called system calibration for a SEM since the traditionally laboratory calibration is not convenient for complex systems where the compensation and the deterioration effect between the different system components are not taken into account. Recently, a landmark-based 3D calibration strategy [Ritter et al., 2006] has been proposed. It considers a 3D micrometer-sized reference structure with the shape of a cascade slope-step pyramid. However, the manufacture of this special 3D reference structure is important and difficult. Since different scales of magnification are needed in some applications, [Malti et al., 2012a] considers the modeling magnification-continuous parameters of the static distortion and the projection of the SEM. [Zhu et al., 2011] has proposed a stereo-vision system under a SEM. The system has been calibrated using distortion-corrected images of a planar object and grid for various orientations [Sutton et al., 2009].

2.2 Geometrical imaging model

In computer vision, in order to describe the position of an object in the 3D world and in the camera, it is necessary to define the *frame* where the coordinates are expressed. Indeed, an object can be expressed to any frame; its coordinates relatively depend on the origin of the defined frame. In computer vision, there are two major frames to be

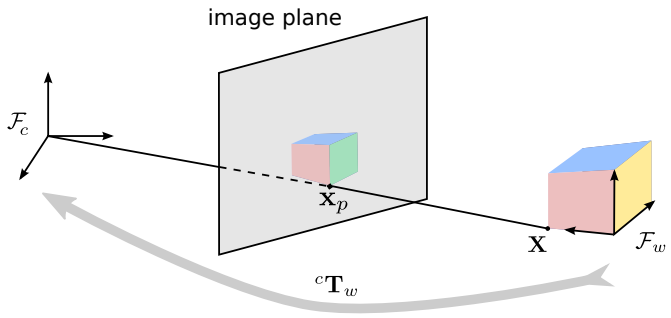


Figure 2.1: Projecting the object expressed in a world frame $\mathcal{F}_w$ to image plan

considered: the sensor (e.g. camera) frame $\mathcal{F}_c$ attached to the vision sensor and the object frame $\mathcal{F}_o$ attached to the object. Sometimes, the world frame $\mathcal{F}_w$ attached to a predefined origin is also widely employed. This frame is usually used when there are multiple objects or both the vision sensor and the object(s) move during the task. In a simple case, the origin of $\mathcal{F}_w$ can be located on the origin of $\mathcal{F}_o$, meaning the two frames are identical. Considering a point of the object ${}^w\mathbf{X} = ({}^wX, {}^wY, {}^wZ)^\top$ expressed in $\mathcal{F}_w$, its coordinates expressed in $\mathcal{F}_c$ can be written as ${}^c\mathbf{X} = ({}^cX, {}^cY, {}^cZ)^\top$. How can we describe the relation between these two coordinates? It is necessary to employ the transformation linking $\mathcal{F}_c$ and $\mathcal{F}_w$.

Denote ${}^c\mathbf{T}_w$ a homogeneous matrix describing the transformation of a point in $\mathcal{F}_w$ to $\mathcal{F}_c$:

$${}^c\mathbf{T}_w = \begin{pmatrix} {}^c\mathbf{R}_w & {}^c\mathbf{t}_w \\ \mathbf{0}_{3\times1} & 1 \end{pmatrix} \tag{2.1}$$

where ${}^c\mathbf{R}_w$ and ${}^c\mathbf{t}_w \in \mathbb{R}^3$ are the rotation matrix and translation vector that define the position of the vision sensor expressed in $\mathcal{F}_w$. ${}^c\mathbf{R}_w$ respects the orthogonality constraints:

$${}^c\mathbf{R}_w \in SO(3) \quad \text{where} \quad SO(3) = \{{}^c\mathbf{R}_w \in \mathbb{R}^{3\times3} | {}^c\mathbf{R}_w{}^\top {}^c\mathbf{R}_w = \mathbf{I}_3, det({}^c\mathbf{R}_w) = 1\} \tag{2.2}$$

Here $SO(3)$ is called Special Orthogonal group. ${}^c\mathbf{T}_w$ belongs to the Special Euclidean group $SE(3)$ defined by:

$${}^c\mathbf{T}_w \in SE(3) \quad \text{where} \quad SE(3) = \{{}^c\mathbf{T}_w = \begin{pmatrix} {}^c\mathbf{R}_w & {}^c\mathbf{t}_w \\ \mathbf{0}_{3\times1} & 1 \end{pmatrix} | {}^c\mathbf{R}_w \in SO(3), {}^c\mathbf{t}_w \in \mathbb{R}^3\}. \tag{2.3}$$

Finally, the point expressed in $\mathcal{F}_w$ can be transformed to $\mathcal{F}_c$:

$${}^c\mathbf{X} = {}^c\mathbf{T}_w {}^w\mathbf{X}. \tag{2.4}$$

This transformation of frames is shown in Figure 2.1.

2.3 Projection models

In this section, we focus on the geometrical calibration of the system projection model. The final objective of our work is to perform visual servoing tasks for the object positioning and manipulations in a SEM. Therefore for simplicity issues classical projection models are considered. Whereas such model has a clear physical meaning when considering optical devices, this is no longer the case with a SEM. Nevertheless, for the targeted applications, considering classical projection models has been proved to be sufficient [Ghosh, 1975] (such assertion may no longer be true for, e.g., structure characterization). It is, however, important to determine the nature of the projection models to be considered [Hemmleb and Albertz, 2000, Howell, 1978]: perspective or parallel models (see Figure 2.2). In this section, both the perspective and parallel projection models including modeling of the image distortion are discussed. Figure 2.3 illustrates the perspective projection and the parallel projection models.

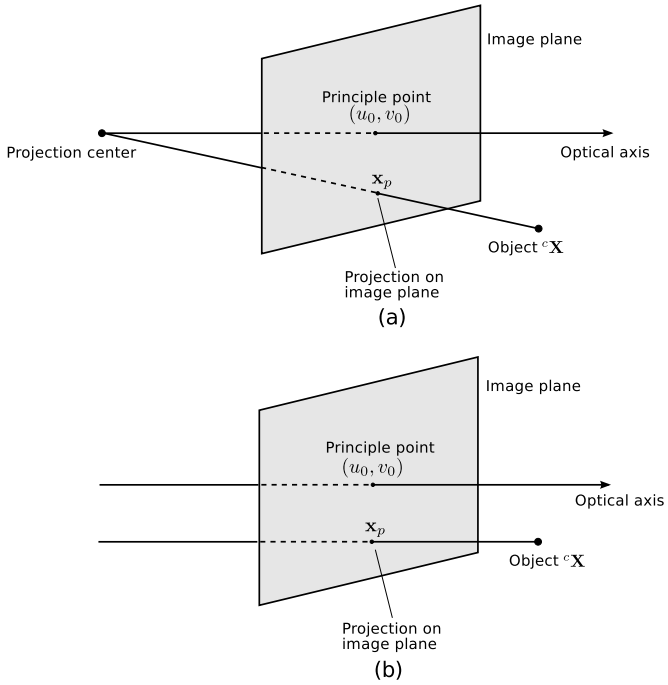


Figure 2.2: Projection models: (a) perspective projection (b) parallel projection

2.3.1 Perspective projection

Let ${}^c\mathbf{X} = ({}^cX, {}^cY, {}^cZ, 1)^\top$ be the homogeneous coordinates of a point on the observed object expressed in the sensor frame $\mathcal{F}_c$ (located on the projection center). $\mathbf{x} = (x, y, 1)^\top$ is the homogeneous coordinates of its projection on the image plane expressed in normalized coordinates (i.e., in meter). It can be expressed by [Ma et al.,

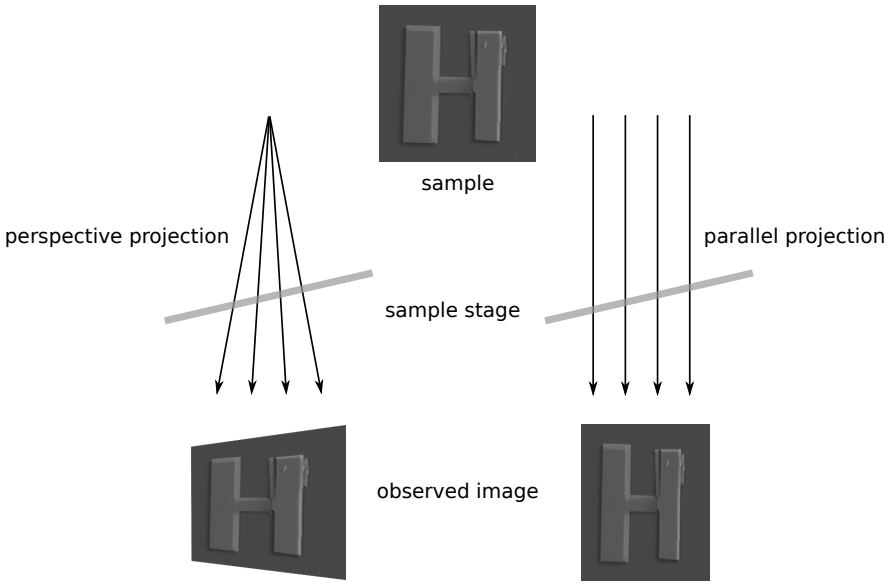


Figure 2.3: Projection models and observed images

2004]

$$\begin{cases} x = \dfrac{{}^cX}{{}^cZ} \\ y = \dfrac{{}^cY}{{}^cZ} \end{cases} \tag{2.5}$$

leading in the actual image coordinates expressed in pixel $\mathbf{x}_p = (u, v)$ on the image plane and given by

$$\begin{cases} u = u_0 + p_x x \\ v = v_0 + p_y y \end{cases} \tag{2.6}$$

where p_x and p_y represent the pixel/meter ratio and u_0, v_0 the principal point coordinates in the image plane. According to equation (2.5) and equation (2.6), the general expression of the perspective projection is:

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \underbrace{\begin{bmatrix} p_x & 0 & u_0 \\ 0 & p_y & v_0 \\ 0 & 0 & 1 \end{bmatrix}}_{\mathbf{K}} \underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}}_{\mathbf{\Pi}} \begin{bmatrix} {}^cX \\ {}^cY \\ {}^cZ \\ 1 \end{bmatrix}. \tag{2.7}$$

For the camera calibration task using perspective projection model, p_x, p_y, u_0 and v_0 are considered as intrinsic parameters. We rewrite equation (2.7) as:

$$\mathbf{x}_p = \mathbf{K}\,\mathbf{\Pi}\,{}^c\mathbf{X} \tag{2.8}$$

As already stated, for calibration issue, we consider a calibration pattern for which the position of some 3D features are known in a reference frame $\mathcal{F}_w$. Let us denote

${}^{w}\mathbf{X} = ({}^{w}X, {}^{w}Y, {}^{w}Z, 1)^{\top}$ the coordinates of a feature expressed in $\mathcal{F}_w$. Its projection in the image plane is then given by

$$\mathbf{x}_p = \mathbf{K}\,\mathbf{\Pi}\,{}^{c}\mathbf{T}_w\,{}^{w}\mathbf{X}. \tag{2.9}$$

2.3.2 Parallel projection

In parallel projection models, the projection rays are parallel. As previously mentioned, the projection center lies at infinite. The coordinates of a 2D point $\mathbf{x} = (x, y)$ corresponds to its 3D coordinates ${}^{c}\mathbf{X}$:

$$\begin{cases} x = {}^{c}X \\ y = {}^{c}Y \end{cases} \tag{2.10}$$

leading to its position expressed in pixel $\mathbf{x}_p = (u, v)$ in the digital image is

$$\begin{cases} u = p_x x \\ v = p_y y \end{cases}. \tag{2.11}$$

According to equation (2.10) and equation (2.11), the general expression of the parallel projection can be written as

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \underbrace{\begin{bmatrix} p_x & 0 & 0 \\ 0 & p_y & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{\mathbf{K}_{\perp}} \underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}}_{\mathbf{\Pi}_{\perp}} \begin{bmatrix} {}^{c}X \\ {}^{c}Y \\ {}^{c}Z \\ 1 \end{bmatrix}. \tag{2.12}$$

Since there is no longer principle point in parallel projections, only p_x and p_y are considered as the intrinsic parameters. As in the previous case, we can rewrite equation (2.12) as:

$$\mathbf{x}_p = \mathbf{K}_{\perp}\,\mathbf{\Pi}_{\perp}\,{}^{c}\mathbf{X}. \tag{2.13}$$

If we consider the 3D coordinates of 3D features in the calibration reference frame $\mathcal{F}_w$, we have:

$$\mathbf{x}_p = \mathbf{K}_{\perp}\,\mathbf{\Pi}_{\perp}\,{}^{c}\mathbf{T}_w\,{}^{w}\mathbf{X}. \tag{2.14}$$

2.4 Image distortion

Some SEM distortions have been presented in Section 1.4. Indeed, the distortion of scanning can be ignored at high magnifications. The distortion of tilt can be modeled into the calibration models. In this section, several distortions considered in the literature [Ghosh, 1975, Brown, 1976, Heikkila and Silven, 1997, Hemmleb and Albertz, 2000] are discussed and modeled.

In classical models [Heikkila and Silven, 1997], the most commonly discussed spatial distortion is radial distortion (see Figure 2.4 (a)). In the perspective projection model,

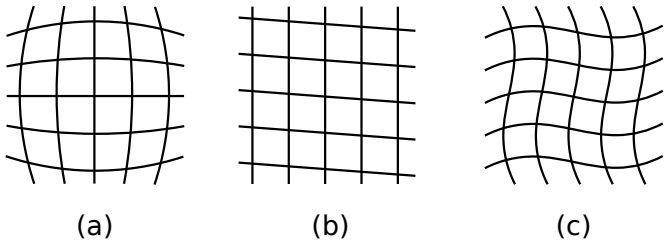


Figure 2.4: Considered distortion models: (a) radial distortion, (b) skewness (c) spiral distortion

instead of using (2.6), the relation between the point position $\mathbf{x}$ and the coordinates in the image plane $\mathbf{x}_p$ in perspective projection is expressed by

$$\begin{cases} u = u_0 + p_x x + \delta_u \\ v = v_0 + p_y y + \delta_v \end{cases}. \tag{2.15}$$

The radial distortion can be approximated using

$$\begin{cases} \delta_u = \tilde{u}(k_1 r^2 + k_2 r^4 + ...) \\ \delta_v = \tilde{v}(k_1 r^2 + k_2 r^4 + ...) \end{cases} \tag{2.16}$$

where $r^2 = \tilde{u}^2 + \tilde{v}^2$, $\tilde{u} = u - u_0$ and $\tilde{v} = v - v_0$. Usually, to compensate the radial distortion, one or two coefficients are enough. Considering the SEM geometry, it has to be noted that in SEM image such distortion appears to be very small. This should be validated by experiments.

Another issue to be considered is the skewness between the x-axis and y-axis (see Figure 2.4 (b)). In this case:

$$\begin{cases} u = u_0 + p_x x + \gamma y \\ v = v_0 + p_y y \end{cases}. \tag{2.17}$$

Typically, γ is null when the pixel in x- and y-axis is exactly rectangular.

Repeated in [Klemperer and Barnett, 1971], the spiral distortion (Figure 2.4 (c)) is caused by the spiral of the electrons within the microscope column. It is usually given by

$$\begin{cases} u = u_0 + p_x(x + \delta_x) \\ v = v_0 + p_y(y + \delta_y) \end{cases} \tag{2.18}$$

where $\delta_x = s_1(x^2 y + y^3)$, $\delta_y = s_2(x^3 + xy^2)$, s_1 and s_2 are spiral coefficients.

In the parallel projection model, the distortion models that replace (2.11) are similar but u_0 and v_0 equal to zero in equation (2.15), (2.17) and (2.18).

2.5 Non-linear calibration process

Calibration is an old research area that received much attention since the early 70's, first in the photogrammetry community (e.g., [Brown, 1971]) then in the computer

vision and robotics communities (e.g., [Faugeras and Toscani, 1987, Tsai, 1987, Weng et al., 1992], etc.). Performing the calibration leads to the estimation of the intrinsic camera parameters (image center, focal length, distortion) but also, as a by-product, extrinsic camera parameters (i.e., the pose). Various techniques exist to achieve the calibration. Among these techniques, full-scale non-linear optimization techniques (introduced within the photogrammetry community, [Brown, 1971]) have proved to be very efficient. They consist in minimizing the error between the observation and the back-projection of the model. Minimization is handled using numerical iterative algorithms such as Newton-Raphson or Levenberg-Marquartd.

2.5.1 Single image calibration

The goal of this method is to minimize the error between the points extracted from the image $\mathbf{x}_p^*$ and the projection of the model of the calibration pattern for given model parameters (both intrinsic parameters and pose) $\mathbf{x}_p(\mathbf{r}, \xi)$.

Denoting ξ the set of intrinsic parameters to be estimated and $\mathbf{r} \in se(3)$ a minimal representation of ${}^c\mathbf{T}_w$ ($\mathbf{r} = ({}^c\mathbf{t}_w, \theta\mathbf{u})^\top$ where θ and $\mathbf{u}$ are the angle and the axis of the rotation ${}^c\mathbf{R}_w$), the problem can be formulated as:

$$(\widehat{\mathbf{r}}, \widehat{\xi}) = \underset{\mathbf{r},\xi}{\operatorname{argmin}} \sum_{i=1}^{N} ({}^i\mathbf{x}_p^* - {}^i\,\mathbf{x}_p(\mathbf{r}, \xi))^2 \tag{2.19}$$

where N is the number of points used in the calibration process. For each point i, ${}^i\mathbf{x}_p(\mathbf{r}, \xi) = \mathbf{K}\mathbf{\Pi}\,{}^c\mathbf{T}_w\,{}^w\mathbf{X}_i$ (for the perspective projection model without considering distortions) and ${}^i\mathbf{x}_p(\mathbf{r}, \xi) = \mathbf{K}_\perp\mathbf{\Pi}_\perp\,{}^c\mathbf{T}_w\,{}^w\mathbf{X}_i$ (for the parallel projection model without considering distortions). The solution of this problem relies on an iterative minimization process such as a Gauss-Newton or a Levenberg-Marquardt method.

Solving equation (2.19) consists in minimizing the cost function $E(\mathbf{r}, \xi) = \|\mathbf{e}(\mathbf{r}, \xi)\|$ defined by:

$$E(\mathbf{r}, \xi) = \mathbf{e}(\mathbf{r}, \xi)^\top \mathbf{e}(\mathbf{r}, \xi), \quad \text{with} \quad \mathbf{e}(\mathbf{r}, \xi) = \mathbf{x}_p(\mathbf{r}, \xi) - \mathbf{x}_p^* \tag{2.20}$$

where $\mathbf{x}_p(\mathbf{r}, \xi) = (..., {}^i\,\mathbf{x}_p(\mathbf{r}, \xi), ...)^\top$ and $\mathbf{x}_p^* = (..., {}^i\,\mathbf{x}^*, ...)^\top$ where ${}^i\mathbf{x}_p(\mathbf{r}, \xi)$ is computed using equation (2.6) or (2.11). To simplify the notation, let us simply denote $\mathbf{e} = \mathbf{e}(\mathbf{r}, \xi)$.

To minimize this cost function, an exponential decrease of the projection error is specified:

$$\dot{\mathbf{e}} = -\lambda\mathbf{e} \tag{2.21}$$

where λ is a proportional coefficient. In equation (2.21), $\dot{\mathbf{e}}$ can be simply computed from the time variation $\dot{\mathbf{x}}_p$ which is given by:

$$\dot{\mathbf{x}}_p = \frac{\partial \mathbf{x}_p}{\partial \mathbf{r}} \frac{d\mathbf{r}}{dt} + \frac{\partial \mathbf{x}_p}{\partial \xi} \frac{d\xi}{dt} \tag{2.22}$$

where $\mathbf{r}$ represent the (virtual) sensor position along the minimization trajectory (translation and rotation), $\mathbf{v} = \dfrac{d\mathbf{r}}{dt}$ is the (virtual) sensor velocity during the minimization.

Rewrite equation (2.22):

$$\dot{\mathbf{x}}_p = \mathbf{J}_p \mathbf{V} \tag{2.23}$$

where $\mathbf{V} = \begin{bmatrix} \mathbf{v} \\ \dot{\xi} \end{bmatrix}$. Matrix $\mathbf{J}_p$ is the image Jacobian, it is given by:

$$\mathbf{J}_p = \begin{bmatrix} \frac{\partial \mathbf{x}_p}{\partial \mathbf{r}} & \frac{\partial \mathbf{x}_p}{\partial \xi} \end{bmatrix}. \tag{2.24}$$

Combining equation (2.23) and equation (2.21), $\mathbf{V}$ can be rewritten as follows:

$$\mathbf{V} = -\lambda \mathbf{J}_p^{+} (\mathbf{x}_p(\mathbf{r}, \xi) - \mathbf{x}_p^*) \tag{2.25}$$

where $\mathbf{J}_p^{+}$ is the pseudo inverse of matrix $\mathbf{J}_p$ and $\mathbf{V}$ being the parameters increment computed at each iteration of this minimization process.

2.5.2 Multi-image calibration

In practice, the intrinsic parameters are usually obtained by different viewpoints of the calibration pattern from the same camera. The optimization scheme then requires the computation of a set of positions of calibration pattern and a common set of intrinsic parameters. In that case the global error to be minimized is given by

$$E = \sum_{i=1}^{n} (\mathbf{e}_i^{\top} \mathbf{e}_i) \tag{2.26}$$

where n is the number of images used in the calibration process and

$$\mathbf{e}_i = \mathbf{x}_p(\mathbf{r}_i, \xi) - \mathbf{x}_p^*. \tag{2.27}$$

Let $\mathbf{x}_p^i$ be a set of images features extracted from the i^{th} image. In multi-image calibration, (2.23) can be rewritten as:

$$\begin{bmatrix} \dot{\mathbf{x}}_p^1 \\ \dot{\mathbf{x}}_p^2 \\ \vdots \\ \dot{\mathbf{x}}_p^n \end{bmatrix} = \mathbf{J}_p \begin{bmatrix} \mathbf{v}^1 \\ \mathbf{v}^2 \\ \vdots \\ \mathbf{v}^n \\ \dot{\xi} \end{bmatrix} \tag{2.28}$$

with

$$\mathbf{J}_p = \begin{bmatrix} \frac{\partial \mathbf{x}_p^1}{\partial \mathbf{r}^1} & 0 & \cdots & 0 & \frac{\partial \mathbf{x}_p^1}{\partial \xi} \\ 0 & \frac{\partial \mathbf{x}_p^2}{\partial \mathbf{r}^2} & 0 & 0 & \frac{\partial \mathbf{x}_p^2}{\partial \xi} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & \frac{\partial \mathbf{x}_p^n}{\partial \mathbf{r}^n} & \frac{\partial \mathbf{x}_p^n}{\partial \xi} \end{bmatrix}. \tag{2.29}$$

2.5.3 Nonlinear optimization

In a nonlinear minimization process, the optimization algorithm is an important issue. The general idea of minimizing a nonlinear function is to successively update the parameters such that the value of the cost function decreases at each iteration, as specified by equation (2.21). The Gauss-Newton method is usually used in nonlinear optimization as presented in equation (2.25).

Particularly, the measured values are small in the SEM imaging (point coordinates are expressed in micrometer (µm) and nanometer (nm)). Several numerical problems are then induced into the optimization algorithms. For example, these tiny values causes rank deficiencies of Jacobian matrix $\mathbf{J}_p$. This is why the Levenberg-Marquardt method is considered, which is numerically more efficient:

$$\mathbf{V} = -\lambda(\mathbf{J}_p^\top \mathbf{J}_p + \mu\mathbf{I})^{-1}\mathbf{J}_p^\top \mathbf{e} \tag{2.30}$$

where $\mathbf{I}$ is an identity matrix and μ is a coefficient whose typical value ranges from 0.001 or 0.0001. By modifying μ, the algorithm is set to adapt the input data and to avoid numerical issues.

2.5.4 Jacobian

In this section, the computation of $\frac{\partial \mathbf{x}_p}{\partial \mathbf{r}}$ and $\frac{\partial \mathbf{x}_p}{\partial \xi}$ in the Jacobian $\mathbf{J}_p$ is presented with the two specified projection models mentioned previously.

The image Jacobian $\frac{\partial \mathbf{x}_p}{\partial \mathbf{r}}$ relates the motion of a point $\mathbf{x}_p$ in the image with respect to the (virtual) sensor motion. It can be expressed by:

$$\frac{\partial \mathbf{x}_p}{\partial \mathbf{r}} = \begin{bmatrix} p_x & 0 \\ 0 & p_y \end{bmatrix} \mathbf{L} \tag{2.31}$$

where $\mathbf{L} = \frac{\partial \mathbf{x}}{\partial \mathbf{r}}$ is the Jacobian which relates the motion of the projection of a point on image plane (coordinates expressed in meter) to the (virtual) sensor motion.

2.5.4.1 Perspective projection

In the perspective projection model, the Jacobian $\mathbf{L}$ is given by [Comport et al., 2006]:

$$\mathbf{L} = \begin{bmatrix} -\frac{1}{Z} & 0 & \frac{x}{Z} & xy & -(1+x^2) & y \\ 0 & -\frac{1}{Z} & \frac{y}{Z} & 1+y^2 & -xy & -x \end{bmatrix}. \tag{2.32}$$

From (2.6), without considering the distortion in the camera model, the deviation of image feature $\mathbf{x}_p$ by intrinsic parameters $\xi = (p_x, p_y, u_0, v_0)$ is:

$$\frac{\partial \mathbf{x}_p}{\partial \xi} = \begin{bmatrix} x & 0 & 1 & 0 \\ 0 & y & 0 & 1 \end{bmatrix}. \tag{2.33}$$

Considering one coefficient k in radial distortion (k_1 in (2.16)), the skew factor γ and spiral coefficient s_1, s_2 as distortion parameters, the deviation of image feature $\mathbf{x}_p$ by intrinsic parameters $\xi = (p_x, p_y, u_0, v_0, k, \gamma, s_1, s_2)$ with distortion factors is:

$$\frac{\partial \mathbf{x}_p}{\partial \xi} = \begin{bmatrix} x + s_1(x^2y + y^3) & 0 \\ 0 & y + s_2(x^3 + xy^2) \\ 1 - k(r^2 + 2\tilde{u}^2) & -2k\tilde{u}\tilde{v} \\ -2k\tilde{u}\tilde{v} & 1 - k(r^2 + 2\tilde{v}^2) \\ \tilde{u}r^2 & \tilde{v}r^2 \\ y & 0 \\ p_x(x^2y + y^3) & 0 \\ 0 & p_y(x^3 + xy^2) \end{bmatrix}^\top . \tag{2.34}$$

2.5.4.2 Parallel projection

In the parallel projection model, the Jacobian is given by:

$$\mathbf{L} = \begin{bmatrix} -1 & 0 & 0 & 0 & -Z & y \\ 0 & -1 & 0 & Z & 0 & -x \end{bmatrix}. \tag{2.35}$$

Comparing with equation (2.32), it is evident that the motion along the z-axis is not observable. Therefore the depth of the calibration pattern cannot be recovered. The deviation $\frac{\partial \mathbf{x}_p}{\partial \xi}$ without distortion for $\xi = (p_x, p_y)$ is given from (2.11):

$$\frac{\partial \mathbf{x}_p}{\partial \xi} = \begin{bmatrix} x & 0 \\ 0 & y \end{bmatrix}. \tag{2.36}$$

With distortion, it is expressed with $\xi = (p_x, p_y, k, \gamma, s_1, s_2)$:

$$\frac{\partial \mathbf{x}_p}{\partial \xi} = \begin{bmatrix} x + s_1(x^2y + y^3) & 0 \\ 0 & y + s_2(x^3 + xy^2) \\ \tilde{u}r^2 & \tilde{v}r^2 \\ y & 0 \\ p_x(x^2y + y^3) & 0 \\ 0 & p_y(x^3 + xy^2) \end{bmatrix}^\top . \tag{2.37}$$

2.6 Experimental results

In the experiments, a Carl Zeiss AURIGA 60 SEM (at FEMTO-ST Institute) has been used to validate the developed calibration method. It provides a wide magnification ranges from 12× to 1,000,000×. Within the SEM a 6-DoF platform is available, including 360° continuous rotation and tilt from -15° to 70°. Another SEM Carl Zeiss

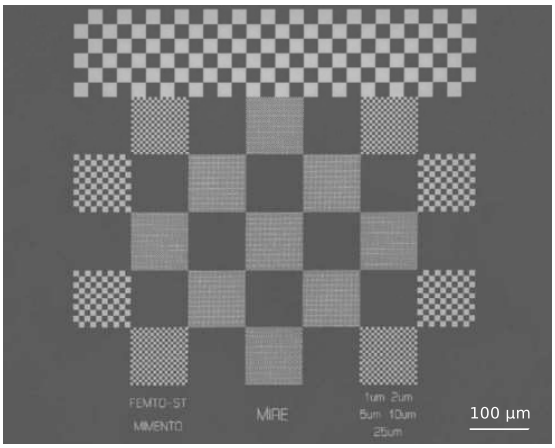


Figure 2.5: Multi-scale calibration planar, square size from 1 µm up to 25 µm

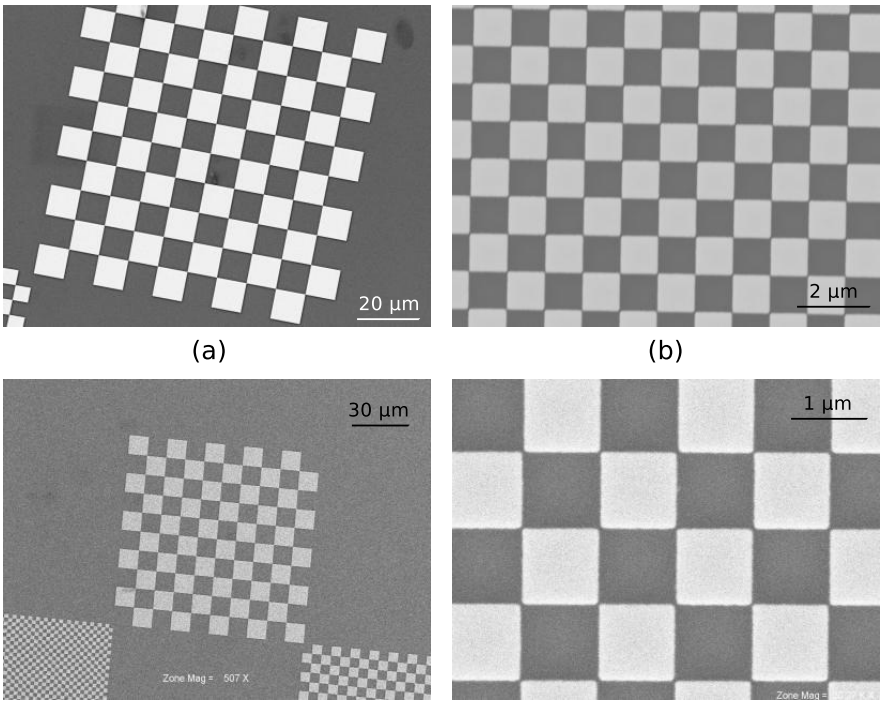


Figure 2.6: Calibration images: (a) 800×, acquired with medium scan speed (b) 10,000× acquired with medium scan speed (c) 507×, acquired with fast scan speed (d) 20,270× acquired with fast scan speed

EVO LS 25 (at ISIR, UPMC) is also employed in the experiments. The magnification of this SEM ranges from 5× to 1,000,000×.

A multi-scale planar calibration pattern[1] (see Figure 2.5) is used in the calibration procedure. It is a hierarchy of chessboard grids where the sizes of each square are 25

[1]fabricated at FEMTO-ST institute, France

µm, 10 µm, 5 µm, 2 µm and 1 µm. Acquired image size is 1024×768 pixels. Several sets of calibration images (Figure 2.6) have been acquired within the SEM with different magnifications ranging from 300× up to 10k×. The images from AURIGA 60 SEM are acquired with a medium scan speed (3.3 µs/pixel) and a fast scan speed (0.25 µs/pixel) respectively. The images from EVO LS 25 have been acquired with a medium scan speed (2.5 µs/pixel). Each group (with a given magnification) contents 7 to 9 images of the pattern acquired from various poses with rotation around z-axis ranging from 0° to 40°, and tilt from 0° to 8°.

The proposed calibration procedure has been implemented with the ViSP library [Marchand and Chaumette, 2005]. Considering the chessboard shape of the calibration pattern, OpenCV chessboard corners detector has been employed in order to obtain a precise localization of each corner. A linear algorithm has been considered to have a first approximation of the calibration parameters [Zhang, 2000]. The proposed multi-image iterative non-linear minimization method for calibration, using both perspective and projection model, is then used. The intrinsic parameters are then computed by minimizing the residual error between the projection of the pattern for the current estimated pose and the observed one.

2.6.1 Minimization process and algorithm behavior

To illustrate the behavior and performances of the proposed algorithm, we consider here the calibration of the SEM using a parallel projection model and without adding any distortion parameters.

AURIGA 60 SEM is employed in this experiment, the SEM magnification has been set to 2000×, and the size of each pattern square is of 5 µm. Eight images of the calibration pattern have been acquired from eight poses with rotation from 0° up to 20° and tilt from 0° up to 8°. The gain λ in equation (2.21) in the algorithm is set to 0.4. Figure 2.7(a) shows the residual error computed at each iteration of the minimization process. The evolution of intrinsic parameters p_x and p_y is shown in Figure 2.7(b). The residual error and the intrinsic parameters converge quickly even though the value is significant at the beginning. Only a few iterations less than 50 are required by the process. Figure 2.8 presents the estimated set of extrinsic parameters (estimated sensor poses) during the minimization process. It can be noted that, as expected, motion along the z-axis is not observable using the parallel projection model (in equation (2.35), the elements in the third column of the Jacobian which corresponds to the Jacobian of translation on z-axis are indeed null).

2.6.2 Projection models

Another experiment aims to test two projection models that can be possibly considered for the calibration of a SEM. To compare the performance with different scales, four magnifications are considered: 500×, 1000×, 2000× and 5000×. Note that it is

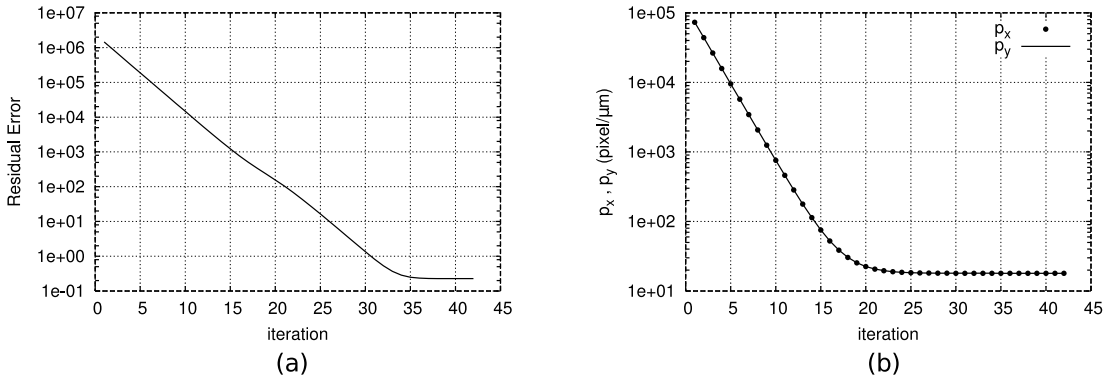


Figure 2.7: Evolution of (a) residual error in pixel and (b) intrinsic parameters p_x and p_y during the minimization process

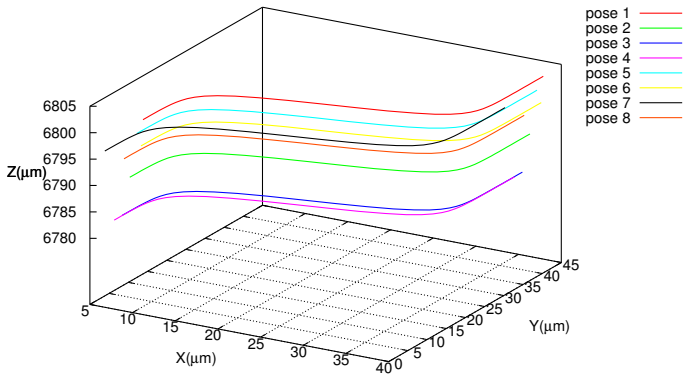


Figure 2.8: Estimated target positions during the minimization process

suggested in the literature [Cornille et al., 2003, Sinram et al., 2002] that perspective projection can be applied for a magnification up to 1000× whereas parallel projection should be considered for higher magnification. The images from AURIGA 60 SEM are firstly used in this experiment. Table 2.1 shows the estimated calibrated intrinsic parameters p_x, p_y, u_0 and v_0, the estimated distance Z_1 between sensor and calibration pattern (for the first image) and the residual error $\|\mathbf{e}\|$ in pixel. In all the cases the algorithm converges and the registration error is less than 0.5 pixel per point which correspond to the noise level in corner extraction. It is quite clear from the estimation of parameters p_x and p_y that, with the perspective projection model, intrinsic parameters are inconsistent. Nevertheless the ratio $p_x/(Z_1 M)$ (M represents the magnification) is almost constant (see Table 2.2) which confirms the fact that the difference between p_x (or p_y) and object depth is not observable. This motivates the choice of the parallel projection model for future visual servoing experiments despite the fact that depth motion are not observable.

Table 2.1: Calibration results in perspective projection

mag. (×)	p_x	p_y	u_0	v_0	$Z_1(\mu m)$	$\|\mathbf{e}\|$
500	70168.0	70058.3	511.4	384.1	15752.7	0.15
1000	201505.3	199729.8	511.6	384.4	22302.1	0.08
2000	122073.3	122312.0	511.5	384.3	6803.4	0.12
5000	103917.4	105067.7	511.5	384.0	2316.2	0.23

Table 2.2: Relation between pixel sizes and depths for various magnification: $p_{(x,y)}/(Z_1 M)$ for perspective projection and $p_{(x,y)}/M$ for parallel projection

mag. M (×)	500	1000	2000	5000
$p_x/(Z_1 M)$	0.00890	0.00903	0.00897	0.00897
$p_y/(Z_1 M)$	0.00889	0.00895	0.00898	0.00907
p_x/M	0.00895	0.00898	0.00898	0.00897
p_y/M	0.00888	0.00895	0.00904	0.00910

Table 2.2 shows $p_{(x,y)}/(Z_1 M)$ for the perspective projection and $p_{(x,y)}/M$ for the parallel projection. These factors are approximately a constant value in the two projection models.

A wide range of magnifications from 300× to 10k× considering the parallel projection model have been tested. The images are acquired by a medium scan speed (see Figure 2.6 (a), (b)). Results are shown in Table 2.3. The intrinsic parameters of AURIGA 60 SEM through magnifications are shown in Figure 2.9. The ratio between the computed intrinsic parameters p_x, p_y and magnification M is almost constant: as expected a quasi linear relation exists between p_x, p_y and magnification. It has to be noted that the residual error $\|\mathbf{e}\|$ is slightly more important for low magnification meaning that parallel projection model is less appropriate at low magnification (300×, 500×) which confirms earlier report [Sinram et al., 2002]. $\|\mathbf{e}\|$ also increases at high magnifications, but the reason is that at low magnification the extraction of corner position on the calibration pattern used in this experiment is far more accurate than that at high magnifications.

To compare the performance of the proposed calibration process within different conditions, another set of images is acquired using fast scan speed (see Figure 2.6 (c), (d)). Results are shown in Table 2.4. It can be seen that the calibration results keep stable while the scan speed has been changed. From the results, $\|\mathbf{e}\|$ increases using fast scan speed due to the noise introduced into the images.

Table 2.5 shows the calibration results on parallel projection using EVO LS 25 SEM. It can be noticed from the table the ratio between p_x, p_y and M is also almost constant as that in Table 2.3. Since the calibration images are acquired from different SEMs

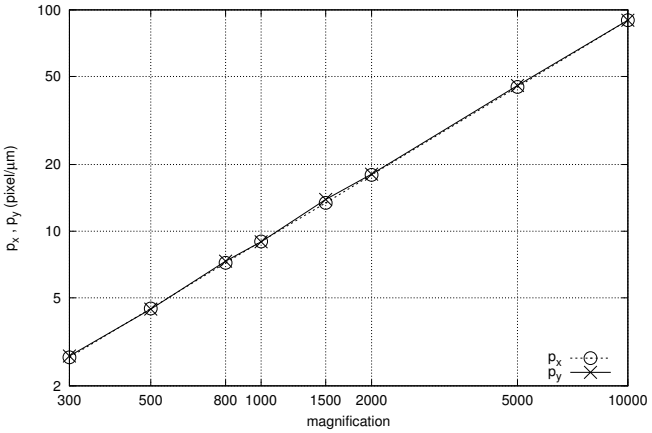


Figure 2.9: Intrinsic parameters from AURIGA 60 SEM with respect to magnification scales (figure in log scale)

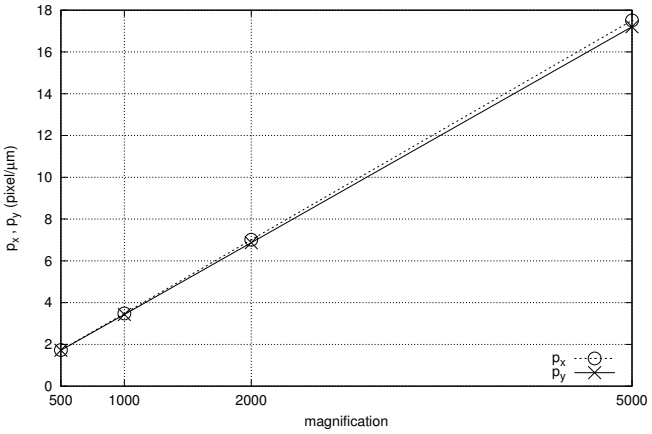


Figure 2.10: Intrinsic parameters from EVO LS 25 SEM with respect to magnification scales

Table 2.3: Calibration results in parallel projection with respect to magnifications with a medium scan speed, using AURIGA 60 SEM

mag. M ($\times$)	p_x (pixel/μm)	p_y (pixel/μm)	$\|\mathbf{e}\|$	p_x/M	p_y/M
300	2.69	2.72	0.51	0.00897	0.00909
500	4.47	4.44	0.38	0.00895	0.00889
800	7.20	7.32	0.27	0.00900	0.00915
1000	8.98	8.96	0.19	0.00895	0.00895
2000	17.96	18.09	0.16	0.00898	0.00904
5000	44.86	45.50	0.24	0.00897	0.00910
10 000	89.81	89.76	0.41	0.00898	0.00897

Table 2.4: Calibration results on parallel projection with respect to magnifications with a fast scan speed, using AURIGA 60 SEM

mag. M (×)	p_x (pixel/μm)	p_y (pixel/μm)	$\|\mathbf{e}\|$	p_x/M	p_y/M
507	4.59	4.54	0.81	0.00906	0.00897
803	7.27	7.20	0.59	0.00906	0.00897
1030	9.31	9.31	0.77	0.00904	0.00904
2000	18.30	17.94	0.87	0.00915	0.00897
5000	45.09	45.46	0.47	0.00902	0.00909
10 000	92.37	89.53	0.94	0.00924	0.00895
20 270	183.62	182.64	1.56	0.00906	0.00901

Table 2.5: Calibration results on parallel projection with respect to magnifications with a medium scan speed, using EVO LS 25 SEM

mag. M (×)	p_x (pixel/μm)	p_y (pixel/μm)	$\|\mathbf{e}\|$	p_x/M	p_y/M
500	1.73	1.72	0.24	0.00346	0.00343
1000	3.48	3.43	0.30	0.00348	0.00343
2000	7.01	6.87	0.55	0.00351	0.00344
5000	17.51	17.21	0.68	0.00350	0.00344

respectively, in Table 2.3 and in Table 2.5 p_x/M and p_y/M are different. Figure 2.10 shows the intrinsic parameters of EVO LS 25 SEM with respect to magnifications.

2.6.3 Distortion issues

Finally, an experiment has been realized to test the potential effects of distortion using AURIGA 60 SEM. Three magnifications are considered in this experiments: 500×, 2000× and 5000×. To compare the performances of calibration with and without distortion parameters, all the factors (gains, coefficients in Levenberg-Marquardt optimization, etc.) in the algorithm are fixed. Table 2.6 shows the calibrated radial distortion parameter k, the skewness parameter γ, the intrinsic parameters (p'_x, p'_y), the residual error $\|\mathbf{e}'\|$ with distortion and the intrinsic parameters (p_x, p_y), and the residual error $\|\mathbf{e}\|$ without distortion. Results are obtained on parallel projection model. It is obvious that introducing distortion parameters does not affect the computation of the main intrinsic parameters (p_x, p_y) and does not improve the residual error. In this case, such spatial distortion could be typically ignored in the calibration process.

Table 2.6: Calibration results with/without distortion

	mag. (×)		
	500	2000	5000
k	-5.65×10^{-9}	-3.67×10^{-10}	-1.15×10^{-10}
γ	0.0024	0.0033	0.0061
s_1	8.64×10^{-7}	-1.28×10^{-7}	-2.87×10^{-7}
s_2	7.19×10^{-6}	2.79×10^{-7}	9.68×10^{-6}
p'_x (pixel/μm)	4.46	17.96	44.87
p'_y (pixel/μm)	4.46	18.00	45.36
$\|\mathbf{e}'\|$	0.57	0.23	0.27
p_x (pixel/μm)	4.46	17.97	44.86
p_y (pixel/μm)	4.46	18.00	45.37
$\|\mathbf{e}\|$	0.57	0.23	0.26

2.7 Conclusion

In this chapter, a simple and efficient method of SEM calibration has been addressed. A global multi-image non-linear minimization process that minimizes the residual error between the projection of the calibration pattern and its observation in the image has been considered. The precise intrinsic parameters, as well as the position of the sensor with respect to the pattern, are computed. Due to the lack of observation of the depth information from a SEM image, the choice of the parallel projection model has been validated for SEM images. The spatial distortion parameters (skewness, radial distortion, and spiral distortion) are insignificant in the experiments and can be eliminated.

Chapter 3

Vision-based control: application in micro- and nano-scale

Robot motion control is an important topic in robotics. As a widely used sensing technology, vision is always indispensable in many robot motion control tasks. Using visual feedback to control a robot is commonly termed visual servoing. The major contribution of this thesis is to employ this technique for 6-DoF automated micro/nano-positioning task in a particular environment. This work is presented in Chapter 3, 4 and 5. The background knowledge on vision-based control and its application in micro/nano-robotic is stated in this chapter. Since a key challenge in micro/nano-robotics is the difficulty on observing the motion along the depth direction at high magnifications in a SEM, in Chapter 4, we propose to use defocus information in visual servoing to control the motion along the depth direction. Based on this technique, a hybrid visual servoing scheme is introduced for 6-DoF automated micro/nano-positioning task in Chapter 5.

This chapter is organized as follows. An introduction on vision-based control is addressed at first. The fundamental knowledge on visual servoing is then presented. The final section provides an overview of the applications of the vision-based control in micro/nano-scale.

3.1 Vision-based control overview

In robotics, vision is one of the most important sensor for automatic control in unknown, complex and dynamics environments. There are two ways to realize the control of the robot via visual information. One way is to apply an open-loop control (described in [Kragic and Christensen, 2002]), where the extraction of information and the control of the robot are conducted separately. A typical example is to estimate the pose (the position and the orientation) of the observed object in the camera coordinate frame at first, and then to move the robot to the target pose directly using the initially estimated absolute pose information. This method does not use the dynamic visual feedback information. The main drawback of the open-loop control is that it is inaccurate and unreliable since no adjustment is considered in the control law.

Alternatively, the closed-loop vision-based control, so-called visual servoing, was introduced in the late 80's [Weiss et al., 1987, Rives et al., 1989, Feddema and Mitchell, 1989] and later in 90's [Corke and Good, 1992, Hutchinson et al., 1996]. This is a multi-discipline research area dealing with robotics, automation, computer vision and image processing. The basic visual servoing task is to control the motion of a robot for a positioning task based on the dynamic feedback information obtained through a vision sensor. It is an efficient approach for vision-based robotic tasks such as positioning and tracking. In recent decades, visual servoing techniques have been widely discussed and applied into different fields [Corke et al., 1996, Sun and Nelson, 2001, Krupa et al., 2003, Metni and Hamel, 2007].

As an active field in robotics and computer vision, a range of visual servoing techniques has been investigated. A general introduction on visual servoing has been presented in [Chaumette and Hutchinson, 2006, Chaumette and Hutchinson, 2007], where the principles and other important issues are introduced. Many visual servoing tasks are performed using 2D or 3D geometrical features extracted from the image. These features include 2D points [Rives et al., 1989], or more complex choices, such as lines, spheres and cylinders [Chaumette and Rives, 1990, Espiau et al., 1992]. These geometrical-features-based visual servoing techniques require an efficient and robust detecting or tracking algorithm to ensure reliable extracted visual features in each frame. Alternatively, a visual servoing task can be performed by homography [Vargas and Malis, 2005, Benhimane and Malis, 2006]. In these methods, instead of extracting a local feature, a template matching method is applied to align the current image to the desired one in each frame. Recently, some novel visual servoing techniques have been proposed to adapt for different issues and imaging conditions. As a direct information, the photometric information [Collewet et al., 2008, Collewet and Marchand, 2011] has been introduced as a visual feature to compute the control law. In [Marchand and Collewet, 2010], image gradient information is introduced to control the camera and light source positions. The mutual information has also been considered to compute the control law [Dame and Marchand, 2009, Dame and Marchand, 2011]. Based on the

information theory [Shannon, 1948], mutual information shows robustness in computing the cost function. All these techniques consider the direct information on the whole image as visual features. In this case, local feature extraction or tracking is no longer required and the problem on the accuracy and reliability of tracking is then solved.

3.2 Classical visual servoing

In this section, the classical visual servoing framework is presented at first, including the eye-in-hand case and the eye-to-hand case. As an important issue in visual servoing, the computation of interaction matrix is addressed.

3.2.1 Modeling

A general visual servoing scheme is illustrated in Figure 3.1. It refers to a closed-loop control: the information (extracted feature) $\mathbf{s}$ is compared with a desired feature $\mathbf{s}^*$. The control law is built in order to minimize the error $\mathbf{e} = \mathbf{s} - \mathbf{s}^*$. This feature can be the visual information from the 2D image or the 3D pose information with respect to a reference coordinate frame. Based on the nature of the visual information, the existing visual servoing approaches can be classified into two main categories: image-based visual servoing (IBVS) and position-based visual servoing (PBVS).

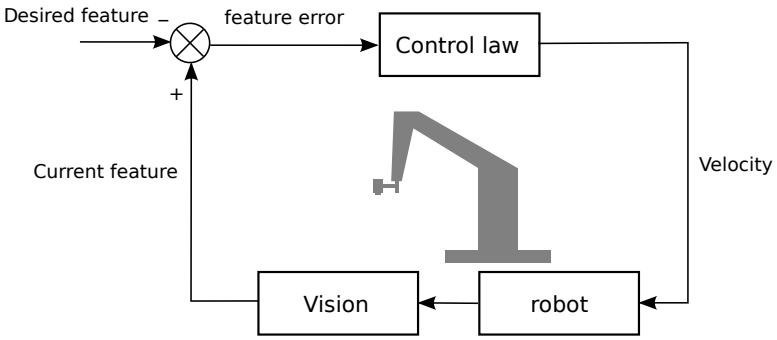


Figure 3.1: Classical visual servoing framework

In IBVS techniques, one or multiple visual features (e.g., points and lines) in the image plane are used to compute the control law. In PBVS techniques, the visual information is used to extract the pose of the robot and the control law is computed from the error between the current and the desired pose.

Suppose that a robot is located at $\mathbf{r}(\mathbf{q})$ ($\mathbf{q}$ is the joint coordinates). Its desired pose is noted $\mathbf{r}^*$. In the general case, depending on the definition of the reference frame, $\mathbf{r}^*$ is usually unknown. In order to achieve the desired position, the general idea of a classical visual servoing task is to move the robot iteratively towards $\mathbf{r}^*$ by minimizing the error $\mathbf{e}$ between the current feature $\mathbf{s}(\mathbf{r})$ and the desired one $\mathbf{s}^*(\mathbf{r}^*)$:

$$\hat{\mathbf{r}} = \underset{\mathbf{r}}{\operatorname{argmin}} \ \|\mathbf{e}(\mathbf{r})\| \quad \text{where} \quad \mathbf{e}(\mathbf{r}) = \mathbf{s}(\mathbf{r}) - \mathbf{s}^* \tag{3.1}$$

When this error $\mathbf{e}$ reaches its minimum, the optimal pose $\hat{\mathbf{r}}$ is obtained. $\hat{\mathbf{r}}$ equals to $\mathbf{r}^*$ when $\mathbf{e}$ is minimized to zero.

In order to compute the robot's velocity from the dynamic feature, it is necessary to find the relation between the time derivative $\dot{\mathbf{s}}$ and the robot joint velocity $\dot{\mathbf{q}}$. This relation can be expressed using the Jacobian $\mathbf{J}_s$:

$$\dot{\mathbf{s}} = \mathbf{J}_s \dot{\mathbf{q}}. \tag{3.2}$$

Considering an exponential decrease of the error $\dot{\mathbf{e}} = -\lambda \mathbf{e}$ during the visual servoing task, with equations (3.1) and (3.2), the control law of a classical visual servoing is:

$$\dot{\mathbf{q}} = -\lambda \mathbf{J}_s^+ \mathbf{e}(\mathbf{r}) \tag{3.3}$$

where λ is the proportional coefficient and $\mathbf{J}_s^+$ is the pseudo-inverse of $\mathbf{J}_s$.

In some real-time applications, it is difficult or time-consuming to obtain the exact value of $\mathbf{J}_s^+$ in each iteration. In order to improve the performance, an approximation $\widehat{\mathbf{J}_s^*}$ could be applied using different approaches [Chaumette and Hutchinson, 2006], e.g., $\widehat{\mathbf{J}_s^+} = \mathbf{J}_{s^*}^+$ or $\widehat{\mathbf{J}_s^+} = 1/2(\mathbf{J}_s + \mathbf{J}_{s^*})^+$.

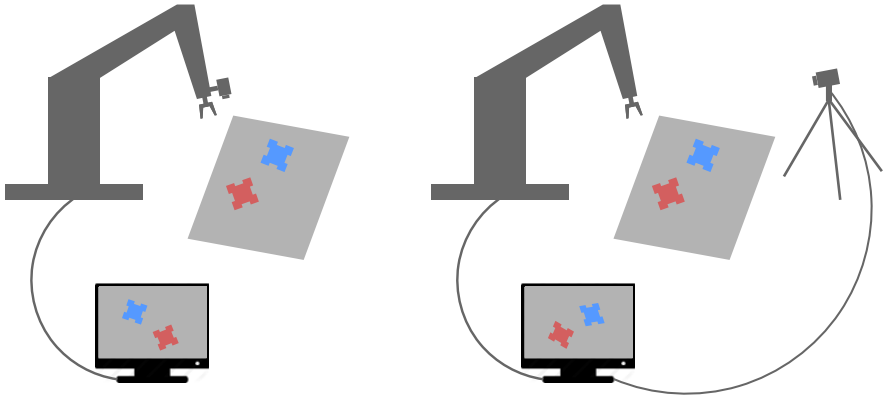


Figure 3.2: Robot-camera configuration in visual servoing: (a) eye-in-hand; (b) eye-to-hand

Two distinct robot-camera configuration cases exist in classical visual servoing: eye-in-hand case (see Figure 3.2 (a)), in which the camera is installed in the end-effector of the robot and the robot's motion results in camera's motion; and alternatively, the eye-to-hand case (see Figure 3.2 (b)), where the camera is fixed and looking towards the end-effector. Thus, the robot's motion does not change the camera's pose. In micro/nano applications, the eye-to-hand case is generally preferred since the sensor (microscope) is usually motionless. Considering the eye-to-hand visual servoing context, the Jacobian $\mathbf{J}_s$ can be expressed as:

$$\mathbf{J}_s = -\mathbf{L}_s {}^c\mathbf{V}_{\mathcal{F}} {}^{\mathcal{F}}\mathbf{J}_n(\mathbf{q}) \tag{3.4}$$

where $\mathbf{L}_s$ represents the interaction matrix, which links the relative camera instantaneous velocity $\mathbf{v}_c$ and the feature motion $\dot{\mathbf{s}}$, ${}^{c}\mathbf{V}_{\mathcal{F}}$ is the motion transform matrix which transforms the velocity expressed in the camera coordinate frame onto the robot coordinate frame, ${}^{\mathcal{F}}\mathbf{J}_n(\mathbf{q})$ is the robot Jacobian in the robot coordinate frame.

3.2.2 Interaction matrix

In order to compute the control law, the interaction matrix of the visual feature $\mathbf{L}_s$ is important in visual servoing. Most of the IBVS techniques compute $\mathbf{L}_s$ from the relation between the velocity and the variation of the point position on the image plane. However, only the perspective projection is considered in the previous research. In this section, we review the computation of the interaction matrix in perspective projection and introduce this computation in the case of parallel projection which is of interest in our application context.

For a 3D point with coordinates $\mathbf{X} = (X, Y, Z)^\top$, let us recall that its projection $\mathbf{x} = (x, y)^\top$ on the image plane (see Section 2.3) in perspective projection model is given by:

$$\begin{cases} x = \dfrac{X}{Z} \\ y = \dfrac{Y}{Z} \end{cases} \tag{3.5}$$

The time derivative of equation (3.5) can be written as:

$$\begin{cases} \dot{x} = \dfrac{\dot{X} - x\dot{Z}}{Z} \\ \dot{y} = \dfrac{\dot{Y} - y\dot{Z}}{Z} \end{cases} \tag{3.6}$$

Relating the velocity of the 3D point to the (relative) camera spatial velocity [Chaumette and Hutchinson, 2006]

$$\dot{\mathbf{X}} = -\boldsymbol{\nu}_c - \boldsymbol{\omega}_c \times \mathbf{X} \Leftrightarrow \begin{cases} \dot{X} = -\nu_x - \omega_y Z + \omega_z Y \\ \dot{Y} = -\nu_y - \omega_z X + \omega_x Z \\ \dot{Z} = -\nu_z - \omega_x Y + \omega_y X \end{cases} \tag{3.7}$$

Injecting equation (3.7) into equation (3.6):

$$\begin{cases} \dot{x} = -\nu_x/Z + x\nu_z/Z + xy\omega_x - (1+x^2)\omega_y + y\omega_z \\ \dot{y} = -\nu_y/Z + y\nu_z/Z - xy\omega_y + (1+y^2)\omega_x - x\omega_z \end{cases} \tag{3.8}$$

This can be rewritten as

$$\dot{\mathbf{x}} = \mathbf{L}_{\mathbf{x}}\mathbf{v}_c \tag{3.9}$$

with $\mathbf{v}_c = (\nu_x, \nu_y, \nu_z, \omega_x, \omega_y, \omega_z)^\top$, where $\mathbf{L}_{\mathbf{x}}$ is the interaction matrix which links the camera velocity and the time derivative of a point on the image plane:

$$\mathbf{L}_{\mathbf{x}} = \begin{bmatrix} -\dfrac{1}{Z} & 0 & \dfrac{x}{Z} & xy & -(1+x^2) & y \\ 0 & -\dfrac{1}{Z} & \dfrac{y}{Z} & 1+y^2 & -xy & -x \end{bmatrix}. \tag{3.10}$$

Most of the image-based visual servoing techniques use this interaction matrix $\mathbf{L_x}$ to compute the Jacobian ($\mathbf{J}_s$ in equation (3.2)), which links the time derivative of the visual feature and the camera velocity.

In parallel projection (details can be found in Section 2.3), the relation between $\dot{\mathbf{X}}$ and $\dot{\mathbf{x}}$ is given by:

$$\begin{cases} \dot{x} = \dot{X} \\ \dot{y} = \dot{Y} \end{cases} \tag{3.11}$$

Injecting equation (3.7) in equation (3.11):

$$\begin{cases} \dot{x} = -\nu_x - \omega_y Z + \omega_z Y \\ \dot{y} = -\nu_y - \omega_z X + \omega_x Z \end{cases} \tag{3.12}$$

In this case, the interaction matrix in equation (3.9) is expressed by:

$$\mathbf{L_x} = \begin{bmatrix} -1 & 0 & 0 & 0 & -Z & y \\ 0 & -1 & 0 & Z & 0 & -x \end{bmatrix}. \tag{3.13}$$

It should be mentioned that in equation (3.13), the third column is null. It means that in parallel projection, the motion along z-axis can no longer be controlled from the variation of the point positions on the image plane. This should be considered as one of the most important issues in visual servoing using the parallel projection model. In fact, at high magnifications under a SEM, it is difficult to observe the motion along the depth direction since the scale of the projection of an object on the image plane can be considered invariable when the position on z-axis changes.

3.3 Vision-based control in micro/nano-scale

The past decade has seen a rapid development of microelectromechanical and microoptoelectromechanical systems (MEMS/MOEMS). These represent a significant potential in the fabrication of smaller components and micro-structures. Hence, they play an important role in several industrial and biomedical areas where the integration of these devices would lead to the development of low-cost and high-performance microsystems [Cohn et al., 1998]. Automatic and reliable handling/assembly of these micro-structures is a very active field [Régnier and Chaillet, 2010, Banerjee and Gupta, 2013]. Moreover, micro/nano-manipulation can be used to operate various objects in micro/nano-scale, such as carbon nanotubes (CNTs) [Yu et al., 1999, Fukuda et al., 2003] and nanowires [Agarwal et al., 2005, Yu et al., 2002], for a dynamic analysis and characterization of the properties of these samples. All the strong requirements lead to the fast development in the automation techniques for micro/nano-manipulation and assembly tasks [Fatikow and Eichhorn, 2008]. In order to perform the manipulation and assembly in micro/nano-scale, some microassembly stations have been realized [Fatikow and Rembold, 1996, Yang et al., 2001, Weck and Peschke, 2004, Probst et al., 2006], especially in a SEM [Fatikow et al., 2007, Eichhorn et al., 2009].

Vision is one of the most important sensing technologies given the constraints in micro/nano-scale, whether through optical or scanning electron microscopy. Visual servoing is hence a necessary tool for automated micro/nano-manipulation. Some early studies on microassembly and handling using visual feedback information can be found in [Koyano and Sato, 1996, Sulzmann et al., 1997, Vikramaditya and Nelson, 1997]. A micromanipulation method using two separated sensing modalities has been proposed [Zhou et al., 1998]. In this method, both the force and the vision feedback are fused for the manipulation. In [Ferreira et al., 2004], automated micromanipulation tasks for teleoperated microassembly assisted by visual servoing and virtual reality techniques have been performed under an optical microscope. A fact in the micromanipulation tasks under a microscope is that the images at high magnifications provide precise measurements but a small field-of-view, while images at low magnification have a large field-of-view but less accurate measurements. To overcome this problem, many researchers have studied a multiview system [Yang et al., 2003, Sun and Chin, 2004, Popa and Stephanou, 2004, Abbott et al., 2007, Probst et al., 2009] or a stereo system [Jähnisch and Fatikow, 2007], and some others have investigated the control of dynamic zoom in the visual control [Tao et al., 2005, Tamadazte et al., 2008].

One of the major issues in micro/nano-manipulation is the accuracy in position and orientation of the object [Ralis et al., 2000, Devasia et al., 2007, Ouarti et al., 2013]. Most approaches on vision-based control in microscale are based on the observation of the object features [Sun et al., 2003, Ogawa et al., 2005, Ru et al., 2011, Gong et al., 2014] or a CAD model from the object features [Feddema and Simon, 1998, Kratochvil et al., 2009]. These features are principally geometrical information (e.g., corners, edges, contours) or markings on the object surface. By estimating the position and orientation of these features, the pose (position and orientation) of the objects in the camera frame can be then detected. A bottleneck in these local-feature-based object localization and servoing methods is the quality of the images from the camera or the microscope. There are still several limitations in the available imaging techniques at micro/nano-scale, including the signal-to-noise ratio (SNR), depth of field, contrast, etc. In some cases, the local-feature-based tracking and servoing approaches could be no longer reliable.

Recently, direct visual servoing techniques have been proposed using image derivative information [Marchand and Collewet, 2010] or photometric information [Collewet et al., 2008, Collewet and Marchand, 2011] as a visual feature in the control law. These direct approaches on visual servoing have been applied in micro/nano-manipulation tasks. [Tamadazte et al., 2012] has proved that an image-intensity-based approach is efficient in micro-positioning with 3 DoFs (translation in x- and y- axes, and rotation around z-axis) under an optical microscope. In [Marturi et al., 2014b], the authors have validated the 2-DoF image-intensity-based approach in a SEM and proposed to estimate the object location from the frequency domain. A set-based direct visual servoing controller for nanopositioning in an AFM has been proposed [Liu et al., 2015]

to avoid the correspondence problem between the two image frames. This method has been evaluated only by some simulations.

It should be mentioned that most of the current visual servoing approaches in SEM only feature the robot motion in 2, 3 or 4 DoFs [Sievers and Fatikow, 2005, Ru et al., 2011, Gong et al., 2014, Marturi et al., 2014b]. Only a few works concerns 6 DoFs, such as the CAD based visual tracking methods [Kratochvil et al., 2009]. A possible reason is that it is difficult to observe the position on the depth direction as well as the rotation around x- and y-axes due to the parallel projection model of a SEM. However, the visual servoing on 6 DoFs is often required in micro/nano-manipulation tasks.

3.4 Conclusion

In this chapter, the basics of vision-based control and an overview of its application in micro/nano-robotics are presented. As a closed-loop control scheme using visual feedback information, visual servoing is performed by minimizing the error between the current visual feature and the desired visual feature. It plays an important role in robotic motion control. However, some difficulties have been found in the application of the traditional visual servoing in a SEM. One of the significant challenges is that it is very difficult to observe the robot's motion from the SEM image at high magnifications because of the parallel projection model. In order to deal with this problem, the visual servoing approach for robot motion along the depth direction is then presented in the next chapter.

Chapter 4

Visual servoing using defocus information

A key challenge in 6-DoF visual servoing tasks in a SEM is the difficulty in observing the motion along the depth direction. Thus, controlling the robot motion along the depth direction is an important issue. This chapter focuses on the visual servoing approach for the robot motion along the depth direction in a SEM. In order to achieve this, the image sharpness functions are evaluated at first to select an appropriate visual feature. The visual servoing control law is then designed using the image gradient as a visual feature. This method is validated by the experimental results of visual servoing along the depth direction.

4.1 Defocus information as a visual feature

As mentioned previously, one difficulty in micro/nano vision while using the parallel projection model is the lack of the observation along the depth direction. The robot motion along the depth direction is uncontrollable through the interaction matrix that links the variation of a point's position in the image and the camera instantaneous velocity with this parallel projection model. Therefore, in order to perform visual servoing along the depth axis, a new suitable visual feature that corresponds to the motion along the depth direction is foremost required.

4.1.1 Depth and focus/defocus

One possible way to control the motion along the depth direction is to estimate the depth information and then perform the visual servoing using this information. Many approaches on depth estimation have been proposed in computer vision. The depth information can be defined as the distance between the camera and the object [Subbarao and Surya, 1994], or the local depth information such as the depth map on 3D object surface [Noguchi and Nayar, 1994, Favaro et al., 2008, Mahmood et al., 2013, Marturi et al., 2013a] or the relative depth of a scene or complex environment [Torralba and Oliva, 2002, Zhuo and Sim, 2009]. One general idea in microscopy application is to employ stereo vision and reconstruct the 3D image from several 2D images [Pouchou et al., 2002, Marinello et al., 2008, Tunnell and Fatikow, 2011, Fan et al., 2014]. In this case, the reliable feature extraction is necessary for the reconstruction. This technique is mostly used to measure the shape and the surface of a sample [Mills and Rose, 2010, Ersoy, 2010, Gavrilenko et al., 2015], but also be applied into the micro/nano-handling [Jähnisch and Fatikow, 2007]. However, both feature extraction and matching could be computationally expensive and unreliable. Moreover, the stereoscopic images are usually obtained by tilting the sample. In real-time automated micro/nano-manipulation tasks, it is impractical to implement this technique to control the motion along the depth direction.

It has been observed that for a sensor with a small depth of field, the image sharpness varies with changes in depth position. This occurs in an optical sensor and also in the electron microscopes. Thus, this can be considered as an important indicator to recover the depth information. Given the relationship between the (optical) sensor focus sets and depth, Depth From Focus (DFF) has been investigated by many authors [Grossmann, 1987, Subbarao, 1988, Ens and Lawrence, 1993, Nayar and Nakagawa, 1994, Subbarao and Choi, 1995] for depth estimation. The underlying principle is to obtain different focus levels by adjusting the camera parameters (i.e., the distance between the lens and image plane, the focal length, and the aperture radius). It involves obtaining many observations for the various camera parameters and estimating the optimal focus using a criterion function. Since various camera parameters and multiple

observations are required, it is not practical in a real-time visual servoing task.

Alternatively, the Depth From Defocus (DFD) approaches have been also widely discussed [Pentland, 1987, Gökstorp, 1994, Watanabe and Nayar, 1998, Schechner and Kiryati, 1999, Favaro et al., 2003] for optical sensors. The main idea of these methods is that the objects at a particular distance from the lens will be focused in an optical system, whereas objects at other distances will be blurred. By measuring the amount of defocus of the object in the image, the depth of the object with respect to the lens can be then recovered with some knowledge of optics. In these methods, the defocus parameters are estimated from the image in the frequency domain [Subbarao and Wei, 1992, Gökstorp, 1994, Rajagopalan and Chaudhuri, 1995, Xiong et al., 1995, Schechner and Kiryati, 1999, Morgan-Mar and Arnison, 2014], by some spatial-domain-based techniques [Lai et al., 1992, Subbarao and Surya, 1994, Ziou and Deschenes, 2001, Favaro et al., 2003, Favaro and Soatto, 2005], or using statistical models [Rajagopalan and Chaudhuri, 1998, Bhasin and Chaudhuri, 2001], etc. Recently, image registration techniques have been introduced into the DFD methods [Ben-Ari, 2014]. In this method, the images are aligned in order to improve the accuracy of the depth estimation. An investigation on DFD/DFF methods and stereo/motion-based methods has been presented in [Schechner and Kiryati, 2000]. The authors have shown that sensitivities of DFF and DFD techniques are not inferior but similar to those of stereo/motion-based methods.

The DFD methods have also been extended to SEM applications [Eichhorn et al., 2008]. The variance of the pixel gray levels of the image has been employed for coarse depth detection in a pick-and-place manipulations task of carbon nanotubes (CNTs) [Eichhorn et al., 2009]. For a visual tracking task in a SEM, the variance has also been considered as a criterion to recover the depth information [Dahmen, 2008, Dahmen, 2011]. In this method, the depth and the corresponding variance of the pixel gray level are recorded into a data set (off-line). During the tracking task (on-line), the variance of the image is computed and the depth position is retrieved by looking up the variance in the data set.

Given the visual servoing application in the parallel projection model, one possible idea is then to estimate the depth information using the DFF and DFD techniques and then perform the servoing task by minimizing the depth error. However, most of the above techniques are based on the geometry of optical camera. The SEM image formation incorporates different dynamics. Another possible solution for SEM application is to employ directly the image sharpness as a visual feature for an IBVS. In this case, the estimation of the depth is not required. One advantage of this method is the sharpness measurement is more reliable compared with the depth estimation. Nevertheless, it is necessary to derive the relation (i.e. the Jacobian) between the variation of the sharpness and the relative camera velocity. Thereby, it is important to select an appropriate sharpness function that describes properly the variation of the sharpness in the image

during the robot motion along the depth direction.

4.1.2 Sharpness function selection

In general, the objective of a visual servoing task along the depth direction is to attain a desired position Z^* (which is unknown) by minimizing the error between the desired visual feature $\mathbf{s}^*$ and the current visual feature $\mathbf{s}$.

$$\hat{Z} = \underset{Z}{\operatorname{argmin}} \left(\mathbf{s}(Z) - \mathbf{s}^*(Z^*)\right)^2 \tag{4.1}$$

4.1.2.1 Sharpness functions

The selection of a suitable visual feature $\mathbf{s}(Z)$ is then an important issue. It is evident that the image sharpness varies when the sample is moved along the depth direction. Some authors have investigated and evaluated the sharpness function for optical microscopy images [Santos et al., 1997, Sun et al., 2005] as well as for SEM images [Rudnaya et al., 2010]. These evaluations have been conducted in different conditions (e.g. noise levels) by various criteria such as accuracy, range, robustness to noise, etc. These studies mostly consider these sharpness functions for an autofocus task. Nevertheless, they can also be considered as a reference for visual feature selection in our case. Our goal is to select an appropriate sharpness function for visual servoing tasks along the depth direction. In order to achieve this, several sharpness functions have been selected and compared using our SEM image sequences with depth position variation.

A. Derivative-based sharpness functions

The derivative-based sharpness functions consider the fact that the intensity differences between the neighboring pixels changes due to the defocus level variation. Consider an edge area with high-frequency pixels, when defocus level increases, the intensity difference between the neighboring pixels decrease consequently. Since the defocus level corresponds to the depth position, it is then possible to employ the derivative-based sharpness functions as a visual feature to control the motion along the depth direction.

Many authors have computed the derivative-based sharpness functions horizontally, since in a SEM the image is generated by line scanning and the noise is correlated horizontally. A general expression of these functions is given by

$$\mathbf{s}_{dx} = \sum_{x=0}^{M}\sum_{y=0}^{N} |\mathbf{I}(x+k,y) - \mathbf{I}(x,y)|^p \quad \text{where} \quad |\mathbf{I}(x+k,y) - \mathbf{I}(x,y)|^p < \theta \tag{4.2}$$

where M, N represent the image dimension, $k \in \mathcal{N}$ is a small integer that represents the horizontal distance between two pixels to be compared, $\theta \in \mathbb{R}^+$ is a threshold to adjust the sensitivity of the sharpness function.

when $k = 1, p = 1$, equation (4.2) is called threshold absolute gradient [Santos et al., 1997], which is given by:

$$\mathbf{s}_{tag} = \sum_{x=0}^{M}\sum_{y=0}^{N} |\mathbf{I}(x+1,y) - \mathbf{I}(x,y)| \tag{4.3}$$

In the case that $k = 1, p = 2$, squared gradient [Santos et al., 1997] is applied using simply the differences between a pixel and its neighbor one:

$$\mathbf{s}_{sg} = \sum_{x=0}^{M}\sum_{y=0}^{N} (\mathbf{I}(x+1,y) - \mathbf{I}(x,y))^2 \tag{4.4}$$

In [Brenner et al., 1976], Brenner proposed to compute the gradient using the derivative between a pixel and its neighbor two points away ($k = 2, p = 2$), called Brenner gradient:

$$\mathbf{s}_{bg} = \sum_{x=0}^{M}\sum_{y=0}^{N} (\mathbf{I}(x+2,y) - \mathbf{I}(x,y))^2 \tag{4.5}$$

It is proved to give a good signal-to-noise ratio (SNR) [Brenner et al., 1976].

Alternatively, other methods compute the derivative-based sharpness functions on both horizontal and vertical directions. In general, it can be expressed by

$$\mathbf{s}_{g} = \sum_{x=0}^{M}\sum_{y=0}^{N} (\nabla {I_x}^2(x,y) + \nabla {I_y}^2(x,y)) \tag{4.6}$$

To compute the horizontal and the vertical gradient image $\nabla I_x(x,y)$ and $\nabla I_y(x,y)$, Tenenbaum gradient (Tenenbrad) is considered in [Krotkov, 1988, Yeo et al., 1993]. The horizontal and vertical gradient image $\nabla I_x(x,y)$ and $\nabla I_y(x,y)$ are computed from the convolution of image $\mathbf{I}(x,y)$ and Sobel operators expressed by:

$$\begin{cases} \nabla I_x(x,y) = S_x * \mathbf{I}(x,y) \\ \nabla I_y(x,y) = S_y * \mathbf{I}(x,y) \end{cases} \text{where} \quad S_x = \begin{pmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{pmatrix}, S_y = \begin{pmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{pmatrix}. \tag{4.7}$$

Another way to obtain a smooth horizontal or vertical gradient is to convolute the image $\mathbf{I}(x,y)$ with a Gaussian filter and a derivative filter. It can be expressed by

$$\begin{cases} \nabla I_x(x,y) = (G_x * D_x) * \mathbf{I}(x,y) \\ \nabla I_y(x,y) = (G_y * D_y) * \mathbf{I}(x,y) \end{cases} \tag{4.8}$$

where G_x, G_y are the vectors which represent the Gaussian filter on x and y directions, respectively; D_x, D_y are the vectors which represent the derivative filter, on x and y directions, respectively.

Other derivative-based techniques are also discussed by previous literature, such as Laplace-based methods [Nayar and Nakagawa, 1994, Subbarao et al., 1993].

B. Statistical sharpness functions

Statistical sharpness functions are generally less sensitive to the noise as compared to the derivative-based ones. They are computed according to the statistical criteria, such as variance and histogram of image intensities.

The variance of an image relates to the image contrast. A small value of variance indicates that the image intensities tend to be very close to their mean value. Image-variance algorithms are based on the fact that the in focus image has higher contrast than the defocused one. Normalized variance shows good performance in many evaluations [Yeo et al., 1993, Groen et al., 1985, Sun et al., 2005]. It is expressed by:

$$\mathbf{s}_{nv} = \frac{1}{MN\mu} \sum_{x=0}^{M} \sum_{y=0}^{N} (\mathbf{I}(x,y) - \mu) \tag{4.9}$$

where $\mu = \bar{\mathbf{I}}(x,y)$ is the mean image intensity.

The correlation-based methods have also been investigated, such as the autocorrelation function [Vollath, 1987, Liu et al., 2007]:

$$\mathbf{s}_{ac} = \sum_{x=0}^{M} \sum_{y=0}^{N} \mathbf{I}(x,y)\mathbf{I}(x+1,y) - \sum_{x=0}^{M} \sum_{y=0}^{N} \mathbf{I}(x,y)\mathbf{I}(x+2,y) \tag{4.10}$$

The standard-deviation-based correlation can be expressed by [Vollath, 1987, Liu et al., 2007]:

$$\mathbf{s}_{sdc} = \sum_{x=0}^{M} \sum_{y=0}^{N} \mathbf{I}(x,y)\mathbf{I}(x+1,y) - MN\mu \tag{4.11}$$

where μ is the mean image intensity.

Histogram-based methods use the histogram to analyze the distribution of the image intensities. Denote the number of pixels with the intensity i by $h(i)$, the range algorithm [Firestone et al., 1991] computes the difference between the highest and the lowest intensity levels:

$$\mathbf{s}_r = \max\{i | h(i) > 0\} - \min\{i | h(i) > 0\} \tag{4.12}$$

Entropy algorithm [Firestone et al., 1991] assumes that a sharp image contains more information. The entropy measures the expected value of the information in the image:

$$\mathbf{s}_e = -\sum_i p_i \log_2(p_i) \tag{4.13}$$

where p_i is the probability of a pixel with intensity i.

Additionally, besides derivative-based and statistical sharpness functions, other sharpness functions such as Wavelet-based functions [Subbarao et al., 1993, Yang et al., 2003], thresholded content [Groen et al., 1985], image power [Santos et al., 1997] are also studied in previous research.

4.1.2.2 Analysis of sharpness function efficiency

Based on the performance of these sharpness functions in accuracy, width of the sharpness peak and robustness to noise in previous studies [Sun et al., 2005, Rudnaya et al., 2010], we evaluate some of them: image gradient (equation (4.6)), normalized variance (equation (4.9)), autocorrelation (equation (4.10)), standard-deviation-based correlation (equation (4.11)), entropy (equation (4.13)). The test SEM image sequence is acquired at 1000× with a medium scan speed (about 0.3 pixel/µs, a visual servoing task can be performed using this scan speed). In these tests, the depth position of a sample varies in a range of 400 µm with a step of 1 µm. The sample images and the evolution of the sharpness scores from the selected sharpness functions are shown in Figure 4.1. In order to compare the performance and the shape of these sharpness functions, the sharpness scores are normalized. We find that entropy cannot be applied to our sample since the sample is simple and has little texture / information. Although in many papers [Sun et al., 2005, Dahmen, 2011, Marturi et al., 2013c] normalized variance shows good performance, it does not perform well in our tests. Figure 4.1 shows that when the sample is close to the in-focus position, it is difficult to recognize the variation of the position on the depth direction from the normalized variance. Comparing the shapes of the sharpness functions (w.r.t. the position on the depth direction), it can be found that image gradient and autocorrelation are more sensitive to depth position variation than the normalized variance and standard-deviation-based correlation. Nevertheless, image gradient is found to be more precise and more robust to noise when the sample is close to the in-focus position. For these reasons, we propose to use the image gradient defined in equation (4.6) for the control of the position on the depth direction in a visual servoing task.

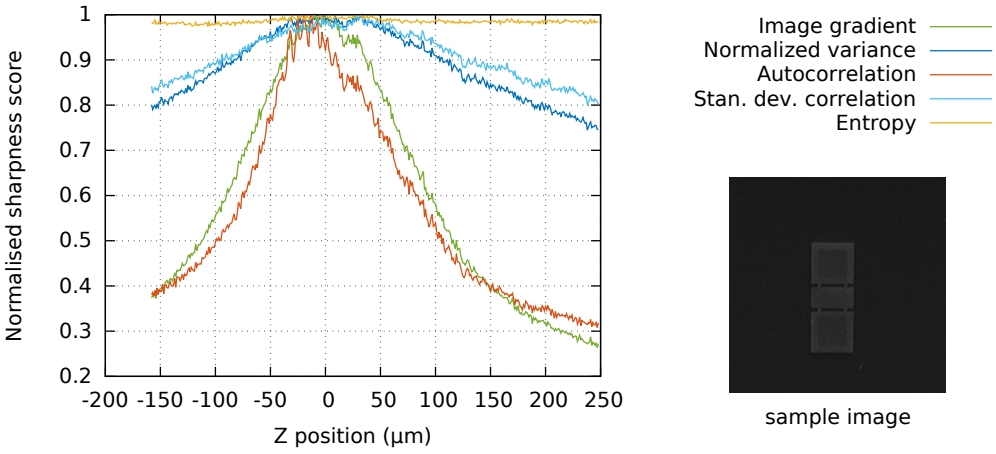


Figure 4.1: Evolution of different sharpness scores with respect to Z position, computed from image sequences at 1000×

4.2 Control of Z using image gradient

In this section, the visual servoing scheme for the motion along the depth direction using image gradient information is presented. The general idea is based on the fact that the image gradient varies when object position on the depth direction changes, by keeping the focal length of the sensor constant. Practically, for a vision sensor (e.g., an optical camera or a SEM) with a small depth of field, the focal length of the sensor can be adjusted in order to acquire a sharp image, this is the so-called autofocus process. Alternatively, one can move the sample position along z-axis to put the scene in focus. Therefore, the sample can be moved to the target position by the error of image gradient between the image at the current position and the image at the desired position.

4.2.1 SEM Image defocus model

The general image formation model, which is commonly used in the case of optical microscopes [Nayar and Nakagawa, 1994], can be extended with a SEM [Nicolls et al., 1997]. Given an extremely small area element $\mathrm{d}x\mathrm{d}y$ centered on (x, y), the secondary electron (SE, see Section 1.2, part of Electron detector) current emitted from this area is

$$\mathrm{d}s(x, y) = \delta(x, y)\varphi(x, y)\mathrm{d}x\mathrm{d}y, \tag{4.14}$$

where $\varphi(x, y)$ is the incident current density at the point (x, y), $\delta(x, y)$ is a yield coefficient which is assigned in a way that δ is the average number of resultant secondary electrons emitted. The total SE current emitted from the specimen s is then given by

$$s = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \delta(x, y)\varphi(x, y)\mathrm{d}x\mathrm{d}y. \tag{4.15}$$

Given a linear relation between emitted SE current s and the result signal i (i.e. $i(x, y) = k \cdot s(x, y)$, k is a constant), the SEM image formation can be seen as a linear convolution of a specimen-dependent component and a system-dependent point-spread function (PSF) [Erasmus and Smith, 1982, Nicolls et al., 1997]. Here, the PSF is the scaled and reflected electron beam current density passing through the origin.

Let Z be the current position of the robot (positioning stage) on the depth direction. The defocus image $\mathbf{I}(x, y, Z)$ at the position Z can be expressed as the convolution of a sharp image $\mathbf{I}^*(x, y, Z^*)$ at the desired pose Z^* and a defocus kernel $f(x, y)$:

$$\mathbf{I}(x, y, Z) = \mathbf{I}^*(x, y, Z^*) * f(x, y) \tag{4.16}$$

Equation (4.16) can be used to model both the SEM and optical image defocus [Nayar and Nakagawa, 1994, Nicolls et al., 1997]. In previous studies, such as [Ens and Lawrence, 1993], the Gaussian kernel is widely used as an approximation of defocus model by many authors. It can be expressed by

$$f(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}. \tag{4.17}$$

where σ is the standard deviation of the Gauss function. In earlier studies on optical camera geometries [Lai et al., 1992, Subbarao and Surya, 1994], the relation between the distance D from a point on the object to the lens and the standard deviation of the Gaussian defocus kernel σ can be described using

$$\sigma = mD^{-1} + c. \tag{4.18}$$

where m, c are constant coefficients which correspond to the optical sensor parameters (focal length, diameter of aperture and distance between the lens and the image plane). The distance D equals to the object position expressed in camera reference frame. For a small displacement δZ on the depth direction, the variation of σ can be approximated using a constant $k \approx -mZ^{-2}$:

$$\delta\sigma = k\delta Z. \tag{4.19}$$

4.2.2 Modeling

For an image $\mathbf{I}(x, y, Z)$ at position Z on z-axis, the square of the norm of the image gradient at a point (x, y) on the image plane is

$$\begin{aligned} g(x, y, Z) &= \|\nabla\mathbf{I}(x, y, Z)\|^2 \\ &= \nabla I_x^2(x, y, Z) + \nabla I_y^2(x, y, Z) \end{aligned} \tag{4.20}$$

Considering the square of the norm of the image gradient for the whole image as the visual feature to be used later in the control law:

$$\begin{aligned} G(Z) &= \sum_{x=0}^{M}\sum_{y=0}^{N} g(x, y, Z) \\ &= \sum_{x=0}^{M}\sum_{y=0}^{N} (\nabla {I_x}^2(x, y, Z) + \nabla {I_y}^2(x, y, Z)) \end{aligned} \tag{4.21}$$

Our goal is then to minimize the error between the current image gradient $G(Z)$ and the desired image gradient $G^*(Z^*)$. The cost function is defined as:

$$e_G(Z) = G(Z) - G^*(Z^*) \tag{4.22}$$

The relation between the relative camera instantaneous linear velocity v_z along z-axis and the time variation of image gradient G is

$$\dot{G} = L_G v_z \tag{4.23}$$

where L_G is the Jacobian (here a scalar) which can be expressed by:

$$L_G = \frac{\partial G}{\partial \sigma}\frac{\partial \sigma}{\partial Z} \tag{4.24}$$

From equation (4.19), we have

$$L_G = k\frac{\partial G}{\partial \sigma}. \tag{4.25}$$

where $\frac{\partial G}{\partial \sigma}$ can be expressed by

$$\frac{\partial G}{\partial \sigma} = \sum_{x=0}^{M}\sum_{y=0}^{N} 2(\nabla I_x(x,y)\frac{\partial \nabla I_x(x,y)}{\partial \sigma} + \nabla I_y(x,y)\frac{\partial \nabla I_y(x,y)}{\partial \sigma}) \tag{4.26}$$

In equation (4.16), the convolution can also be written as:

$$\mathbf{I}(x,y) = \sum_u \sum_v \mathbf{I}^*(x-u, y-v) f(u,v). \tag{4.27}$$

From equation (4.17), compute the derivative

$$\frac{\partial f(u,v)}{\partial \sigma} = \frac{1}{2\pi}(u^2+v^2-2\sigma^2)\sigma^{-5}e^{-\frac{u^2+v^2}{2\sigma^2}}. \tag{4.28}$$

According to equations (4.27) and (4.28):

$$\frac{\partial \nabla I_x(x,y)}{\partial \sigma} = \sum_u \sum_v \nabla(I_x^*(x-u,y-v)\cdot\frac{1}{2\pi}(u^2+v^2-2\sigma^2)\sigma^{-5}e^{-\frac{u^2+v^2}{2\sigma^2}}) \tag{4.29}$$

and

$$\frac{\partial \nabla I_y(x,y)}{\partial \sigma} = \sum_u \sum_v \nabla(I_y^*(x-u,y-v)\cdot\frac{1}{2\pi}(u^2+v^2-2\sigma^2)\sigma^{-5}e^{-\frac{u^2+v^2}{2\sigma^2}}) \tag{4.30}$$

Injecting equation (4.29) and (4.30) in equation (4.26), L_G can be finally computed. The control law is then expressed by

$$v_z = -\lambda L_G^{-1}(G(Z) - G^*) \tag{4.31}$$

where λ is the gain of the control law.

4.2.3 Experimental validations

It was supposed that the proposed method can work under an optical microscope as well as a SEM. In order to validate it, we have conducted some experiments at first under an optical microscope (Basler acA1600-60gm) using a microchip (measures 10 mm × 5 mm, see Figures 4.2(a)) as a sample. This work has been conducted at ISIR, UPMC. The images are acquired at 60× magnification. In this visual servoing task, only motion along the depth direction is controlled keeping the other DoFs fixed (the robot motion is perpendicular to the image plane). In this experiment, the sample is placed at an initial position first and is guided to move toward the desired position. The distance (on z-axis) from the initial position to the desired position is 2 mm, which is quite important according to the depth of field of the microscope in this experiment. Figure 4.2(b), 4.2(c) and 4.2(d) show the evolution of image gradient error per pixel, velocity and the distance between the desired position and the current position. The image gradient error converges quickly and the desired position is attained with an accuracy around 2 µm.

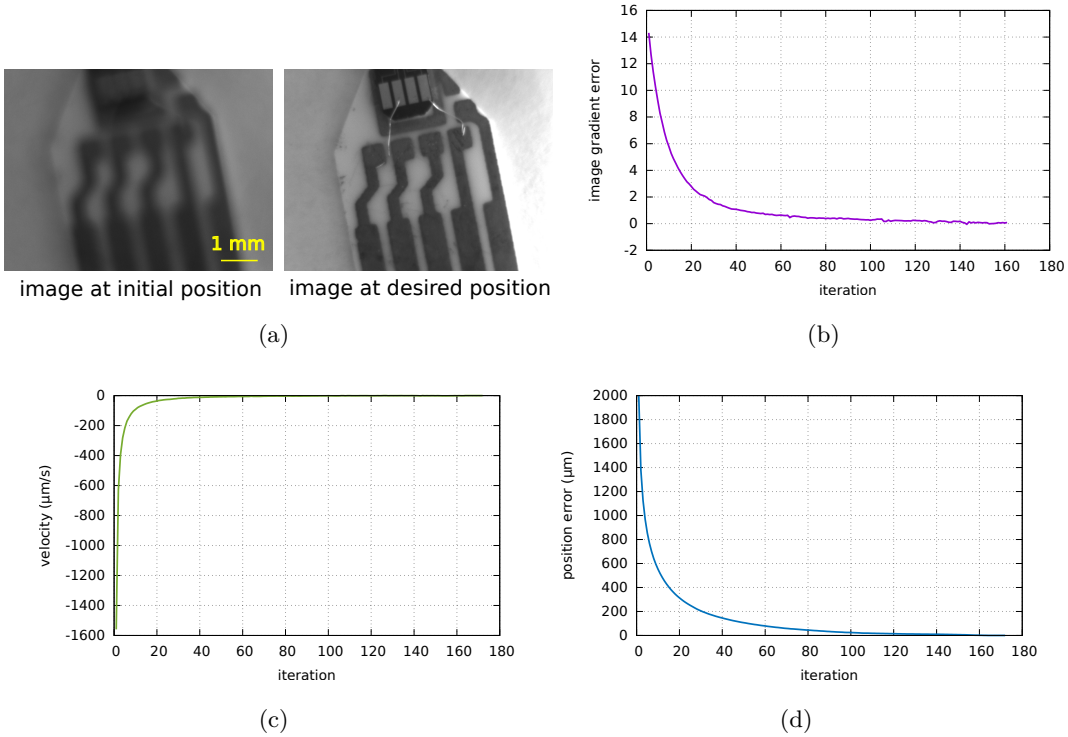


Figure 4.2: Test sample images (a) and visual servoing results with an optical microscope: evolution of image gradient error (b), velocity along the depth direction (c) and position error (d) (distance between the current position and the desired position) with respect to iterations, respectively

A similar experiment has also been performed in a SEM (Zeiss EVO LS 25) at 1000× using a planar calibration pattern (which has been used in the experiments in Chapter 2, see Figure 2.5). The images (360×360 pixels) are acquired by a medium scan speed (around 3.3 µs/pixel). In this chapter, a 3×3 medium filter has been used to reduce the SEM image noise. Similar to the previous experiment, the visual servoing task has been performed only along the depth direction. Both the images at the initial position and the desired position can be found in Figure 4.3(a). The distance (on z-axis) between the initial position and the desired position, where the sample is considered to be in-focus, is 340 µm. The experimental results are shown in Figure 4.3. It can be seen that the desired depth position is attained within a few iterations. The obtained depth position can be attained with an accuracy of 1 µm. This accuracy depends on the SEM image quality as well as (the texture of) the sample. It should be pointed out that the quality of an optical image is much better than that of a SEM image. In our different trials on various conditions using different samples under a SEM, the accuracy on the depth direction is normally less than 10 µm at 1000×.

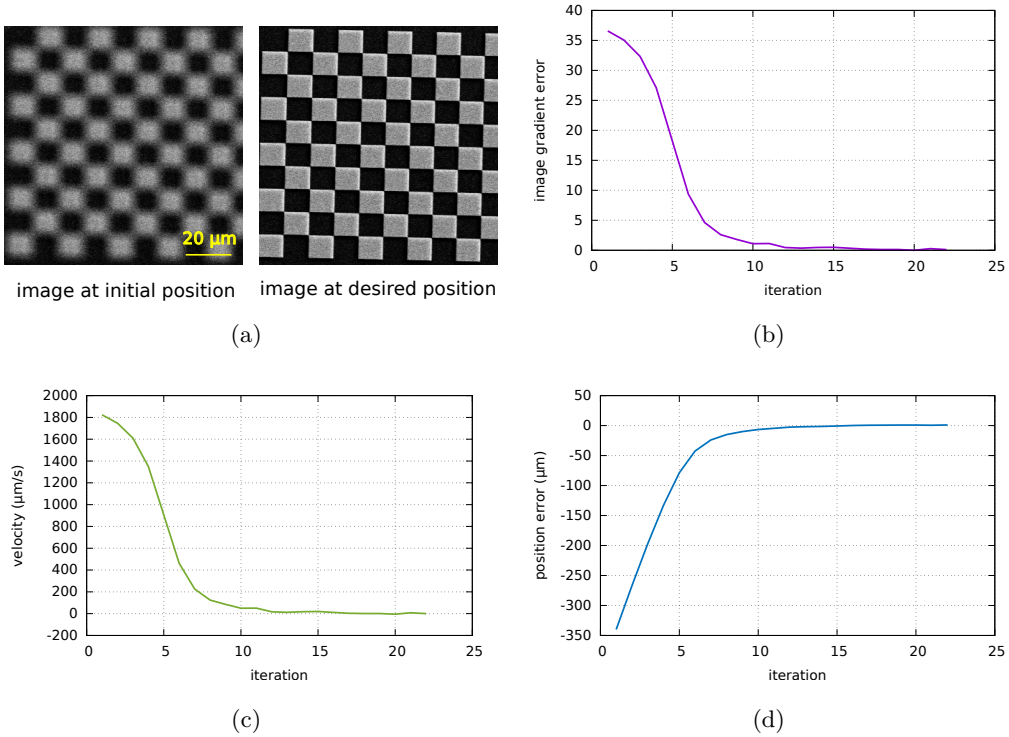


Figure 4.3: Test sample images (a) and visual servoing results with a SEM: evolution of image gradient error (b), velocity along the depth direction (c) and position error (d) (distance between the current position and the desired position) with respect to iterations, respectively

4.2.4 Dynamic approximation of the Jacobian

In the previous section, we have shown analytically the Jacobian linking the variation of image gradient G to the time derivative of the depth position Z. However, since the convolution has been employed during the computation, this algorithm could be time-consuming for large images. Moreover, in order to compute the control law accurately, the standard deviation σ should be estimated dynamically. One solution for these problems is to use an approximation such as $\widehat{\mathbf{J}_s^+} = \mathbf{J}_{s^*}^+$ [Chaumette and Hutchinson, 2006] to avoid the computation in each iteration. Alternatively, another way to compute the control law, without any a priori information (e.g., training data) or tracking during the visual servoing procedure, is proposed in this section. The general idea of this method is that the relation between the image gradient G and the Z position can be approximated by a rational function. In each iteration of the visual servoing task, by estimating the coefficients of this rational function using the data (G and Z) obtained in previous iterations, the Jacobian L_G can then be approximated.

4.2.4.1 Modeling

We should pay attention to the relation between the image gradient and the position Z. Instead of computing the Jacobian analytically, we can also approximate this relation using a given function from statistical methods. By testing various functions (e.g., Gaussian, polynomial, etc.) and fitting them with the data from a range of SEM image sequences (varying Z from defocus position to focused position), we found that the quadratic rational function has the best fitting performance (see Figure 4.4). It is given by

$$f(x) = \frac{p_0 + p_1 x + p_2 x^2}{q_0 + q_1 x + x^2}, \quad p_2 \neq 0. \tag{4.32}$$

where p_0, p_1, p_2, q_0, q_1 are the coefficients of the model.

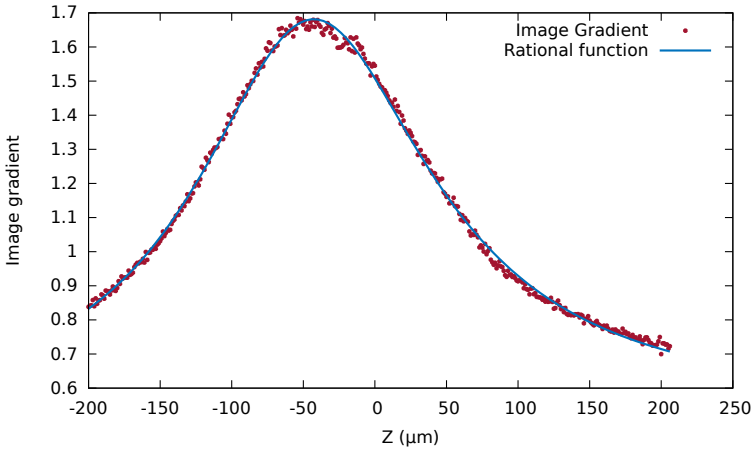


Figure 4.4: Image gradient and its approximation using rational function with respect to depth position, respectively

It is then possible to model the relation between the image gradient G and depth position Z using a quadric rational function:

$$G(Z) = \frac{p_0 + p_1 Z + p_2 Z^2}{q_0 + q_1 Z + Z^2} + \varepsilon, \quad p_2 \neq 0. \tag{4.33}$$

where ε is an error term that can be considered as the noise in the measurement. In this case, the Jacobian L_G can be approximated by

$$L_{app} = -\frac{(p_1 - p_2 q_1)Z^2 + 2(p_0 - p_2 q_0)Z - p_1 q_0 + p_0 q_1}{(q_0 + q_1 Z + Z^2)^2} \tag{4.34}$$

In order to estimate the model coefficients, considering n different measurements of G and Z, equation (4.33) can be rewritten as a linear system:

$$\underbrace{\begin{pmatrix} G_1 Z_1^2 \\ G_2 Z_2^2 \\ \vdots \\ G_n Z_n^2 \end{pmatrix}}_{\mathbf{b}} = \underbrace{\begin{pmatrix} Z_1^2 & Z_1 & 1 & -G_1 Z_1 & -G_1 \\ Z_2^2 & Z_2 & 1 & -G_2 Z_2 & -G_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ Z_n^2 & Z_n & 1 & -G_n Z_n & -G_n \end{pmatrix}}_{\mathbf{A}} \underbrace{\begin{pmatrix} p_2 \\ p_1 \\ p_0 \\ q_1 \\ q_0 \end{pmatrix}}_{\mathbf{p}} + \underbrace{\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}}_{\mathbf{e}} \tag{4.35}$$

According to the Gauss-Markov theorem, in a linear model where the errors have a zero expectation, have equal variances and are uncorrelated, the best linear unbiased estimator of the coefficients is given by the ordinary least squares estimator [Lehmann, 1951]. It minimizes the sum of squared residuals:

$$\hat{\mathbf{p}} = \underset{\mathbf{p}}{\operatorname{argmin}} \ \|\mathbf{b} - \mathbf{A}\mathbf{p}\|, \tag{4.36}$$

the coefficients $\mathbf{p}$ can be estimated by

$$\hat{\mathbf{p}} = (\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top \mathbf{b} \tag{4.37}$$

The least-squares solution, that minimizes the sum of squared residuals, gives the maximum-likelihood values of the parameters. However, in several cases (e.g., with a small number of measurements, correlated parameters, etc.), the matrix $\mathbf{A}$ could be ill-conditioned that leads to the difficulty in estimating the parameters. In this case, so-called Tikhonov regularization [Tikhonov et al., 2013, Marroquin et al., 1987] (similar to the Levenberg-Marquardt algorithm in non-linear optimizations), an estimator which is no longer unbiased, but has considerably less variance than the least-squares estimator.

$$\hat{\mathbf{p}} = \underset{\mathbf{p}}{\operatorname{argmin}} \ \|\mathbf{b} - \mathbf{A}\mathbf{p}\| + \|\lambda \mathbf{p}\| \tag{4.38}$$

the coefficients $\mathbf{p}$ can be estimated by

$$\hat{\mathbf{p}} = (\mathbf{A}^\top \mathbf{A} + \lambda \mathbf{I})^{-1} \mathbf{A}^\top \mathbf{b} \tag{4.39}$$

where $\mathbf{I}$ is an identity matrix.

4.2.4.2 Simplification of the model

One reason of using equation (4.32) to fit the measurement is that it shows the best performance for a wide range of depth positions. However, in this model there are 5 parameters to be estimated. In a linear system with a large number of parameters, a lot of measurements are required for a reliable estimation, and the estimation could be sensitive to noise. In this case, a simplification of the model is necessary. In our visual servoing scheme, the desired position is set to be the maximum of the image gradient. Assuming that the approximation of the Jacobian is computed using the data (Z position and its corresponding image gradient G) obtained from the initial

position to the desired position (a reduced range), we only need to fit the data using a function to be estimated. In this reduced range, instead of using (4.33), we find that the relation can be simplified using

$$G(Z) = \frac{1}{q_0 + q_1 Z + q_2 Z^2} + \varepsilon. \tag{4.40}$$

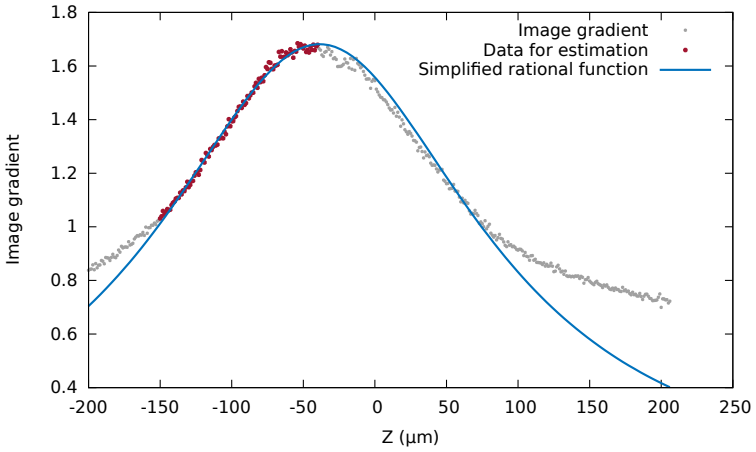


Figure 4.5: Image gradient and simplified rational function using partial data with respect to depth position, respectively. Gray points represents the image gradient measurement in a wide range. The parameters of the simplified rational function (blue line) are estimated from partial data (red points).

Figure (4.5) shows the simplified rational function using partial image gradient data with respect to the depth position. The simplified rational function is estimated using the image gradient and the depth position data in a reduced range (from $Z = -150$ µm to $Z = -40$ µm, we assume that the visual servoing task is performed in this range). In this range, the data is excellently fitted using the simplified rational function. Even though this function is not well fitted in other areas (gray points on the figure), we find that using partial data is enough for our visual servoing task.

As in the previous case, equation (4.40) can be rewritten as a linear system:

$$\underbrace{\begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}}_{\mathbf{b}} = \underbrace{\begin{pmatrix} G_1 Z_1^2 & G_1 Z_1 & G_1 \\ G_2 Z_2^2 & G_2 Z_2 & G_2 \\ \vdots & \vdots & \vdots \\ G_n Z_n^2 & G_n Z_n & G_n \end{pmatrix}}_{\mathbf{A}} \underbrace{\begin{pmatrix} q_2 \\ q_1 \\ q_0 \end{pmatrix}}_{\mathbf{p}} + \underbrace{\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}}_{\mathbf{e}} \tag{4.41}$$

With this simplification, there remains 3 parameters to be estimated. Thus, the approximation of the Jacobian linking image gradient variation to the velocity along

the depth direction is given by

$$L_{app} = -\frac{2q_2Z + q_1}{(q_0 + q_1Z + q_2Z^2)^2} \tag{4.42}$$

4.2.4.3 Validations by simulation

The visual servoing task for motion along the depth direction can be performed similarly as that in the previous section. Instead of computing the Jacobian analytically, in this method the Jacobian is approximated using equation (4.42). In should be noted that, at the beginning of the visual servoing task, Z position should be updated in several iterations with a small fixed displacement in order to obtain enough data (G and Z) to compute equation (4.37). After this step, in each iteration i, the parameters of the rational function (4.40) can be then estimated dynamically from the observed image gradient $\{G_0, G_1, \ldots, G_{i-1}\}$ and the corresponding depth position $\{Z_0, Z_1, \ldots, Z_{i-1}\}$ from previous iterations.

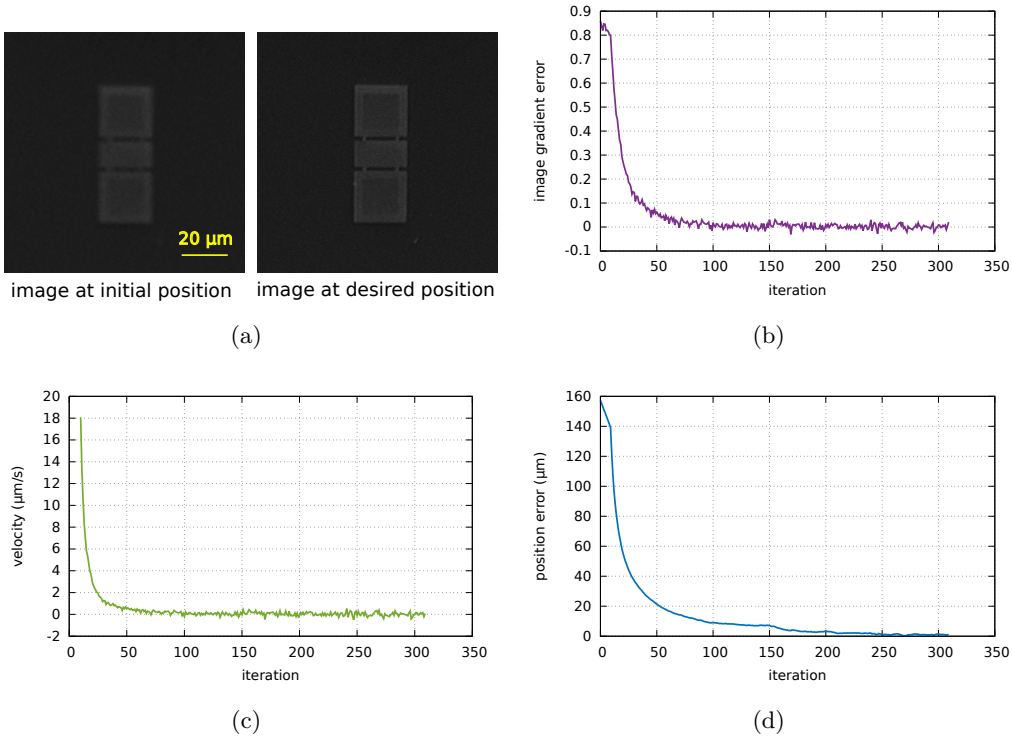


Figure 4.6: Test sample images (a) and simulation results using approximated-Jacobian-based approach: evolution of estimated defocus level and image gradient error (b), velocity along the depth direction (c) and position error (d) (distance between the current position and the desired position) with respect to iterations, respectively

This method has been validated by a simulation using the data extracted from the SEM image sequences. In the simulation, the image gradient for a given depth position

is obtained from a database using look-up table method. In the database, a wide range of depth positions and the corresponding image gradient values of the sample image at these positions are recorded with a small step. In order to simulate the real case, a Gaussian random noise with zero mean and 0.01 standard deviation is added to the obtained image gradient (note that the image gradient per pixel ranges from 0.8 to 1.7 in this simulation). During the simulation, the image gradient is acquired from a given initial position. The obtained results are shown in Figure 4.6. The final position is obtained with an accuracy of 1.5 µm.

4.3 Control of Z by Fourier transform

We have presented in previous sections visual servoing scheme using defocus information in the spatial domain. In fact, instead of studying the relation between the sharpness function and the depth position, the visual servoing can also be performed by estimating the defocus level and its corresponding depth. In this section, a visual servoing scheme for the motion along the depth direction by estimating the defocus level in the frequency domain is proposed.

4.3.1 Determining defocus level in frequency domain

Instead of using a sharpness measurement in the spatial domain, the standard deviation σ of Gaussian Kernel in (4.17) can be considered as an important factor of defocus level. In order to determine the value of σ from a given image and a desired image, we can take the Fourier transform of the linear image defocus model given by equation (4.16):

$$\mathbf{I}_f(u,v) = \mathbf{I}^*_f(u,v)F(u,v), \tag{4.43}$$

where (u,v) is the pixel position in the frequency image. $\mathbf{I}_f(u,v), \mathbf{I}^*_f(u,v)$ and $F(u,v)$ are Fourier transforms of $\mathbf{I}(x,y), \mathbf{I}^*(x,y)$ and $f(x,y)$, respectively. $F(u,v)$ is expressed in the frequency domain by:

$$F(u,v) = e^{-2\pi^2(u^2+v^2)\sigma^2}. \tag{4.44}$$

Applying logarithm to equation (4.44):

$$-2\pi^2(u^2+v^2)\sigma^2 = \ln(\mathbf{I}(u,v)) - \ln(\mathbf{I}^*(u,v)). \tag{4.45}$$

Finally, we can get the square of σ as:

$$\sigma^2 = \frac{\ln(\mathbf{I}^*(u,v)) - \ln(\mathbf{I}(u,v))}{2\pi^2(u^2+v^2)}. \tag{4.46}$$

With equation (4.46), the defocus level σ can be then estimated from the desired (sharp) image and the current image.

It should be noticed that, σ can be obtained from a single pixel in the frequency domain. Nevertheless, since the noise is introduced in the image, the computation from a single pixel could not be accurate. In this case, we propose to compute the average of the estimations from a confidence region. We experimentally find that the pixels at low frequency are more robust to the noise than that at high frequency. The confidence region can be selected as $R_c = \{\mathbf{I}(u,v) | u^2 + v^2 < threshold\}$.

4.3.2 Control law

Given σ as the visual feature, we aim to minimize the error between σ at the current position and σ^* at the desired position. Assume the desired image is located at a position where the sensor is well focused, i.e., $\sigma^* = 0$, the goal is to minimize σ in order to obtain the image at the focused position:

$$\hat{Z} = \underset{Z}{\operatorname{argmin}} \ (\sigma) \tag{4.47}$$

The relationship between the time derivative $\dot{\sigma}$ and the relative camera instantaneous linear velocity $\dot{Z}$ is given by:

$$\dot{\sigma} = L_\sigma \dot{Z} \tag{4.48}$$

where L_σ is the Jacobian (in the case of 1 DoF, it is a scalar). From the inverse of the linear relation described in equation (4.18), L_σ can be expressed by

$$L_\sigma = -\frac{m}{Z^2} \tag{4.49}$$

Given an exponential decay of velocity along z-axis, i.e. $\dot{\sigma} = -\lambda\sigma$, the control law is:

$$v_z = -\lambda L_\sigma^{-1} \sigma \tag{4.50}$$

4.3.3 Experimental validations

The evolution of defocus level σ with regard to depth position is tested with a real SEM image sequence at first. Figure 4.7 shows the evolution of defocus level σ with respect to depth position. As a comparison, the image gradient per pixel with respect to depth position is also shown in the figure. It can be seen from (the left part of) the figure that the estimated defocus level σ decreases when the depth position increases, and σ reaches its minimum when the sample is located at the in-focus position. Therefore, this causes the fact that the estimated σ can be employed as a visual feature to perform visual servoing task. However, it is found that, in the experiments under a SEM, the estimated σ never reaches 0 even though the sample is in-focus. This is mainly caused by the noise on the image which leads to the error on the estimation of σ. The SEM noise should be considered as an important influence to the SEM image. It could be amplified in the frequency domain. Applying denoising filters could reduce the noise,

but it is always difficult to remove the noise by keeping all the "original" information. Actually, even with a slow scan speed (that decreases the noise level), the pixel gray level always varies slightly for each acquired image $\mathbf{I}$. This leads to significant changes in pixel intensities in each frequency-domain image $\mathbf{I}_f$. We assume the error occurs as an addition to the estimated defocus level: $\hat{\sigma} = \sigma + e$. A simple solution for this uncontrolled estimation error is to measure the error by taking a set of images at the same depth position. Let $\{\mathbf{I}_Z^0, \mathbf{I}_Z^1, \ldots, \mathbf{I}_Z^n\}$ be the acquired images at position Z, take $\mathbf{I}_Z^0$ as reference image $\mathbf{I}^*$, for an image $\mathbf{I}_Z^i$, the estimation error e_Z^i can be expressed by the estimated $\hat{\sigma}_Z^i$ using equation (4.46) since the theoretical σ should be 0 for the images at the same depth position. Thus, the estimated error can be modeled and σ can be recovered by $\sigma = \hat{\sigma} - e$.

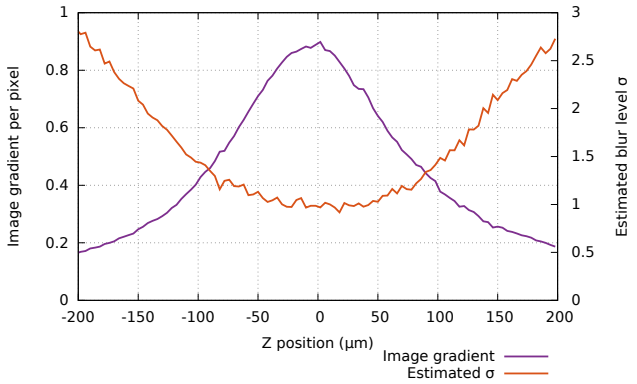


Figure 4.7: Evolution of image gradient per pixel and estimated defocus level σ with respect to depth position

Experiments using Fourier-transform-based visual servoing method have been performed in a SEM (EVO LS 25) at 1000×. The sample used in the experiments is a membrane (see Figure 4.8(a)). The evolution of the defocus level σ with respect to the iterations is shown in Figure 4.8(b). As a comparison, the image gradient at each iteration is also shown in Figure 4.8(b). The velocity computed from the defocus level σ is shown in Figure 4.8(c). Given the noise in the estimation of σ, the resultant oscillation can be seen in the figure.

In order to compare the performance, another experiment using image gradient (see Section 4.2) has been performed on the same condition. The results are shown in Figure 4.9. Similarly, the desired depth position has been achieved. From both Figure 4.8(b) and 4.9(b), image gradient shows more robustness to the image noise compared with the defocus level σ. Less oscillation on velocity is found in the later experiment. Therefore, for the future visual servoing task under a SEM, although the visual servoing task can also be performed by the proposed Fourier-transform-based visual servoing scheme, we propose to use the image gradient as the visual feature for the motion along the depth direction.

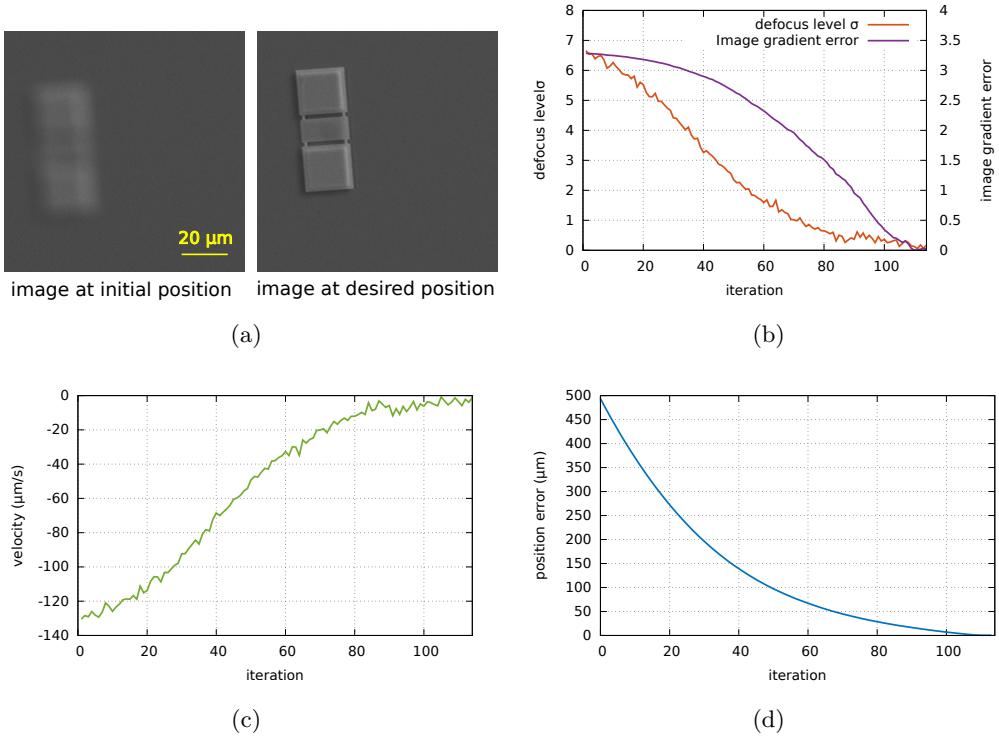


Figure 4.8: Test sample images (a) and visual servoing results with a SEM using Fourier-transform-based approach: evolution of estimated defocus level and image gradient error (b), velocity along the depth direction (c) and position error (d) (distance between the current position and the desired position) with respect to iterations, respectively

4.4 Conclusion

In this chapter, we focus on the design of visual servoing control law for the robot motion along the depth direction. Among various existing sharpness functions, the image gradient is selected as the visual feature in visual servoing. Different approaches have been proposed to perform the visual servoing task. The first approach is to minimize the image gradient error between the desired image and the current image in the spatial domain. The control law can be analytically computed or be approximated using a rational function. Alternatively, the visual servoing for the motion along the depth direction can also be conducted in the frequency domain. The standard deviation in the Gaussian kernel is modeled into the cost function. The visual servoing scheme is conducted by minimizing the estimated standard deviation. The experimental results show that the first approach is robust and accurate. Due to the high noise level in a SEM image, the estimation in the second approach is inaccurate when the current position is close to the in-focus position. Therefore, we propose the spatial domain approach for a 6-DoF micro/nano-positioning task in a SEM.

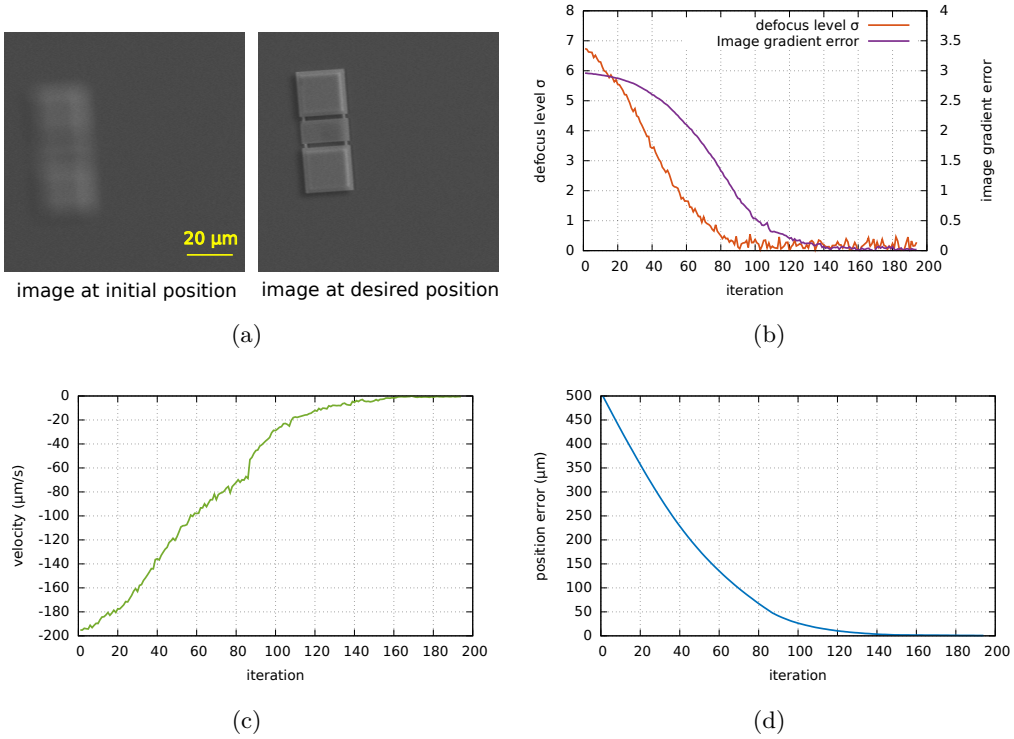


Figure 4.9: Test sample images (a) and visual servoing results with a SEM using image gradient: evolution of estimated defocus level and image gradient error (b), velocity along the depth direction (c) and position error (d) (distance between the current position and the desired position) with respect to iterations, respectively

Chapter 5

Micro/nano-positioning by visual servoing

This chapter presents the automated micro/nano-positioning in a SEM. Since we are able to control the robot motion along the depth direction (presented in Chapter 4), a hybrid visual servoing scheme is proposed in this chapter to perform the micro/nano-positioning task in 6 DoFs. In this approach, the image gradient is considered as a visual feature to compute the control law along the depth direction. Meanwhile, the image intensity is used to derive the control law on other five DoFs. This visual servoing scheme has been validated using an optical microscope as well as a SEM. The experimental results show the good performance of this automated micro/nano-positioning method. The content of this chapter has been partially published in IEEE Int. Conf. on Robotics and Automation, ICRA 2015 [C2] and a minor part of this work has been published in Int. Conf. on Advanced Intelligent Mechatronics, AIM 2014 [C3].

5.1 Hybrid visual servoing

In order to perform the micro/nano-positioning task, a hybrid visual servoing scheme has been proposed. The motion along the depth direction is controlled by the image gradient information using the visual servoing scheme proposed in Section 4.2. Since the image intensity has shown good performance [Tamadazte et al., 2012, Marturi et al., 2014b], the other degrees of freedoms are controlled by the image intensity using a photometric visual servoing scheme.

5.1.1 Image intensity as a visual feature

Considering the intensity of all the pixels I from the pure image $\mathbf{I}$ at current pose $\mathbf{r}(\mathbf{q}) = (t_x, t_y, r_x, r_y, r_z)^\top$ (where $\mathbf{q}$ is the joint coordinates) as the main visual feature, the error between the current visual feature and the desired visual feature is defined as [Collewet et al., 2008, Collewet and Marchand, 2011]:

$$\mathbf{e}_I(\mathbf{r}) = \mathbf{I}(\mathbf{r}) - \mathbf{I}^*(\mathbf{r}^*) \tag{5.1}$$

where $\mathbf{I}^*(\mathbf{r}^*)$ represents the image at desired pose $\mathbf{r}^*$.

For a point $\mathbf{x} = (x, y)$ in the image plane, the time deviation of $\mathbf{x}$ can be expressed by

$$\dot{\mathbf{x}} = \mathbf{L}_\mathbf{x}\mathbf{v}. \tag{5.2}$$

where $\mathbf{v} = (\boldsymbol{v}, \boldsymbol{w})$ contains the relative camera instantaneous linear velocity $\boldsymbol{v} = (\boldsymbol{v}_x, \boldsymbol{v}_y)^\top$ along x- and y-axes and angular velocity $\boldsymbol{w} = (\boldsymbol{w}_x, \boldsymbol{w}_y, \boldsymbol{w}_z)^\top$ around x-, y- and z-axes. $\mathbf{L}_\mathbf{x}$ is the interaction matrix, in parallel projection model, it can be expressed by:

$$\mathbf{L}_\mathbf{x} = \begin{bmatrix} -1 & 0 & 0 & -Z & y \\ 0 & -1 & Z & 0 & -x \end{bmatrix}. \tag{5.3}$$

Let $I(\mathbf{x}, t)$ be the intensity of the pixel $\mathbf{x}$ at time t, then

$$\nabla I = \begin{bmatrix} \frac{\partial I}{\partial x} & 0 \\ 0 & \frac{\partial I}{\partial y} \end{bmatrix}, \tag{5.4}$$

the total deviation of the intensity $I(\mathbf{x}, t)$ can be written as

$$\dot{I}(\mathbf{x}, t) = \nabla I \dot{\mathbf{x}} + \dot{I}, \tag{5.5}$$

where $\dot{I} = \frac{\partial I}{\partial t}$ represents the time variation of I. According to [Horn and Schunck, 1981] based on the temporal luminance constancy hypothesis, $\dot{I}(\mathbf{x}, t) = 0$. In this case,

$$\dot{I} = -\nabla I \mathbf{L}_\mathbf{x}\mathbf{v} = \mathbf{L}_I\mathbf{v}. \tag{5.6}$$

Considering the entire image, $\mathbf{I} = (I_{00}, I_{01}, \cdots, I_{MN})$, where M, N represent the image size:

$$\dot{\mathbf{I}} = \begin{pmatrix} \mathbf{L}_{I_{00}} \\ \vdots \\ \mathbf{L}_{I_{MN}} \end{pmatrix} \mathbf{v} = \mathbf{L_I v} \tag{5.7}$$

where $\dot{\mathbf{I}}$ is the variation of the intensities of the whole image. $\mathbf{L_I}$ is a $MN \times 5$ matrix that theoretically allows the control law to compute the velocity in the 5 DoFs.

5.1.2 Control law for hybrid visual servoing

For our hybrid visual servoing scheme, both the image intensity and the image gradient are considered as visual features $\mathbf{s} = (\mathbf{I}(\mathbf{r}(\mathbf{q})), G(Z))^\top$. Using the image intensity as a visual feature, the velocities (the linear and angular velocities on x- and y-axes, and the angular velocity around z-axis) of the end-effector can be computed from:

$$\dot{\mathbf{q}} = -\lambda \mathbf{J}_I^+ \mathbf{e}_I. \tag{5.8}$$

Considering eye-to-hand visual servoing in our context, the Jacobian is given by

$$\mathbf{J}_I = -\mathbf{L}_I {}^c\tilde{\mathbf{V}}_F {}^F\tilde{\mathbf{J}}_n(\mathbf{q}) \tag{5.9}$$

where ${}^c\tilde{\mathbf{V}}_F$ is a special motion transform matrix which transforms velocity (in 5 DoFs) expressed in camera reference frame onto the robot frame, ${}^F\tilde{\mathbf{J}}_n(\mathbf{q})$ is the robot Jacobian (in 5 DoFs) in the robot reference frame.

When computing the control law, rank deficiency of Jacobian matrix may occur if some values are negligible, because of the specificities of measurement in micro-scale. This leads to the difficulties in correctly computing equation (5.8). To improve the robustness of algorithm, the Levenberg-Marquardt-like method is considered:

$$\dot{\mathbf{q}} = -\lambda(\mathbf{H} + \mu \cdot diag(\mathbf{H}))^{-1} \mathbf{J}_I^\top \mathbf{e}_I \tag{5.10}$$

where μ is a coefficient whose typical value ranges from 0.001 to 0.0001. $diag(\mathbf{H})$ represents a diagonal matrix of the matrix $\mathbf{H} = \mathbf{J}_I^\top \mathbf{J}_I$.

Similarly, using the image gradient as a visual feature, the linear velocity along z-axis is (details can be found in Section 4.2.2):

$$\dot{Z} = -\lambda_z J_G^{-1} e_G(Z) \tag{5.11}$$

where λ_z is an exponential coefficient. $J_G = -L_G {}^c\tilde{V}_F {}^F\tilde{J}_n(Z)$ is the Jacobian where ${}^F\tilde{J}_n(Z)$ represents the robot Jacobian along the depth direction, ${}^c\tilde{V}_F$ (from the motion transform matrix) transforms the velocity along the depth direction from camera frame to robot frame. During the visual servoing process, the control laws for motion along z-axis and the other 5 DoFs are computed respectively.

An observed fact in 6-DoF visual servoing is that both the motion along the depth direction and the motion on other DoFs affect the sharpness of the observed image. In this case, we consider that the variation on the image gradient can be modeled by the motion in all the DoFs:

$$\dot{G} = J_G \dot{Z} + \mathbf{J}_g \mathbf{q}. \tag{5.12}$$

where $\mathbf{J}_g$ is the Jacobian links the variation of image gradient and the velocities on other DoFs. The control law along the depth direction can be then expressed by:

$$\dot{Z} = -\lambda_z J_G^{-1}(\lambda(G - G^*) + \mathbf{J}_g \mathbf{q}) \tag{5.13}$$

With this hybrid control law, the 6-DoF micro/nano-positioning task can be performed at high magnifications. It should be noted that, in the case that the robot motion along the depth direction can be obviously observed by the image (in perspective projection model), the 6-DoF visual servoing can be performed using only the image intensity. The visual servoing framework is same as described in [Collewet et al., 2008, Collewet and Marchand, 2011].

5.2 Experimental validation using optical microscope

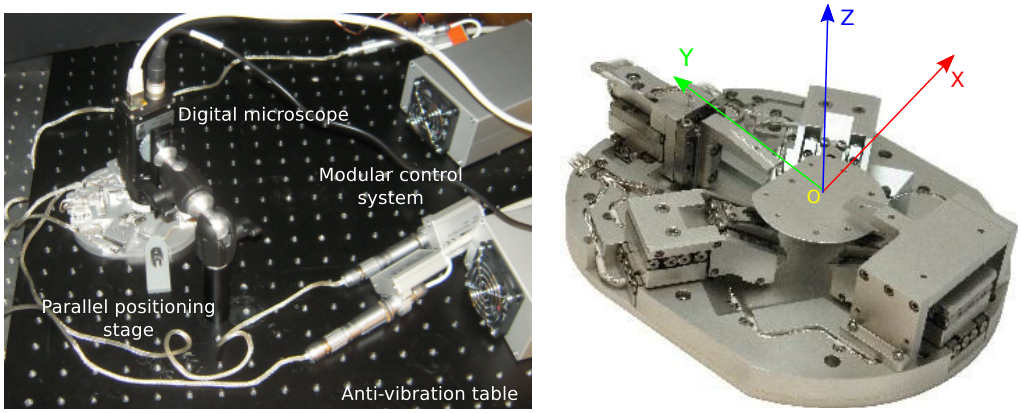
The proposed visual servoing scheme has been first validated using an optical microscope. The hybrid visual servoing scheme, as well as the visual servoing using only the image intensity, is conducted using a parallel robotic positioning stage. The experimental results are illustrated and discussed.

5.2.1 Experimental setup

Experiments have been performed on a micropositioning workcell installed on an antivibration table shown in Figure 5.1(a). It contains a 6-DoF positioning-kinematics micro-stage (SmarPod 70.42-S-HV made by SmarAct with its positioner SLC 17.20-S-HV) as well as its modular control system and a digital microscope (Basler acA1600-60gm) with an aperture-adjustable lens towards the top-plate of positioning stage. Experiments are realized with an optical magnification of 60×.

The SmarPod positioning stage is a parallel robot (hexapod) that provides three positioners supporting a top-plate. The top-plate can be moved in three directions and rotated around three axes by the positioners' motion. The hexapod and the reference frame are shown in Figure 5.1(b). Table 5.1 describes its specifications.

The specimen is a microchip which measures 10 mm×5 mm, with 0.5 mm in thickness. The resolution of acquired image in our experiments from the digital microscope is 659 × 494 pixels.



(a) (b)

Figure 5.1: (a) The micropositioning workcell; (b) parallel positioning stage

Table 5.1: Positioning stage specifications

	Travel range	Closed-loop resolution
X	+/-6 mm	1 nm
Y	+/-6 mm	1 nm
Z	+/-3 mm	1 nm
θ_X	about +/-8°	1 μrad
θ_Y	about +/-8°	1 μrad
θ_Z	about +/-15°	1 μrad

5.2.2 Validation of the method

First, the positioning stage is moved from -2.3 mm to 2.1 mm along the z-axis to evaluate the variation of the image gradient with respect to z position. Images are acquired at each 40 μm step. By computing the image gradient per pixel for each image, the relation between the image gradient and z position is shown in Figure 5.2. It can be seen from this figure that the depth of field is small enough for an accurate positioning task and a single optimum is found in the image gradient.

In the positioning experiments, the stage is first set to an initial pose and then moved to a predefined desired pose iteratively by comparing the image at the desired pose with the image at the current pose. The focus of the microscope is adjusted so that the image is focused at the desired position.

To validate the method, the initial pose of the positioning stage is set to 500 μm in x-axis and 1 mm in y-axis, 2 mm in z-axis; 0.1° around x-axis, 2° around z-axis away from the desired pose to test the performance of the proposed method. The initial image and desired image after image processing is shown in Figure 5.4(a) and Figure 5.4(b), respectively. Figures 5.4(c) and 5.4(d) show the evolution of the image intensity

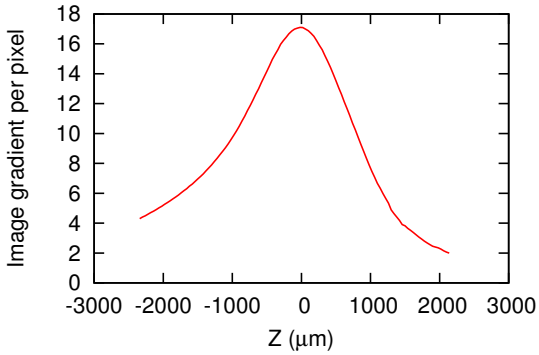


Figure 5.2: Estimated image gradient per pixel of the object images with respect to Z position, using camera Basler acA1600-60gm by varying the Z position

error $\mathbf{e}_I(\mathbf{q}) = \mathbf{I}(\mathbf{q}) - \mathbf{I}^*$ until the end of the visual servoing procedure. The velocities converge fast to 0. As a consequence of the optimization, the error image is almost null.

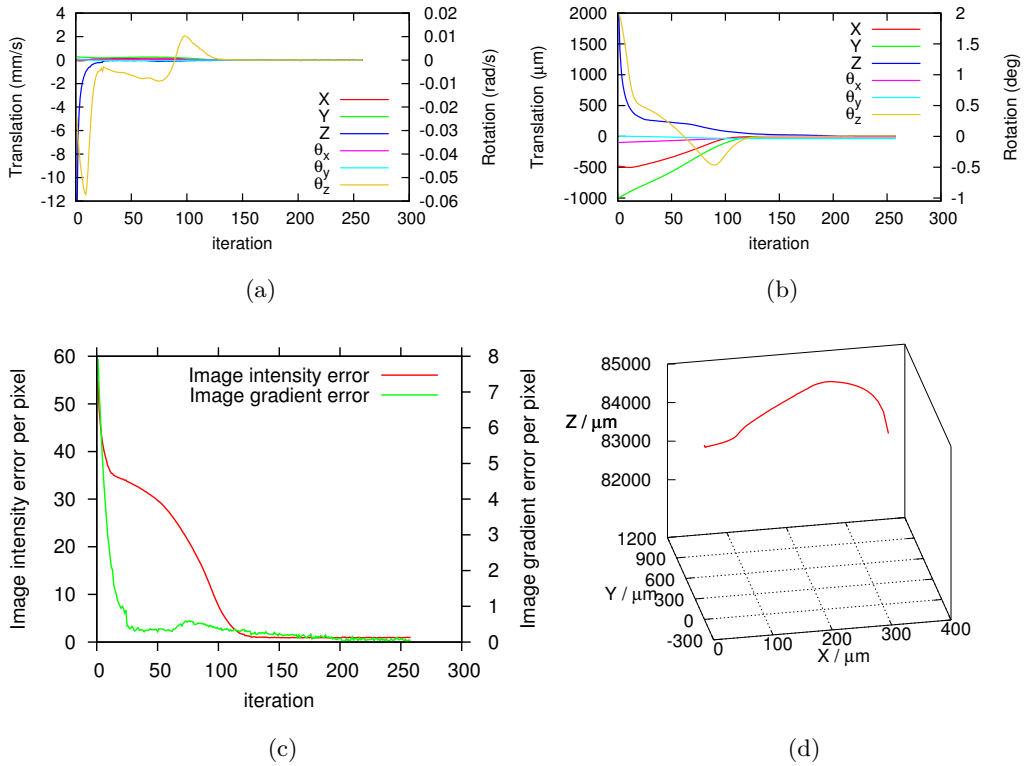


Figure 5.3: 6-DoF positioning using hybrid visual servoing (a) Evolution of joint velocity (in mm/s and rad/s). (b) Evolution of object pose error (in μm/s and degree). (c) Evolution of the image intensity error and the image gradient error per pixel. (d) Object trajectory in camera frame

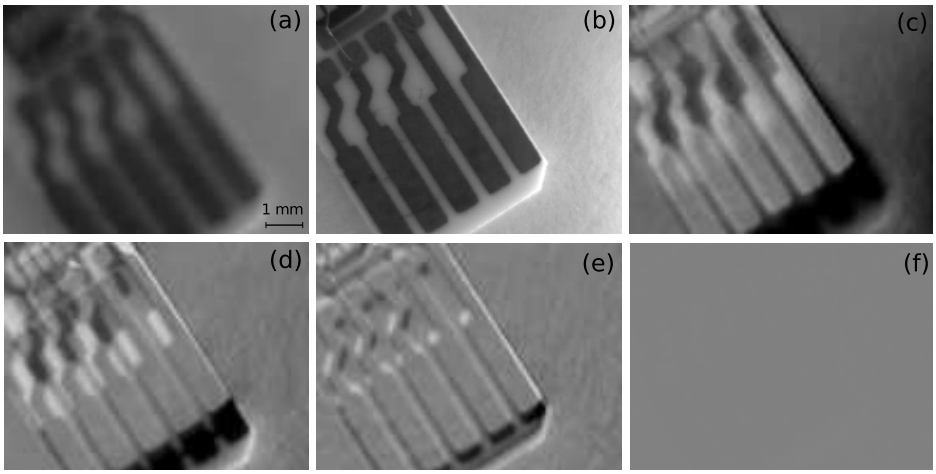


Figure 5.4: Progress of 6-DoF positioning using hybrid visual servoing (a) Initial image, (b) desired image, (c) to (f) show the image intensity error $\mathbf{e}_I(\mathbf{q})$ at $1st$, $16th$, $82th$ and last iteration.

The experimental results are shown in Figure 5.3. Since visual servoing is robust to calibration errors [Espiau, 1993], the positioning task without explicit calibration also performs quite well. The image intensity error and the image gradient error per pixel decrease to negligible values when the velocities converge. The object pose errors between the final pose and the desired pose reach 0.65 μm, 0.47 μm and less than 1 μm in translation along x-, y- and z-axes; 0.027°, 0.036° and 0.003° in rotation around x-, y- and z-axes, respectively. It is mentioned that in Figure 5.3(c), the image gradient error increases around the 70th iteration. It is mainly because the sample is too large to be presented in the whole image. When the positioning stage is moving, details of the sample on the image vary, which causes the computation of the image gradient to be slightly disturbed. In experiments, the proposed visual servoing scheme shows robustness to such situations.

5.2.3 Hybrid visual servoing vs. visual servoing using image intensity

Experiments have been performed to evaluate the proposed hybrid visual servoing by comparing it to an approach using only the image intensity [Cui et al., 2014] in the same conditions. The latter method uses the perspective interaction matrix to compute the control law in 6 DoFs. In fact, in the experiments the variation of the object scale due to the motion along the depth direction can be observed (at magnification 60×). In this case, the perspective projection model can be de facto applied to this optical sensor. The initial pose of the positioning stage is set to be 2 mm on z-axis and 2° around z-axis away from the desired pose. The positioning task based on the proposed hybrid visual servoing and the image intensity based visual servoing are accomplished respectively. The evolutions of the joint velocities are illustrated in Figure 5.5 and Figure 5.6. The

positioning error on translation along z-axis using hybrid visual servoing is less than 1 μm, which is smaller than the error using the image intensity based visual servoing (1.66 μm), in which more iterations are needed for the convergence. On other DoFs, the performances of these two methods are equivalent, where the pose errors are less than 0.6 μm in translation along x- and y-axes, and less than 0.01° in rotation around x-, y- and z-axes.

Furthermore, because of the limited travel range of the positioning stage on z-axis, the initial pose on z cannot be extremely far away from the desired pose. Indeed, in that case, the image-intensity-only method fails to converge because few details can be extracted from the initial blurred image. However, the hybrid visual servoing performs well since the motion on z-axis can be conducted even the image is heavily blurred.

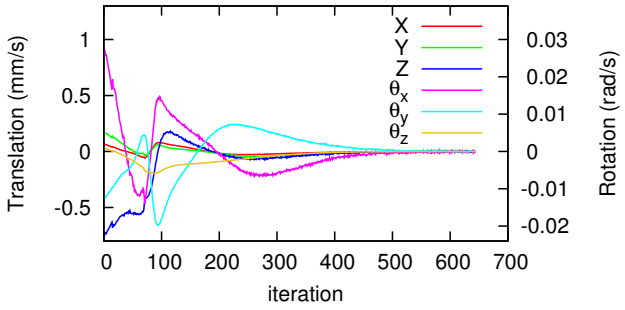


Figure 5.5: Positioning using only image-intensity-based visual servoing: evolution of joint velocity (in mm/s and rad/s)

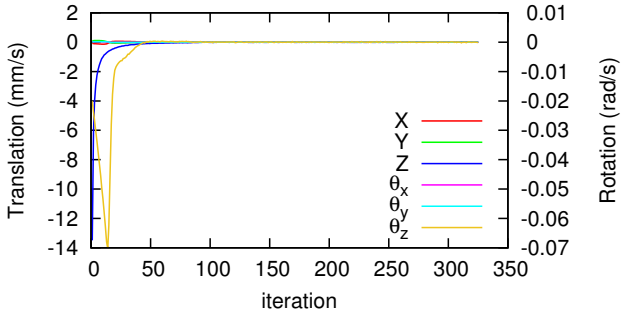


Figure 5.6: Positioning using hybrid visual servoing: evolution of joint velocity (in mm/s and rad/s)

5.2.4 Robustness to light variations

As the proposed method uses photometric information, the sensitivity to variable light conditions is an important issue. Therefore, the robustness to light variations of the proposed method is tested. The initial pose of the stage is also set to be 2 mm on z-axis

and 2° around z-axis. Figure 5.7 shows the evolution of joint velocity. The luminance of the environment light is changed suddenly at the 8th iteration. Oscillations in velocities appear, caused by the lighting changes. However, the convergence and the accuracy are unaffected in spite of the changing light. The system remains stable to a small perturbation occurring during the positioning task.

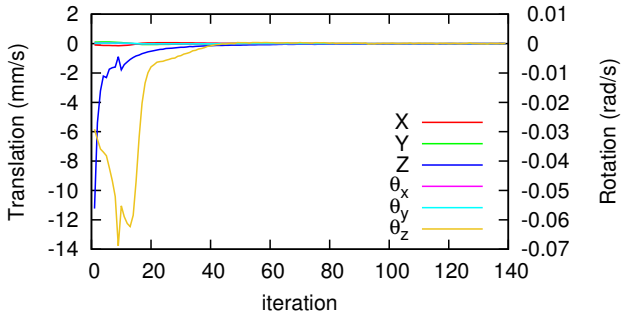


Figure 5.7: Positioning using hybrid visual servoing with lightning perturbation: Evolution of joint velocities (in mm/s and rad/s)

5.3 Experimental validation in SEM

Since the proposed hybrid visual servoing scheme has been validated using an optical microscope, it is necessary to evaluate its performance in SEM environment which is our achieved goal. These experiments have been conducted at ISIR, UPMC. In this section, the experimental setup and the experimental results are presented.

5.3.1 Experimental setup

A robotic platform has been established to perform the micro-positioning experiments. The positioning stage is the same parallel robot (see Figure 5.1(b)) that has been employed in the previous experiments. The SEM in our platform is Carl Zeiss EVO LS 25. The magnification of this SEM ranges from 5× to 1,000,000×. The setup inside the SEM chamber is shown in Figure 5.8. There are two CCD cameras installed inside the vacuum chamber to monitor the positioning stage (see Figure 5.9). During the experiments, the acceleration voltage is 17.14 kV and the probe current is 1.789 A.

Three samples have been employed in our experiments:

- Calibration rig (gold and silicon, Figure 5.10 (a))
- Membrane (indium phosphide and silicon, Figure 5.10 (b))
- MEMS (silicon and oxide, Figure 5.10 (c))

Considering the texture and the size of the sample, the membranes are used in our positioning experiments.

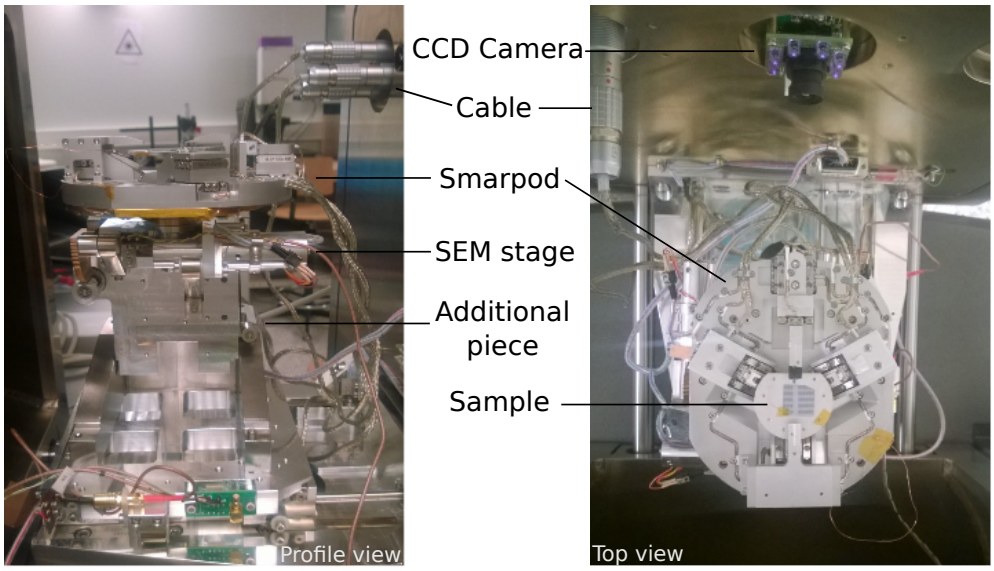


Figure 5.8: Experimental setup inside SEM

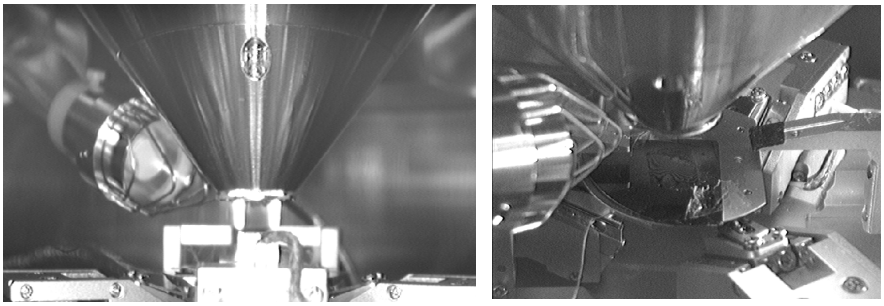


Figure 5.9: Views of positioning stage from CCD cameras inside the SEM

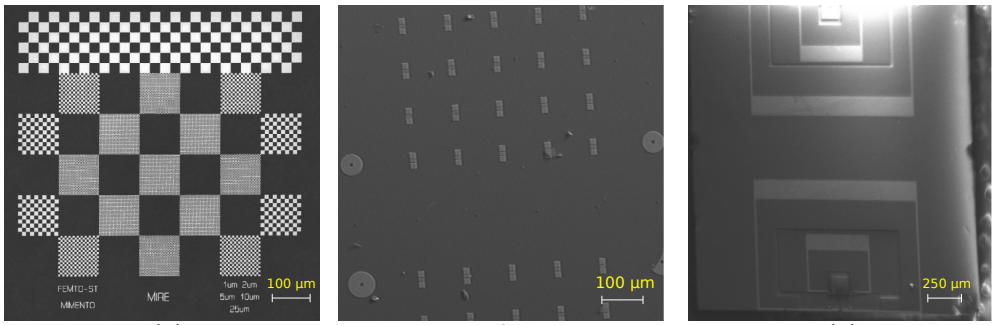


Figure 5.10: Samples used in experiments: (a) Calibration rig, (b) membranes, (c) MEMS

5.3.2 SEM Image quality issues for vision-based control

5.3.2.1 Drift

In the experiments, the drift (see Section 1.4) occurs, potentially due to the mechanical instability of the column or the sample support, thermal expansion and contraction

of the microscope components, accumulation of the charges in the column, mechanical disturbances etc. In our experiments, the drift has been observed by recording a serial of images of the planar calibration pattern (see Figure 5.10 (a)) at 1000× during a long time. The positions (x, y) of corners on pixel in each image have been extracted (see Figure 5.11) using openCV corner detection algorithm. Taking the first image as a reference image, the root mean of squared error (RMSE) for the kth image has been computed using

$$\begin{cases} RMSE_x^k = \sqrt{\frac{\sum_{i=1}^{n}(x_i^k - x_i^*)^2}{n}} \\ RMSE_y^k = \sqrt{\frac{\sum_{i=1}^{n}(y_i^k - y_i^*)^2}{n}} \end{cases} \tag{5.14}$$

where n is the number of extracted corner in the image. (x_i^k, y_i^k) and (x_i^*, y_i^*) are the positions of the ith point on kth image and that on the reference image, respectively.

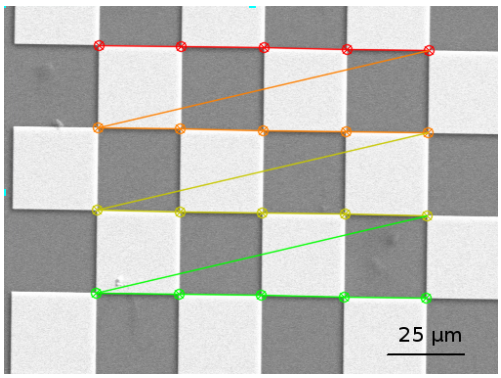
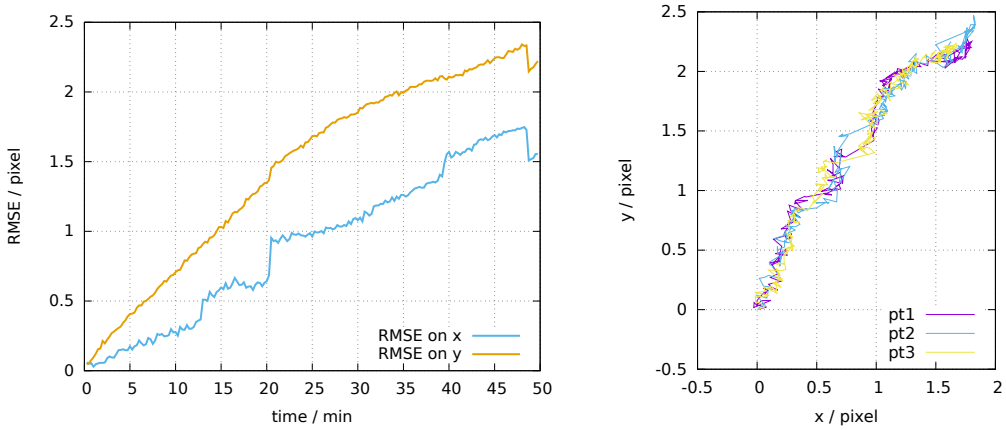


Figure 5.11: Corner detection on calibration pattern

Figure 5.12(a) shows the computed RMSE on both horizontal and vertical positions with respect to time. Figure 5.12(b) shows the evolution of 3 points' (among all the observed points) positions (on pixel) with respect to the reference image. It can be noticed that the time-dependent drift is not a regular variation in the image which can be expressed analytically. By observing the evolution of the detected points, we conclude that the drift is negligible in our positioning process (which usually takes less than 10 minutes using fast or medium scan speed).

5.3.2.2 Noise

In our experiments, three different denoising methods-the median filter, the Gaussian filter and a non-local mean filter [Buades et al., 2005] have been applied into the positioning task. The performance of these denoising methods is validated using a real SEM image. Three images (Figure 5.13) with different textures, contrasts and scan speeds are selected in the experiments. Image-1 (size: 506 × 376 pixels) and Image-2 (size: 360 × 360 pixels) are from the calibration rig and Image-3 (size: 506 × 376



(a) RMSE of all the points position w.r.t. time

(b) Displacement of 3 points w.r.t. reference image

Figure 5.12: Experimental results on drift in SEM

pixels) is from the membrane.

(a) Image-1 (b) Image-2 (c) Image-3

Figure 5.13: SEM images used for denoising filter evaluation

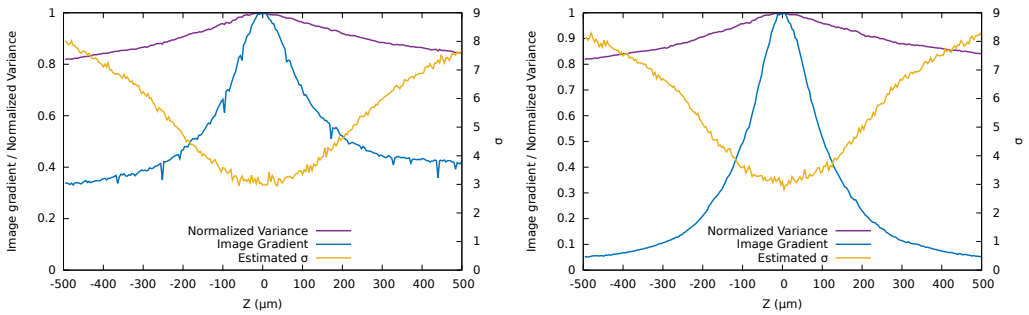
Since no original image (image without noise) can be employed as a reference, we use a noise level measurement method proposed in [Liu et al., 2013] to evaluate the performance from a single image. Table 5.2 shows the estimated noise level and computing time using Gaussian filter, median filter, non-local mean filter and without filters. It should be noted that the parameters of the Gaussian filter and the non-local mean filter are modified in each experiment to achieve a good compromise of denoising and sharpness. It can be seen from the table that the non-local mean method has a good performance for noise level but time-consuming for computing. Since the computing time of a Gaussian filter and a median filter is much less than the non-local mean method, they can also be applied when the noise level of the filtered image is acceptable.

Table 5.2: Noise level and computing time (ms) of denoising filters

Filter	Image-1		Image-2		Image-3	
	Noise level	Time	Noise level	Time	Noise level	Time
No filter	3.375	-	0.940	-	0.813	-
Gaussian	1.259	0.708	0.479	0.552	0.221	0.685
3×3 Median	0.378	0.175	0.359	0.155	0.190	0.169
Non-local Mean	0.350	288.219	0.291	214.644	0.0496	288.183

5.3.3 Experimental results

An evaluation of the sharpness functions using both noisy image and denoised image are first performed. An image sequence of the sample (see Figure 5.13 (c)) obtained by varying the depth position is used to test the performance of the image gradient, the normalized variance and the estimated standard deviation of the focus kernel σ (which corresponds to the method described in Section 4.2, 4.3).



(a) (b)

Figure 5.14: Normalized variance, image gradient (on normalized scale) and estimated σ with respect to the Z position using (a) original noisy images and (b) denoised images by a non-local mean filter

Figure 5.14(a) and 5.14(b) show the comparison of the three considered sharpness functions with respect to the position on the depth direction using the acquired original noisy images and the denoised images by a non-local mean filter. In these figures, the normalized variance and the image gradient are demonstrated on a unified scale. From Figure 5.14, it is seen that image gradient is sensitive to depth position changes but not robust to SEM noise. The image gradient shows good performance when the denoising procedure is applied. The normalized variance is robust to SEM noise but not sensitive to small movements along the depth direction. Due to the SEM noise, the estimated σ is not accurate, especially when the sample is close to the focused position. Since the acquired image texture in each image varies, small oscillation of the estimated σ is

found. σ value remains unchanged when the filter is applied.

In the experiment of 6-DoF visual servoing, the magnification of the SEM is set to be 1000×. The SEM images are acquired with a medium scan speed (about 3.3 µs/pixel). At the first time, the membrane sample (see Figure 5.10 (b)) is located at an initial pose, and then it is moved towards the desired pose by comparing the current image of the sample and the desired image. In the first experiment, the initial pose is set to be 10 µm on x-axis, 3 µm on y-axis, 120 µm on z-axis; $-0.2°$ around x-axis, $0.5°$ around y-axis, $-3°$ around z-axis from the desired pose. Since the initial pose error on z-axis is much greater than other DoFs, in order to reduce the influence between the motion along z-axis and the motion along other DoFs and to reduce the noise, the visual servoing task is performed by the following steps:

(1) Perform visual servoing only on z-axis for coarse positioning on the depth direction using the image gradient;

(2) Perform visual servoing on other 5 DoFs for coarse positioning using image intensity;

(3) Perform visual servoing on all the 6 DoFs for precise positioning.

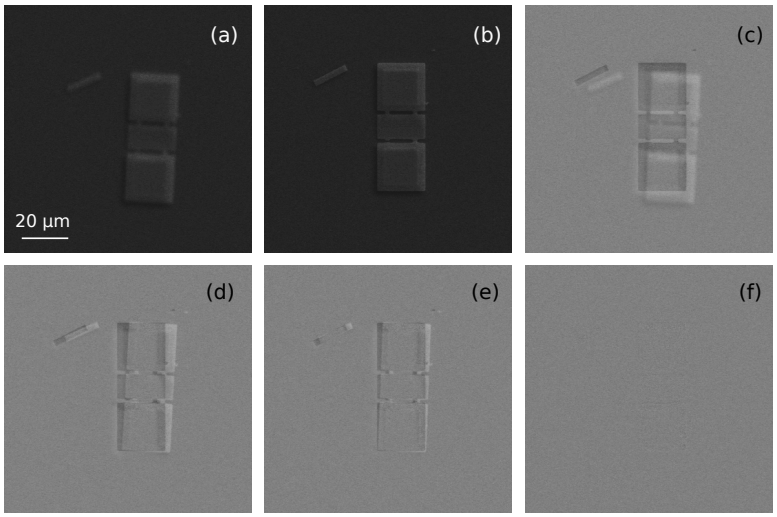


Figure 5.15: Snapshots of 6-DoF positioning using hybrid visual servoing in a SEM (a) Initial image, (b) desired image, (c) to (f) show the image intensity error at $1st$, $40th$, $90th$ and last iteration.

Figure 5.15 shows the initial image, the desired image, and the image differences during the visual servoing task in this experiment. Figure 5.16 shows the experimental results. It can be seen that the residual error on the image intensity and the image gradient converge in 200 iterations. In our experiments, the computing time of each iteration is about 1s, including about 400 ms for image acquisition and about 600 ms for computing the control law. In this experiment, the error between the final pose and the desired pose is 0.56 µm on x-axis, 0.02 µm on y-axis, 3.3 µm on z-axis; $0.002°$ around x-axis, $0.01°$ around y-axis and $0.01°$ around z-axis, respectively.

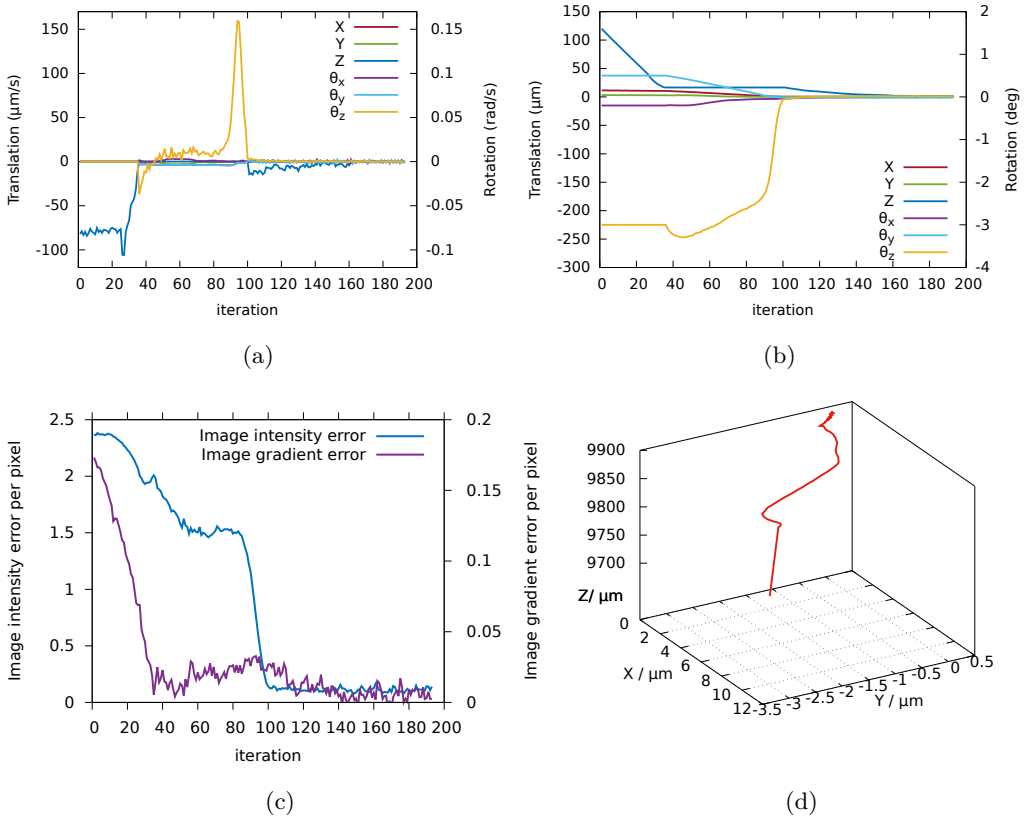


Figure 5.16: 6-DoF positioning using hybrid visual servoing in a SEM (a) Evolution of joint velocity (in µm/s and rad/s). (b) Evolution of object pose error (in μm/s and degree). (c) Evolution of the image intensity error and the image gradient error per pixel. (d) Object trajectory in camera frame

The brightness and the contrast of the SEM image change result from changes in accelerating voltage, spot size (probe current) and tilt angle, each of which changes the secondary electron to backscattered electron ratio. In the experiments, we found that they may also change due to the motion of the sample and the variation of SEM vacuum environmental conditions. Another experiment has been performed in a different condition. In this experiment, the noise level is superior to the previous experiments and the contrast is inferior. The initial pose is set at 10 µm on x-axis, 5 µm on y-axis, 150 µm on z-axis; -1° around x-axis, 2° around y-axis, -5° around z-axis away from the desired pose.

The snapshots of this experiment are shown in Figure 5.17. The experimental results are shown in Figure 5.18. It can be seen from Figure 5.18(c) that the image gradient varies heavily when the current pose is close to the desired pose. The main reason is that the sample that we have does not contain complex texture. In this case, the motion along the depth direction does not change the image gradient obviously. Since the noise level on SEM image is significant, the estimated image gradient varies

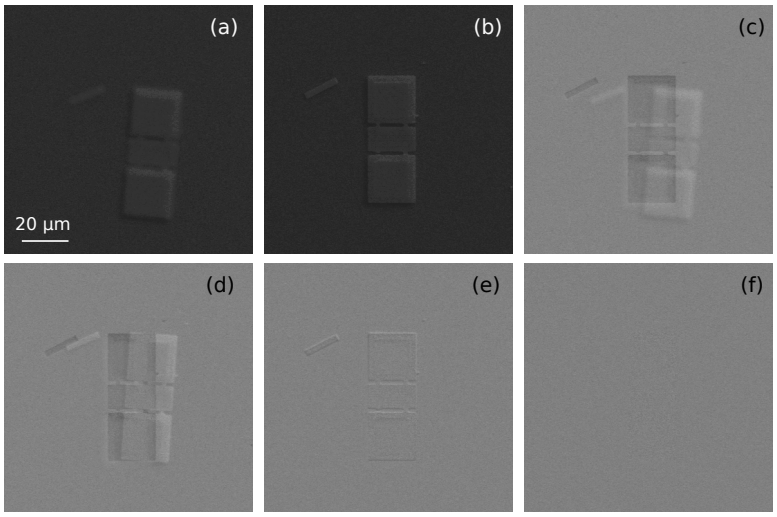


Figure 5.17: Snapshots of 6-DoF positioning using hybrid visual servoing in a SEM (a) Initial image, (b) desired image, (c) to (f) show the image intensity error at $1st$, $50th$, $100th$ and last iteration.

heavily relative to the image contrast in this experiment, even when there is only slight motion along the depth direction. To solve this problem, the gain for depth motion is reduced when the estimated image gradient error is inferior to a given threshold. In this experiment, the error between the final pose and the desired pose is 0.34 µm on x, 0.33 µm on y-axis, 6.8 µm on z-axis; 0.08° around x-axis, 0.11° around y-axis and 0.04° around z-axis, respectively.

5.3.4 Discussion

It is evident that the accuracy on the depth direction is inferior to other DoFs. To improve the accuracy, the depth of field should be reduced. In fact, according to equations (1.2) and (1.3), depth of field decreases at high magnifications and when the sample is close to the objective lens of the SEM. Considering the available sample that can be used in our experiments, the images are acquired at 1000×. Actually, a higher magnification such as 10,000× is more appropriate since the images sharpness variations are more obvious at high magnifications. Moreover, the visual servoing could be more robust if a sample with complex texture is employed.

In our experiments, we find that when we use the membrane sample, if the initial pose is far away from the desired pose (on the $x-y$ plane), the proposed intensity-based visual servoing scheme could not work well. The main reason is that the membrane sample does not contain complex textures. The cost function could be no longer unimodal around the initial pose and the robot could move to a local optimum. To avoid this problem, it is suggested to perform a coarse visual servoing on the $x-y$ plane before applying the proposed visual servoing scheme for an accurate positioning. This

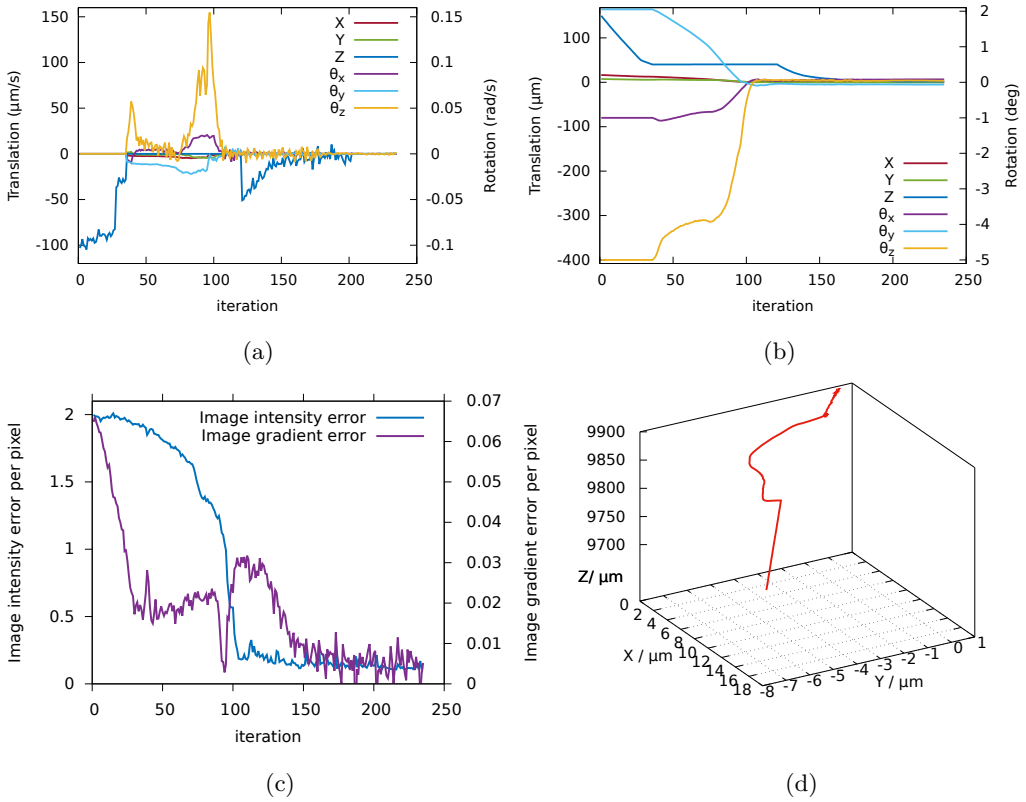


Figure 5.18: 6-DoF positioning using hybrid visual servoing in a SEM, in noisy condition (a) Evolution of joint velocity (in µm/s and rad/s). (b) Evolution of object pose error (in μm/s and degree). (c) Evolution of the image intensity error and the image gradient error per pixel. (d) Object trajectory in camera frame

can be achieved by computing the displacement between the initial and the desired object position on the $x-y$ plane by matching the two images or alternatively, by a global method on visual servoing such as using the histogram of the intensities as a visual feature [Bateux and Marchand, 2015].

5.4 Conclusion

In this chapter, a hybrid visual servoing scheme is proposed for an automated micro/nano-positioning task in 6 DoFs. Different from traditional visual tracking and object localization approach, only pure image appearance information is required in this method. The image intensity information is employed to control the linear motion along x- and y-axes and the angular motion around x-, y- and z-axes. Based on the research in Chapter 4, the image gradient is introduced as a visual feature according to the variation of image sharpness due to the motion along z-axis. This method is validated by experiments on a 6-DoF parallel positioning stage and an optical microscope at first.

The performance of the hybrid visual servoing scheme and that of the visual servoing using only image intensity are evaluated and compared. Considering their performance, we suggest the hybrid method for SEM-based applications. The latter method can be applied when the depth of field of the sensor is large and the magnification is very low. In this case, the motion along the depth direction can be obviously observed from the image and the perspective projection model can be applied. Finally, the hybrid visual servoing scheme is validated using the same robot in a SEM at 1000× for 6-DoF micropositioning. As discussed previously, future works could be the validation of this approach at a higher magnification using different experimental setups and different samples and improving the robustness of the hybrid visual servoing scheme.

Chapter 6

SEM Autofocusing

For high accuracy in manipulation tasks or micro/nano-scale measurements under a scanning electron microscope (SEM), high-quality and sharp images are always required. For this purpose, an efficient and reliable SEM autofocusing task has to be performed before the manipulation process. Based on the study presented in Chapter 4, here we propose a closed-loop control scheme for SEM autofocusing. The proper value of SEM focal length (working distance) is obtained by maximizing the image gradient. The experimental results from a SEM in various conditions validate this method. The content of this chapter has been published in Int. Symp. on Optomechatronics Technology, ISOT 2015 [C1].

6.1 SEM Autofocusing overview

Autofocusing is a process of maximizing the image sharpness by regulating the device focus sets. There are two types of autofocusing techniques [Baina and Dublet, 1995]: active methods, which use a different subsystem to modify the lens position and passive methods, which solely rely on the image sharpness information. Out of the two, passive methods are commonly employed for microscopic devices.

Most of the autofocusing methods are based on evaluating the image sharpness score i.e., the score should reach a single optimum of a selected sharpness function at the in-focus image. Therefore, many sharpness criteria such as image variance, autocorrelation, wavelets, Fourier transform were discussed [Groen et al., 1985, Vollath, 1987, Krotkov, 1988, Firestone et al., 1991]. Considering microscopic applications, some authors have evaluated the available sharpness functions[1] [Yeo et al., 1993, Santos et al., 1997, Sun et al., 2005]. A comparison of these criteria regarding electron microscopy was discussed in [Rudnaya et al., 2010].

In this chapter, we use a SEM as a reference application for our autofocus method. To perform the passive autofocusing process with SEM, a first method is to obtain a sequence of images within a given defocus range and to compute their sharpness scores. The optimal SEM focal length that corresponds to the maximum of sharpness score is then obtained [Erasmus and Smith, 1982]. The main drawback in this approach is that it requires the acquisition of many images, which is time-consuming in a wide focus range of SEM. Alternatively, a second method is to start with an initial set of SEM imaging parameters that correspond to a defocus image. Then an iterative algorithm is used to search for the best focus position [Batten, 2000, Rudnaya et al., 2009]. Even though these methods are effective, they are highly dependent on the search history. Rudnaya [Rudnaya et al., 2011] has proposed to use Nelder-Mead method for searching the optimum of image variance. An alternative method has been proposed in [Rudnaya et al., 2012], based on fitting the sharpness function to a quadratic polynomial approximately using some initial measurements. In [Marturi et al., 2013c], the autofocusing has been achieved by computing the derivative of sharpness function numerically. In the frequency domain, an autofocusing method based on Fourier transform has been proposed [Ong et al., 1997] and improved [Ong et al., 1998a]. In [Ong et al., 1998b] the authors have proposed to use autocorrelation as a sharpness function to perform the autofocusing task. Moreover, statistical learning-based autofocusing methods were studied for SEM [Nicolls et al., 1997], but were never implemented.

6.2 Background on SEM focusing

In order to develop an efficient autofocusing scheme, it is necessary to study the background knowledge on SEM focusing. As stated previously (see Section 1.3, 2.3), the

[1] note that these criteria are also considered in the problem of visual servoing in Section 4.1.2.

image formation and projection model of a SEM are different to optical devices [Kratochvil et al., 2009, Cui and Marchand, 2015]. In this section, SEM Focusing geometry and SEM image formation are detailed.

In general, the SEM images are formed by raster scanning a sample surface by means of a focused beam of high-energy electrons. Different sets of electromagnetic lenses that are present in the SEM electron column are responsible for performing the focusing task (SEM components are illustrated in Figure 1.3). The first are the condenser lenses that control the beam diameter and the second are the objective lenses that focus the spot sized beam onto the sample surface. Apart from them, an objective aperture (diameter A) is present in between them to filter out the non-directional electrons. The distance that is measured electronically between the final pole piece of the objective lens and the focal plane is the electronic working distance W (focal length), which plays a vital role in the focusing process. This distance depends on two factors: the beam acceleration voltage and the current passing through the objective lens. In this work, we assume the former remains constant and the main focusing is performed only using the latter. The total focusing process is illustrated in the Figure 6.1(a). For any selected magnification, at a distance $D/2$ on both sides of the focal plane, the beam diameter is two times the length of the pixel diameter. This results in images that look to be acceptably in-focus.

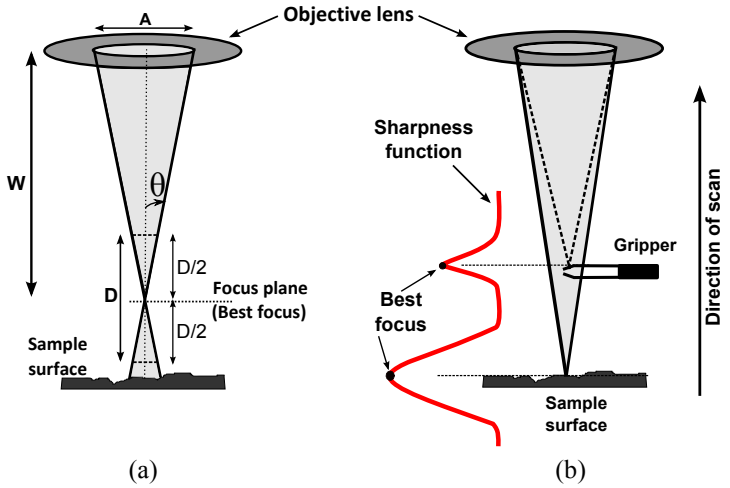


Figure 6.1: (a) SEM focusing geometry (b) sharpness function variation.

Similar to Section 4.2.1 (where the image sharpness depends on the sample position on the depth direction), the image sharpness varies when the electronic working distance changes. Let W be the current working distance, the defocus image $I(x, y, W)$ acquired at W can be expressed as the convolution of a sharp image $I^*(x, y, W^*)$ at the desired working distance W^* and a defocus kernel $f(x, y)$:

$$I(x, y, W) = I^*(x, y, W^*) * f(x, y) \tag{6.1}$$

Using the Gaussian kernel as an approximation of the defocus model, its point spread function (PSF) is given by

$$f(x,y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}. \tag{6.2}$$

where σ is the standard deviation of the Gaussian kernel.

6.3 Closed-Loop autofocus scheme

As mentioned before, autofocusing can be achieved by scanning along the focal axis and by computing the maximum value of the sharpness function (see Figure 6.1(b)). In this Chapter, we consider the autofocusing issue as a control problem and propose a direct closed-loop control scheme to solve it. The objective is to control the device focal length (i.e. working distance) iteratively based on the time variation of the gradient information of acquired image. An analytical formulation of the relation between the displacement of the working distance and the variation of the gradient information is proposed. In this section, we will first show how to use the image gradient for a closed-loop control scheme (in the case of parallel projection) and later we derive the control law to perform the autofocusing task.

6.3.1 Sharpness function and Jacobian

In Chapter 3, it was shown that the gradient-based sharpness measures perform well with the electronic imaging. One of the underlying reasons to use the image gradient is that it shows a good compromise in the case of unstable image contrast, which is an important fact to be considered with SEM [Cornille, 2005]. Moreover, it was proven in previous chapters that the image gradient shows good performance as a visual feature for controlling the motion along the depth direction. For these reasons, the image gradient is considered as a good sharpness function for our SEM autofocusing task. Instead of varying the sample position in the visual servoing task, in the autofocusing scheme, the sample is motionless and the best focus is obtained by varying the working distance of the SEM. In this case, the working distance is then the parameter to be optimized.

Figure 6.2 shows the variation of the image gradient for a series of SEM electronic working distances. Considering the fact that the image gradient varies when the image focus changes i.e., when the working distance varies, we aim to update the working distance by a closed-loop control law to obtain the maximum of the image gradient. For an acquired image $I(x,y)$, recall the image gradient:

$$G = \sum_{x=0}^{M} \sum_{y=0}^{N} (\nabla {I_x}^2(x,y) + \nabla {I_y}^2(x,y)). \tag{6.3}$$

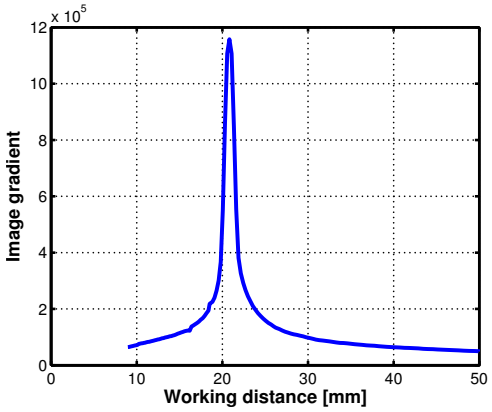


Figure 6.2: Evolution of the image gradient with respect to the working distance.

where, $\nabla I_x^2(x,y)$ and $\nabla I_y^2(x,y)$ represent the squares of gradient in x and y directions, respectively.

In order to use the image gradient information as the sharpness function for full scale in a SEM, the relation between the temporal variations of working distance W and the image gradient G are considered:

$$\dot{G} = J_G \dot{W}. \tag{6.4}$$

The Jacobian J_G in equation (6.4), which links the variation of the image gradient to the time derivative of the working distance, can be expressed by

$$J_G = \frac{\partial G}{\partial \sigma} \frac{\partial \sigma}{\partial W} \tag{6.5}$$

The details on computation of the Jacobian can be found in Section 4.2, where the process of computation is quite similar.

6.3.2 Control law

The objective of our approach is to maximize the image gradient G by controlling the working distance W to obtain an optimized focus of SEM. In order to maximize G, we aim to minimize a cost function given by

$$\varepsilon(W) = \alpha e^{-\beta G(W)} - \gamma \tag{6.6}$$

where $\alpha, \beta \in \mathbf{R}^+$ are adaptive gains that control the variation of working distance and the speed of convergence. γ is a small value that can be considered as a threshold to determine if the optimal focus is reached. An illustration of this cost function is shown in Figure 6.3. α can be considered as the gain of the cost function. β controls the shape of the cost function. The speed of convergence of the error ε will be increased when β is increased. The optimal working distance W^* is obtained when the error ε is inferior to γ.

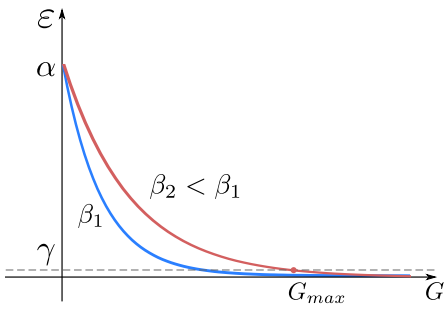


Figure 6.3: Cost function (equation (6.6)) to compute the control law

Considering an exponential decrease of the error i.e., $\dot{\varepsilon} = -\lambda\varepsilon$, the control law is:

$$\xi = -\lambda J_{\varepsilon}^{-1}\varepsilon \tag{6.7}$$

where, ξ is the velocity along the focal axis and J_{ε} is the Jacobian and can be expressed by

$$\begin{aligned} J_{\varepsilon} &= \frac{\partial \varepsilon}{\partial W} \\ &= -(\varepsilon + \gamma)\beta J_G. \end{aligned} \tag{6.8}$$

Rewriting equation (6.7) using equation (6.8), leads to

$$\xi = \frac{\lambda\varepsilon}{(\varepsilon + \gamma)\beta J_G} \tag{6.9}$$

Subsequently, the W displacement (working distance) to be set with the SEM has been computed as follows

$$\Delta W = \xi \Delta t \tag{6.10}$$

where Δt is the time between two image acquisitions. For each iteration, the working distance is updated as given by

$$W_{new} = \begin{cases} W_{prev} - |\Delta W| & \text{if } W_0 \text{ close to } W_{max} \\ W_{prev} + |\Delta W| & \text{if } W_0 \text{ close to } W_{min} \end{cases} \tag{6.11}$$

where W_{new} is the working distance to be updated, W_{prev} and W_0 are previous and initial working distances, respectively, $|\Delta W|$ is the magnitude of ΔW, $W_{max} = 50$ and $W_{min} = 9$ are the factory provided maximum and minimum values for the electronic working distance (in mm) of the employed SEM, respectively. In our experiments, equation (6.11) is used to control the direction of the displacement computed by the control law. For the initial working distance close to a middle value between 1 and 50, according to the single maximum in the evolution of the image gradient with respect to the working distance (see Figure 6.2), the direction can be obtained by comparing $G(W_0)$ with $G(W_0 + \mathrm{d}W)$, where $\mathrm{d}W$ is a small change in the working distance.

6.4 Experimental validations in SEM

6.4.1 Experimental setup

In order to validate the proposed method, different experiments have been realized at FEMTO-ST Institute. Figure 6.4 shows the experimental setup architecture used for this work. The SEM used is a Jeol JSM 820 tungsten gun SEM that is equipped with a conventional Everhart-Thornley SE detector. Its electron column is equipped with different sets of electromagnetic lenses and an objective aperture strip containing 4 changeable apertures of different diameters. The magnification of the SEM varies from $10\times$ to $100,000\times$ and the maximum allowable electronic working distance is 50 mm. A beam control and image acquisition system, DISS5 (from point electron GmbH) has been interfaced with the microscope. It is mainly responsible for sending the scan parameters to SEM and to acquire the data coming from SE detector. Later this data is amplified, digitized and saved as an image in the computer to which the DISS5 is connected.

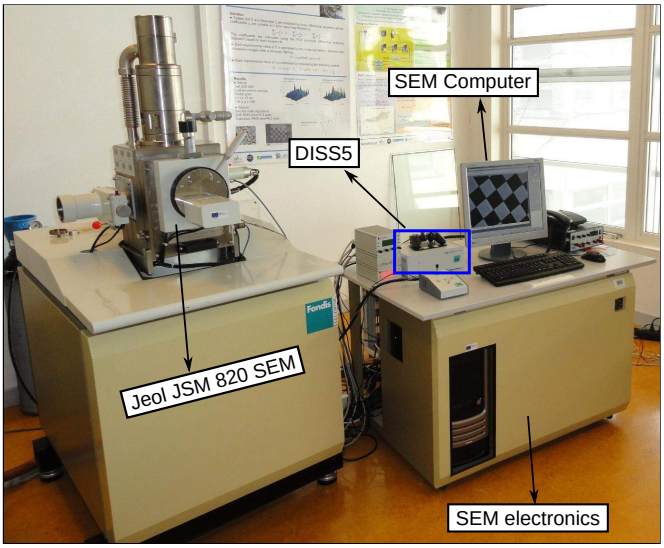


Figure 6.4: Experimental setup architecture.

All the autofocus experiments are performed using the SE images of size 512×512 pixels and are monitored using the developed special purpose graphical user interface program. Besides, DISS5 provides a user interface control for the device focus by linking the working distance with a range of focus steps, i.e., each step corresponds to a specific working distance. The relation between these two parameters is illustrated in Figure 6.5. The focus steps value is obtained by varying the working distance from 9 mm to 50 mm. The experiments are performed in this range where the optimal focus is obtained (see Figure 6.2). For experiments with this system, a model given by equation (6.12) has been obtained by approximating the curve using least squares fitting. This

Table 6.1: Coefficients for the focus step to working distance model.

Coefficient	5 kV	10 kV
p_1	-5.0964e-06	-2.0953e-05
p_2	0.00080911	0.0022916
p_3	-0.052185	-0.10777
p_4	1.8515	3.0379
p_5	-45.185	-65.817
p_6	1259.7	1783.4

model will be used to compute the corresponding focus step for a working distance given by equation (6.11) to modify the device focus.

$$F = \begin{cases} \sum_{j=1}^{C=6} p_j W^{C-j} & \text{if } 9 < W < 50 \\ 586 & \text{if } W \geq 50 \\ 973 & \text{if } W \leq 9 \end{cases} \tag{6.12}$$

where the focus step $F = 586$ and $F = 973$ correspond to the maximum and minimum working distance in our experiments, respectively. $p_{i=1...6}$ are the coefficients of the model and F is the focus step of the SEM. The computed coefficients using least squares fitting at different acceleration voltages used for the experiments are summarized in Table 6.1. As the acceleration voltage used to excite the electrons vary the focusing model, for each voltage used in this work, a corresponding model has been derived. However, for the experiments, the voltage is fixed for all the tests performed with a specific sample.

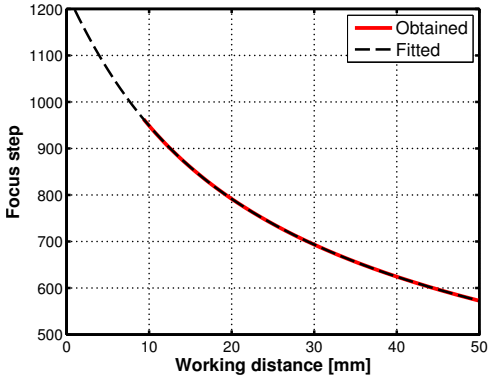


Figure 6.5: Relationship between working distance and focus steps in Jeol SEM using an acceleration voltage of 5 kV.

6.4.2 Validation of the method

An initial test is performed to validate the performance of the proposed method. The sample used for the experiment is a silicon micropart (Figure 6.6(b)) whose dimensions are $10 \times 500 \times 20$ µm^3. The acceleration voltage used to generate the electron beam is 5 kV and has been fixed through all the experiments performed with this sample. The magnification used for this test is 300× and the images are acquired with a raster scan speed of 0.72 µs/pixel, which provides a frame rate of 2.2 frames per second. The brightness and the contrast are set to optimal values for the image acquisition process. The evolution of the focus step and the image gradient are shown in Figure 6.7(a) and the variations of velocity and working distance are shown in Figure 6.7(b). From the obtained results, it is evident that the velocity decreases to 0 when the image gradient reaches its maximum, which points out that the best focus has been accomplished successfully.

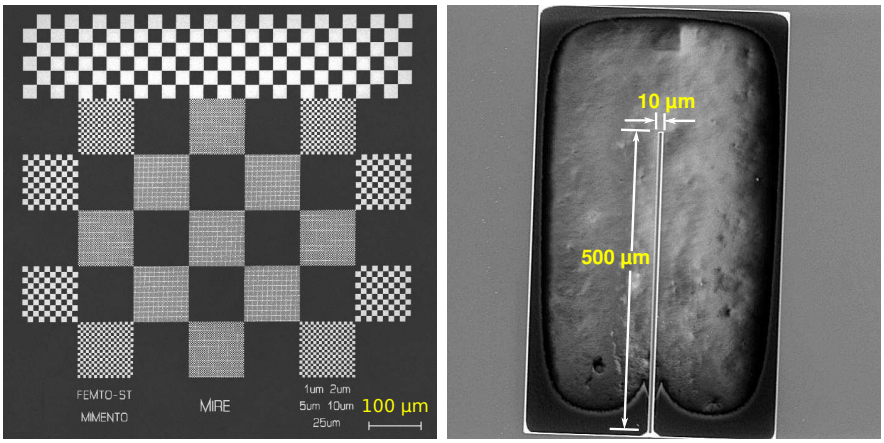
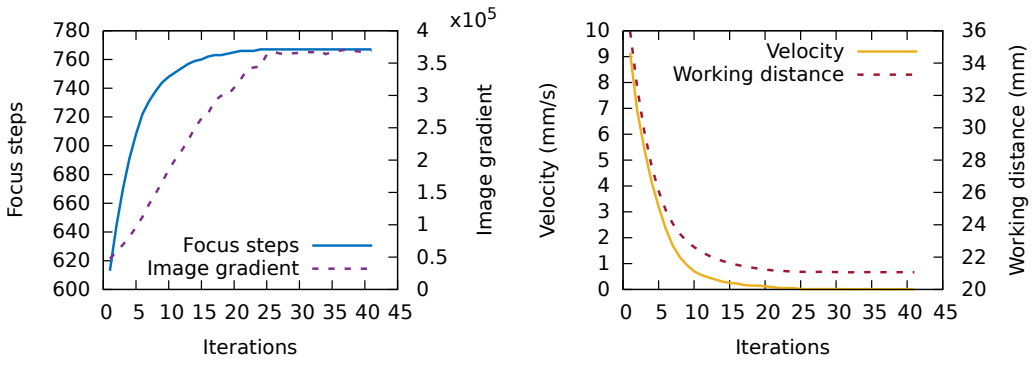


Figure 6.6: The samples[3] used for the experiments: (a) sample-1: microscale calibration rig (b) sample-2: Silicon micropart.

6.4.3 Validation under different conditions

Several experiments have been conducted to validate the proposed method at various experimental conditions that include the variation in scan speed and magnification. Usually with SEM, usage of higher scan speeds degrades the useful image information by increasing the level of random noise [Marturi et al., 2014a], which slightly affects the image gradient. However, any such influence can be readily compensated by the closed-loop control scheme. Apart from that, the performance of the method has also been evaluated by comparing it with an iterative search-based method [Batten, 2000]. It is a three-fold technique that operates in three different iterations by varying the step

[3]Both two samples are fabricated at the clean room facility of FEMTO-ST Institute.



(a) (b)

Figure 6.7: Validation of the method at a magnification of 300×: Evolution of (a) focus step and image gradient (b) absolute velocity and working distance during the proposed process.

size (distance between working distances) to search for the best focus position that provides the maximum image sharpness. A normalized variance sharpness function has been used with this method. For the experiments, the step sizes used are 50, 5 and 1, respectively in each iteration. In both cases i.e., for the proposed and the search-based methods, the optimum working distance estimated by a skilled human operator has been used as the reference in computing the error.

As a pre-processing step, the images were filtered using a Gaussian filter of size 5×5 to reduce the level of noise. Two samples used for these experiments are a microscale calibration rig containing chessboard patterns (Figure 6.6(a)), for which the magnification varies from 300× to 1200× with a step change of 300 and the silicon microparts (Figure 6.6(b)), for which the magnification varies from 100× to 400× with a step change of 100. For simplicity, hereafter we call calibration rig as sample-1 and silicon microparts as sample-2. The acceleration voltages used for the sample-1 and sample-2 are 10 kV and 5 kV, respectively.

Table 6.2 and Table 6.3 summarize the obtained results with sample-1 at different magnifications using scan speeds of 0.72 µs/pixel (optimal) and 0.18 µs/pixel (high), respectively. From these results, it can be noticed that the accuracy of proposed method is better than search-based method under both conditions, with an improved average accuracy of 60% in comparison with the search-based method. This is mainly due to the fact that the proposed method is not affected by the lens hysteresis.

Similar experiments are performed with the sample-2 that contains comparatively fewer textures than sample-1. The obtained results at different magnifications using the optimum and the high scan speeds are presented in Table 6.4 and Table 6.5, respectively. Similar to the previous case, the proposed method has performed better than the search-based method under all the experimental conditions with a comparatively better accuracy ($> 27\%$).

Table 6.2: Autofocus results with sample-1 using the optimal scan speed.

Mag (×)	Obtained working distance (mm)			Error (mm)	
	proposed	manual	search	proposed	search
300	20.984	20.957	21.119	0.027	0.162
600	20.785	20.83	21.014	-0.045	0.184
900	20.864	20.83	20.811	0.034	-0.019
1200	21.037	21.114	21.012	-0.077	-0.102
RMSE				0.049	0.133

Table 6.3: Autofocus results with sample-1 using the high scan speed.

Mag (×)	Obtained working distance (mm)			Error (mm)	
	proposed	manual	search	proposed	search
300	20.953	20.891	20.817	0.062	-0.074
600	21.028	20.934	21.110	0.094	0.176
900	21.017	21.000	21.122	-0.070	0.122
1200	20.831	20.875	20.706	-0.044	-0.115
RMSE				0.055	0.127

Table 6.4: Autofocus results with sample-2 using the optimal scan speed.

Mag (×)	Obtained working distance (mm)			Error (mm)	
	proposed	manual	search	proposed	search
100	21.267	21.279	21.247	0.012	-0.032
200	21.290	21.268	21.017	0.022	-0.251
300	20.799	21.017	20.983	-0.218	-0.034
400	21.130	21.000	21.154	0.13	0.154
RMSE				0.127	0.149

Table 6.5: Autofocus results with sample-2 using the high scan speed.

Mag (×)	Obtained working distance (mm)			Error (mm)	
	proposed	manual	search	proposed	search
100	21.655	21.594	21.437	0.061	-0.157
200	21.235	21.260	21.359	-0.025	0.099
300	21.899	21.718	22.001	0.181	0.283
400	21.530	21.621	21.527	-0.091	-0.094
RMSE				0.106	0.175

From the analysis, the obtained results clearly show the efficiency and the repeatability of the proposed method of autofocus regardless of the sample surface as well as the experimental conditions. Some of the images acquired during different experiments

using sample-1 and sample-2 are shown in the Figure 6.8 and Figure 6.9, respectively. Important comments from the result images relate to the rotation and zoom effect in the defocused images. The former is due to the helical path followed by the electron beam in the presence of magnetic field. The latter is because the diameter of the beam (that interacts with the sample) is high when the object is not in-focus and the pixels appears to be wider and intersected with each other, which looks like a "zoom" image. It has been found that the proposed method is robust to these phenomena during the autofocus process.

6.4.4 Speed test

The final experiments are conducted to evaluate the computing time by the proposed method at different scan speeds. Here the overall time includes the computing time to acquire an image along with the processing time. Later it has been compared to the iterative search-based method. The same defocus range is applied for both methods. The three scan speeds (in µs/pixel) used for the tests are: 0.72 (speed-1), 0.36 (speed-2) and 0.18 (speed-3). This experiment has been performed using sample-1. The obtained results are summarized in the Table 6.6 and they clearly prove the rapidity of the proposed method in converging to the best focus.

Table 6.6: Time taken by both methods at different conditions.

Mag (×)	Scan speed	Proposed		Search	
		images acquired	time (seconds)	images acquired	time (seconds)
600	speed-1	19	9.31	39	19.11
	speed-2	12	4.20	39	13.65
	speed-3	14	3.78	39	10.53
1200	speed-1	21	10.29	39	19.60
	speed-2	13	4.55	39	13.39
	speed-3	11	2.97	39	10.71

6.4.5 Discussion

The obtained experimental results clearly show the accuracy and efficiency of the proposed method in the case of various real world scenarios. Since the Jacobian is computed analytically, the autofocus procedure is proved to be efficient. However, there are few limitations where the performance of the method will be affected. It should be mentioned that in our experiments, both the two samples are flat. If the sample is far from perpendicular to the vision sensor, there is a risk that the sharpness function could get multimodal. Generally in a SEM, the support of sample can be set to be perpendicular to the electron gun by the SEM software. In this case, the tilt is normally smaller than

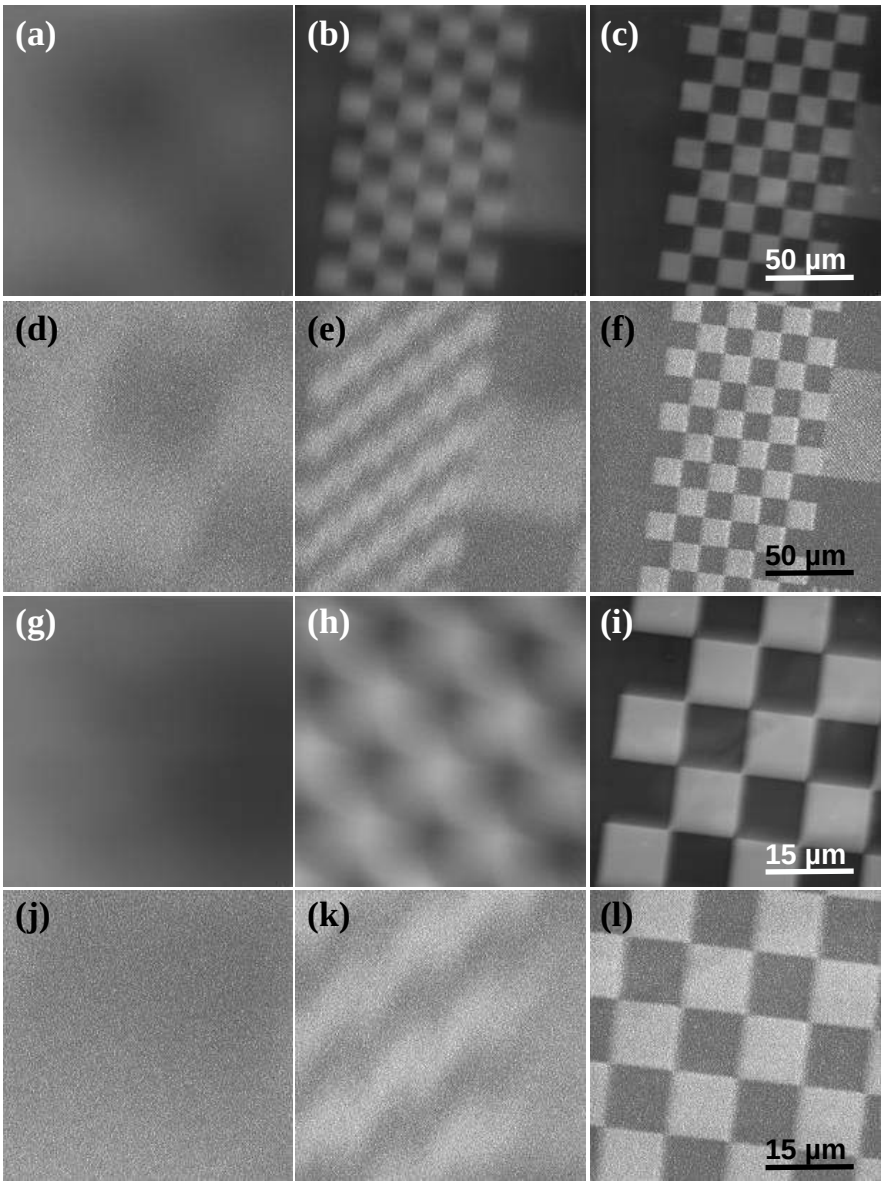


Figure 6.8: Screenshots obtained during the autofocus process using sample-1: (a) to (c) with optimal scan speed at 300× magnification; (d) to (f) with high scan speed at 300× magnification; (g) to (i) with optimal scan speed at 900× magnification; (j) to (l) with high scan speed at 900× magnification. Last column depicts the *in-focus* images.

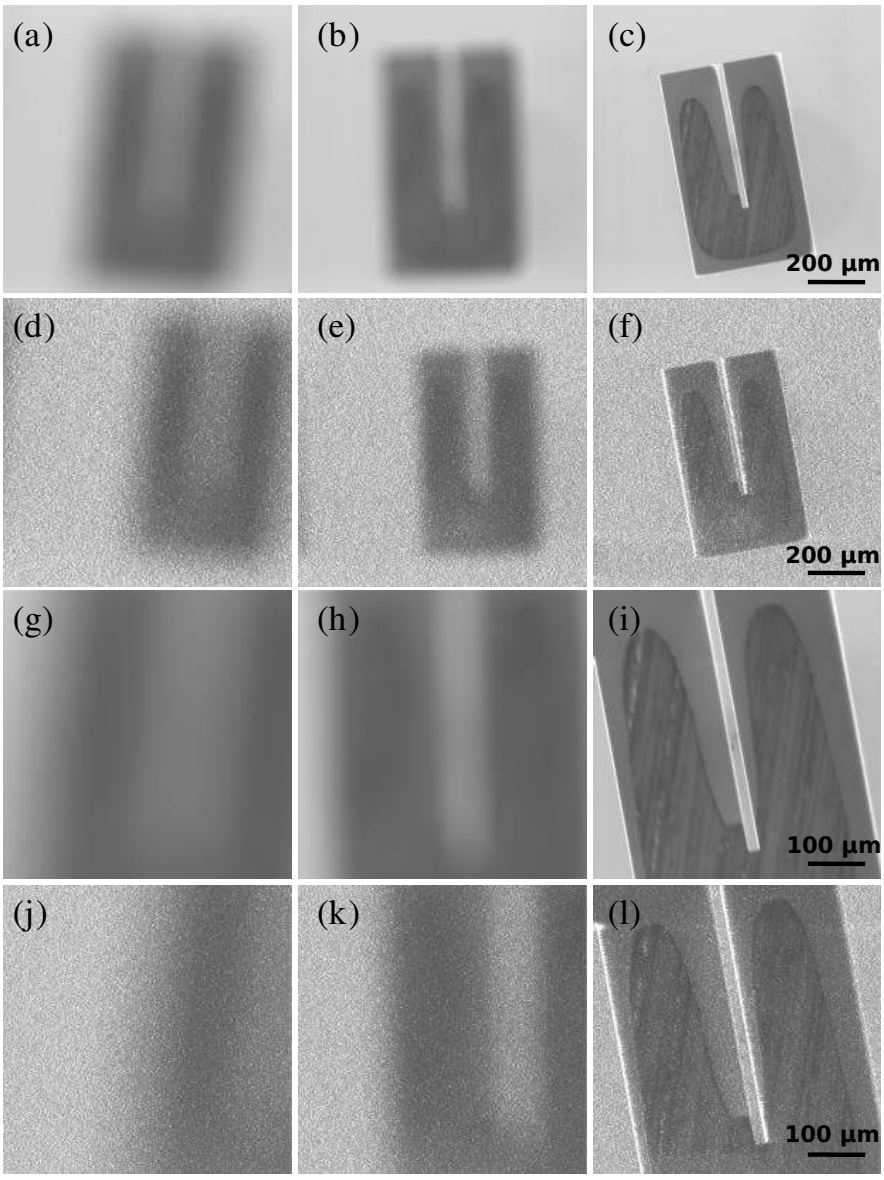


Figure 6.9: Screenshots obtained during the autofocus process using sample-2: (a) to (c) with optimal scan speed at 200× magnification; (d) to (f) with high scan speed at 200× magnification; (g) to (i) with optimal scan speed at 400× magnification; (j) to (l) with high scan speed at 400× magnification. Last column depicts the *in-focus* images.

the field of view, keeping the sample in-focus when the autofocus procedure achieves the optimization. Similar to other autofocus techniques (using any imaging device), the proposed method also requires the objects with sufficient texture information. Due to this requirement, the computation of the image gradient for plain texture-less scenes is obviously impossible.

6.5 Conclusions

In this chapter, a closed-loop control scheme has been proposed for a full scale autofocusing of SEM. It uses the image gradient information as the sharpness score in designing the vision-based control law. The optimum of focus i.e., the maximal image sharpness has been obtained by updating the device working distance iteratively. Since the designed cost function decreases exponentially, the proposed new method quickly converges to the optimal value. Unlike the conventional search-based methods, the proposed method directly reaches the optimal focus position, which makes it robust to the electromagnetic lens hysteresis. The method has been validated for different experimental conditions in terms of performance and speed and the obtained results clearly show the method's efficiency.

Chapter 7

Visual tracking and pose estimation in SEM

Visual tracking and estimation of the 3D pose of the observing object are important to perform visual guidance for the micro/nano-manipulation task. The positioning task can also be achieved by tracking the object and then performing classical visual servoing. In this chapter, we propose a template-based visual tracking method to estimate the 3D pose of the micro-scale object. This method is validated by the experiments in 4 DoFs in a SEM. It is also shown that by applying particle filter in our framework, the accuracy on depth position estimation can be significantly improved.

7.1 Visual tracking in SEM

Generally, visual tracking involves the estimation of the pose or the trajectory of an object by detecting the visual features. Many of the current tracking algorithms are based on the extraction of geometrical features from the image. These features include points of interest [Harris and Stephens, 1988, Hager and Belhumeur, 1998], straight lines [Deriche and Faugeras, 1990], contours or silhouettes [Berger, 1994, Blake and Isard, 1998, Drummond and Cipolla, 2002, Yilmaz et al., 2004], segments [Boukir et al., 1998, Hager and Belhumeur, 1998], etc. The appearance features have been studied by many researchers, such as probability densities of object appearance [Zhu and Yuille, 1996], templates [Lucas and Kanade, 1981, Baker and Matthews, 2004] and active appearance models [Cootes et al., 2001]. However, in a complex environment, the extraction of the features could be affected by occlusions, noises in the image, the complex shape of the object or the loss of information from 3D representations to 2D images. To enhance the feature extraction, Scale-Invariant Feature Transform (SIFT) has been proposed [Lowe, 2004] as a descriptor. Partly inspired from SIFT, authors [Bay et al., 2006] have introduced Speeded Up Robust Features (SURF) for robust and fast local feature detection.

Numerous visual tracking techniques perform by matching the representation of the target model built from the previous frame(s). For example, Kanade-Lucas-Tomasi (KLT) tracker [Lucas and Kanade, 1981, Baker and Matthews, 2004] finds the geometric transformation match between the current frame and a reference template by minimizing (or maximizing) the similarity (or dissimilarity) function. These functions can be the sum of squared differences (SSD) [Shi and Tomasi, 1994], the normalized cross-correlation (NCC) [Irani et al., 1992] and the mutual information [Dame and Marchand, 2010]. Alternatively, other tracking methods are based on the distinction of the target foreground against the background. Some classifiers are built to distinguish target pixels from the background pixels, and updates the classifier by new samples coming in, such as foreground-background tracker [Nguyen and Smeulders, 2006], Hough-based tracking [Godec et al., 2013] and super pixel tracking [Wang et al., 2011]. Instead of using 2D features on image, 3D model of an object has received much attention in visual tracking. The markerless model-based tracking methods have been studied by many researchers [Lowe, 1991, Marchand et al., 1999, Marchand et al., 2001, Drummond and Cipolla, 2002, Comport et al., 2006]. Incorporating both the features and the models, a hybrid visual tracking method has been proposed [Pressigout and Marchand, 2007]. In this method, the points on the visible faces of the model have been taken into account in the visual tracking task to improve the robustness.

Considering the applications in micro-electromechanical systems (MEMSs), many authors focus on the visual guidance techniques for micro-manipulation or micro-assembly [Feddema and Simon, 1998, Zhou et al., 1998, Sun and Chin, 2004]. Some authors have investigated CAD-based tracking algorithms for observing the interested

object during the micromanipulation or microassembly process [Yesin and Nelson, 2005, Lee and Cho, 2009, Tamadazte et al., 2010]. In [Lee et al., 2001], 3D-shaped micro parts have been recognized and tracked using multiple visions for micromanipulation. In microscopy field, some authors have studied the tracking of (fast) moving objects (e.g., cells and bacteria), such as [Teunis et al., 1992, Ogawa et al., 2005]. Visual tracking has also been employed for the estimation of the interaction of the other sensors with the environment. For example, a vision-based tracking approach has been proposed to estimate the forces acting on a cantilever during nanomanipulation [Greminger et al., 2004]. Similar methods have also been proposed by [Liu et al., 2009] for nano-Newton force sensing.

Over the last decade, visual tracking played an important role in automated (or semi-automated) micro/nano-manipulation tasks in a SEM. Nevertheless, only a few tracking algorithms have been actually implemented inside a SEM. An active-contours-based and correlation-based pattern matching method for nanohandling in a SEM has been proposed [Sievers and Fatikow, 2006], in which the pose on 3 DoFs (translation along x- and y-axes, rotation around z-axis) has been estimated. This method has been improved and applied to a microrobot system inside a SEM [Fatikow et al., 2007, Fatikow et al., 2008] for semi-automatic nanohandling. In [Jasper and Fatikow, 2010], instead of acquiring the whole image, dedicated line scans are used to detect the movement of a nano-object or reference pattern. This approach can be applied into a closed-loop positioning task. An advantage of these template-matching-based methods is that its simple implementation and robustness to additive noise on the SEM image. However, these methods highly depend on the template and could be sensitive to clutter. Alternatively, the model-based tracking method has been proposed and implemented for precise automated manipulation and measurement in a SEM [Kratochvil et al., 2009]. The 3D model-based approaches have good performances to estimate the 3D pose of the object, although they show less robustness to additive noise and highly depend on the model and the feature extraction. Recently, [Tamadazte et al., 2010] has proposed a visual tracking framework using CAD model and 3D visual-based control for MEMS microassembly. In this method, a microscale part assembly task is realized by tracking the 3D model of these microscale parts under an optical microscope. Additionally, an improved template matching based contour model was proposed for the tracking task in a SEM [Ru et al., 2012] and was applied into vision-guided nanomanipulation of nanowires using four nanoprobes [Ru et al., 2011]. In this method, a gradient based subpixel method has been introduced to the nanoprobe contour tracking task to improve the accuracy. Moreover, tracking of nanoprobe tips along x and y directions has been implemented into a robotic nanoprobing task [Gong et al., 2014].

It should be noticed that most of the current visual tracking methods ignore a particular fact in SEM that the image sharpness varies when the sample moves along the depth direction, especially at high magnifications. In this case, the acquired SEM image could be blurred due to the defocus. This leads to inaccuracy on the feature

extraction or the template matching process. When the image is significantly blurred, the detection of points or lines in the model-based tracking task could be highly affected and the visual tracking task could fail. Moreover, the previous template-based methods consider only the motion along x and y directions and possibly the rotation around z-axis. In order to estimate the 3D position of the object with high accuracy, the defocus can be considered as an important issue to recover the depth information. In the previous literature, [Dahmen, 2008, Dahmen, 2011] have proposed to employ the normalized variance to recover the position on the depth direction for a 3D position estimation. In this method, the position on the depth direction and the corresponding normalized variance are previously recorded in a data set. During the tracking task, the position on the depth direction is recovered by estimating the normalized variance of the current image and looking up the corresponding position on the depth direction in the data set. This method highly depends on the data set and the estimation could be affected by the random image noise.

7.2 Visual tracking in presence of defocus blur

The sharpness of the image varies with the sample's motion along the depth direction in a SEM. To deal with this problem, it is necessary to model the defocus blur in the observed image into the visual tracking framework. The template-based tracking method is considered in our case, where the appearance of the image is employed.

7.2.1 Template registration for visual tracking

The Kanade-Lucas-Tomasi visual tracking approach [Lucas and Kanade, 1981, Shi and Tomasi, 1994, Baker and Matthews, 2004] is one of the most popular algorithms to determine the displacement of an object by minimizing the differences between a reference template and a given image.

Considering the appearance of the object is learned from a reference template I^* with pixels position $\mathbf{x} \in W$, the idea of this template registration is to look for a new location of these pixels $w(\mathbf{x}, \mathbf{u})$ in the current image I (where $\mathbf{u}$ is the displacement parameters) by minimizing the dissimilarity between the reference image and the current image. The sum of squared differences (SSD) is usually considered as this dissimilarity function:

$$\hat{\mathbf{u}} = \underset{\mathbf{u}}{\operatorname{argmin}} \sum_{\mathbf{x} \in W} (I(w(\mathbf{x}, \mathbf{u})) - I^*(\mathbf{x}))^2 \tag{7.1}$$

Using the Gauss-Newton optimization method to solve this non-linear problem, for each pixel, the first order Taylor expansion of the error $C(\mathbf{u}) = I(w(\mathbf{x}, \mathbf{u})) - I^*(\mathbf{x})$ is given by:

$$C(\mathbf{x}, \mathbf{u} + \delta\mathbf{u}) \approx I(w(\mathbf{x}, \mathbf{u})) + \nabla I \frac{\partial w(\mathbf{x}, \mathbf{u})}{\partial \mathbf{u}} \delta\mathbf{u} - I^*(\mathbf{x}) \tag{7.2}$$

where $\delta\mathbf{u}$ is the increment of the displacement parameters, $\Delta I = (\frac{\partial I}{\partial x}, \frac{\partial I}{\partial y})^\top$ is the gradient of the image evaluated at $w(\mathbf{x}, \mathbf{u})$ and $\frac{\partial w(\mathbf{x},\mathbf{u})}{\partial \mathbf{u}}$ is the Jacobian of the warp.

Injecting equation (7.2) into (7.1):

$$C(\mathbf{x}, \mathbf{u} + \delta\mathbf{u}) = \sum_{\mathbf{x}\in W} (I(w(\mathbf{x}, \mathbf{u})) + \nabla I \frac{\partial w(\mathbf{x}, \mathbf{u})}{\partial \mathbf{u}} \delta\mathbf{u} - I^*(\mathbf{x}))^2 \tag{7.3}$$

The partial derivative of equation (7.3) with respect to $\delta\mathbf{u}$ is:

$$\frac{\partial C(\mathbf{x}, \mathbf{u} + \delta\mathbf{u})}{\partial \delta\mathbf{u}} = 2 \sum_{\mathbf{x}\in W} (\nabla I \frac{\partial w(\mathbf{x}, \mathbf{u})}{\partial \mathbf{u}})^\top (I(w(\mathbf{x}, \mathbf{u})) + \nabla I \frac{\partial w(\mathbf{x}, \mathbf{u})}{\partial \mathbf{u}} \delta\mathbf{u} - I^*(\mathbf{x})). \tag{7.4}$$

It is evident that when the cost function C reaches its minimum, equation (7.4) equals zero. In this case, the increment of the displacement can be then estimated using:

$$\delta\mathbf{u} = H^{-1} \sum_{\mathbf{x}\in W} (\nabla I \frac{\partial w(\mathbf{x}, \mathbf{u})}{\partial \mathbf{u}})^\top (I^*(\mathbf{x}) - I(w(\mathbf{x}, \mathbf{u}))), \tag{7.5}$$

where H is the Gauss-Newton approximation of the Hessian matrix:

$$H = \sum_{\mathbf{x}\in W} (\nabla I \frac{\partial w(\mathbf{x}, \mathbf{u})}{\partial \mathbf{u}})^\top (\nabla I \frac{\partial w(\mathbf{x}, \mathbf{u})}{\partial \mathbf{u}}). \tag{7.6}$$

The displacement parameters $\mathbf{u}$ can be then updated by $\delta\mathbf{u}$ in each iteration during the non-linear optimization process until the convergence.

7.2.2 Warp functions

To express the displacement of an object in the given image with respect to a reference template, the warp functions $w(\cdot)$ are commonly employed. In our case, depending on the degrees of freedom, two warp functions are considered: the sRt transformation (4 DoFs) and the homography (6 DoFs).

The sRt warp function considers the translations along x, y axes, the rotations around z-axis and the scale. $\mathbf{u} = (s, \theta, t_x, t_y)$ between two pixel locations can be modeled as:

$$\mathbf{x}_2 = s\mathbf{R}\mathbf{x}_1 + \mathbf{t} \tag{7.7}$$

where s is a scale factor which corresponds to the depth in general term. However, since parallel projection is applied in our context, it refers to the magnification of the SEM. Since the magnification is fixed in our experiments, here we consider that $s = 1$ and then $\mathbf{u} = (\theta, t_x, t_y)$. In equation (7.7), $\mathbf{t} = (t_x, t_y)^\top$ is a translation vector and $\mathbf{R}$ is a rotation matrix:

$$\mathbf{R} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

The Jacobian of warp $\frac{\partial w(\mathbf{x}, \mathbf{u})}{\partial \mathbf{u}}$ is given by (in case $s = 1$):

$$\frac{\partial w(\mathbf{x}, \mathbf{u})}{\partial \mathbf{u}} = \begin{pmatrix} -x\sin\theta - y\cos\theta & 1 & 0 \\ x\cos\theta - y\sin\theta & 0 & 1 \end{pmatrix}. \tag{7.8}$$

The homography is widely used in computer graphics and computer vision [Baker and Matthews, 2004, Benhimane and Malis, 2006]. It refers to a 2D transformation representing a 3D motion. Given a matrix $\mathbf{H}$ parameterized by $\mathbf{u} = (h_0, h_1, \ldots, h_7)$, such that:

$$\mathbf{x}_2 = \mathbf{H}\mathbf{x}_1 \quad \text{with} \quad \mathbf{H} = \begin{pmatrix} 1+h_0 & h_2 & h_4 \\ h_1 & 1+h_3 & h_5 \\ h_6 & h_7 & 1 \end{pmatrix}. \tag{7.9}$$

The Jacobian of the warp $\dfrac{\partial w(\mathbf{x}, \mathbf{u})}{\partial \mathbf{u}}$ is given by:

$$\frac{\partial w(\mathbf{x}, \mathbf{u})}{\partial \mathbf{u}} = \begin{pmatrix} x_1 & 0 & y_1 & 0 & 1 & 0 & -x_1x_2 & -y_1x_2 \\ 0 & x_1 & 0 & y_1 & 0 & 1 & -x_1y_2 & -y_1y_2 \end{pmatrix}. \tag{7.10}$$

7.2.3 Visual tracking using defocus information

In order to perform a visual tracking task for three-dimensional motions of a micro-scale object in a SEM, the variation of the sharpness of the image caused by the motion of the sample along the depth direction is considered. There are two possible cases concerning the sharpness of the current image and reference template: the template is in-focus while the current image is out of focus, or conversely, the current image is sharper than the template. To simplify the formulation, in this chapter, we assume that the reference template is in-focus during the visual tracking task.

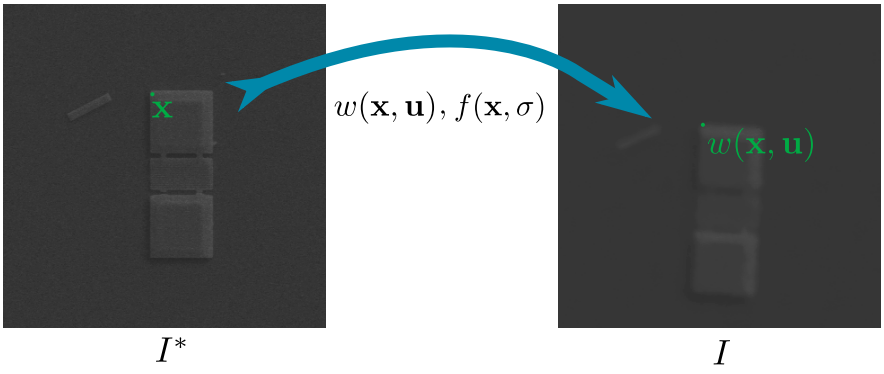


Figure 7.1: Visual tracking based on minimizing the dissimilarity of both displacements and blur level

The general idea is to determine the defocus level σ and the displacement parameters $\mathbf{u} = (\theta, t_x, t_y)$ by minimizing the dissimilarity between the warped and (artificially) blurred image and the reference image using a non-linear optimization process (see Figure 7.1). This problem can be written as:

$$\hat{\mathbf{u}} = \underset{\mathbf{u}}{\operatorname{argmin}} \sum_{\mathbf{x} \in W} \left(I(w(\mathbf{x}, \mathbf{u}), \sigma) - I^*(\mathbf{x}, \sigma^*) \right)^2 \tag{7.11}$$

and

$$\hat{\sigma} = \underset{\sigma}{\operatorname{argmin}} \sum_{\mathbf{x} \in W} (G(w(\mathbf{x}, \mathbf{u}), \sigma) - G^*(\mathbf{x}, \sigma^*))^2 \tag{7.12}$$

where G is the image gradient of image I defined by:

$$\begin{aligned} G &= \sum_{x=0}^{M} \sum_{y=0}^{N} \|\nabla I(x, y)\|^2 \\ &= \sum_{x=0}^{M} \sum_{y=0}^{N} (\nabla I_x^2(x, y) + \nabla I_y^2(x, y)), \end{aligned} \tag{7.13}$$

and G^* is the image gradient of the reference template.

Let us recall the SEM image blur model (see Section 4.2.1). A blurred image $I(x, y)$ can be expressed as the convolution of a sharp image $I^*(x, y)$ and the Gaussian kernel:

$$I(x, y) = I^*(x, y) * f(x, y) \tag{7.14}$$

where the Gaussian kernel $f(x, y)$ can be expressed by:

$$f(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}. \tag{7.15}$$

where σ is the standard deviation of the Gaussian kernel. Since we assume that the reference template is in-focus, in our visual tracking scheme the reference template is blurred artificially using equation (7.14). In this case, σ is considered as the blur level to be optimized.

The Jacobian $J_\sigma = \frac{\partial G}{\partial \sigma}$ linking σ and the gradient G is obtained by (details can be found in Section 4.2.2):

$$\frac{\partial G}{\partial \sigma} = \sum_{x=0}^{M} \sum_{y=0}^{N} 2(\nabla I_x(x, y) \frac{\partial \nabla I_x(x, y)}{\partial \sigma} + \nabla I_y(x, y) \frac{\partial \nabla I_y(x, y)}{\partial \sigma}). \tag{7.16}$$

With the Gauss-Newton optimization method, the minimization problem is solved by updating $\mathbf{u}$ and σ alternatively in each iteration:

$$\partial \mathbf{u} = -\mathbf{J_u}^+ (I(w(\mathbf{x}, \mathbf{u}), \sigma) - I^*) \tag{7.17}$$

and

$$\partial \sigma = -J_\sigma^{-1} (G(w(\mathbf{x}, \mathbf{u}), \sigma) - G^*) \tag{7.18}$$

where the Jacobian $\mathbf{J_u}$ is defined as $\mathbf{J_u} = (\ldots, \nabla I \frac{\partial w(\mathbf{x}, \mathbf{u})}{\partial \mathbf{u}}, \ldots)^\top$.

7.3 Experimental validations of visual tracking

The first experiment evaluates the performance of the proposed visual tracking framework in the presence of blur. The sample is the membrane that has been used in

previous experiments. The images (size 360×360 pixels) are acquired in the SEM Zeiss EVO 25 LS (at ISIR, UPMC) with a medium scan speed (about 3.3 µs/pixel). The sample was positioned on 4 DoFs (translations along x-, y-, and z-axes, rotations around z-axis) and as mentioned before, the magnification is fixed at 1000× during the visual tracking task. Figure 7.2 shows the snapshots of some frames in the experiments. It can be seen that the sample becomes blurred since its position varies along the depth direction. The estimated parameters, including translation along x- and y-axes, rotation around z-axis and the blur level σ are shown in Figure 7.3. Although the rotation around z-axis varies slightly (about 0.04 degree/frame), the evolution of the angle can still be estimated during the visual tracking task. From Figure 7.3(b) one can find that the evolution of σ with respect to the frames can be estimated since the blur is obviously observed.

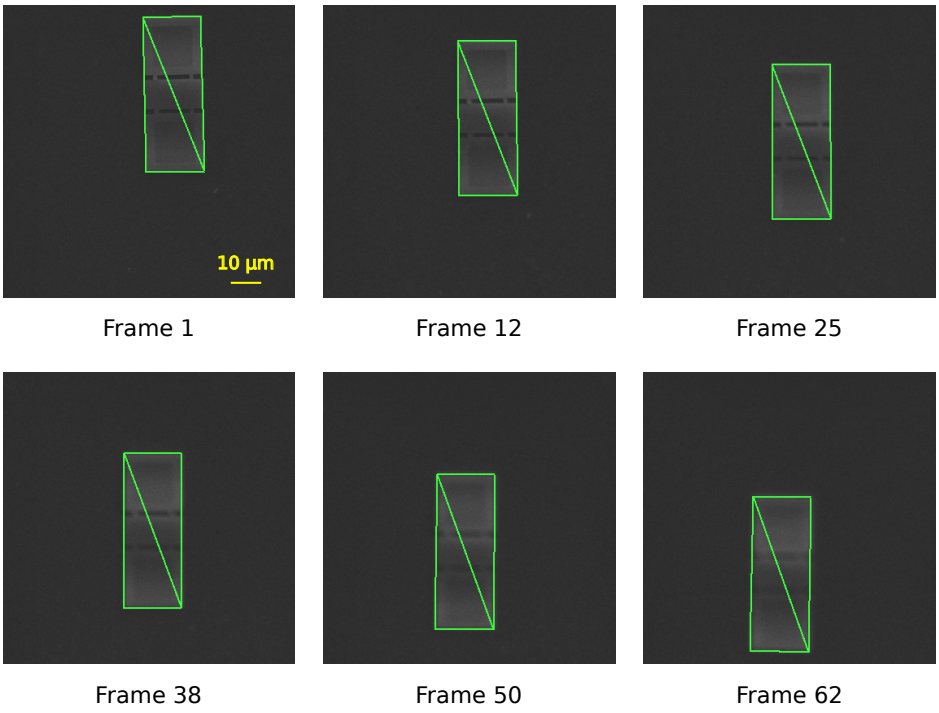
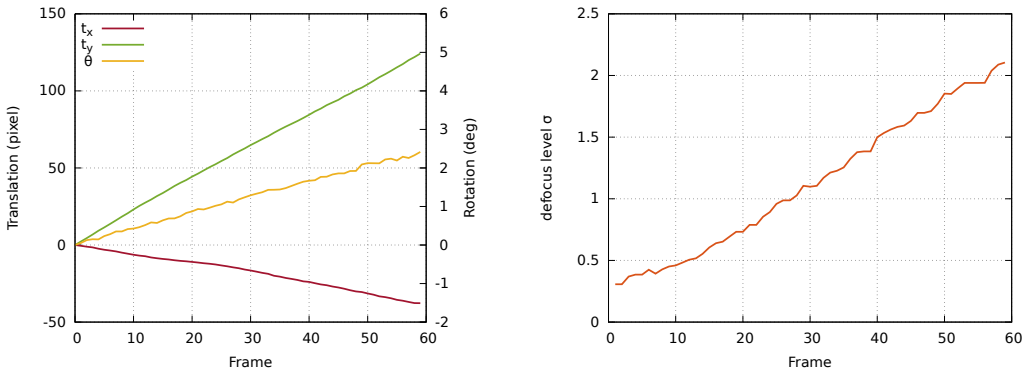


Figure 7.2: Snapshots in visual tracking using proposed method, with medium scan speed

In order to evaluate the proposed approach with respect to the traditional SSD approach and the correlation-based approach (Zero mean Normalized Cross-Correlation, ZNCC) in noisy condition, an experiment has been performed at a high scan speed (about 0.72 µs/pixel) using the same sample at the same magnification. A comparison of zoomed image acquired at the medium and the high scan speed is shown in Figure 7.4. The snapshots of the experiments using the three methods above are shown in Figure 7.5, 7.6 and 7.7, respectively. It is found in the figures that the tracking task could fail using traditional SSD method and ZNCC method if the image is highly



(a) (b)

Figure 7.3: Parameters estimation: (a) translation along x, y, rotation around z; (b) blur level σ

degraded from blur and noise. A reason is that traditional SSD-based or ZNCC-based template matching methods consider only the geometrical transformation of the object on geometry. When the blur is presented in the images, the dissimilarity function could be no longer unimodal. One possible solution for this problem is to detect if the tracking is lost and reinitialize the tracking task. Alternatively, the proposed method shows robustness since the image blur is modeled in the minimization process of the dissimilarity function.

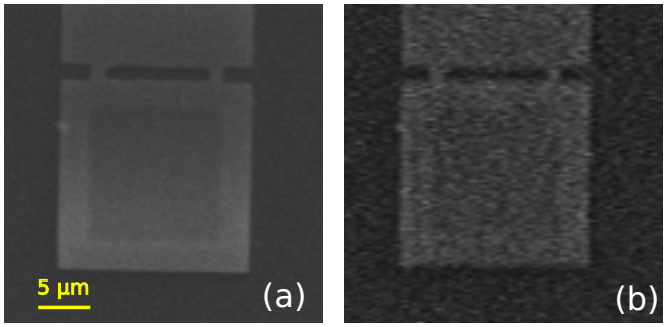


Figure 7.4: SEM images with different scan speed: (a) medium scan speed; (b) high scan speed

7.4 Position and orientation estimation

In the visual tracking process, the parameters in the warp function and the blur level σ are estimated. With these parameters, the pose of the object in the camera coordinate frame or in the world coordinate frame can be then recovered.

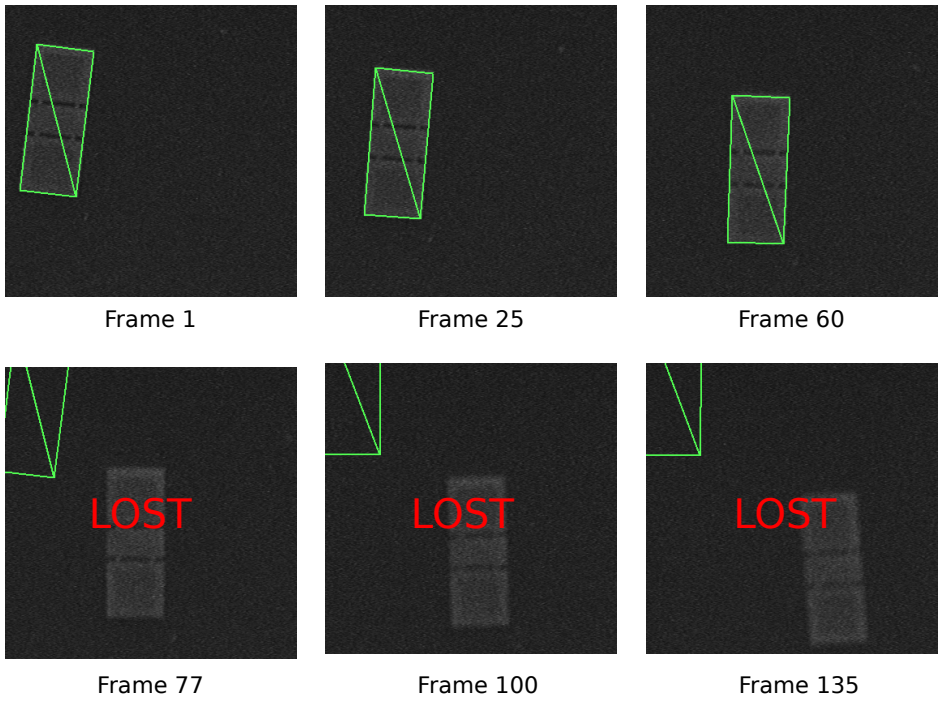


Figure 7.5: Snapshots in visual tracking using traditional SSD method, with high scan speed

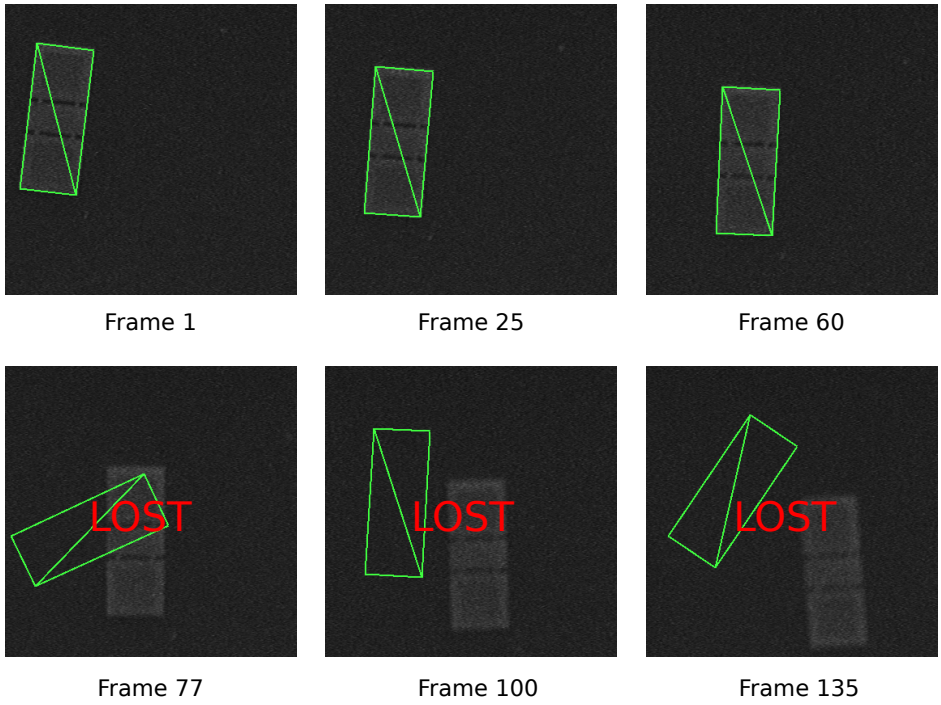


Figure 7.6: Snapshots in visual tracking using traditional ZNCC method, with high scan speed

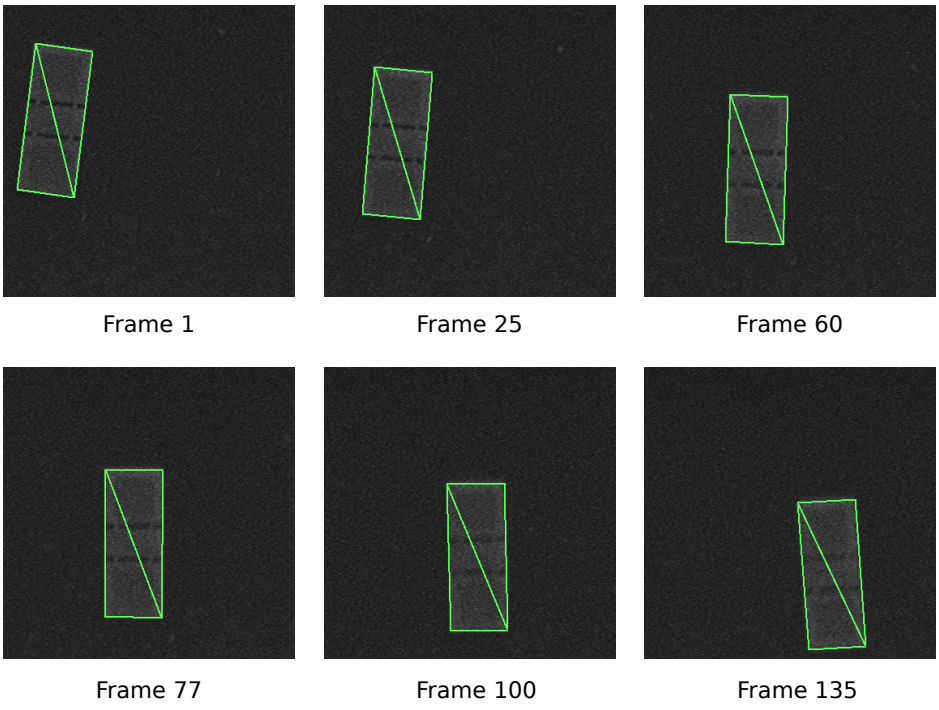


Figure 7.7: Snapshots in visual tracking using proposed method, with high scan speed

7.4.1 Estimating positions and orientation from 3D registration

Considering a 3D point ${}^{w}\mathbf{X} = ({}^{w}X, {}^{w}Y, {}^{w}Z, 1)^{\top}$ in an object reference frame, its projection on the image plane (expressed in pixels) $\mathbf{x}_p = (u, v, 1)^{\top}$ can be model by

$$\mathbf{x}_p = \mathbf{K}\mathbf{\Pi}\,{}^{c}\mathbf{T}_{w}\,{}^{w}\mathbf{X} \tag{7.19}$$

where $\mathbf{K} = \begin{pmatrix} p_x & 0 & 0 \\ 0 & p_y & 0 \\ 0 & 0 & 1 \end{pmatrix}$, $\mathbf{\Pi} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$ and ${}^{c}\mathbf{T}_{w} = \begin{pmatrix} {}^{c}\mathbf{R}_{w} & {}^{c}\mathbf{t}_{w} \\ \mathbf{0}_{3\times 1} & 1 \end{pmatrix}$ is an homogeneous matrix that describes the relation between the object frame and the camera frame. In general, the pixel/meter ratio p_x, p_y can be easily obtained from the SEM software, from calibration procedure (see Chapter 2) or simply computed from a known object measurement in meter and in pixel in the image.

Since the pixel position of a point ${}^{i}\mathbf{x}$ on the image can be estimated from the tracking task, we are able to obtain its 3D pose $\mathbf{r}$ of the object by minimizing the registration error between the re-projected pixel position ${}^{i}\mathbf{x}_p(\mathbf{r})$ and the tracked pixel position ${}^{i}\mathbf{x}_p^*$ using a non-linear optimization. The problem can be written as:

$$\hat{\mathbf{r}} = \underset{\mathbf{r}}{\operatorname{argmin}} \sum_{i=1}^{N} ({}^{i}\mathbf{x}_p(\mathbf{r}) - {}^{i}\mathbf{x}_p^*)^2 \tag{7.20}$$

where N is the number of points used to estimate the pose.

The update in each iteration using the Gauss-Newton optimization method is:

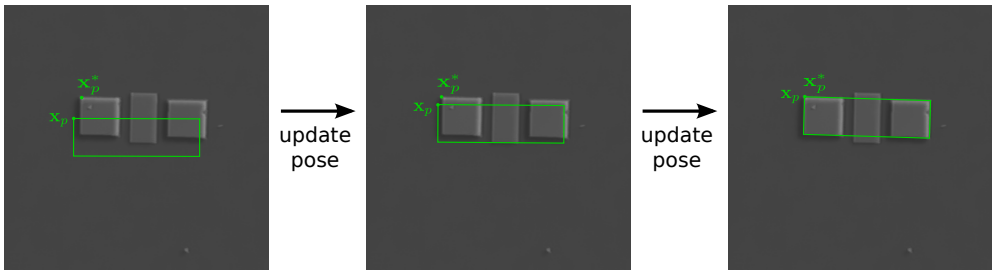


Figure 7.8: Pose estimation by 3D registration

$$\delta \mathbf{r} = -\lambda \mathbf{J}^{+}(\mathbf{x}_p(\mathbf{r}) - \mathbf{x}_p^*) \tag{7.21}$$

where $\mathbf{J}$ is a Jacobian linking the variation of the pose $\mathbf{r}$ and the pixel location $\mathbf{x}_p$ on image. At low magnification of the SEM, if the perspective project model is considered, it can be expressed by:

$$\mathbf{J} = \begin{pmatrix} -\frac{1}{Z} & 0 & \frac{x}{Z} & xy & -(1+x^2) & y \\ 0 & -\frac{1}{Z} & \frac{y}{Z} & 1+y^2 & -xy & -x \end{pmatrix}. \tag{7.22}$$

In this case, the object pose on 6 DoFs can be then obtained by the 3D registration.

However, at high magnifications, the parallel projection model should be considered since the scale of the sample remains unchanged while the robot moves along the depth direction. In this case, the Jacobian is given by:

$$\mathbf{J} = \begin{pmatrix} -1 & 0 & 0 & -Z & y \\ 0 & -1 & Z & 0 & -x \end{pmatrix}. \tag{7.23}$$

As mentioned previously, in the parallel projection model, the depth motion is unobservable from the variation of the pixel position in the image of the sample (or the scale of the sample that is projected on the image plane). In this case, the depth information can no longer be recovered from the 3D registration and only 5 DoFs are considered in the Jacobian. To track the robot motion along the depth direction, alternative methods should be employed. The following sections focus on this problem.

7.4.2 Estimating depth position from defocus model

An observed fact is that the defocus level varies when the object moves along the depth direction. This enables us to recover the depth information from the blur level σ. Taking Z_0 as this in-focus position, it should be noted that image blur level at the defocus position $Z_1 = Z_0 + \Delta Z$ and $Z_2 = Z_0 - \Delta Z$ could be identical due to a symmetric relation. To avoid the ambiguity, we consider only the case that $Z_i > Z_0$ or $Z_i < Z_0$ for all the images I_i during the tracking stage. In this case, the estimation range of the position along the depth direction is reduced.

In early studies on optical cameras, the relation between the depth position and the defocus can be modeled using the sensor parameters [Pentland, 1987, Subbarao and Surya, 1994, Ziou and Deschenes, 2001]:

$$\begin{cases} Z = \dfrac{Fv}{v - F - kf\sigma} & \text{if } Z > u \\ Z = \dfrac{Fv}{v - F + kf\sigma} & \text{if } Z < u \end{cases}, \tag{7.24}$$

where u is the distance between the lens and the focused position, v the distance between the lens and the image plan, F the focal lens of the lens system, f the f-number of the lens system, and k the proportionality coefficient between the blur circle radius and σ. In this model, f, F, v and k are camera intrinsic parameters that are independent of the pixel locations. This model can be simplified as

$$\sigma = mZ^{-1} + c, \tag{7.25}$$

where m, c are constants depending on these intrinsic parameters. In this case, by determining the optical sensor parameters, the depth information can be then recovered with the estimated σ.

Since SEM has different image formation process, it is difficult to use directly these intrinsic parameters from optical image formation models. Instead of applying equation (7.24), we propose to train the data before the on-line tracking. It can be simply performed by varying the sample depth position in a given range and recording the images and corresponding depth position before the tracking task. We denote the ith image in the training stage by ${}^tI_i$. Taking the in-focus image as a reference, for each given image ${}^tI_i$, the defocus blur level ${}^t\sigma_i$ is obtained from equation (7.18) and is recorded into the training data along with the corresponding depth position ${}^tZ_i$.

In the on-line tracking stage, the estimated defocus blur level $\hat{\sigma}_i$ for an image I_i can be computed by comparing it with the in-focus reference image I^*. Assuming that the SEM configurations and the image conditions (brightness, noise level, etc...) remain unchanged during the tracking stage, the same reference template could be employed in both the training stage and the tracking stage. The corresponding depth Z_i can be recovered by looking up a closed ${}^t\sigma_j$ value in the training data. This can be simply attained by a look-up-table method or an interpolation based method. A framework illustrating the whole visual tracking and pose estimation is shown in Figure 7.9.

Considering this case, we experimentally find that the relation between Z and σ can be fitted by a rational function and an error ε:

$$\sigma(Z) = \frac{1}{q_0 + q_1 Z + q_2 Z^2} + \varepsilon. \tag{7.26}$$

It should be noticed that, when q_2 approximates zero, equation (7.26) could be very similar to equation (7.25) which is derived from an optical image formation model.

In case the sharpness of reference image I^* used in the tracking stage is different from the reference image ${}^tI^*$ in the training stage, the relative blur level in training

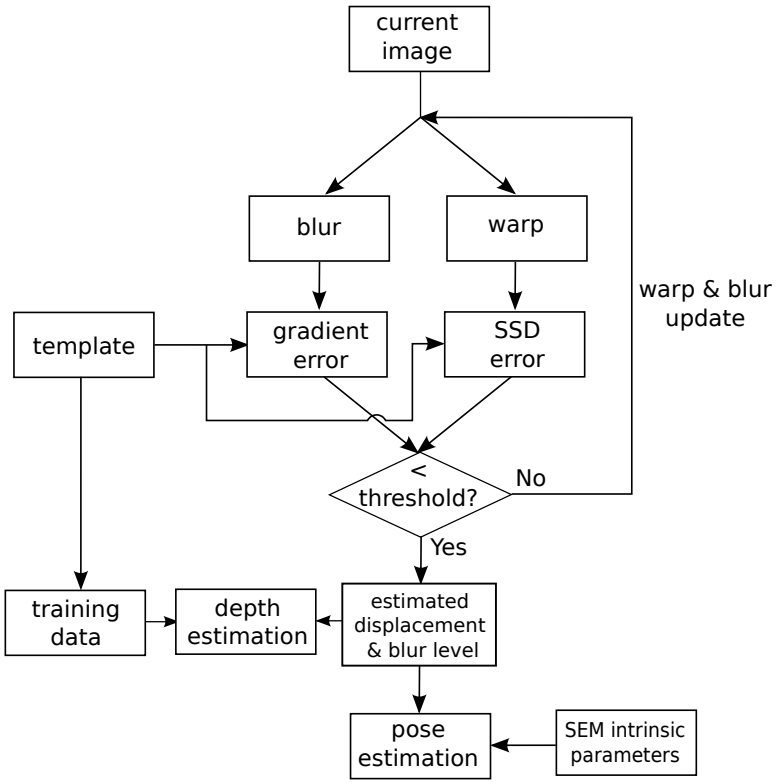


Figure 7.9: Framework of visual tracking and pose estimation using defocus information

stage ${}^t\sigma$ and in the tracking stage $\hat{\sigma}$ can be different and the depth information can no longer be determined accurately through σ. By assuming that the SEM environment is stable, the position on the depth direction can be recovered from the sharpness of the image (i.e. the image gradient G in our case). By experimentally testing more than 20 SEM image sequences, we consider the quadric rational function to approximate the relation between the image gradient and the position on the depth direction:

$$G(Z) = \frac{p_0 + p_1 Z + p_2 Z^2}{q_0 + q_1 Z + Z^2} + \varepsilon, \quad p_2 \neq 0. \tag{7.27}$$

The coefficients $\mathbf{p} = (p_2, p_1, p_0, q_1, q_0)$ can be obtained by fitting the training data of the depth Z_i with the corresponding gradient G_i using a linear method. Equation (7.27) can be rewritten as a linear system:

$$\underbrace{\begin{pmatrix} G_1 Z_1^2 \\ G_2 Z_2^2 \\ \vdots \\ G_n Z_n^2 \end{pmatrix}}_{\mathbf{b}} = \underbrace{\begin{pmatrix} Z_1^2 & Z_1 & 1 & -G_1 Z_1 & -G_1 \\ Z_2^2 & Z_2 & 1 & -G_2 Z_2 & -G_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ Z_n^2 & Z_n & 1 & -G_n Z_n & -G_n \end{pmatrix}}_{\mathbf{A}} \underbrace{\begin{pmatrix} p_2 \\ p_1 \\ p_0 \\ q_1 \\ q_0 \end{pmatrix}}_{\mathbf{p}} + \underbrace{\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}}_{\mathbf{e}}. \tag{7.28}$$

Applying the ordinary least square method which minimizes the sum of squared residuals:

$$\hat{\mathbf{p}} = \underset{\mathbf{p}}{\operatorname{argmin}} \ \|\mathbf{b} - \mathbf{A}\mathbf{p}\| \tag{7.29}$$

the coefficients $\mathbf{p}$ can be estimated by

$$\hat{\mathbf{p}} = (\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top \mathbf{b}. \tag{7.30}$$

In the visual tracking stage, an approximation of depth $\hat{Z}$ can then be computed from (7.27) by observing the image gradient G and using non-linear optimization method or look-up table method.

7.4.3 Estimating depth position using particle filter

In practice, the depth position estimation method described above could be less reliable than the estimation on other degrees of freedom using 3D registration due to some errors. These errors are categorized by two different terms. One error term can be considered as system noise (e.g., ε in equation (7.27)), which describes the inaccuracy of the model. Another error term comes from the observation, e.g., the resulting noise (caused in the SEM image formation process) on the SEM image and the uncontrollable variation of brightness and contrast of the SEM image during the tracking process. All these noises could lead to inaccurate image sharpness estimations which play an important role in the estimations of the depth position.

Alternative techniques should be employed to perform robustly and accurate pose estimating tasks. This problem involves the estimation of the state of a given system, from measurements of the input and output of the system. One of the popular methods is to apply the Kalman filter [Kalman, 1960]. Known as linear quadratic estimation, Kalman filter provides an efficient computational (recursive) means to estimate the state of a system, in a way that minimizes the mean of the squared error. It provides an effective solution to a linear dynamic system when the noise has a Gaussian distribution. For the systems that are non-linear and non-Gaussian, the particle filter [Gordon et al., 1993, Carpenter et al., 1999] could also be considered to solve the estimation problem.

Particle filters are Bayesian-based methods for performing inference in state-space models for a dynamic system via noisy measurements (observations). They comprise a broad family of sequential Monte Carlo algorithms for approximate inference in partially observable Markov chains. The general idea of particle filter techniques is to represent the required posterior density function by a set of random samples (particles) with associated weights and to estimate the internal state in dynamic systems based on these samples and weights [Arulampalam et al., 2002]. The general model is illustrated in Figure 7.10.

A particle filter is based on a system dynamics model that describes the time-dependent evolution of the state:

$$\mathbf{S}_k = \mathbf{F}(\mathbf{S}_{k-1}, \boldsymbol{\nu}_{k-1}) \tag{7.31}$$

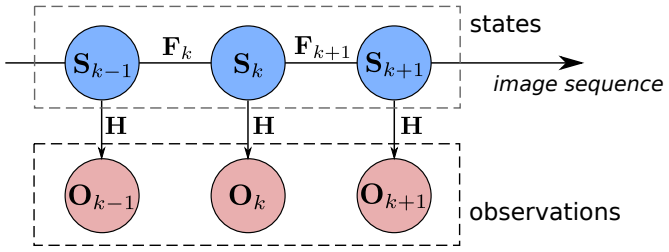


Figure 7.10: Particle filter: estimating the state via the observations

where $\mathbf{S}_k$ is the state vector at kth frame in the tracking, $\mathbf{F}$ is a possibly nonlinear function of the state $\mathbf{S}_{k-1}$. $\boldsymbol{\nu}$ is an independent and identically distributed (i.i.d.) system noise sequence. Equation (7.31) represents the evolution of a state vector $\mathbf{S}$ from frame $k-1$ to frame k. In our tracking framework, we denote the state vector by $\mathbf{S}_k = (Z_k, \dot{Z}_k)^\top$. Using Langevin motion model [Langevin, 1908], equation (7.31) can be rewritten as:

$$\begin{pmatrix} Z_k \\ \dot{Z}_k \end{pmatrix} = \begin{pmatrix} 1 & \Delta t \\ 0 & \alpha \end{pmatrix} \begin{pmatrix} Z_{k-1} \\ \dot{Z}_{k-1} \end{pmatrix} + \begin{pmatrix} 0 \\ \beta \end{pmatrix} \nu_{k-1} \tag{7.32}$$

where $\dot{Z}_k$ is the velocity along the depth direction, Δt is the time interval between k and $k-1$, α and β are system parameters and $\nu \in \mathcal{N}(0, \sigma_\nu)$ is the stochastic velocity disturbance.

The objective of a tracking task is to recursively estimate the state $\mathbf{S}_k$ from the observation $\mathbf{O}_k$ defined by:

$$\mathbf{O}_k = \mathbf{H}(\mathbf{S}_k, \boldsymbol{\varepsilon}_k). \tag{7.33}$$

where $\mathbf{O}_k$ represents an observation vector at frame k. $\mathbf{H}$ is a possibly nonlinear function and vector $\boldsymbol{\varepsilon}$ is an i.i.d. observation noise sequence. In our case, the observation is the image gradient, so $\mathbf{O}_k = G_k$. Equation (7.33) can be approximated using equation (7.27) in our tracking framework. The distribution and the variance of the noise ε can be estimated during the training stage.

The posterior predictive distribution of the state $\mathbf{S}_k$ conditional on the observations $\mathbf{O}_{1:k-1} = \{\mathbf{O}_1, \mathbf{O}_2, \ldots, \mathbf{O}_{k-1}\}$ up to frame $k-1$ can be computed recursively:

$$p(\mathbf{S}_k|\mathbf{O}_{1:k-1}) = \int p(\mathbf{S}_k|\mathbf{S}_{k-1}) p(\mathbf{S}_{k-1}|\mathbf{O}_{1:k-1}) \mathrm{d}\mathbf{S}_{k-1} \tag{7.34}$$

According to Bayes' theory, at kth frame the posterior can be updated with the observation $\mathbf{O}_k$:

$$p(\mathbf{S}_k|\mathbf{O}_{1:k}) = \frac{p(\mathbf{O}_k|\mathbf{S}_k) p(\mathbf{S}_k|\mathbf{O}_{1:k-1})}{p(\mathbf{O}_k|\mathbf{O}_{1:k-1})} \tag{7.35}$$

where the normalization constant $p(\mathbf{O}_k|\mathbf{O}_{1:k-1})$ depends on the observation likelihood $p(\mathbf{O}_k|\mathbf{S}_k)$ defined by the observation model (7.33). Applying sequential importance sampling, the posterior density $p(\mathbf{S}_k|\mathbf{O}_{1:t})$ is then approximated using a set of weighted

particles (random samples) $\{\mathbf{S}_k^i, \omega_k^i\}$ where ω_k^i represents the weight of $\mathbf{S}_k^i$:

$$p(\mathbf{S}_k|\mathbf{O}_{1:k}) \approx \sum_{i=1}^{N_p} \omega_k^i \delta(\mathbf{S}_k - \mathbf{S}_k^i) \tag{7.36}$$

where N_p is number of particles. Usually, the weighted particles can be updated using [Arulampalam et al., 2002]:

$$\omega_k^i \propto \omega_{k-1}^i p(\mathbf{O}_k|\mathbf{S}_k^i) \tag{7.37}$$

In our tracking framework, we model the observation likelihood $p(\mathbf{O}_k|\mathbf{S}_k)$ using a registration error $\epsilon_k = \| \mathbf{O}_k - \mathbf{H}(\hat{\mathbf{S}}_k) \|^2$:

$$p(\mathbf{O}_k|\mathbf{S}_k) \propto e^{-\tau\epsilon_k} \tag{7.38}$$

where $\tau \in \mathbb{R}^+$ is a constant.

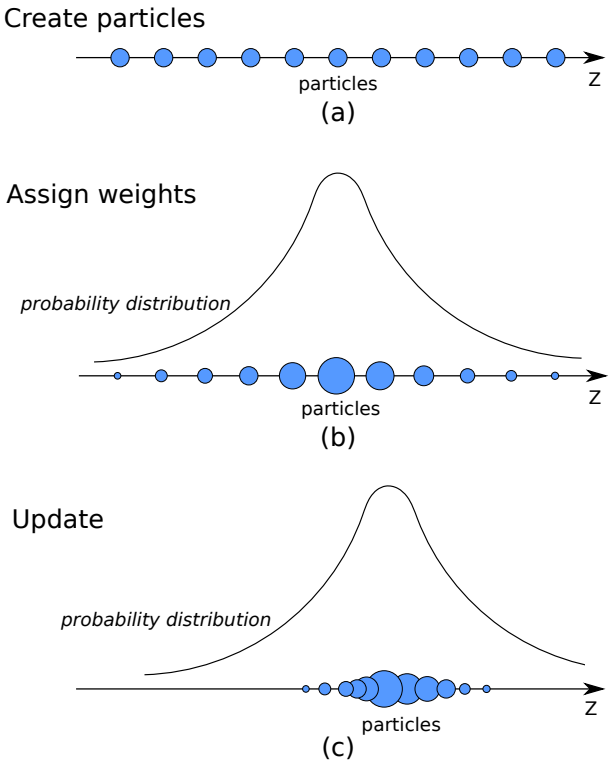


Figure 7.11: Particle filter: (a) create particles (b) assign weight to each particle, weights are computed from the possibility distribution (c) update particles according to system dynamics model

In our tracking and position estimation framework, a range of particles (with depth position and velocity along the depth direction) are generated randomly (in a given range) and assigned the same weight at first. For each frame in the tracking stage, the image gradient of the current image is computed and the particles are updated

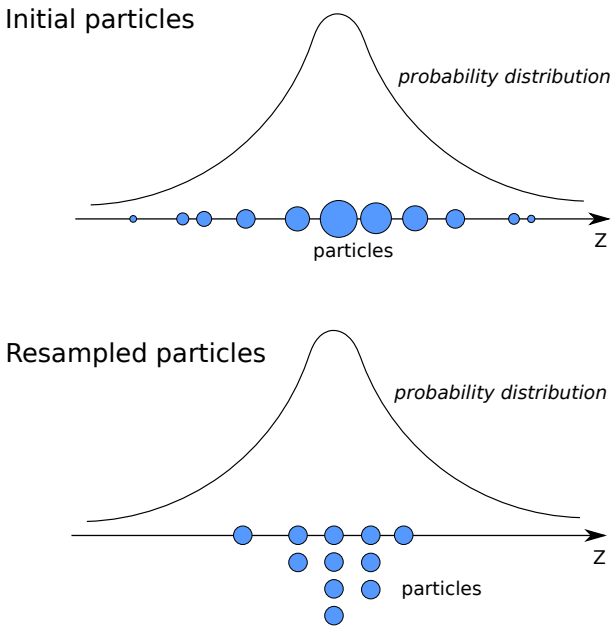


Figure 7.12: Resampling in particle filter

using the system dynamics model (equation (7.32)). The weight of each particle is then recomputed according to equation (7.37). The estimation of the state is then computed through equation (7.36). The overall algorithm for estimating the position on the depth direction using the particle filter is described in Algorithm 1. An illustration of the process of a particle filter is shown in Figure 7.11.

It should be noticed that as the general case, resampling is necessary to avoid the degeneracy. The resampling process is realized by replacing the weakly weighted particles by a number of important weighted particles according to the probability of all the possible weights. The algorithm of resampling is shown in Algorithm 2.

We stated above estimating the position on the depth direction by a particle filter. Actually, this particle filter can also be extended to estimate the pose on all the DoFs. In this case, the state vector should be modified to describe the 3D pose (e.g. $\mathbf{S} = (X, Y, Z, \theta_x, \theta_y, \theta_z)^\top$) and the other observations (e.g. estimated warp parameters) should be added into the observation vector. More particles are potentially needed when more DoFs are estimated. However, in our experiments, the estimation using 3D registration shows very good performance. In this case, we use particle filter only for depth position estimation.

7.5 Experimental results on pose estimation

Experiments have been performed to evaluate the estimation of position and orientation of the object. These experiments are performed using the same experimental setup as that in Section 7.3. The images are acquired with a medium scan speed at 1000×. An

Algorithm 1 Particle filter for estimating depth position

for $k := 1$ to N_{frame} **do**
  Measure image gradient: $\mathbf{O}_k = G_k$
  for $i := 1$ to $N_{particle}$ **do**
    Evolve sample using $\mathbf{S}_k^i \sim p(\mathbf{S}_k|\mathbf{S}_{k-1}^i, \mathbf{O}_k)$
    Assign the particle $\mathbf{S}_k^i$ a weight ω^i using equation (7.37)
  end for
  Compute the sum of the weights: $\omega^{sum} \leftarrow \sum_{i=1}^{N_p} \omega^i$
  for $i := 1$ to $N_{particle}$ **do**
    Normalize the weights of the samples: $\omega^i \leftarrow \dfrac{\omega^i}{\omega^{sum}}$
  end for
  if number of effecitve particles $<$ threshold **then**
    Resample using Algorithm 2
  end if
  Estimate the state: $\widehat{\mathbf{S}_k} = \sum_{i=1}^{N_p} \omega^i \mathbf{S}_k^i$
end for

Algorithm 2 Resampling algorithm

Create array: $\{l_1, l_2, \ldots, l_{N_p}\}$
$l_1 \leftarrow 0$
for $i := 1$ to N_p **do**
  Assign array value: $l_i \leftarrow l_{i-1} + \omega^i$
end for
for $i := 1$ to N_p **do**
  Generate an uniform random value between 0 and 1: $r \sim U(0, 1)$
  while $r < l_j$ **do**
    $j \leftarrow j + 1$
  end while
  Assign new weight: $\omega_{new}^i \leftarrow \omega^j$
  Assign new particle: $\mathbf{S}^i \leftarrow \mathbf{S}^j$
end for
Compute the sum of the weights: $\omega_{new}^{sum} \leftarrow \sum_{i=1}^{N_p} \omega_{new}^i$
for $i := 1$ to N_p **do**
  Normalize the weights of the samples: $\omega_{new}^i \leftarrow \dfrac{\omega_{new}^i}{\omega_{new}^{sum}}$
end for

image sequence is acquired in the same condition of the SEM by varying the position on the depth direction to provide the training data. In this experiment, the sample moves on 4 DoFs as previous experiments. It should be noted that the velocities can be easy to be modeled in a particle filter or a Kalman filter if they are constant. In this case,

the tracking can be conducted faultlessly if the system dynamics model is well defined. To evaluate the performance of the pose estimation methods in complex conditions, in our experiment, the velocities (along all the DoFs) as well as the accelerations are variable.

7.5.1 Estimation from 3D registration

To compute the pose of the object from 3D registration, the calibration results of the SEM (see Table 2.5 in Chapter 2) is used to provide the SEM intrinsic parameters. Figure 7.13 shows the evolution of the position on x- and y-axes and rotation around z-axis estimated by 3D registration. Here we use the pose at the first image in the tracking process as a reference to visualize the displacement of these pose measurements. Small oscillations are found in the estimation of the rotation around z-axis (yellow curve in the figure). Actually, since the increment of this rotation is about 0.02° in each, corresponding less than 0.1 pixel displacement in the image, it is very difficult to determine this tiny displacement from a blurred image.

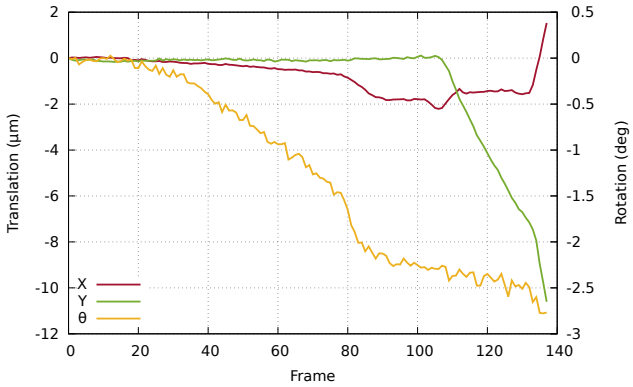
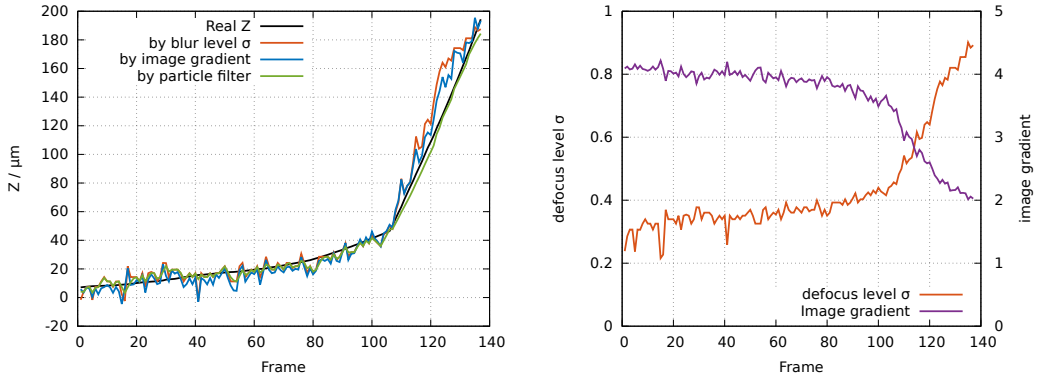


Figure 7.13: Estimated position on x, y and orientation around z

7.5.2 Estimation of depth position

In this experiment, the position on the depth direction is estimated using three methods. The first one uses the estimated blur level σ in the tracking stage and computes Z from equation (7.26). In the second method, instead of using the estimation of σ, the image gradient is considered and Z is computed from equation (7.27). The last method observes the image gradient in each frame and uses particle filter (see Section 7.4.3).

Figure 7.14 shows the results of the estimation of the sample position on the depth direction. Actually, in the visual tracking task, the optimization process of both blur level and displacement are performed simultaneously. Considering the high noise level on the SEM image, the cost function computation for the blur level estimation could be affected by the noise and the variation of the displacement during the warp process.



(a) (b)

Figure 7.14: Evolution of estimated position on the depth direction Z : (a) estimation of Z using different methods and (b) evolution of defocus level and image gradient, with respect to frames, respectively

Since the image gradient is computed directly from the tracked zone, it is more reliable than the blur level that is estimated using optimization process. In the experiment using particle filter, the number of particles is set to 200. In the experiments, we find that this number represents a good compromise between the performance and the time consumption in our experiments. A very large number of particles does not obviously improve the performance in our experiments. It can be seen from the Figure 7.14(a) that particle filter shows robustness to the observation (image gradient) variation.

7.5.3 Discussion

Nevertheless, the estimation of the position on the depth direction can be performed only in the case that the image sharpness varies. This means that the image should be acquired out of focus. This depends on the depth of field of the SEM (see Section 1.3). Actually, depth of field decreases at high magnifications and in the case that the sample is close to the objective lens of the SEM (see equations (1.2) and (1.3)). Moreover, in comparison with the estimation of X and Y, the estimation of Z could be less accurate since the image sharpness could be less reliable than the pixel position in a noisy environment. However, the image sharpness is still the most important visual feature that can recover the position on the depth direction.

In our experiments, the proposed visual tracking scheme has been validated for 4 DoFs motion of the sample. Considering the parallel projection model in a SEM, it is difficult to achieve an accurate estimation of the rotation around x- and y-axes since the rotation variation around x- and y-axes in the image is relatively slight. Our positioning stage in the experiment provides a rotation range around x- and y-axes from -5° to 5°. In this case, assuming that we have a sample of size L, the difference between an image without a tilt and that with a tilt of 5° is $0.5L \times (1 - \cos(5°)) = 0.0019L$. It is hard

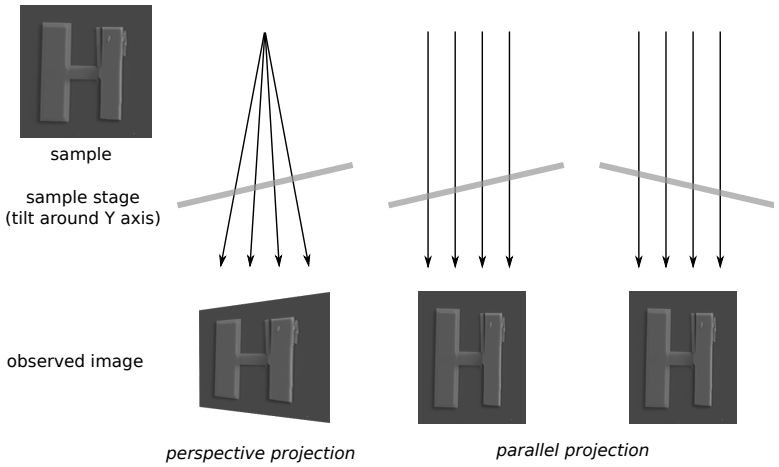


Figure 7.15: Tilt of sample and the observed images in perspective and parallel projections

to estimate accurately this small value. For this reason, observing the rotation around x- or y-axis during the visual tracking is difficult since the angular increment between each frame is always less than 1°. Moreover, for a parallel projection model, an image with a tilted sample can correspond to two different rotations (see Figure 7.15). In this case, the displacement on rotation around x- and y-axes can no longer be distinguished from the object position in the image.

Considering the available sample that can be used in our experiments, the images are acquired at 1000×. Actually, a higher magnification such as 10,000× is more appropriate since the images sharpness variation is more obvious at high magnifications. The appearance of our sample is plain. The tracking and pose estimation tasks could be more robust if a sample with complex texture is employed.

7.6 Conclusion

In this chapter, a template-based approach for tracking a micro-scale object is proposed. We consider the variation of the defocus information as an important issue to recover the object position on the depth direction. In this case, the pose of the object is estimated in three dimensions. Our method is validated by testing the image sequence in a SEM at 1000× in 4 DoFs. However, it is difficult to estimate the rotation around x- and y-axes accurately at high magnifications since the SEM projection model is parallel. The further work could be looking for a solution to realize the 6-DoF pose estimation.

Conclusion and perspectives

As a new research topic, micro/nano-techniques receive much attention over the last couple of decades. They are expected to be applied in many fields, such as electronics, aeronautics/astronautics or health care industry. However, some bottlenecks have been discovered in this fast developing field. Particularly in micro/nano-robotics, one of the challenges is to realize automated robust and reliable manipulation and assembly tasks in micro/nano-scale. Vision is a feasible way to observe the object in micro/nano-scale. It provides direct information to perform the automation of micro/nano-handling and assembly. As a necessary tool in robot motion control, visual servoing plays an important role. The objective of this thesis was then to analyze the problems on micro/nano-vision and to propose solutions using visual servoing to perform robust and reliable micro/nano-positioning tasks.

Microscopes are indispensable to observe the micro/nano-world. One of the most common microscopes for micro/nano-robotics is the SEM. It generates images by scanning the surface of the sample using an electron beam in a vacuum chamber. Since the SEM structure and the SEM image formation are quite different from optical microscopes, some particular issues in SEM vision were studied at first. Based on this study, a non-linear optimization process for SEM calibration has been addressed. Both the perspective projection and parallel projection models have been considered in this method. Image distortions have also been modeled in the proposed method. It is found in this work that one key challenge in SEM vision is that it is difficult to observe the robot motion along the depth direction through the SEM images. In a SEM vision system, instead of the perspective projection model, the parallel projection model should be adopted.

In order to solve this problem and to perform the visual servoing task along the depth direction, the image sharpness information is considered as a visual feature for visual servoing tasks. Among various sharpness functions, the image gradient is selected according to its experimental performance. In order to perform 6-DoF micro-positioning task in a SEM, a direct hybrid visual servoing framework using image appearance information has been proposed. In this method, the image gradient is considered as a visual feature to control the motion along the depth direction, while the image intensity is used to control the motion along the other 5 DoFs. This visual servoing scheme has been validated using a parallel robot under an optical microscope and

in a SEM. Since image sharpness is an important factor in autofocus, based on similar techniques, a new SEM autofocus approach has also been introduced. In this method, the SEM autofocusing task is considered as a closed-loop control problem. This method has been validated by experiments in a SEM.

In order to realize the visual guidance for micro/nano-manipulation in a SEM, a visual tracking and 3D pose estimation approach has been proposed. The variation of defocus information is considered and modeled into a template-based visual tracking scheme. The depth position is then recovered by the observed defocus information. The position on x- and y-axes as well as the orientation around z-axis are computed from the geometric transformation. Experimental results validate the proposed visual tracking scheme.

Perspectives

Due to system limitations and experimental setup maintenances, several works could still be realized or be completed in the future. First, the proposed hybrid visual servoing scheme for 6-DoF micro-positioning as well as the visual tracking approach have been validated at 1000×. Because of the limitation of the sample and of the SmarPod, the validation has not been performed at higher magnifications. Although the experimental condition could be similar at higher magnifications (the image quality could be degraded), it is always interesting to test the performance of the proposed approaches at higher magnifications in nano-scale. Additionally, since the structure of the sample that is used in the experiments is simple, the estimation of image gradient could be sensible to random noise. Testing a sample with more complex textures could improve the performance. In this case, for the further work, the experiments using different samples at multi-magnifications should be conducted.

The proposed dynamic approximation of the Jacobian in the control of the motion along the depth direction has been tested in simulation. Due to the maintenance of the experimental instruments, this method has not been validated in real-time in a SEM. For similar reasons, the proposed visual tracking method has not been validated by moving the sample in three dimensions while varying the magnification (that changes the scale of the sample). Since the magnification can also be considered as a degree of freedom, it can be tested in further work.

The proposed visual tracking and pose estimation scheme has been validated by experiments where the object moves with 4 DoFs at a high magnification. As stated previously, a great challenge in the estimation of 6 DoFs is how to estimate the tilt (rotation around x and y axes) accurately in the parallel projection model. A possible solution is to detect the image sharpness in local areas and compute a "depth map" of the observed sample. The tilt could be computed from different depth positions of interest points. However, this required a very small depth of field of the SEM. In this case, it could be only performed at a very high magnification. Experiments could be

conducted to test this solution.

It should be considered that the proposed vision-based control method could also be extended into any automation task by a visual sensor with an orthographic view, such as a camera with telecentric lens. In this case, the depth of field could be changed by the focal length of the camera. This has to be tested by further experiments.

For the long term perspective, it could be interesting to apply the visual servoing scheme in the real-time micro/nano-manipulation and assembly tasks. In this case, some particular cases, such as the occlusion should be considered. Another possible application is in biological and biomedical domain, such as cell and sub-cell manipulation. In this case, the time-dependent deformation of the cell should be considered in the vision-based control framework.

Bibliography

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

[Agarwal et al., 2005] Agarwal, R., Ladavac, K., Roichman, Y., Yu, G., Lieber, C., and Grier, D. (2005). Manipulation and assembly of nanowires with holographic optical traps. *Optics Express*, 13(22):8906–8912.

[Arulampalam et al., 2002] Arulampalam, M. S., Maskell, S., Gordon, N., and Clapp, T. (2002). A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking. *Signal Processing, IEEE Transactions on*, 50(2):174–188.

[Baina and Dublet, 1995] Baina, J. and Dublet, J. (1995). Automatic focus and iris control for video cameras. In *Image Processing and its Applications, 1995., Fifth International Conference on*, pages 232–235. IET.

[Baker and Matthews, 2004] Baker, S. and Matthews, I. (2004). Lucas-kanade 20 years on: A unifying framework. *Int. Journal of Computer Vision*, 56(3):221–255.

[Banerjee and Gupta, 2013] Banerjee, A. G. and Gupta, S. K. (2013). Research in automated planning and control for micromanipulation. *Automation Science and Engineering, IEEE Transactions on*, 10(3):485–495.

[Bateux and Marchand, 2015] Bateux, Q. and Marchand, E. (2015). Direct visual servoing based on multiple intensity histograms. In *IEEE Int. Conf. on Robotics and Automation, ICRA'15*, pages 6019–6024, Seattle, WA.

[Batten, 2000] Batten, C. F. (2000). Autofocusing and astigmatism correction in the scanning electron microscope. Master's thesis, Citeseer.

[Bay et al., 2006] Bay, H., Tuytelaars, T., and Van Gool, L. (2006). Surf: Speeded up robust features. In *Computer vision–ECCV 2006*, pages 404–417. Springer.

[Ben-Ari, 2014] Ben-Ari, R. (2014). A unified approach for registration and depth in depth from defocus. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 36(6):1041–1055.

[Benhimane and Malis, 2006] Benhimane, S. and Malis, E. (2006). Homography-based 2d visual servoing. In *IEEE Int. Conf. on Robotics and Automation, ICRA'06*, Orlando, Fl.

[Berger, 1994] Berger, M.-O. (1994). How to track efficiently piecewise curved contours with a view to reconstructing 3D objects. In *Int. Conf on Pattern Recognition, ICPR'94*, pages 32–36, Jerusalem.

[Bhasin and Chaudhuri, 2001] Bhasin, S. and Chaudhuri, S. (2001). Depth from defocus in presence of partial self occlusion. In *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*, volume 1, pages 488–493 vol.1.

[Binnig and Rohrer, 1983] Binnig, G. and Rohrer, H. (1983). Scanning tunneling microscopy. *Surface science*, 126(1):236–244.

[Blake and Isard, 1998] Blake, A. and Isard, M. (1998). *Active Contours*. Springer Verlag.

[Boukir et al., 1998] Boukir, S., Bouthemy, P., Chaumette, F., and Juvin, D. (1998). A local method for contour matching and its parallel implementation. *Machine Vision and Application*, 10(5/6):321–330.

[Boyde, 1970] Boyde, A. (1970). Practical problems and methods in the three-dimensional analysis of scanning electron microscope images. *Scanning electron microscopy*, 1970:105–112.

[Boyde, 1973] Boyde, A. (1973). Quantitative photogrammetric analysis and qualitative stereoscopic analysis of sem images. *Journal of Microscopy*, 98(3):452–471.

[Brenner et al., 1976] Brenner, J. F., Dew, B. S., Horton, J. B., King, T., Neurath, P. W., and Selles, W. D. (1976). An automated microscope for cytologic research a preliminary evaluation. *Journal of Histochemistry & Cytochemistry*, 24(1):100–111.

[Brisset et al., 2012] Brisset, F. et al. (2012). *Microscopie électronique à balayage et microanalyses*. EDP sciences.

[Brown, 1971] Brown, D. (1971). Close-range camera calibration. *Photogrammetric Engineering*, 4(2):127–140.

[Brown, 1976] Brown, D. C. (1976). The bundle adjustment-progress and prospects. *Int. Archives Photogrammetry*, 21(3):1–1.

[Buades et al., 2005] Buades, A., Coll, B., and Morel, J.-M. (2005). A non-local algorithm for image denoising. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 2, pages 60–65. IEEE.

[Carpenter et al., 1999] Carpenter, J., Clifford, P., and Fearnhead, P. (1999). Improved particle filter for nonlinear problems. *IEE Proceedings-Radar, Sonar and Navigation*, 146(1):2–7.

[Chaumette and Hutchinson, 2006] Chaumette, F. and Hutchinson, S. (2006). Visual servo control, Part I: Basic approaches. *IEEE Robotics and Automation Magazine*, 13(4):82–90.

[Chaumette and Hutchinson, 2007] Chaumette, F. and Hutchinson, S. (2007). Visual servo control, Part II: Advanced approaches. *IEEE Robotics and Automation Magazine*, 14(1):109–118.

[Chaumette and Rives, 1990] Chaumette, F. and Rives, P. (1990). Vision-based-control for robotic tasks. In *IEEE Int. Workshop on Intelligent Motion Control*, pages 395–400, Istanbul, Turquie.

[Chen et al., 2014] Chen, Z., Liao, H., and Zhang, X. (2014). Telecentric stereo micro-vision system: Calibration method and experiments. *Optics and Lasers in Engineering*, 57:82–92.

[Cizmar et al., 2008] Cizmar, P., Vladár, A. E., Ming, B., and Postek, M. T. (2008). Artificial sem images for testing resolution-measurement methods. *Microscopy and Microanalysis*, 14(S2):910–911.

[Cizmar et al., 2011] Cizmar, P., Vladár, A. E., and Postek, M. T. (2011). Real-time scanning charged-particle microscope image composition with correction of drift. *Microsc. Microanal.*, 17(4):302–308.

[Cohn et al., 1998] Cohn, M. B., Boehringer, K. F., Noworolski, J. M., Singh, A., Keller, C. G., Goldberg, K. A., and Howe, R. T. (1998). Microassembly technologies for mems. In *Micromachining and Microfabrication*, pages 2–16. International Society for Optics and Photonics.

[Collewet and Marchand, 2011] Collewet, C. and Marchand, E. (2011). Photometric visual servoing. *IEEE Trans. on Robotics*, 27(4):828–834.

[Collewet et al., 2008] Collewet, C., Marchand, E., and Chaumette, F. (2008). Visual servoing set free from image processing. In *IEEE Int. Conf. on Robotics and Automation, ICRA'08*, pages 81–86, Pasadena, CA.

[Comport et al., 2006] Comport, A., Marchand, E., Pressigout, M., and Chaumette, F. (2006). Real-time markerless tracking for augmented reality: the virtual visual servoing framework. *IEEE Trans. on Visualization and Computer Graphics*, 12(4):615–628.

[Cootes et al., 2001] Cootes, T. F., Edwards, G. J., and Taylor, C. J. (2001). Active appearance models. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (6):681–685.

[Corke and Good, 1992] Corke, P. and Good, M. (1992). Dynamics effects in high performance visual servoing. In *IEEE Int. Conf. on Robotics and Automation*, volume 2, pages 1838–1843, Nice, France.

[Corke et al., 1996] Corke, P. I. et al. (1996). *Visual Control of Robots: high-performance visual servoing*. Research Studies Press Baldock.

[Cornille, 2005] Cornille, N. (2005). *Accurate 3D shape and displacement measurement using a scanning electron microscope*. PhD thesis, École des Mines d'Albi, France.

[Cornille et al., 2003] Cornille, N., Garcia, D., Sutton, M. A., McNeill, S., and Orteu, J.-J. (2003). Automated 3-d reconstruction using a scanning electron microscope. In *SEM annual conf. & exp. on experimental and applied mechanics*.

[Cui and Marchand, 2015] Cui, L. and Marchand, E. (2015). Scanning electron microscope calibration using a multi-image non-linear minimization process. *Int. Journal of Optomechatronics*, 9(2):151–169.

[Cui et al., 2014] Cui, L., Marchand, E., Haliyo, S., and Régnier, S. (2014). 6-dof automatic micropositioning using photometric information. In *IEEE/ASME Int Conf. on Advanced Intelligent Mechatronics, AIM'14*.

[Dahmen, 2008] Dahmen, C. (2008). Focus-based depth estimation in the sem. In *International Symposium on Optomechatronic Technologies*, pages 72661O–72661O. International Society for Optics and Photonics.

[Dahmen, 2011] Dahmen, C. (2011). Defocus-based threedimensional tracking in sem images. In *Informatics in Control Automation and Robotics*, pages 243–254. Springer.

[Dame and Marchand, 2009] Dame, A. and Marchand, E. (2009). Entropy-based visual servoing. In *IEEE Int. Conf. on Robotics and Automation, ICRA'09*, pages 707–713, Kobe, Japan.

[Dame and Marchand, 2010] Dame, A. and Marchand, E. (2010). Accurate real-time tracking using mutual information. In *IEEE Int. Symp. on Mixed and Augmented Reality, ISMAR'10*, Seoul, Korea.

[Dame and Marchand, 2011] Dame, A. and Marchand, E. (2011). Mutual information-based visual servoing. *IEEE Trans. on Robotics*, 27(5):958–969.

[Deriche and Faugeras, 1990] Deriche, R. and Faugeras, O. (1990). Tracking lines segments. *Image and Vision Computing*, 8(4):261–270.

[Devasia et al., 2007] Devasia, S., Eleftheriou, E., and Moheimani, S. (2007). A survey of control issues in nanopositioning. *Control Systems Technology, IEEE Transactions on*, 15(5):802–823.

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

[Egerton, 2005] Egerton, R. (2005). *Physical Principles of Electron Microscopy: An Introduction to TEM, SEM, and AEM*. Springer US.

[Eichhorn et al., 2008] Eichhorn, V., Fatikow, S., Wich, T., Dahmen, C., Sievers, T., Andersen, K. N., Carlson, K., and Bøggild, P. (2008). Depth-detection methods for microgripper based cnt manipulation in a scanning electron microscope. *Journal of Micro-Nano Mechatronics*, 4(1-2):27–36.

[Eichhorn et al., 2009] Eichhorn, V., Fatikow, S., Wortmann, T., Stolle, C., Edeler, C., Jasper, D., Sardan, O., Boggild, P., Boetsch, G., Canales, C., et al. (2009). Nanolab: A nanorobotic system for automated pick-and-place handling and characterization of cnts. In *Robotics and Automation, 2009. ICRA'09. IEEE International Conference on*, pages 1826–1831. IEEE.

[El Ghazali, 1984] El Ghazali, M. (1984). System calibration of scanning electron microscopes. *International Archives of Photogrammetry and Remote Sensing*, 25:258–266.

[Ens and Lawrence, 1993] Ens, J. and Lawrence, P. (1993). An investigation of methods for determining depth from focus. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 15(2):97–108.

[Erasmus and Smith, 1982] Erasmus, S. J. and Smith, K. C. A. (1982). An automatic focusing and astigmatism correction system for the sem and ctem. *Journal of Microscopy*, 127(2):185–199.

[Ersoy, 2010] Ersoy, O. (2010). Surface area and volume measurements of volcanic ash particles by {SEM} stereoscopic imaging. *Journal of Volcanology and Geothermal Research*, 190(3-4):290 – 296.

[Espiau, 1993] Espiau, B. (1993). Effect of camera calibration errors on visual servoing in robotics. In *Int. Symposium on experimental Robotics, ISER'93*, Kyoto.

[Espiau et al., 1992] Espiau, B., Chaumette, F., and Rives, P. (1992). A new approach to visual servoing in robotics. *IEEE Trans. on Robotics and Automation*, 8(3):313–326.

[Fan et al., 2014] Fan, S., Yu, M., Wang, Y., and Jiang, G. (2014). A depth estimation method based on geometric transformation for stereo light microscope. *Bio-medical materials and engineering*, 24(6):2743–2749.

[Fatikow and Eichhorn, 2008] Fatikow, S. and Eichhorn, V. (2008). Nanohandling automation: trends and current developments. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, 222(7):1353–1369.

[Fatikow et al., 2008] Fatikow, S., Eichhorn, V., Stolle, C., Sievers, T., and Jähnisch, M. (2008). Development and control of a versatile nanohandling robot cell. *Mechatronics*, 18(7):370–380.

[Fatikow and Rembold, 1996] Fatikow, S. and Rembold, U. (1996). An automated microrobot-based desktop station for micro assembly and handling of micro-objects. In *Emerging Technologies and Factory Automation, 1996. EFTA'96. Proceedings., 1996 IEEE Conference on*, volume 2, pages 586–592. IEEE.

[Fatikow et al., 2007] Fatikow, S., Wich, T., Hülsen, H., Sievers, T., and Jähnisch, M. (2007). Microrobot system for automatic nanohandling inside a scanning electron microscope. *Mechatronics, IEEE/ASME Transactions on*, 12(3):244–252.

[Faugeras and Toscani, 1987] Faugeras, O. and Toscani, G. (1987). Camera calibration for 3D computer vision. In *Proc Int. Workshop on Machine Vision and Machine Intelligence*, pages 240–247, Tokyo.

[Favaro et al., 2003] Favaro, P., Mennucci, A., and Soatto, S. (2003). Observing shape from defocused images. *Internat. J. Comput. Vision,*, 52(1):25–43.

[Favaro and Soatto, 2005] Favaro, P. and Soatto, S. (2005). A geometric approach to shape from defocus. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 27(3):406–417.

[Favaro et al., 2008] Favaro, P., Soatto, S., Burger, M., and Osher, S. J. (2008). Shape from defocus via diffusion. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 30(3):518–531.

[Feddema and Mitchell, 1989] Feddema, J. and Mitchell, O. (1989). Vision-guided servoing with feature-based trajectory generation. *IEEE Trans. on Robotics and Automation*, 5(5):691–700.

[Feddema and Simon, 1998] Feddema, J. and Simon, R. (1998). Visual servoing and cad-driven microassembly. *Robotics & Automation Magazine, IEEE*, 5(4):18–24.

[Ferreira et al., 2004] Ferreira, A., Cassier, C., and Hirai, S. (2004). Automatic microassembly system assisted by vision servoing and virtual reality. *Mechatronics, IEEE/ASME Transactions on*, 9(2):321–333.

[Firestone et al., 1991] Firestone, L., Cook, K., Culp, K., Talsania, N., and Preston, K. (1991). Comparison of autofocus methods for automated microscopy. *Cytometry*, 12(3):195–206.

[Freundlich, 1963] Freundlich, M. M. (1963). Origin of the electron microscope. *Science*, 142(3589):185–188.

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

[Gavrilenko et al., 2015] Gavrilenko, V., Karabanov, D., Kuzin, A., Mityukhlyaev, V., Mikhutkin, A., Todua, P., Filippov, M., Baimukhametov, T., and Vasilév, A. (2015). Three-dimensional reconstruction of the surfaces of relief structures from stereoscopic images obtained in a scanning electron microscope. *Measurement Techniques*, 58(3):256–260.

[Ghosh, 1975] Ghosh, S. K. (1975). Photogrammetric calibration of a scanning electron microscope. *Photogrammetria*, 31(3):91 – 114.

[Godec et al., 2013] Godec, M., Roth, P. M., and Bischof, H. (2013). Hough-based tracking of non-rigid objects. *Computer Vision and Image Understanding*, 117(10):1245–1256.

[Gökstorp, 1994] Gökstorp, M. (1994). Computing depth from out-of-focus blur using a local frequency representation. In *Pattern Recognition, 1994. Vol. 1-Conference A: Computer Vision & Image Processing., Proceedings of the 12th IAPR International Conference on*, volume 1, pages 153–158. IEEE.

[Goldstein et al., 2003] Goldstein, J., Newbury, D., Joy, D., Lyman, C., Echlin, P., Lifshin, E., Sawyer, L., and Michael, J. (2003). *Scanning Electron Microscopy and X-ray Microanalysis: Third Edition*. Springer US.

[Gong et al., 2014] Gong, Z., Chen, B. K., Liu, J., and Sun, Y. (2014). Robotic probing of nanostructures inside scanning electron microscopy. *Robotics, IEEE Transactions on*, 30(3):758–765.

[Gonzalez and Woods, 2008] Gonzalez, R. and Woods, R. (2008). *Digital Image Processing*. Pearson/Prentice Hall.

[Gordon et al., 1993] Gordon, N. J., Salmond, D. J., and Smith, A. F. (1993). Novel approach to nonlinear/non-gaussian bayesian state estimation. In *IEE Proceedings F (Radar and Signal Processing)*, volume 140, pages 107–113. IET.

[Greminger et al., 2004] Greminger, M., Nelson, B. J., et al. (2004). Vision-based force measurement. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 26(3):290–298.

[Groen et al., 1985] Groen, F. C., Young, I. T., and Ligthart, G. (1985). A comparison of different focus functions for use in autofocus algorithms. *Cytometry*, 6(2):81–91.

[Grossmann, 1987] Grossmann, P. (1987). Depth from focus. *Pattern Recognition Letters*, 5(1):63–69.

[Hafner, 2007] Hafner, B. (2007). Scanning electron microscopy primer. *Characterization Facility, University of Minnesota-Twin Cities*, pages 1–29.

[Hager and Belhumeur, 1998] Hager, G. and Belhumeur, P. (1998). Efficient region tracking with parametric models of geometry and illumination. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 20(10):1025–1039.

[Harris and Stephens, 1988] Harris, C. and Stephens, M. (1988). A combined corner and edge detector. In *Alvey Conference*, pages 147–151, Manchester.

[Heikkila and Silven, 1997] Heikkila, J. and Silven, O. (1997). A four-step camera calibration procedure with implicit image correction. In *Proceedings of the IEEE Computer Society Conference Computer Vision and Pattern Recognition*, pages 1106–1112.

[Hemmleb and Albertz, 2000] Hemmleb, M. and Albertz, J. (2000). Microtopography-the photogrammetric determination of friction surfaces. *International Archives of Photogrammetry and Remote*, 33:56–63.

[Horn and Schunck, 1981] Horn, B. and Schunck, B. (1981). Determining optical flow. *Artificial Intelligence*, 17(1-3):185–203.

[Howell, 1978] Howell, P. G. T. (1978). A theoretical approach to the errors in sem photogrammetry. *Scanning*, 1(2):118–124.

[Hutchinson et al., 1996] Hutchinson, S., Hager, G., and Corke, P. (1996). A tutorial on visual servo control. *IEEE Trans. on Robotics and Automation*, 12(5):651–670.

[IBM Research, 2009] IBM Research (2009). A boy and his atom: The world's smallest movie. `http://www.research.ibm.com/articles/madewithatoms.shtml`. [Online; accessed 31-August-2015].

[Irani et al., 1992] Irani, M., Rousso, B., and Peleg, S. (1992). Detecting and tracking multiple moving objects using temporal integration. In *ECCV'92*, pages 282–287.

[Jähnisch and Fatikow, 2007] Jähnisch, M. and Fatikow, S. (2007). 3-D Vision Feedback for Nanohandling Monitoring in a Scanning Electron Microscope. *Int. J. Optomechatronics*, 1(February 2015):4–26.

[Jasper and Fatikow, 2010] Jasper, D. and Fatikow, S. (2010). Line scan-based high-speed position tracking inside the sem. *International Journal of Optomechatronics*, 4(2):115–135.

[Kalman, 1960] Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of Fluids Engineering*, 82(1):35–45.

[Klemperer and Barnett, 1971] Klemperer, O. and Barnett, M. E. (1971). *Electron optics*. Cambridge University Press, Cambridge, U.K.

[Koyano and Sato, 1996] Koyano, K. and Sato, T. (1996). Micro-object handling system with concentrated visual fields and new handling skills. In *Photonics East'96*, pages 130–140. International Society for Optics and Photonics.

[Kragic and Christensen, 2002] Kragic, D. and Christensen, H. (2002). Survey on visual servoing for manipulation. *Comput. Vis. Act. Percept. ...*.

[Kratochvil et al., 2009] Kratochvil, B. E., Dong, L., and Nelson, B. J. (2009). Real-time rigid-body visual tracking in a scanning electron microscope. *The International Journal of Robotics Research*, 28(4):498–511.

[Krotkov, 1988] Krotkov, E. (1988). Focusing. *International Journal of Computer Vision*, 1(3):223–237.

[Krupa et al., 2003] Krupa, A., Gangloff, J., Doignon, C., de Mathelin, M., Morel, G., Leroy, J., Soler, L., and Marescaux, J. (2003). Autonomous 3D positioning of surgical instruments in robotized laparoscopic surgery using visual servoing. *IEEE Trans. on robotics and automation*, 19(5):842 –853.

[Lacey et al., 1996] Lacey, A. J., Thacker, N. A., and Yates, R. B. (1996). Surface approximation from industrial sem images. In *British Machine Vision Conference*, pages 725–734.

[Lai et al., 1992] Lai, S.-H., Fu, C.-W., and Chang, S. (1992). A generalized depth estimation algorithm with a single image. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (4):405–411.

[Langevin, 1908] Langevin, P. (1908). Sur la théorie du mouvement brownien. *C. R. Acad. Sci. Paris*, 146:530–533.

[Lee and Cho, 2009] Lee, D. and Cho, H. (2009). Control point adaptation algorithm for cad-based visual tracking of micro mems parts using active stereo vision system. In *Asian Control Conference, 2009. ASCC 2009. 7th*, pages 495–500. IEEE.

[Lee et al., 2001] Lee, S. J., Kim, K., Kim, D.-H., Park, J.-O., and Park, G.-T. (2001). Recognizing and tracking of 3d-shaped micro parts using multiple visions for micromanipulation. In *Micromechatronics and Human Science, 2001. MHS 2001. Proceedings of 2001 International Symposium on*, pages 203–210.

[Lehmann, 1951] Lehmann, E. L. (1951). A general concept of unbiasedness. *Ann. Math. Statist.*, 22(4):587–592.

[Li and Tian, 2013] Li, D. and Tian, J. (2013). An accurate calibration method for a camera with telecentric lenses. *Optics and Lasers in Engineering*, 51(5):538 – 541.

[Liu et al., 2009] Liu, X., Kim, K., Zhang, Y., and Sun, Y. (2009). Nanonewton force sensing and control in microrobotic cell manipulation. *The international journal of robotics research*.

[Liu et al., 2013] Liu, X., Tanaka, M., and Okutomi, M. (2013). Single-image noise level estimation for blind denoising. *IEEE Transactions on Image Processing*, 22(12):5226–5237.

[Liu et al., 2007] Liu, X., Wang, W., and Sun, Y. (2007). Dynamic evaluation of autofocusing for automated microscopic analysis of blood smear and pap smear. *Journal of microscopy*, 227(1):15–23.

[Liu et al., 2015] Liu, Z., Wang, J., and Poh, E. K. (2015). Set-based direct visual servoing for nanopositioning. In *Control and Automation (MED), 2015 23th Mediterranean Conference on*, pages 772–776. IEEE.

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

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

[Lucas and Kanade, 1981] Lucas, B. and Kanade, T. (1981). An iterative image registration technique with an application to stereo vision. In *Int. Joint Conf. on Artificial Intelligence, IJCAI'81*, pages 674–679.

[Ma et al., 2004] Ma, Y., Soatto, S., Košecká, J., and Sastry, S. (2004). *An invitation to 3-D vision*. Springer.

[Mahmood et al., 2013] Mahmood, M. T., Shim, S.-O., Alshomrani, S., and Choi, T.-S. (2013). Depth from image focus methods for micro-manufacturing. *The International Journal of Advanced Manufacturing Technology*, 67(5-8):1701–1709.

[Malti et al., 2012a] Malti, A. C., Dembélé, S., Le Fort-Piat, N., Rougeot, P., and Salut, R. (2012a). Magnification-continuous static calibration model of a scanning-electron microscope. *Journal of Electronic Imaging*, 21(3):033020–1.

[Malti et al., 2012b] Malti, A. C., Dembélé, S., Le Piat, N., Arnoult, C., and Marturi, N. (2012b). Toward fast calibration of global drift in scanning electron microscopes with respect to time and magnification. *International Journal of Optomechatronics*, 6(1):1–16.

[Marchand et al., 2001] Marchand, E., Bouthemy, P., and Chaumette, F. (2001). A 2D-3D model-based approach to real-time visual tracking. *Image and Vision Computing, IVC*, 19(13):941–955.

[Marchand et al., 1999] Marchand, E., Bouthemy, P., Chaumette, F., and Moreau, V. (1999). Robust real-time visual tracking using a 2D-3D model-based approach. In *IEEE Int. Conf. on Computer Vision, ICCV'99*, volume 1, pages 262–268, Kerkira, Greece.

[Marchand and Chaumette, 2005] Marchand, E. and Chaumette, F. (2005). Feature tracking for visual servoing purposes. *Robotics and Autonomous Systems*, 52(1):53–70. special issue on "Advances in Robot Vision", D. Kragic, H. Christensen (Eds.).

[Marchand and Collewet, 2010] Marchand, E. and Collewet, C. (2010). Using image gradient as a visual feature for visual servoing. In *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, IROS'10*, Taipei, Taiwan.

[Marinello et al., 2008] Marinello, F., Bariani, P., Savio, E., Horsewell, A., and De Chiffre, L. (2008). Critical factors in sem 3d stereo microscopy. *Measurement Science and Technology*, 19(6):065705.

[Marroquin et al., 1987] Marroquin, J., Mitter, S., and Poggio, T. (1987). Probabilistic solution of ill-posed problems in computational vision. *Journal of the american statistical association*, 82(397):76–89.

[Marturi et al., 2013a] Marturi, N., Dembélé, S., and Piat, N. (2013a). Depth and shape estimation from focus in scanning electron microscope for micromanipulation. In *Control, Automation, Robotics and Embedded Systems (CARE), 2013 International Conference on*, pages 1–6. IEEE.

[Marturi et al., 2013b] Marturi, N., Dembélé, S., and Piat, N. (2013b). Fast image drift compensation in scanning electron microscope using image registration. In *Automation Science and Engineering (CASE), 2013 IEEE International Conference on*, pages 807–812. IEEE.

[Marturi et al., 2014a] Marturi, N., Dembélé, S., and Piat, N. (2014a). Scanning electron microscope image signal-to-noise ratio monitoring for micro-nanomanipulation. *Scanning*, 36(4):419–429.

[Marturi et al., 2013c] Marturi, N., Tamadazte, B., Dembélé, S., and Piat, N. (2013c). Visual servoing-based approach for efficient autofocusing in scanning electron microscope. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 2677–2682. IEEE.

[Marturi et al., 2014b] Marturi, N., Tamadazte, B., Dembélé, S., and Piat, N. (2014b). Visual servoing schemes for automatic nanopositioning under scanning electron microscope. In *Robotics and Automation (ICRA), 2014 IEEE International Conference on*, pages 981–986. IEEE.

[Maune, 1976] Maune, D. F. (1976). Photogrammetric self-calibration of scanning electron microscopes. *Photogrammetric Engineering and Remote Sensing*, 42(9).

[Metni and Hamel, 2007] Metni, N. and Hamel, T. (2007). A uav for bridge inspection: Visual servoing control law with orientation limits. *Automation in construction*, 17(1):3–10.

[Mills and Rose, 2010] Mills, O. P. and Rose, W. I. (2010). Shape and surface area measurements using scanning electron microscope stereo-pair images of volcanic ash particles. *Geosphere*, 6(6):805–811.

[Minnich et al., 1999] Minnich, B., Leeb, H., Bernroider, E., and Lametschwandtner, A. (1999). Three-dimensional morphometry in scanning electron microscopy: a technique for accurate dimensional and angular measurements of microstructures using stereopaired digitized images and digital image analysis. *Journal of Microscopy*, 195(1):23–33.

[Mizuno et al., 1997] Mizuno, F., Shimizu, M., Sasada, K., and Mizuno, T. (1997). Evaluation of the long-term stability of critical-dimension measurement scanning electron microscopes using a calibration standard. *Journal of Vacuum Science & Technology B*, 15(6):2177–2180.

[Morgan-Mar and Arnison, 2014] Morgan-Mar, D. and Arnison, M. R. (2014). Depth from defocus using the mean spectral ratio. In *IS&T/SPIE Electronic Imaging*, pages 90230H–90230H. International Society for Optics and Photonics.

[Mulapudi and Joy, 2003] Mulapudi, S. and Joy, D. (2003). Is sem noise gaussian. *Microscopy and Microanalysis*, 9(S02):982–983.

[Nayar and Nakagawa, 1994] Nayar, S. K. and Nakagawa, Y. (1994). Shape from focus. *Pattern analysis and machine intelligence, IEEE Transactions on*, 16(8):824–831.

[Nguyen and Smeulders, 2006] Nguyen, H. T. and Smeulders, A. W. (2006). Robust tracking using foreground-background texture discrimination. *International Journal of Computer Vision*, 69(3):277–293.

[Nicolls et al., 1997] Nicolls, F., de Jager, G., and Sewell, B. (1997). Use of a general imaging model to achieve predictive autofocus in the scanning electron microscope. *Ultramicroscopy*, 69(1):25 – 37.

[Noguchi and Nayar, 1994] Noguchi, M. and Nayar, S. (1994). Microscopic shape from focus using active illumination. *Proc. 12th Int. Conf. Pattern Recognit.*, 1:147–152.

[Ogawa et al., 2005] Ogawa, N., Oku, H., Hashimoto, K., and Ishikawa, M. (2005). Microrobotic visual control of motile cells using high-speed tracking system. *Robotics, IEEE Transactions on*, 21(4):704–712.

[Ong et al., 1997] Ong, K., Phang, J., and Thong, J. (1997). A robust focusing and astigmatism correction method for the scanning electron microscope. *Scanning*, 19(8):553–563.

[Ong et al., 1998a] Ong, K., Phang, J., and Thong, J. (1998a). A robust focusing and astigmatism correction method for the scanning electron microscope-part iii: An improved technique. *Scanning*, 20(5):357–368.

[Ong et al., 1998b] Ong, K., Phang, J., and Thong, J. (1998b). A robust focusing and astigmatism correction method for the scanning electron microscopeâĂŤpart ii: Autocorrelation-based coarse focusing method. *Scanning*, 20(4):324–334.

[Ouarti et al., 2013] Ouarti, N., Sauvet, B., Haliyo, S., and Régnier, S. (2013). Robposit, a robust pose estimator for operator controlled nanomanipulation. *Journal of Micro-Bio Robotics*, 8(2):73–82.

[Pentland, 1987] Pentland, A. P. (1987). A new sense for depth of field. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, (4):523–531.

[Popa and Stephanou, 2004] Popa, D. O. and Stephanou, H. E. (2004). Micro and mesoscale robotic assembly. *Journal of manufacturing processes*, 6(1):52–71.

[Postek et al., 1993] Postek, M., Vladar, A., Jones, S., and Keery, W. (1993). Interlaboratory study on the lithographically produced scanning electron microscope magnification standard prototype. *Journal of research of the National Institute of Standards and Technology*, 98:447–447.

[Pouchou et al., 2002] Pouchou, J.-L., Boivin, D., BeauchÃłne, P., Besnerais, G. L., and Vignon, F. (2002). 3d reconstruction of rough surfaces by sem stereo imaging. *Microchimica Acta*, 139(1-4):135–144.

[Pratt, 2013] Pratt, W. (2013). *Introduction to Digital Image Processing*. Taylor & Francis.

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

[Probst et al., 2009] Probst, M., Hürzeler, C., Borer, R., and Nelson, B. (2009). A microassembly system for the flexible assembly of hybrid robotic mems devices. *International Journal of Optomechatronics*, 3(2):69–90.

[Probst et al., 2006] Probst, M., Vollmers, K., Kratochvil, B. E., and Nelson, B. J. (2006). Design of an advanced microassembly system for the automated assembly of bio-microrobots. In *Proc. 5th International Workshop on Microfactories*.

[Rajagopalan and Chaudhuri, 1995] Rajagopalan, A. N. and Chaudhuri, S. (1995). A block shift-variant blur model for recovering depth from defocused images. In *Image Processing, 1995. Proceedings., International Conference on*, volume 3, pages 636–639. IEEE.

[Rajagopalan and Chaudhuri, 1998] Rajagopalan, A. N. and Chaudhuri, S. (1998). Optimal recovery of depth from defocused images using an mrf model. In *Computer Vision, 1998. Sixth International Conference on*, pages 1047–1052. IEEE.

[Ralis et al., 2000] Ralis, S., Vikramaditya, B., and Nelson, B. (2000). Micropositioning of a weakly calibrated microassembly system using coarse-to-fine visual servoing strategies. *Electronics Packaging Manufacturing, IEEE Transactions on*, 23(2):123–131.

[Régnier and Chaillet, 2010] Régnier, S. and Chaillet, N., editors (2010). *Microrobotics for Micromanipulation*. Wiley-ISTE.

[Reimer, 1998] Reimer, L. (1998). *Scanning Electron Microscopy: Physics of Image Formation and Microanalysis*. Springer.

[Ritter et al., 2006] Ritter, M., Hemmleb, M., Lich, B., Faber, P., and Hohenberg, H. (2006). Sem/fib stage calibration with photogrammetric methods. In *ISPRS Commission V Symp. 2006 (Int. Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences)*, volume 36.

[Rives et al., 1989] Rives, P., Chaumette, F., and Espiau, B. (1989). Visual servoing based on a task-function approach. In *1st Int. Symposium on Experimental Robotics*, Montréal, Canada.

[Ru et al., 2012] Ru, C., Zhang, Y., Huang, H., and Chen, T. (2012). An improved visual tracking method in scanning electron microscope. *Microscopy and Microanalysis*, 18(03):612–620.

[Ru et al., 2011] Ru, C., Zhang, Y., Sun, Y., Zhong, Y., Sun, X., Hoyle, D., and Cotton, I. (2011). Automated four-point probe measurement of nanowires inside a scanning electron microscope. *Nanotechnology, IEEE Transactions on*, 10(4):674–681.

[Rudnaya et al., 2009] Rudnaya, M., Mattheij, R., and Maubach, J. (2009). Iterative autofocus algorithms for scanning electron microscopy. *Microscopy and Microanalysis*, 15(2):1108–1109.

[Rudnaya et al., 2010] Rudnaya, M., Mattheij, R., and Maubach, J. (2010). Evaluating sharpness functions for automated scanning electron microscopy. *Journal of microscopy*, 240(1):38–49.

[Rudnaya et al., 2012] Rudnaya, M., Ter Morsche, H., Maubach, J., and Mattheij, R. (2012). A derivative-based fast autofocus method in electron microscopy. *Journal of Mathematical Imaging and Vision*, 44(1):38–51.

[Rudnaya et al., 2011] Rudnaya, M., Van den Broek, W., Doornbos, R., Mattheij, R., and Maubach, J. (2011). Defocus and twofold astigmatism correction in haadf-stem. *Ultramicroscopy*, 111(8):1043–1054.

[Santos et al., 1997] Santos, A., Ortiz de Solorzano, C., Vaquero, J. J., Pena, J., Malpica, N., and Del Pozo, F. (1997). Evaluation of autofocus functions in molecular cytogenetic analysis. *Journal of microscopy*, 188(3):264–272.

[Schechner and Kiryati, 1999] Schechner, Y. Y. and Kiryati, N. (1999). The optimal axial interval in estimating depth from defocus. In *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on*, volume 2, pages 843–848. IEEE.

[Schechner and Kiryati, 2000] Schechner, Y. Y. and Kiryati, N. (2000). Depth from defocus vs. stereo: How different really are they? *International Journal of Computer Vision*, 39(2):141–162.

[Schreier et al., 2004] Schreier, H. W., Garcia, D., and Sutton, M. A. (2004). Advances in light microscope stereo vision. *Experimental mechanics*, 44(3):278–288.

[Shannon, 1948] Shannon, C. (1948). A mathematical theory of communication. *Bell system technical journal*, 27:379–423, 623–656.

[Shi and Tomasi, 1994] Shi, J. and Tomasi, C. (1994). Good features to track. In *IEEE Int. Conf. on Computer Vision and Pattern Recognition, CVPR'94*, pages 593–600, Seattle, Washington.

[Sievers and Fatikow, 2005] Sievers, T. and Fatikow, S. (2005). Visual servoing of a mobile microrobot inside a scanning electron microscope. *2005 IEEE/RSJ Int. Conf. Intell. Robot. Syst.*, pages 1350–1354.

[Sievers and Fatikow, 2006] Sievers, T. and Fatikow, S. (2006). Real-time object tracking for the robot-based nanohandling in a scanning electron microscope. *Journal of Micromechatronics*, 3(3):267–284.

[Sim et al., 2013] Sim, K., Nia, M., and Tso, C. P. (2013). Noise variance estimation using image noise cross-correlation model on sem images. *Scanning*, 35(3):205–212.

[Sim et al., 2004] Sim, K., Thong, J., and Phang, J. (2004). Effect of shot noise and secondary emission noise in scanning electron microscope images. *Scanning*, 26(1):36–40.

[Sinram et al., 2002] Sinram, O., Ritter, M., Kleindick, S., Schertel, A., Hohenberg, H., and Albertz, J. (2002). Calibration of an sem, using a nano positioning tilting table and a microscopic calibration pyramid. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 34(5):210–215.

[Snella, 2010] Snella, M. T. (2010). Drift Correction for Scanning-Electron Microscopy by. Master's thesis, Massachusetts Institute of Technology.

[Subbarao, 1988] Subbarao, M. (1988). Parallel depth recovery by changing camera parameters. In *ICCV*, pages 149–155.

[Subbarao and Choi, 1995] Subbarao, M. and Choi, T. (1995). Accurate recovery of three-dimensional shape from image focus. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(3):266–274.

[Subbarao et al., 1993] Subbarao, M., Choi, T., and Nikzad, A. (1993). Focusing techniques. *Journal of Optical Engineering*, 32:2824–2836.

[Subbarao and Surya, 1994] Subbarao, M. and Surya, G. (1994). Depth from defocus: A spatial domain approach. *International Journal of Computer Vision*, 13(3):271–294.

[Subbarao and Wei, 1992] Subbarao, M. and Wei, T.-C. (1992). Depth from defocus and rapid autofocusing: a practical approach. In *Computer Vision and Pattern Recognition, 1992. Proceedings CVPR'92., 1992 IEEE Computer Society Conference on*, pages 773–776. IEEE.

[Sulzmann et al., 1997] Sulzmann, A., Breguet, J.-M., and Jacot, J. (1997). Micromotor assembly using high accurate optical vision feedback for microrobot relative 3d displacement in submicron range. In *Solid State Sensors and Actuators, 1997. TRANSDUCERS'97 Chicago., 1997 International Conference on*, volume 1, pages 279–282. IEEE.

[Sun and Chin, 2004] Sun, W. and Chin, T. (2004). Image-based visual servo for micromanipulation: a multiple-view and multiple-scale approach. In *Micro-Nanomechatronics and Human Science, 2004 and The Fourth Symposium Micro-Nanomechatronics for Information-Based Society, 2004. Proceedings of the 2004 International Symposium on*, pages 341–346.

[Sun et al., 2005] Sun, Y., Duthaler, S., and Nelson, B. (2005). Autofocusing algorithm selection in computer microscopy. In *IEEE/RSJ International Conference onIntelligent Robots and Systems(IROS)*, pages 70–76.

[Sun et al., 2003] Sun, Y., Greminger, M., Potasek, D., and Nelson, B. (2003). A visually servoed mems manipulator. In *Experimental Robotics VIII*, pages 255–264. Springer.

[Sun and Nelson, 2001] Sun, Y. and Nelson, B. J. (2001). Autonomous injection of biological cells using visual servoing. In *Experimental Robotics VII*, pages 169–178. Springer.

[Sutton et al., 2009] Sutton, M., Orteu, J.-J., and Schreier, H. (2009). *Image Correlation for Shape, Motion and Deformation Measurements: Basic Concepts,Theory and Applications*. Springer Publishing Company, Incorporated, 1st edition.

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

[Sutton et al., 2007] Sutton, M. A., Li, N., Garcia, D., Cornille, N., Orteu, J. J., McNeill, S. R., Schreier, H. W., Li, X., and Reynolds, a. P. (2007). Scanning Electron Microscopy for Quantitative Small and Large Deformation Measurements Part II: Experimental Validation for Magnifications from 200 to 10,000. *Exp. Mech.*, 47(6):789–804.

[Tamadazte et al., 2008] Tamadazte, B., Dembélé, S., Fortier, G., Fort-Piat, L., et al. (2008). Automatic micromanipulation using multiscale visual servoing. In *Automation Science and Engineering, 2008. CASE 2008. IEEE International Conference on*, pages 977–982. IEEE.

[Tamadazte et al., 2012] Tamadazte, B., Le-Fort Piat, N., and Marchand, E. (2012). A direct visual servoing scheme for automatic nanopositioning. *Mechatronics, IEEE/ASME Transactions on*, 17(4):728–736.

[Tamadazte et al., 2010] Tamadazte, B., Marchand, E., Dembélé, S., and Le Fort-Piat, N. (2010). Cad model-based tracking and 3d visual-based control for mems microassembly. *The International Journal of Robotics Research*.

[Tao et al., 2005] Tao, X., Cho, H., and Cho, Y. (2005). Microassembly of peg and hole using active zooming. In *Optomechatronic Technologies 2005*, pages 605204–605204. International Society for Optics and Photonics.

[Teunis et al., 1992] Teunis, P., Bretschneider, F., and Machemer, H. (1992). Real-time three-dimensional tracking of fast-moving microscopic objects. *Journal of Microscopy*, 168(3):275–288.

[Tikhonov et al., 2013] Tikhonov, A., Goncharsky, A., Stepanov, V., and Yagola, A. G. (2013). *Numerical methods for the solution of ill-posed problems*, volume 328. Springer Science & Business Media.

[Timischl et al., 2012] Timischl, F., Nemoto, S., et al. (2012). A statistical model of signal–noise in scanning electron microscopy. *Scanning*, 34(3):137–144.

[Torralba and Oliva, 2002] Torralba, A. and Oliva, A. (2002). Depth estimation from image structure. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(9):1226–1238.

[Triggs et al., 2000] Triggs, B., McLauchlan, P. F., Hartley, R. I., and Fitzgibbon, A. W. (2000). Bundle adjustment-a modern synthesis. In *Vision algorithms: theory and practice*, pages 298–372. Springer.

[Tsai, 1987] Tsai, R. (1987). 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.

[Tunnell and Fatikow, 2011] Tunnell, R. and Fatikow, S. (2011). 3d position detection with an fib-sem dual beam system. In *Proceedings of the 10th WSEAS international conference on communications, electrical & computer engineering, and 9th WSEAS international conference on Applied electromagnetics, wireless and optical communications*, pages 128–133. World Scientific and Engineering Academy and Society (WSEAS).

[Van Helden et al., 2010] Van Helden, A., Dupré, S., and van Gent, R. (2010). *The Origins of the Telescope*. Geschiedenis Van De Wetenschap in Nederland. KNAW Press.

[Vargas and Malis, 2005] Vargas, M. and Malis, E. (2005). Visual servoing based on an analytical homography decomposition. In *Decision and Control, 2005 and 2005 European Control Conference. CDC-ECC'05. 44th IEEE Conference on*, pages 5379–5384. IEEE.

[Vikramaditya and Nelson, 1997] Vikramaditya, B. and Nelson, B. (1997). Visually guided microassembly using optical microscopes and active vision techniques. *IEEE Int. Conf. on Robotics and Automation, ICRA'97*, 4:3172–3177.

[Vollath, 1987] Vollath, D. (1987). Automatic focusing by correlative methods. *Journal of Microscopy*, 147(3):279–288.

[Wang et al., 2011] Wang, S., Lu, H., Yang, F., and Yang, M.-H. (2011). Superpixel tracking. In *Computer Vision (ICCV), 2011 IEEE International Conference on*, pages 1323–1330. IEEE.

[Watanabe and Nayar, 1998] Watanabe, M. and Nayar, S. K. (1998). Rational filters for passive depth from defocus. *International Journal of Computer Vision*, 27(3):203–225.

[Weck and Peschke, 2004] Weck, M. and Peschke, C. (2004). Equipment technology for flexible and automated micro-assembly. *Microsystem technologies*, 10(3):241–246.

[Wei and Ma, 1994] Wei, G.-Q. and Ma, S. (1994). Implicit and explicit camera calibration: Theory and experiments. *IEEE Trans. on Pattern Analysis and Machine intelligence*, 16(5):469–480.

[Weiss et al., 1987] Weiss, L., Sanderson, A., and Neuman, C. (1987). Dynamic sensor-based control of robots with visual feedback. *IEEE Journal of Robotics and Automation*, 3(5):404–417.

[Weng et al., 1992] Weng, J., Cohen, P., and Rebiho, N. (1992). Motion and structure estimation from stereo image sequences. *IEEE Trans. on Robotics and Automation*, 8(3):362–382.

[Wergin, 1985] Wergin, W. (1985). Three-dimensional imagery and quantitative analysis in sem studies of nematodes. *Agriculture, ecosystems & environment*, 12(4):317–334.

[Xiong et al., 1995] Xiong, Y., Shafer, S., et al. (1995). Moment filters for high precision computation of focus and stereo. In *Intelligent Robots and Systems 95.'Human Robot Interaction and Cooperative Robots', Proceedings. 1995 IEEE/RSJ International Conference on*, volume 3, pages 108–113. IEEE.

[Yang et al., 2001] Yang, G., Gaines, J., and Nelson, B. (2001). A flexible experimental workcell for efficient and reliable wafer-level 3d micro-assembly. In *IEEE Int. Conf. on Robotics and Automation*, volume 1, pages 133–138. IEEE.

[Yang et al., 2003] Yang, G., Gaines, J. a., and Nelson, B. J. (2003). A supervisory wafer-level 3D microassembly system for hybrid MEMS fabrication. *J. Intell. Robot. Syst. Theory Appl.*, 37:43–68.

[Yeo et al., 1993] Yeo, T., Ong, S., Sinniah, R., et al. (1993). Autofocusing for tissue microscopy. *Image and Vision Computing*, 11(10):629–639.

[Yesin and Nelson, 2005] Yesin, K. and Nelson, B. (2005). A cad-model based tracking system for visually guided microassembly. *Robotica*, 23:409–418.

[Yilmaz et al., 2004] Yilmaz, A., Li, X., and Shah, M. (2004). Contour-based object tracking with occlusion handling in video acquired using mobile cameras. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 26(11):1531–1536.

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

[Yu et al., 2002] Yu, M.-F., Wagner, G. J., Ruoff, R. S., and Dyer, M. J. (2002). Realization of parametric resonances in a nanowire mechanical system with nanomanipulation inside a scanning electron microscope. *Physical Review B*, 66(7):073406.

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

[Zhou et al., 1998] Zhou, Y., Nelson, B., and Vikramaditya, B. (1998). Fusing force and vision feedback for micromanipulation. In *IEEE Int. Conf. on Robotics and Automation*, volume 2, pages 1220–1225. IEEE.

[Zhu and Yuille, 1996] Zhu, S. C. and Yuille, A. (1996). Region competition: Unifying snakes, region growing, and bayes/mdl for multiband image segmentation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 18(9):884–900.

[Zhu et al., 2011] Zhu, T., Sutton, M., Li, N., Orteu, J.-J., Cornille, N., Li, X., and Reynolds, A. (2011). Quantitative stereovision in a scanning electron microscope. *Experimental Mechanics*, 51(1):97–109.

[Zhuo and Sim, 2009] Zhuo, S. and Sim, T. (2009). On the recovery of depth from a single defocused image. In *Computer Analysis of Images and Patterns*, pages 889–897. Springer.

[Ziou and Deschenes, 2001] Ziou, D. and Deschenes, F. (2001). Depth from defocus estimation in spatial domain. *Computer Vision and Image Understanding*, 81(2):143 – 165.

Abstract

With the development of nanotechnology, it became possible to design and assemble nano-objects. For robust and reliable automation processes, handling and manipulation tasks at the nanoscale is increasingly required over the last decade. In this thesis, we address the issue of micro- and nano-positioning by visual servoing in a Scanning Electron Microscope (SEM). The SEM vision geometry models are studied at first. A nonlinear optimization process for SEM calibration has been presented considering both perspective and parallel projection model. In order to solve the problem that the motion along the depth direction is not observable in a SEM, the image defocus information is considered as a visual feature to control the motion along the depth direction. A hybrid visual servoing scheme has been proposed for 6-DoF micropositioning task. It has been validated using a parallel robot in a SEM. Based on the similar idea, a closed-loop control scheme for SEM autofocusing task has been introduced. In order to achieve the visual guidance in a SEM, a visual tracking and 3D pose estimation framework has been proposed.

Keywords: Visual servoing, visual tracking, scanning electron microscope, micro-robotics, defocus information

Résumé

Avec le développement de les nanotechnologies, il est devenu possible et souhaitable de créer et d'assembler des nano-objets. Afin d'obtenir des processus automatisés robustes et fiables, la manipulation à l'échelle nanométrique est devenue, au cours des dernières années, une tâche primordiale. Dans cette thèse, nous abordons la problématique du micro- et nano-positionnement par asservissement visuel via l'utilisation d'un microscope électronique à balayage (MEB). Dans un premier temps, les modèles géométriques de la vision appliqués aux MEB sont étudiés afin de présenter, par la suite, une méthode l'étalonnage de MEB par l'optimisation non-linéaire considérant les modèles de projection perspective et parallèle. Afin de résoudre le problème de la non-observabilité du mouvement dans l'axe de la profondeur du MEB les informations de défocalisation d'image sont considérées comme caractéristiques visuelles pour commander le mouvement sur cet axe. Une méthode d'asservissement visuelle hybride est alors proposée pour effectuer le micro-positionnement en 6 degrés de liberté. Cette méthode est ensuite validée via l'utilisation d'un robot parallèle dans un MEB. Finalement, un système de contrôle en boucle fermée pour l'autofocus du MEB est introduit, et une méthode de suivi visuel et d'estimation de pose 3D est proposée afin de réaliser le guidage visuel dans un MEB.

Mot clé: Asservissement visuel, suivi visuel, microscope électronique à balayage, micro-robotique, défocalisation

Résume

Avec le développement des nanotechnologies, il est devenu possible et souhaitable de créer et d'assembler des nano-objets. Afin d'obtenir des processus automatisés robustes et fiables, la manipulation à l'échelle nanométrique est devenue, au cours des dernières années, une tâche primordiale. Dans cette thèse, nous abordons la problématique du micro- et nano-positionnement par asservissement visuel via l'utilisation d'un microscope électronique à balayage (MEB). Dans un premier temps, les modèles géométriques de la vision appliqués aux MEB sont étudiés afin de présenter, par la suite, une méthode l'étalonnage de MEB par l'optimisation non linéaire considérant les modèles de projection perspective et parallèle. Afin de résoudre le problème de la non-observabilité du mouvement dans l'axe de la profondeur du MEB les informations de défocalisation d'image sont considérés comme caractéristiques visuelles pour commander le mouvement sur cet axe. Une méthode d'asservissement visuel hybride est alors proposée pour effectuer le micro-positionnement en 6 degrés de liberté. Cette méthode est ensuite validée via l'utilisation d'un robot parallèle dans un MEB. Finalement, un système de contrôle en boucle fermée pour l'autofocus du MEB est introduit, et une méthode de suivi visuel et d'estimation de pose 3D est proposée afin de réaliser le guidage visuel dans un MEB.

Ce travail a été réalisé dans le cadre du projet ANR Nanorobust de l'Agence Nationale de la recherche. It est intitulé "Caractérisation multiphasique de nano-objets et manipulation robotisée sous environnement MEB". Ce projet porte sur deux thèmes de recherche: (1) les manipulations d'objets en petite échelle par une approche de commande afin de les mettre sur une base pour les transporter vers le système de mesures; (2) l'analyse des propriétés structurales de ces objets sous un MEB, sans endommager ou de contaminer les objets.

Cette thèse porte sur la commande par la vision MEB dans ce projet. La motivation de cette thèse est de réaliser le micro / nano-positionnement robust dans un MEB. L'un des défis liés à ces tâches est que le MEB produit des images différemment d'un microscope optique. Dans ce cas, le processus d'imagerie MEB doit être étudié, en particulier sur le processus d'étalonnage du MEB compte tenu des distorsions. En effet, à fort grossisse-

ment, le modèle de projection géométrique du MEB est différent de celle d'une caméra optique. Au lieu d'utiliser un modèle de projection en perspective, les modèles de projection parallèles doivent être considérés à fort grossissement. Il est difficile d'observer le mouvement sur la direction de la profondeur. Afin d'effectuer une tâche de positionnement 6-DDL(Degré de liberté) dans un MEB, la commande du mouvement sur la direction de la profondeur doit être étudiée de manière adéquate. Dans cette thèse, l'objectif envisagé est de proposer une solution fiable et robuste pour le micro / nano-positionnement par asservissement visuel en utilisant les informations d'images.

Le chapitre 1 présent le contexte sur l'imagerie MEB. Parmi les nombreuses microscopies, le MEB est important dans le travail de cette thèse. Le MEB est l'une des techniques les plus importantes dans l'observation des objets en micro / nano-échelle. Différent des autres microscopes optiques, la MEB produit des images en balayant la surface de l'échantillon avec un faisceau focalisé d'électrons à haute énergie. Dans ce chapitre, nous avons présenté des connaissances de base fondamentale sur l'imagerie MEB. La structure et les composants d'un MEB sont présentés, y compris le canon à électrons, des lentilles, des détecteurs d'électrons, etc. Comme une question importante dans notre travail sur l'asservissement visuel dans un MEB, le processus de formation d'image MEB et d'autres facteurs sont détaillés. Avec cette connaissance de base, dans le chapitre suivant, nous traiterons de la méthode d'étalonnage pour un MEB.

Le chapitre 2 est consacré à un problème fondamental pour la vision de MEB: étalonnage. Comme une tâche nécessite le calcul de l'information métrique à partir des images 2Ds acquises, l'étalonnage de la MEB est un problème important. Dans ce chapitre, un aperçu de l'étalonnage d'un capteur optique et un MEB sont d'abord affirmés. Comme les connaissances de base, le modèle d'imagerie géométrique de la caméra et les modèles de projection sont présentés. Nous proposons d'utiliser un processus d'optimisation non linéaire pour l'étalonnage du MEB. Les paramètres intrinsèques et extrinsèques est calculés par un algorithme d'optimisation non linéaire itérative qui minimise l'erreur d'alignement entre la position actuelle estimée et sa position observée. Les résultats expérimentaux de deux MEB différentes ont prouvé l'efficacité de l'approche proposée.

Chapitre 3 présent la commande par la vision et un aperçu de son application dans les micros / nano-robotique. Comme un système de commande en boucle fermée à l'aide des informations visuelles, l'asservissement visuel est effectué en minimisant l'erreur entre la fonction visuelle actuelle et la fonction visuelle souhaitée. Il joue un rôle important dans le contrôle des mouvements robotiques. Cependant, certaines difficultés ont été trouvées dans l'application de l'asservissement visuel traditionnel dans un MEB. L'un des défis importants est qu'il est très difficile d'observer le mouvement du robot de l'image MEB à des grossissements élevés en raison du modèle de projection parallèle. Afin de résoudre ce problème, l'approche de l'asservissement visuel pour le mouvement du robot sur la direction de la profondeur est ensuite présentée dans le chapitre suivant.

Nous nous concentrons sur la conception de la loi de commande par l'asservissement visuel afin de bouger le robot sur la direction de la profondeur dans Chapitre 4. Parmi les différentes fonctions existant de netteté d'image, le gradient de l'image est sélectionné comme la fonction visuelle dans l'asservissement visuel. Différentes approches ont été proposées pour effectuer la tâche de l'asservissement visuel. La première approche consiste à minimiser l'erreur image de gradient entre l'image souhaitée et l'image courante dans le domaine spatial. La loi de commande peut être analytiquement calculé ou être estimés à l'aide d'une fonction rationnelle. En variante, l'asservissement visuel pour le mouvement sur la direction de la profondeur peut également être effectué dans le domaine fréquentiel. L'écart-type dans le noyau gaussien est modélisé dans la fonction de coût. Le régime d'asservissement visuel est réalisé en minimisant l'écart type estimé. Les résultats expérimentaux montrent que la première approche est robuste et précis. En raison du niveau de bruit élevé dans une image MEB, l'estimation de la seconde approche est inexacte lorsque la position actuelle est proche de la position de mise au point. Par conséquent, nous proposons l'approche de domaine spatial pour un micro / nano-positionnement en 6 DDL dans un MEB.

Dans Chapitre 5, l'asservissement visuel hybride est proposé pour une tâche du micro / nano-positionnement automatisé en 6 DDL. Différent de suivi visuel traditionnel, seulement l'information image de pure apparence est nécessaire dans cette méthode. Les informations d'intensité d'image

sont utilisées pour contrôler le mouvement linéaire sur x- et y- axes de et le mouvement angulaire autour de x-, y- et z-axes. Basé sur la recherche dans le chapitre IV, le gradient d'image est présenté comme un élément visuel en fonction de la variation de netteté de l'image par le mouvement sur z-axis. Cette méthode est validée par des expériences sur une étape de positionnement parallèle à 6 DDL et un microscope optique au premier abord. La performance du système hybride asservissement visuel et celui de l'asservissement visuel en utilisant uniquement l'intensité de l'image sont évaluées et comparées. Compte tenu de leur performance, nous suggérons la méthode hybride pour les applications basées MEB. Cette dernière méthode peut être appliquée lorsque la profondeur de champ du capteur est grande et le grossissement est très faible. Dans ce cas, le mouvement le long de la direction de la profondeur peut évidemment être observé à partir de l'image et le modèle de projection en perspective peut-être appliquée. Enfin, le régime hybride asservissement visuel est validé en utilisant le même robot dans une MEB à 1000x pour 6-DDL micropositionnements. Comme indiqué précédemment, les travaux futurs pourraient être la validation de cette approche à un grossissement plus élevé en utilisant différentes configurations expérimentales et différentes échantillons et l'amélioration de la robustesse du système hybride asservissement visuel.

Dans Chapitre 6, un système de commande en boucle fermée a été proposé pour une mise au point automatique à pleine échelle de la MEB. Il utilise les informations de gradient d'image que le score de netteté dans la conception de la loi de commande basée sur la vision. L'optimum de la netteté de l'image a été obtenu en mettant à jour la distance de travail du MEB par une manière itérative. Le procédé proposé converge rapidement vers la valeur optimale. Contrairement aux méthodes basées sur la recherche classique, la méthode proposée atteint directement la position de focalisation optimale. La méthode a été validée par les expériences différentes et les résultats obtenus montrent clairement l'efficacité des méthodes.

Le suivi visuel et l'estimation de la pose 3D de l'objet sont importants pour effectuer un guidage visuel pour la micro / nano-manipulation. La tâche de positionnement peut également être réalisée par le suivi de l'objet, puis d'effectuer l'asservissement visuel classique. Dans Chapitre 7, nous proposons une méthode de suivi visuel basé sur un modèle pour estimer la pose

3D de l'objet à micro-échelle. Cette méthode est validée par les expériences en 4 DDL dans un MEB. Il est également montré que, en appliquant un filtre particules dans notre cadre, la précision de la position de profondeur estimation peut être considérablement améliorée.