EIDGENOSSISCHE TECHNISCHE HOCHSCHULE LAUSANNE
POLITECNICO FEDERALE DI LOSANNA
SWISS FEDERAL INSTITUTE OF TECHNOLOGY

INSTITUT DE PRODUCTION ROBOTIQUE (IPR)
LABORATOIRE DE SYSTEMES ROBOTIQUES (LSRO)

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Technische Universität München
Fakultät für Elektro- und Informationstechnik
FG Mikrostrukturierte mechatronische Systeme

The optimal use of vision as part of the manipulation of micron-sized objects

Diploma Thesis by
Thomas Wimböck

September 17th, 2004

Contents

Chapter 1

Introduction

1.1 Motivation

The demand for tools to handle micron-sized objects with high precision gives rise to the CTI TOP NANO21 project *Manipulating Microscale Objects with Nanoscale Precision*. Applications of ultra-high precision manipulation can be found in the field of optics, telecommunications, medical engineering and the automotive industry where a higher integration density can lead to better device performance and smaller size. The manipulation of micron-sized parts ($< 1mm$ and $> 1\mu m$) parts can be considered as micromanipulation (MM) with a positioning resolution being significantly smaller than the dimension of the parts.

In particular, the assembly of micro-opto-electro-mechanical systems (MOEMS) can be improved. MOEMS are characterized to be highly integrated, flexible, batch-processed, microscale devices. However, many constraints have to be considered during the development: manufacturing processes, product specifications, materials, interfaces to the environment. The packaging costs of a MOEMS dominate the manufacturing costs [8]. There are different approaches to decrease the packaging costs. For instance, techniques like flip-chip packaging create a mechanical and an electrical connection at one time. Most of the MOEMS are designed either monolithic or hybrid. Monolithic devices integrate all functions in one substrate and hence little packaging is necessary. The drawback of this concept is that the design, because of technological constraints, is complicated and that the corresponding product cannot be modified easily. Hybrid system can be designed very fast and flexible since modules being manufactured by different technologies (for example different materials) could be integrated easily using micromanipulation [1].

A second big field of application is the manipulation of cells. Automated cell manipulation can be for example used for full-automated microinjections, used by molecular biologists for cell research or used for new approaches to gene therapies. Currently, the repetitive task of microinjection is tele-operated by a human operator.

This work is part of a first step towards the manipulation with nanoscale resolution within the mentioned project. The aim of this diploma thesis is to develop an interface that allows a human operator to manipulate objects in a semi-automatic fashion: The operator can select between detected objects which are manipulated full-automatically. The goal is to

obtain a positioning repeatability of less than $1\mu m$ after releasing the object. Therefore a vision system, calibration procedures and a control strategy for micromanipulation will be integrated and a high-precision robot and an end-effector suitable to handle micron-sized parts will be employed.

1.2 State-of-the-art

Micromanipulation (MM) can be categorized either as manual, tele-operated or automated. Manual MM is not desirable because it is slow and the repetitive process is fatiguing the operator. Tele-operated MM shows the highest positioning resolution (high enough to probe surfaces at atomic scale), requires however an operator as well [1]. An overview of the developments in tele-operated MM can be found in [1]. The category this work is seeking at is automatic manipulation. The following section gives a short summary of related work in this category considering especially the vision-related issues. All projects have in common that the main visual sensor for lateral positioning is a light microscope and that almost all groups present an autofocus method to obtain position information along the optical axis of the microscope.

A research group at the university of Oldenburg [4] [10] detects the manipulator and the object of interest with the Matrox Imaging Library (MIL) geometrical model finder, which is a robust contour-based tool for pattern recognition. The correlation result for the end-effector is also used to be able to detect the end-effector height when it is close to the plane in which the objects are.

In [1] region growing is used for the detection of objects and the manipulator and the autofocus is calculated based on the Brenner function. A relative lateral positioning accuracy before the object release of $\pm 1\mu m$ is achieved.

The authors of [3] detect the position and orientation of objects using the probalistic generalised Hough transform. For objects with non-rigid shapes active contours are employed. The autofocus is based on the Tenengrad criterion, however information about object height is obtained through a laser triangulation system. An object tracking system based on template matching was implemented.

In [9] and [11] an object tracker based on an optical flow technique called Sum-of-Squared-Differences (SSD) is employed which is improved by a pyramidal search scheme. A histogram-based autofocus technique is employed.

The authors of [6] use SSD tracking under the XVision software package that is initialized by the user for objects and by blob analysis for the end-effector. In order to detect the end-effector height relative to the substrate, the end-effector is lowered till the substrate is moved. This motion is registered by an optical flow technique. The end-effector is then lifted in steps of $1\mu m$ until the substrate does not move anymore.

The authors of [7] employ template matching. Because of the high computational effort a look-and-move control strategy is used. A template of the object to be manipulated is selected by the user at the beginning of the manipulation. Using a stereo-light microscope and an

extended Kalman filter to update the camera calibration during operation, a relative lateral position precision in the order of $100nm$ is obtained.

1.3 Notation and definitions

This section describes the notation and some definitions used in the following.

As described in [22] the temporally- and spatially-sampled images coming from the data acquisition can be considered as matrices $I(\mathbf{x}_i) \in \mathbb{R}^M \times \mathbb{R}^N$ whose elements represent the image intensities at the corresponding index $\mathbf{x}_i = [x_i, y_i]^T$. M and N are integers and correspond to the height and the width of the image correspondingly. The image coordinates x_i, y_i are integers and run from 0 to $N-1$ and $M-1$, respectively. Each pixel $\mathbf{x}_i$ corresponds to the area of each array element of the camera sensor. Hence the averaged intensity level of illumination over the sensor element area is represented as $I(\mathbf{x}_i)$. The origin of the coordinate system is situated in the upper left corner of the image, where the x_i-axis points to the right upper image corner and the y_i-axis to the lower left image corner. For example the coordinates (u, v) determine the pixel location as shown in figure 1.1. $\mathcal{I}$ represents the set of all image coordinates $\mathbf{x}_i$. Some image processing algorithms return their results with *sub-pixel* precision so that in this case $\mathbf{x}_i \in \mathbb{R}^2$. Variables with a hat, for example $\hat{\mathbf{x}}$, indicate estimated values. The frame j in which a vector $\mathbf{x}$ is defined, is also written as $\{\mathbf{x}\}_j$.

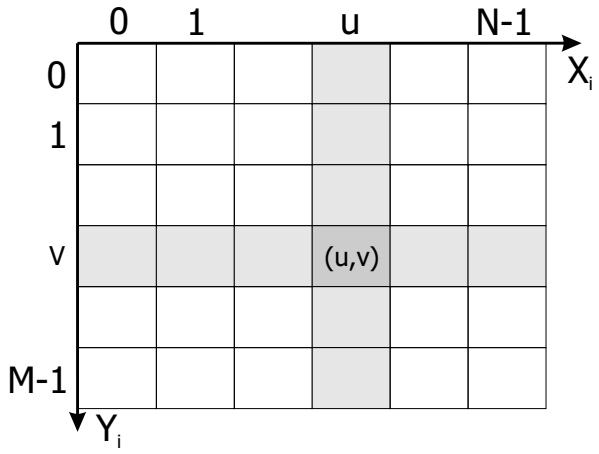

Figure 1.1: Representation of an image.

The image size of the high-resolution microscope camera is 1024×768 pixel. If nothing else is stated image operations are performed on this image format.

The software is running on the Windows XP operating system on a PC with two Pentium III 801MHz processors. The execution times that will be presented are measured during the normal operation of the whole system. Due to the multi-tasking behavior of Windows XP the execution times vary. Hence, delays of background processes like displaying video streams are included in the measurement. Because of that the execution time is averaged over 10 measurements.

The operator $\mathcal{E}\{\cdot\}$ is used to represent the calculation of the expectation value and $\mathcal{V}\{\cdot\}$ is the operator to calculate the variance over all elements that are passed as argument.

1.4 Outline

Chapter 2 starts with an introduction to micromanipulation. An overview of the used experimental setup is given. The different manipulation steps and their sequence are also presented. The section ends with the software design and graphical user interface description. Machine vision and its application to micromanipulation is stressed in chapter 3. Beginning with an analysis of the vision system and the selection of an image processing library, section-wise the acquisition, calibration, the properties of the images are discussed accompanied by experimental results. The last two sections of the machine vision chapter treat the detection of the relative distance of the substrate to the focal plane and lateral position detection in the field of view (FoV) of the microscope. In chapter 4 the calibration of the relative distance between an end-effector and the focal plane and a relative positioning law are discussed. Chapter 5 presents the results of the overall system performance to manipulate micron-sized objects. The last chapter summarizes the presented work and gives an outlook of the steps which could be followed.

Chapter 2

Micromanipulation - General concept

The basic idea of micromanipulation, the description of the setup and the sequence of manipulation steps are topics of this chapter. At the end, the software design and the developed graphical user interface (GUI) is discussed briefly.

As already stated in the introduction, micromanipulation is known as the manipulation of micron-sized parts with even higher precision. The high precision leads to very demanding requirements for the positioning of the manipulator. Considering the manipulation of *macroscaled* objects in an industrial application, it is often sufficient to utilize a well-calibrated robot in feed-forward mode. However, in the micro-domain effects like thermal extension and errors in the robot kinematic cannot be neglected. These effects limit the positioning accuracy of a robot in general. The *positioning accuracy* is the precision with which a certain point can be reached by the robot from any start position at any time in the work space, whereas the *positioning repeatability* is the ability of a robot to return to the same position within a short time interval. Certainly, if there is thermal expansion of the robot the start position is already shifted and hence this introduces and error in the positioning accuracy. To enhance the manipulation performance with respect to accuracy and manipulation duration, additional sensor information is needed. A microscope combined with a camera is used to detect objects and the gripper and their relative position to each other. This enables to close the relative positioning control loop locally in the field of view (FoV) of the microscope so that a high relative positioning precision is obtained.

Even more important towards the overall manipulation performance is the *repeatability of the object placement*. Figure 2.1 depicts the possible linear offsets in the plane that can occur during a manipulation. The displacements are measured after fixing the object to see if the manipulator was able to align an object or if maybe the object moved relative to the manipulator. The displacement is measured a second time after releasing the object which shows how well an object can be aligned with a reference frame. Furthermore, information about the fixing and releasing process can be obtained. The main objective of this work is to be able to detect the relative position of the manipulator to the object and to minimize these offsets, hence to obtain a high repeatability in object positioning.

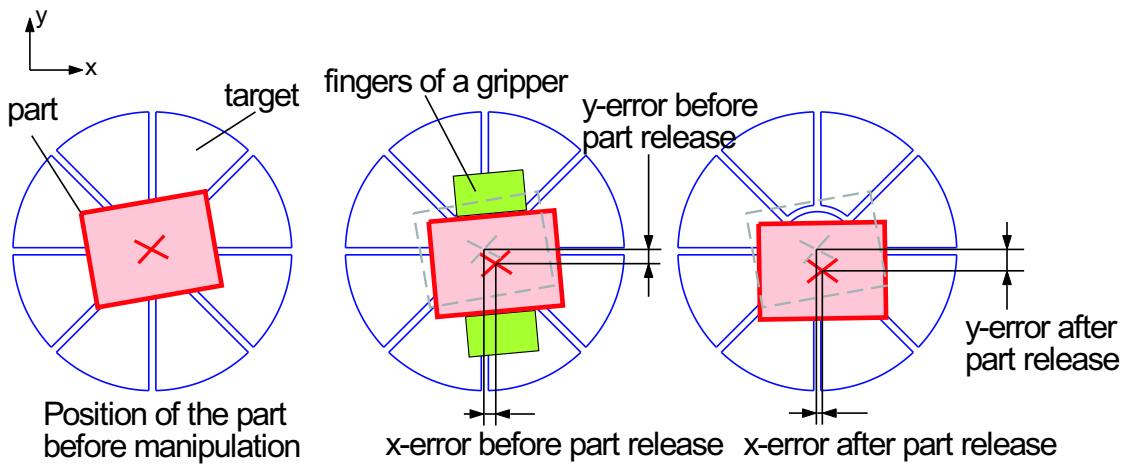

Figure 2.1: Measuring the repeatability to position objects.

2.1 Setup

This section describes the micromanipulation setup. Figure 2.2 gives an schematic overview of the system structure.

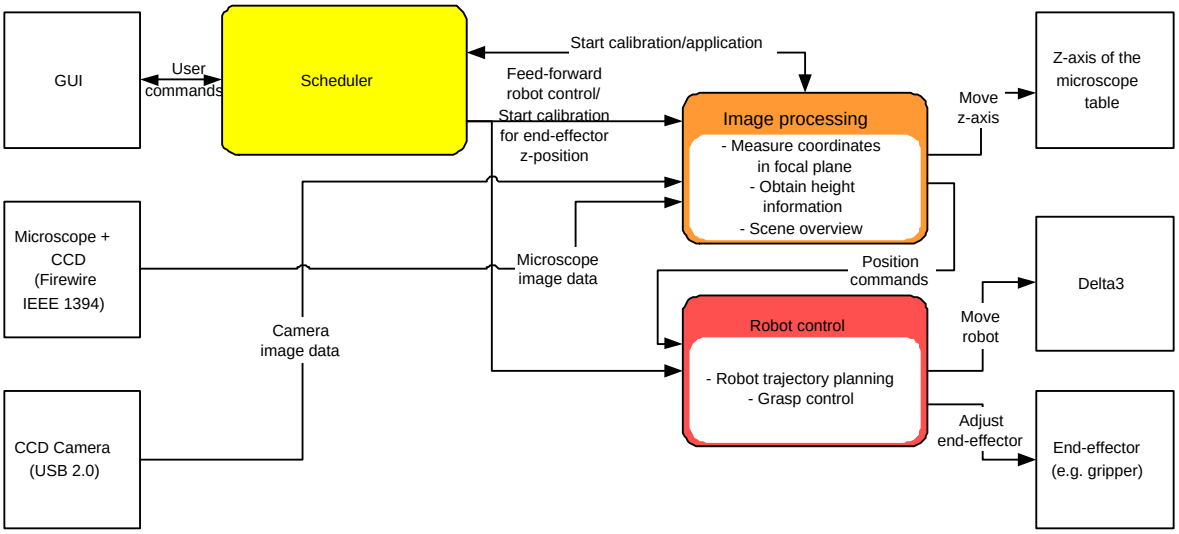

Figure 2.2: Schematic system overview.

The following elements are visualized:

- The user interacts with the system over a graphical user interface (GUI).
- A scheduler works as an intermediate layer between the user input and the image processing and the robot control.

- Two cameras are employed: A Sony XCD-X710 high-resolution CCD camera is mounted on top of the microscope to obtain the microscope images. The camera is connected to a PC via FireWire IEEE1394. A ToUcam PRO II from Philips camera is used to give a coarse view onto the workspace of the robot.
- The employed manipulator Δ^3 *version 2* provides three DOF of linear motion. This robot provides a range of $\pm 2mm$, a resolution of $50nm$, a repeatability of $\pm 50nm$ and an accuracy of $\approx 10\mu m$ before calibration and a maximum velocity of $30\frac{mm}{s}$. An calibration procedure is currently developed at the LSRO to increase the accuracy to $100nm$, which is not used in this application.
- 2D images taken by the camera are processed to determine (x_i, y_i)-coordinates in the focal plane of the microscope. Height information of the position of the substrate, respectively the z-coordinates along the optical axis, is evaluated using autofocus techniques and the controllable height position of the microscope table. Moving the substrate up and down, the microscope images at several positions are used to defer height information[1].
- The manipulation of micron-sized objects requires a special end-effector: In this work pipettes as they are used for manipulation of biological cells, as well as two-jaw grippers are employed. The end-effectors are mounted towards the focal plane under an angle β (compare figure 2.3). In case of the two-jaw gripper at each finger two edges at its tip are considered: The *lower* edge refers to the one which is closer to the substrate surface. The other edge is called *upper* edge in the following.

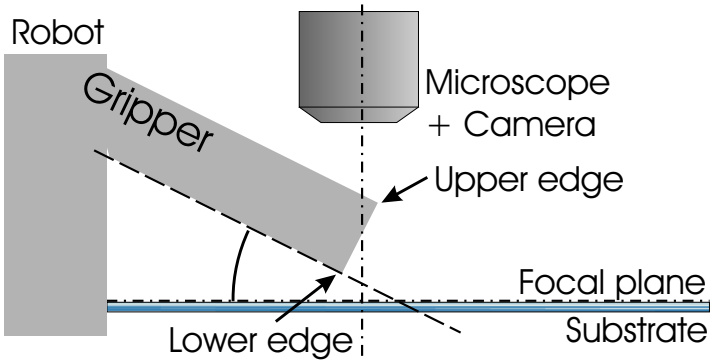

Figure 2.3: Definition of the angle between the orientation of the end-effector and the focal plane.

Figure 2.4 presents the setup alternative 1 in which the microscope is looking from above onto the substrate: With a 20× magnification objective the glass substrate below is observed. The substrate is positioned temporarily on a beam splitter which is used in combination with the light source, that can be seen on the left side of the image, to obtain backlighting of the substrate. The beam splitter stands on a manual XYZ positioning table. A glass pipette is mounted on the Δ^3 *version 1*.

Figure 2.5 shows a photograph of the setup in which the microscope is looking from below the

[1]Note that the substrate positioning was not yet motorized and that the substrate was moved into focus manually.

substrate. The main differences are to the configuration above: Using a mirror the optical path is redirected so that the glass substrate is observed from below. The Δ^3 *version 2* [2] is used and a two-jaw gripper which was developed at the LSRO is mounted. The gripper is manufactured by laser-cutting leading to high surface roughness of a few μm. It is actuated pneumatically and the pressure is supplied through the blue tube shown in the photograph. The gripper thickness is $50\mu m$ and the mounting angle $\beta = 20^o$. Due to a misalignment of about $7\mu m$ between the two fingers in height only the edge of one finger can be in focus.

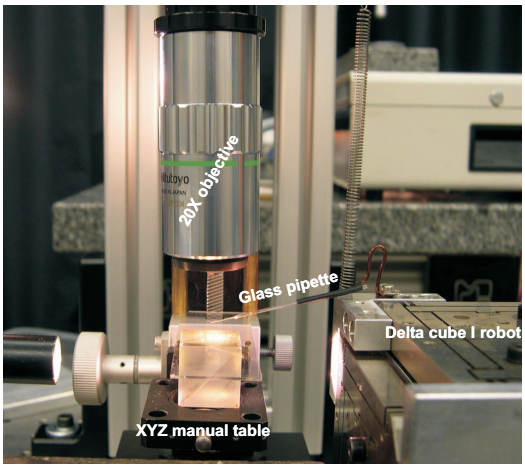

Figure 2.4: Image of the experimental setup: Alternative 1; Configuration vertical microscope.

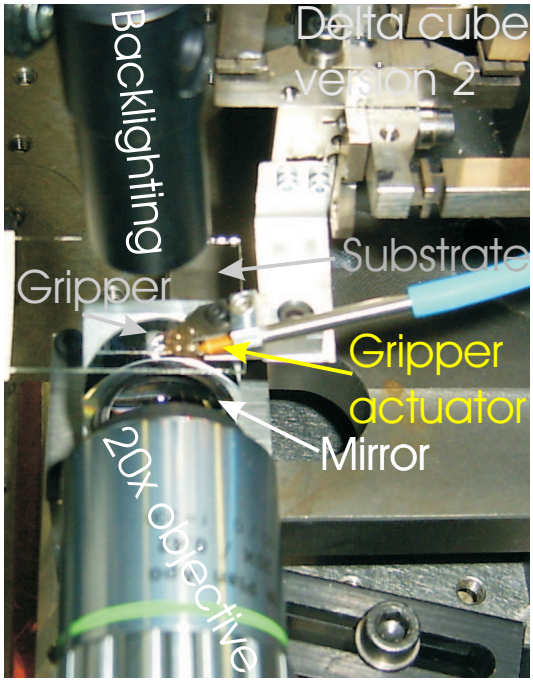

Figure 2.5: Image of the experimental setup: Alternative 2; Configuration horizontal microscope.

2.2 Manipulation sequence

In a *macroworld* application a manipulation sequence is easily described: Locate the object and the gripper in the workspace. Track the position of both and move the end-effector to the object and manipulate the object. In the used setup the limited field of view and limited depth of focus of the microscope objective must be considered. Figure 2.6 shows the manipulation sequence that is divided into basically three blocks: *Calibration* of the cameras and of the height of the end-effector (e.g. gripper) with respect to the substrate, *Application initialization* and *Visual servoing*.

The user calibration or the application is started by the user. During the *Application initialization*

- the microscope table is adjusted such that the objects are in the focal plane to which

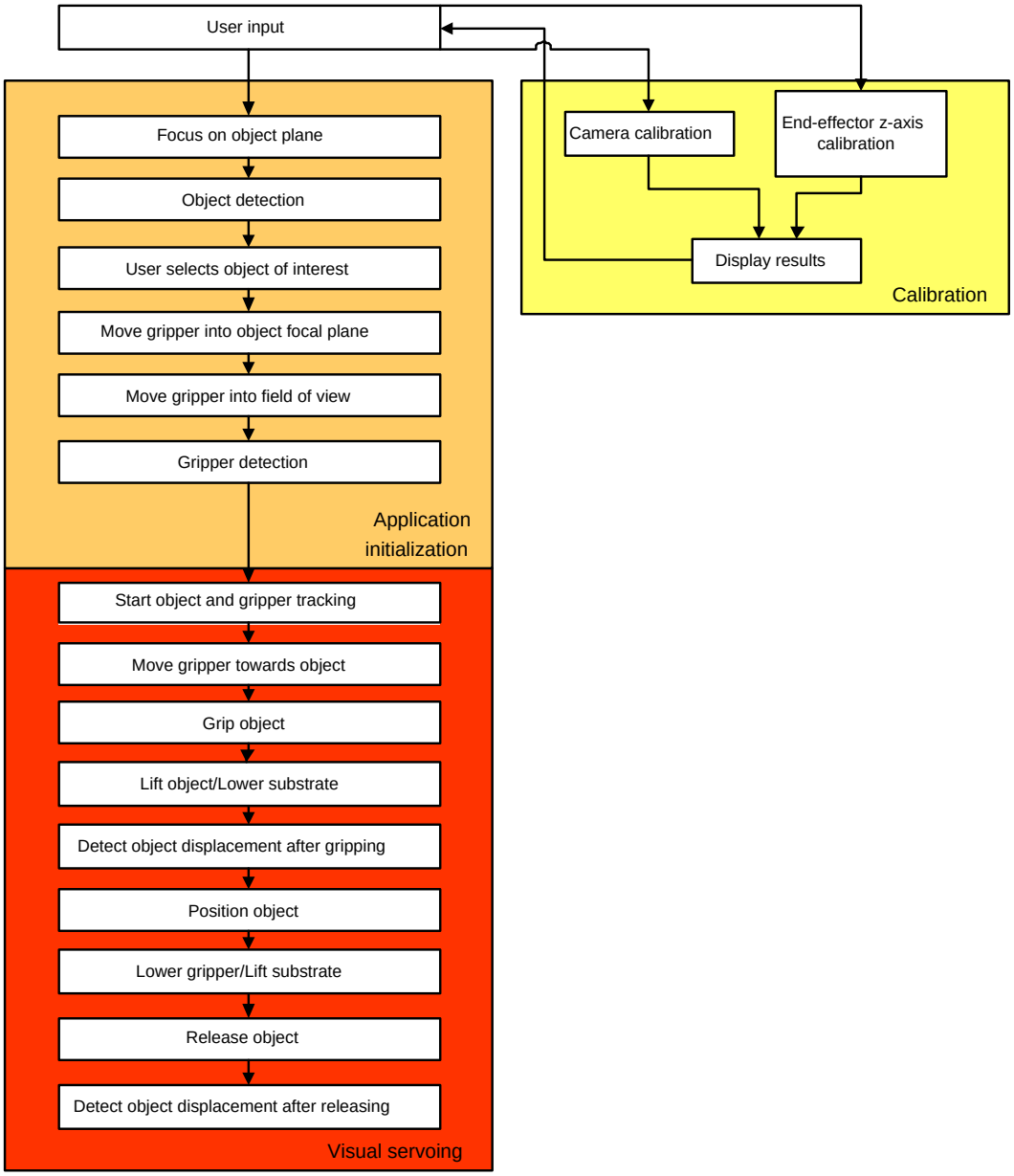

Figure 2.6: Sequence of the whole manipulation task.

will be referred in the following as object plane.

- Assuming there are objects in the field of view, the objects appear sharp in the microscope images which enables object detection and their characterization.
- The user receives a choice of detected objects from which one has to be selected.
- Afterwards, the end-effector is moved into the focal plane and into the field of view. This is possible because the relative displacement between the field of view (FoV) and the robot zero position is known from previous experiments and the accuracy of the robot is high enough to return into the FoV.

- Detecting the end-effector position in the FoV and knowing the object position,
- tracking of the end-effector and the object can be initiated and
- the end-effector can be moved towards the object.
- Once the manipulation position is reached the object is fixed and
- the substrate is moved down[2]. This means that the object and the gripper are still in the focal plane such that the following detection of the object displacement can be performed in the focal plane.
- The object is then positioned at a desired position. Therefore the object position is estimated while being attached to the end-effector. This estimate is subtracted from the desired position to generate a relative positioning error that is used to position the object,
- The substrate is lifted for the same amount it was lowered, and
- the object is released.
- Finally, the object displacement is determined to see how repeatable the positioning of the objects is.

2.3 Software design and graphical user interface

The basic specifications of a software system, that controls a manipulation process as described above, can be formulated as:

- Interfacing several different hardware systems,
- Image processing,
- Calibration,
- End-effector control,
- Human-machine interface,
- Data logging and exchange.

To meet these specifications, a modular object-oriented structure was developed to insure maintainability, re-useability and readability of the source code. Therefore the system was divided into the following objects:

- Camera object from which the microscope camera and the coarse/overview camera are derived,
- Video recorder object,
- Autofocus object,
- Image processing object,

[2]Since the substrate positioning was not yet motorized. Hence the gripper was lifted.

- Robot object,
- End-Effector object from which the Gripper object is derived and
- Visualization and control objects

The development environment Microsoft Visual C++ 6.0 was used. The GUI was developed using the Microsoft Foundation Class (MFC). The objects of the GUI correspond to the visualization and control objects. Figure 2.7 presents the GUI. Several buttons are provided to start the

- calibration of the microscope (*Fine camera calibration)* or of the overview camera (*Coarse camera calibration*),
- detection of the $z-$position of the substrate using autofocus (*Determine substrate z-position*),
- calibration of the $z-$position of the tip of the end-effector (*Calibrate z(EE)*),
- initialization routine of the robot (*Init Robot*),
- detection of objects in the field of view followed by a selection of one of the found objects (*Find objects*),
- combined detection of the end-effector and a selected object(*Find gripper and object*),
- detection of the gripper position (*Find gripper*),
- pick&place operation as shown in chapter 5 (*Start Manipulation*),
- pausing or stopping of the manipulation (*Pause, STOP*),
- manual robot motion dialog box that allows global and relative motion commands (*Move Robot*),
- definition of the gripper template needed for the template matching described in section 3.7.2 (*Define Gripper Template*),
- displaying and optional saving of the current microscope image in full resolution (1024×768)
- testing of routines (*Test*).

In the menu bar→Camera Settings the framerate and the properties of the microscope camera can be set. Furthermore, the partial scanning options of the microscope camera can be accessed and the recording of the microscope video stream can be started and stopped. In the menu bar→Measurements the image signal-to-noise ratio (SNR) can be determined as well a routine to obtain image statistics can be started.

There are several outputs in the GUI:

- current manipulation and gripping status,
- current position of the robot in world coordinates (compare also fig. 3.12),
- image processing results are shown: estimated object and gripper tip position in FoV coordinates (compare fig. 3.12) are displayed,

- live streaming of the overview and of the microscope camera data (the blue cross is aligned with the the center of gravity of the selected object; the other two crosses are aligned with the corners at the gripper tip (compare section 3.7.2)),
- text output used for longer user messages and debugging,
- the gripping force is currently not available.

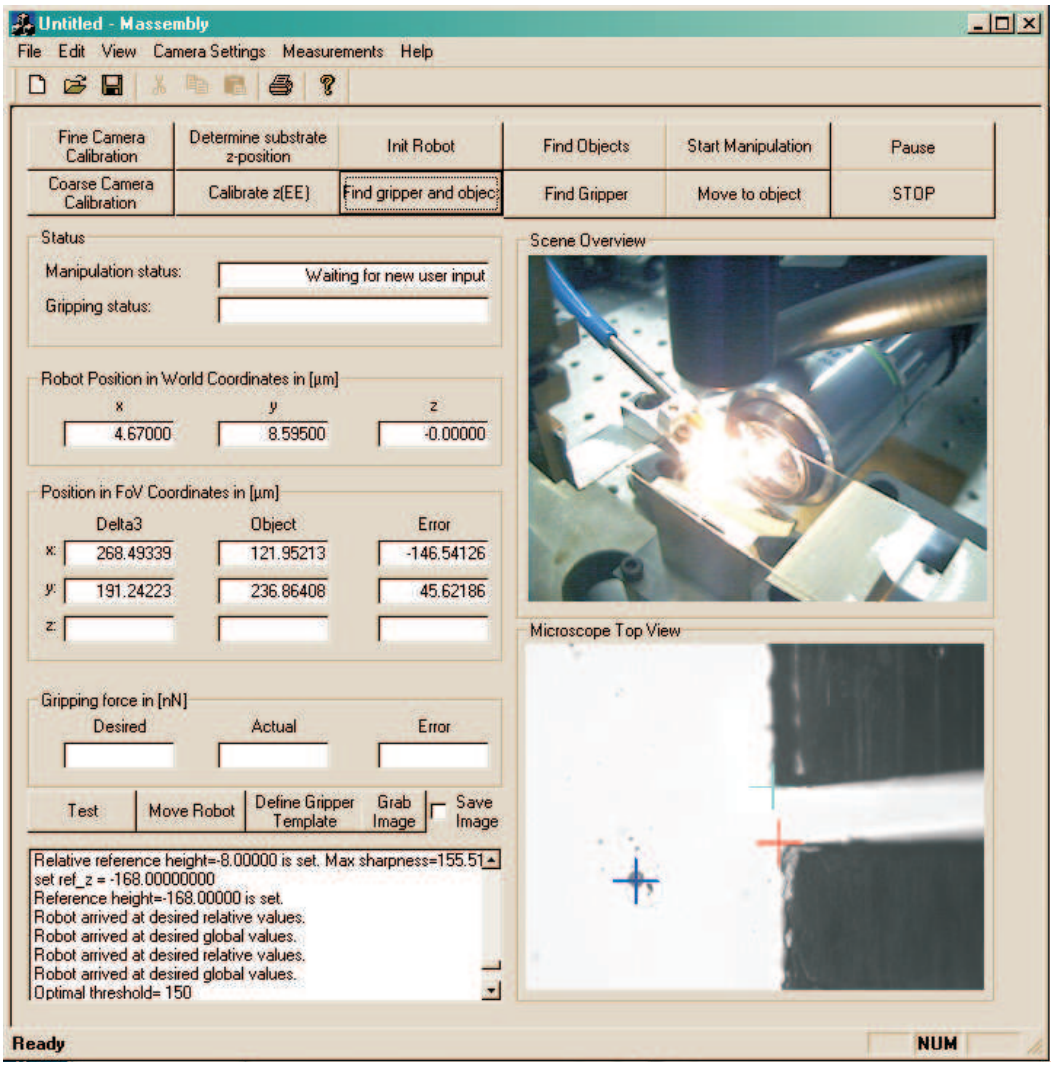

Figure 2.7: Graphical user interface to control and overview the manipulation.

Chapter 3

Machine vision

Visually-guided micromanipulation requires a complex sensor system. The main sensor employed is the high precision Sony XCD-X710 CCD camera that is mounted on top of the microscope. From this sensor lateral and height measurements are obtained. Certainly there are different sensors suitable for this task. A good comparison of different systems considering resolution, range, measurement speed and their limitations can be found in [1]. Another camera is used to obtain an overview of the whole scene. This chapter analyzes first the properties of the microscope vision system. To facilitate the implementation of the analysis of the obtained images two image processing libraries were compared and one of them selected. The process of transferring the image data from the cameras to the PC and making it accessible to the software is known as image acquisition. There is a short paragraph about this topic. The images are obtained in pixel coordinates. For positioning of the robot however it is necessary to translate positions from image coordinates into a physical scale. This mapping is done by the so-called camera calibration. The properties of the images were examined. The limited depth of focus of a microscope requires that the objects of interest appear sharp in the image. Therefore an autofocusing procedure is presented. Since the characteristics of the objects and the gripper are different, different techniques for their detection are considered. The description of the detection algorithm will conclude this chapter.

3.1 Microscope and cameras

The properties of the microscope and the employed camera is discussed in the following. Especially how the optical effects, mechanical properties of the system and the behavior of the CCD camera influence the sensor resolution.

In order to achieve a relative positioning reliability of less than $1\mu m$ a light microscope is used. Its advantages are high availability and short time delays (video rates of $30Hz$). With modern microscopes it is possible to have a resolution down to $0.2\mu m$ in terms of the Rayleigh criterion. An objective *M Plan Apo 20x* from Mitutoyo is employed.

Optical effects

- *Diffraction:* This wave optics phenomenon occurs when the light waves have to pass an aperture of similar dimensions as their wave length. Modern microscope lenses can be modelled as a linear shift-invariant system with a point spread function (PSD) followed by a constant gain including the objective magnification [16]. The PSD is a first-order Bessel function parameterized by the numerical aperture of the objective NA and the wave length of the illuminating light λ. The PSD has a bright region in the center surrounded by rings. The bright region in the middle is also know as Airy disk [16]. Two different points can be distinguished if the corresponding Airy disks have at least a displacement of Δx from each other. There are different definitions for thresholds indicating the maximal resolution Δx to separate two points in the object plane. Using the Rayleigh criterion it can be calculated that

$$\Delta x = 0.61 \frac{\lambda}{NA}, \tag{3.1}$$

  Furthermore, derived from wave optics the depth of focus can be written as

$$\Delta z = \frac{\lambda}{4n(1 - \sqrt{1 - (\frac{NA}{n})^2})}, \tag{3.2}$$

  where n is the index of refraction [16]. Figure 3.1 depicts the relationship between the numerical aperture NA and lateral resolution Δx as well as the effect on the depth of focus Δz with $n = 1.0$ and $\lambda = 555nm$. With increasing numerical aperture the lateral resolution improves and the depth of focus decreases. The results for objectives with different magnifications from Mitutoyo *M Plan Apo* series are displayed marked by circles.

  Considering the lateral resolution of the objective with the 20× magnification $\Delta x = 0.8\mu m$ and the dimensions of one pixel ($4.65 \times 4.65\mu m^2$) it can be seen that Δx corresponds to 3.44 pixels. Note that if lines or contours consisting of many points the resolution can be much higher.

- *Illumination:* There are several possibilities to illuminate the object plane. The two most common in microscopy are transillumination (light is coming from behind the object aligned with the optical axis of the objective) and epi-illumination (in-line illumination using a beam splitter to introduce the light along the path through the objective). The image brightness F is dependent on the numerical aperture of the objective NA and its magnification M. Using transillumination the image brightness has a proportional dependency on the numerical aperture and the magnification:

$$F_{transillumination} \propto \frac{NA^2}{M^2}. \tag{3.3}$$

  The relationship for epi-illumination is dependent to the fourth power of the numerical aperture and the magnification squared:

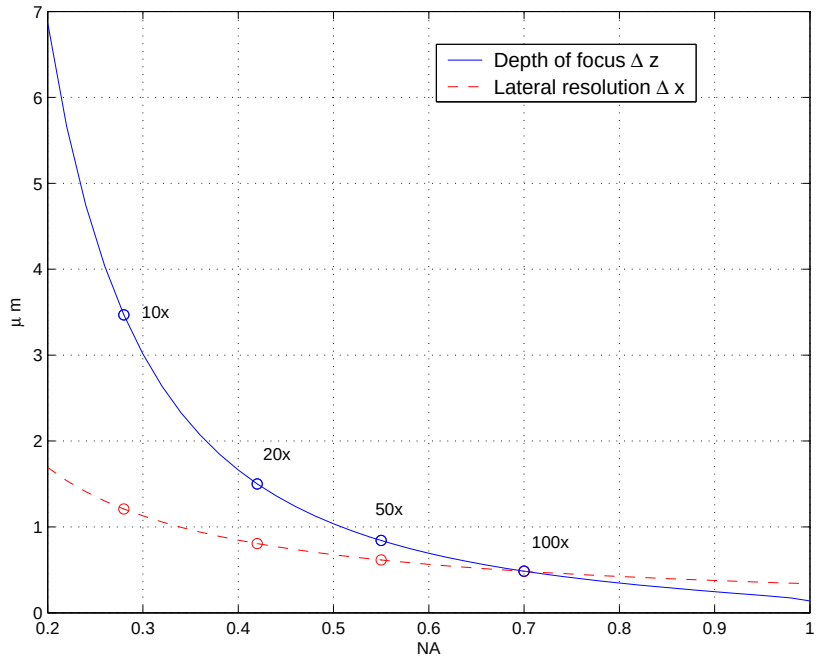

Figure 3.1: With increasing numerical aperture NA and a constant wave length of $\lambda = 555nm$ and $n = 1.0$ the lateral resolution Δx improves and the depth of focus (DoF) Δz decreases.

$$F_{epi-illumination} \propto \frac{NA^4}{M^2}. \tag{3.4}$$

For small magnification and corresponding numerical apertures the transillumination results in higher image brightness. Oil immersion objectives have $NA > 1.0$, so that in this case the epi-illumination leads to a better image brightness. In general, to maximize image brightness the maximal numerical aperture of the corresponding magnification should be chosen [14].

The choice of illumination is also dependent on the objects that appear in the image. Since reflecting objects, like the gripper, are of interest it is necessary to account for that. In order to avoid hot spots and blooming effects diffuse illumination should be used. Additionally, polarization filter in front of the video lens and in front of a light source can be put to reduce glare. The polarization axes of the polarizers should be perpendicularly orientated such that glare is minimized and an even illumination with high contrast is obtained. Depending on the impact of glare it could be sufficient to use only one polarizer. The disadvantage of the use of polarizers is the attenuation of the illumination by $60 - 65\%$ for each polarization [18].

There are different illumination techniques which are presented briefly.

- *Backlighting* can be used to show the silhouette of objects [18].
- *Directional lighting* can be employed such that the reflected beams do not go directly into the microscope.

– Shadow effects and high-lighted edges could be employed to gather more information about the object geometry [18].

– A directional light source could be mounted on the gripper illuminating its tip. Using a geometric model of the shadows, additional information to locate the object can be obtained. Instead of detecting the object itself the shadow is observed (compare fig. 3.2)

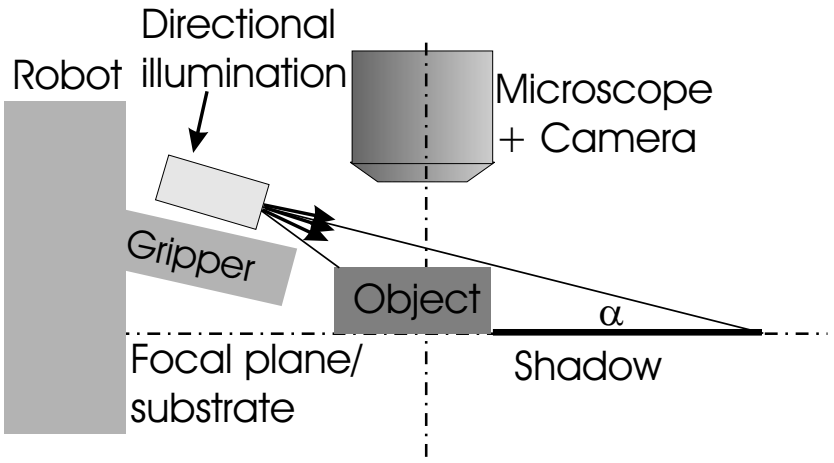

Figure 3.2: Using directional light mounted on the gripper to increase the lateral detection resolution of the object.

However this is only possible if the object geometry is known precisely. It is more interesting to use this light for the height calibration of the gripper since the shadow thrown by the gripper tip onto the substrate will be minimal when the tip is touching the substrate (compare fig. 3.3).

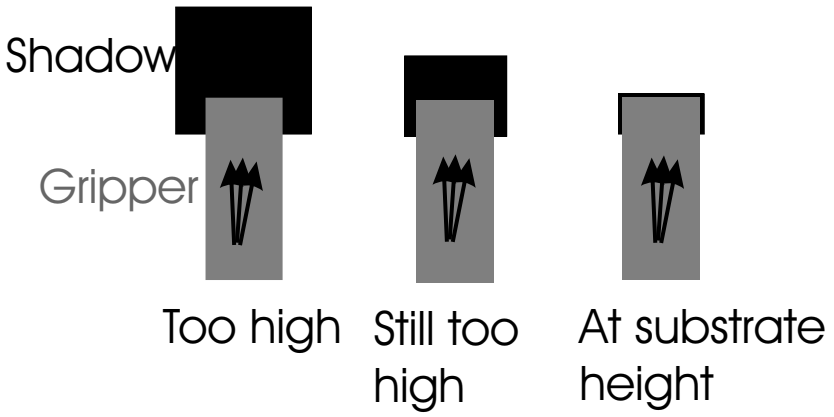

Figure 3.3: Top view onto the substrate. Using directional light mounted on the gripper, the shadow that is thrown by the gripper onto the substrate can be used to determine when the gripper is close to the substrate.

– As already seen in the case of the polarizer, filters can be used to change illumination conditions. In general, optical filter can be divided in two classes:

Filters attenuating light across most of the spectrum evenly and filters attenuating wavelength dependent. E.g. polarizers and neutral density filters, which are used to protect from overexposure or blooming, count to the first class. Examples for the wavelength-dependent filters are interference, dichroic and color filters. Wavelength-dependent filters can be used to generate monochromatic light in combination with a light source with a suitable spectrum. For instance using a tungsten lamp with its spectrum depicted in figure 3.4: only very little intensity for the violet color (wave length $\lambda \approx 350nm$) would be obtained when trying to filter out this wavelength whereas a high intensity could be obtained when filtering with a red filter ($\lambda \approx 630nm$).

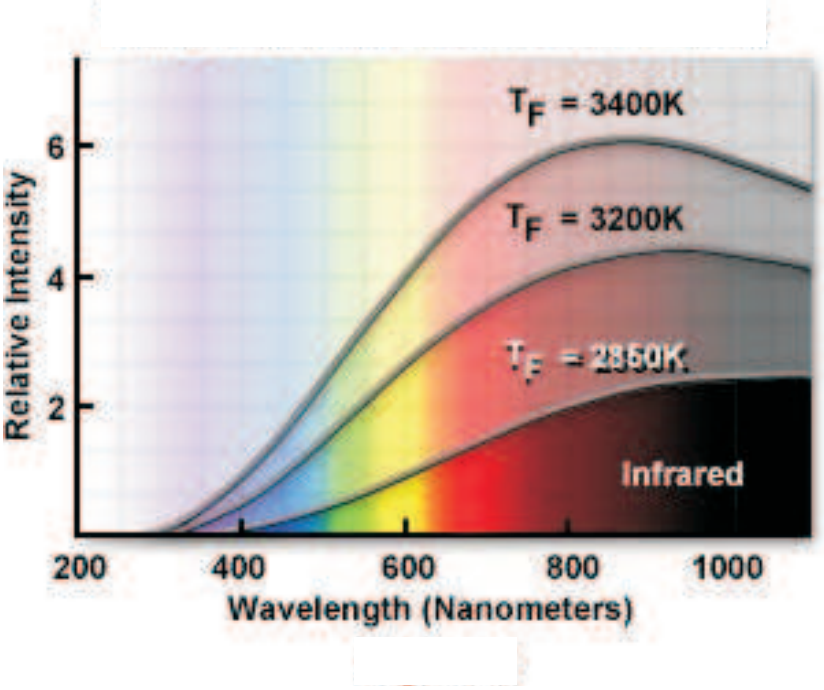

Figure 3.4: Tungsten lamp emission spectrum [12].

- *Optical aberrations:* The used objectives are of the category *M Plan Apo*. *Apo* stands for apochromatic and means that the objective is corrected for chromatic aberrations (colors: red, blue and yellow). *Plan* is a quality measure for the field flatness which means that 95% of the field of view appears flat when observed through the objective. There are two more levels of correction: Achromatic (65%) and semi-plan (80%). Figure 3.5 illustrates the difference between the different classes of field flatness.

  This flatness includes compensation for field curvature, astigmatism and distortions. Figure 3.6 illustrates the effect of field curvature. The elliptical specimen is projected onto the curved Petzval surface. By replacing this surface by a planar CCD camera, blurring effects close to the optical axis occur while having a focused image close to the borders and vice versa. These aberrations cannot be completely cancelled out, however their effects are negligible [13].

  This can be verified when using a regular test pattern covering the whole field of view.

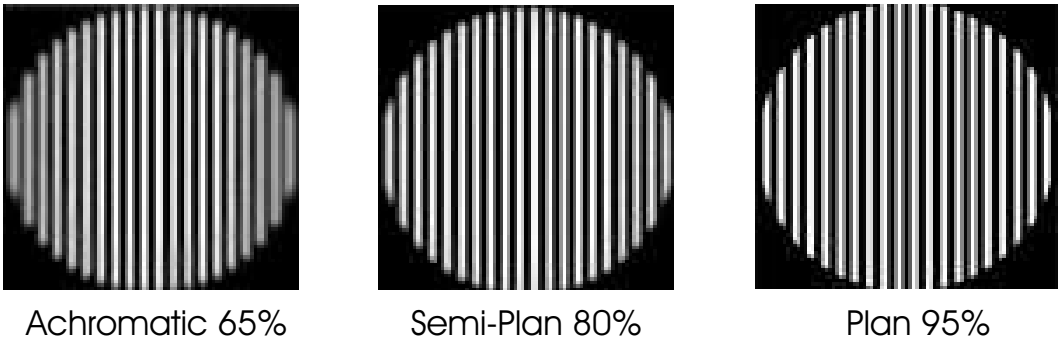

Figure 3.5: Image of a test pattern showing the effects of field flatness [18].

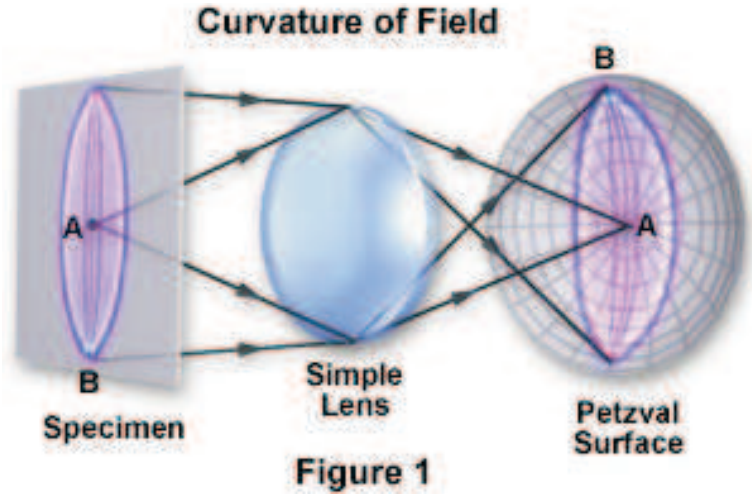

Figure 3.6: The curvature of field introduces blur in certain areas of the image [13].

In the first step the microscope camera calibration is done using the features close to the optical axis, and the calibration is tested estimating the position of the features close to the border of the field of view using some error measure. In the second step the features close to the optical axis are estimated based on a calibration matrix obtained by using features far away from the optical axis. The difference in error for these two cases can be interpreted as measure for the impact of aberrations on the image quality. However this test is not suitable to show the existence of astigmatism. Since a high-quality objective is used its effect is expected to be negligible.

Alternatively, the transformation of a grid test pattern can be observed and analyzed.

- *Statistical behavior of photons and the creation of photoelectrons:* The production of photons is governed by the laws of quantum physics and hence a statistical component is introduced. Modern CCD cameras can be so sensitive being able to detect single photons. These fluctuations cannot be removed and limit the overall signal-to-noise ratio (SNR). This behavior introduces noise into the image. The higher the radiometric intensity in Watt per meter squared the higher the variance of the photon numbers that arrive at the CCD chip [27]. The higher the variance the lower the SNR (compare

equation 3.13).

Mechanical effects

- *Assembly precision:* Limited assembly precision introduces small angular and linear displacements. The effects of linear displacements can be compensated through calibration. However, the calibration needs not to be extremely precise since the relative positioning of the gripper to the object is the main objective. The angular displacements have more impact on the system precision. Assuming there is a tilt angle $\Delta\Theta$ between the focal plane of the microscope and the workspace/microscope table, objects that are on the table at the limit of the field of view at $x = \frac{w}{2}$ will become blurry being observed through the microscope when the displacement to the focal plane becomes greater than half the depth focus Δz (compare fig. 3.7).

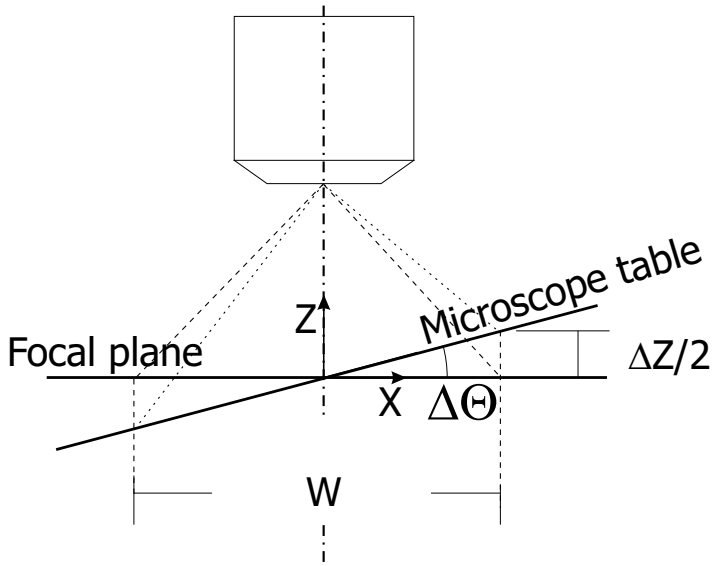

Figure 3.7: The microscope table is tilted by $\Delta\Theta$ with respect to the focal plane of the microscope.

If w the width of the field of view and an objective with magnification M is used the following expression can be derived using a trigonometric relationship:

$$tan(\Delta\Theta) = \frac{\frac{\Delta z}{2}}{\frac{w}{2}}. \tag{3.5}$$

Since only small angular displacements are considered the small-angle approximation can be applied.

If $\tilde{w}$ is the width of the CCD sensor ($4.8mm$), the width of the field of view can be calculated using the magnification of the objective:

$$w = \frac{\tilde{w}}{M} \tag{3.6}$$

Inserting this expression into equation 3.5, a maximum angular displacement

$$\Delta\Theta_{max} = \frac{\Delta z \; M}{\tilde{w}} \tag{3.7}$$

in which the whole field of view is perceived sharp in the microscope image can be formulated. Table 3.1 shows the maximum allowed angular displacements for Mitutoyo M Plan Apo objectives for different magnifications. For the 20× magnification objective the highest precision of 0.38^o is needed.

M	Δz **in** μm	$\Delta\Theta_{max}$ **in** ^{o}
10x	3.5	0.42
20x	1.6	0.38
50x	0.9	0.54
100x	0.6	0.72

Table 3.1: $\Delta\Theta_{max}$ in ^{o} for Mitutoyo M Plan Apo objectives for different magnifications.

It is assumed that the alignment of the CCD camera with the objective is much more precise compared to the alignment of the objective relative to the microscope table so that the effects of a misalignment between the camera and the objective are expected to be negligible.

CCD camera effects

- *Spectral characteristics:* CCD sensors are normally more sensitive to longer wavelengths (maximum $\approx 800nm$) [27]. However, long wavelengths reduce the lateral resolution proportionally (compare equation 3.1). This results in a trade-off between the sensitivity of the camera, respectively the signal-to-noise ratio (SNR), and the lateral resolution. For intermediate magnification (10x, 20x) the illumination is expected to be sufficient to obtain a good SNR.
- *Image digitization:* The microscope image is spatially sampled. Depending on the Rayleigh criteria 3.1, a suitable magnification of the objective should be chosen to fulfill the Nyquist criteria to sample the images. If the magnification is chosen properly the introduced error is negligible. The intensity is usually quantized using 8 or 10bit introducing granular noise. This quantization noise can be neglected compared to other sources of error [27]. The theory about image digitization is well described in [22].
- *Exposure time and motion blur:* A long exposure time T_e enables the CCD sensor to gather more light and to produce a higher SNR. However, if features in the scene are moving they appear blurry in the image which is called motion blur. Using this blur it is possible to determine the velocity of an object or if the velocity is known, the position can be determined well. Certainly, a short exposure time reduces motion blur [27].
- *Sensitivity:* Sensitivity is the DC gain of the incident illumination. Cameras provide very often automatic gain control (AGC) to control the average gray-level of the images. For little illumination, respectively low SNR the gain is very high and the weak signal is amplified heavily but the noise as well [27].

- *Dark current and noise:* Dark current is generated by random electron-hole pairs generated thermally or by tunnelling affecting the sensor output. Using a reference on the chip the average dark current can be determined and its effects are compensated within the camera system. Dark current is one source of noise. Others are the statistical behavior of electrons, cross-talk between pixel, the quantum efficiency of the sensor, sensor amplification and readout noise. More details can be found in [27] and [16].

Mounting of the microscope

Two options were developed to mount the microscope in the experimental setup (see also fig. 3.8): 1. The microscope is mounted above the workspace and the optical axis is perpendicular to the workspace looking down onto the microscope table. The setup in figure 2.4 shows this alternative. 2. The microscope is mounted horizontally below the workspace. The optical axis is parallel to the workspace. A beam splitter or a mirror with an angle of 45^o towards the workspace redirects the optical axis to be perpendicular to the workspace so that the microscope table is observed from below.

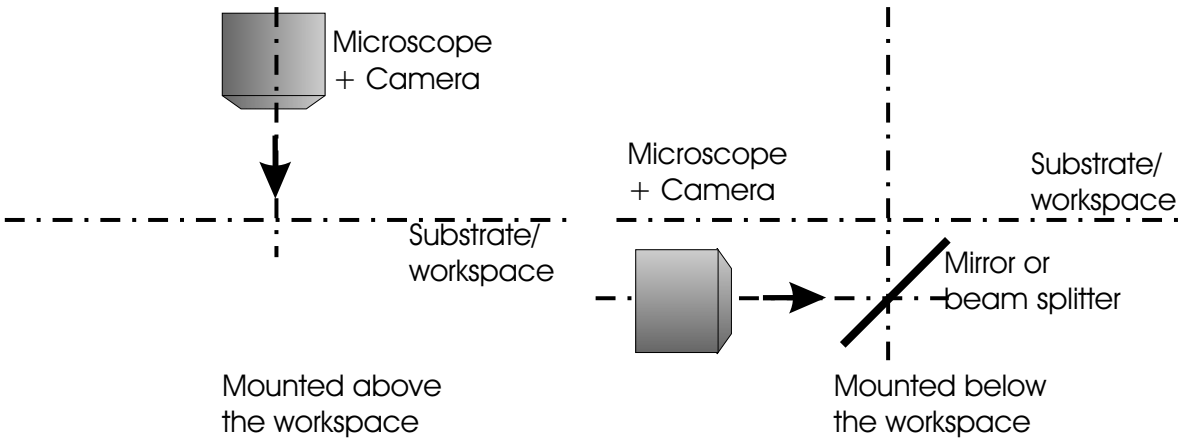

Figure 3.8: Two alternatives to mount the microscope.

The following table 3.2 will discuss the advantages and disadvantages of both approaches assuming that the application is to manipulate an object, which lies on a substrate, with the two-jaw gripper.

Criteria	Mounted above workspace	Mounted below workspace
Alignment of workspace with optical axis	+ As described above the tilt angle $\Delta\Theta$ has to be considered	o the angular position of the beam splitter needs the double precision because of the law of reflection; If high precision to position the beam-splitter is available it could be possible to compensate $\Delta\Theta$
Occlusion/ Blur effects	- the blur and shadow of the upper edge of the gripper can affect the resolution to detect an object; - the location of the lower edge of the gripper can only be determined indirectly (see also fig. 3.9)	+ the object and the lower edge of the gripper can be localized in the same focal plane; shadow can occlude the objects partially (compare fig. 3.43)
Mechanical interferences	- the workspace is shared by the microscope, the manipulator and the objects restricting the possibilities to employ the manipulator; the workspace becomes more limited for objectives with high magnification since their working distance is less	+ the objective does not limit the workspace; - for high magnification the working distance of the objectives is too short in order to re-direct the optical path
Sensitivity to vibrations	- Since the microscope is suspended rather far above the base plate, a very high stiffness is required for the frame connecting the microscope with the base plate	+ The microscope is mounted directly to the base plate leading to a minimum of vibrations
Miscellaneous		- Object height cannot be determined; - Microscope table must be transparent

Table 3.2: Advantages and disadvantages of the different mounts of the microscope relative to the workspace.

Comparing the two configurations both alternatives have advantages and disadvantages: The main disadvantages are that alternative 2 cannot be employed in the presented setup for magnifications greater than 20×, that it is more sensitive to misalignment between the optical axis and the workspace and that the substrate has to be transparent. The main disadvantages of alternative 1 are: The indirect localization of the lower edge, the introduced blur reducing the performance of the object detection. The occlusion of objects can be avoided by first examining a scene and moving the manipulator afterwards such that the object of interest is not occluded. The optimal configuration is application-dependent:

- *From above:* The substrate is opaque and/or higher magnifications are required. Note that the results of the camera calibration (see section 3.4) from alternative 2 can be used here as well[1].
- *From below:* The resolution of the 20× objective is sufficient and a transparent substrate is employed. Also, if an opaque substrate is used and objects that stick to it are manipulated from below.

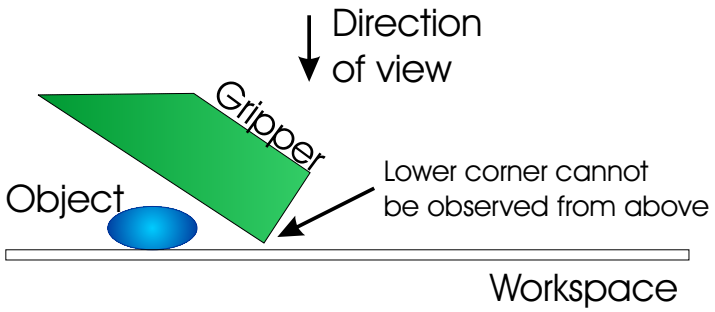

Figure 3.9: The microscope view from above the workspace cannot detect the position of the lower end of the gripper. Objects can be occluded by the gripper.

Summarizing the presented effects the parameters with high impact on the resolution are:

- the wavelength of the used light λ,
- the numerical aperture NA, the magnification M,
- the index of refraction n,
- the brightness of the image F,
- the exposure time T_e,
- the mounting of the microscope

NA, n and M are properties of the objective. High numerical aperture leads to high lateral resolution, less depth of focus and better illumination. A small depth of focus has the advantage that it can improve the autofocus results such that the resolution to determine object heights can be in the range of the wave length [9].

With decreasing wave length of the illuminating light λ the lateral resolution improves and the depth of focus increases proportionally. Assuming there is enough illumination short wave

[1] This requires that the substrate is removed out of the optical path when calibrating.

lengths combined with a short exposure time lead to better results. If the illumination is not sufficient another wave length should be considered that is optimal for the spectral sensitivity of the CCD camera.

The exposure time can be increased as well. However if there is relative motion between the camera and objects in the scene this will introduce motion blur.

If possible the microscope should be mounted horizontally looking on the substrate from below.

3.2 Image processing libraries

Image processing is the treatment of images with algorithms. The number of algorithms that can be applied to an image is vast and their implementation can be very easy (e.g. setting all pixels which have a certain intensity to zero) or very complex (e.g. a pyramidal segmentation algorithm described in [22]). Since many of the used functions are standard functions they are already implemented and collected in libraries. There are more than 100 commercial as well as free libraries available. However there are few libraries that are free and provide execution-speed optimized image processing functions, image acquisition tools and visualization. Therefore only Intel's OpenCV and the TLib (Virtual Reality and Active Interfaces (VRAI) of the Swiss Federal Institute of Technology) were considered to be used. Besides their optimized functions, both libraries can run either on Windows or Linux. The TLib is written such that it is easy to port to other operating systems.

OpenCV is used by a big community ([24]) and provides more than 300 image processing functions that are on middle-to-high level including also complex functions like e.g. pyramid linking. The used data structures are however very complex which complicates the use of the provided functions. Additionally, mathematical routines, especially linear algebra, are included. TLib is mostly designed to act as an easy-to-use introduction to computer vision especially for students which usually work on short-term projects and hence have not enough time to study a complex API like OpenCV.

TLib	**Intel's OpenCV**
Basic image processing functions	$>$ 300 functions including many mid- and high-level routines
Easy-to-use API	Complex API with complex data structures
	Includes mathematical functions (e.g. linear least-squares problems)

Table 3.3: Properties of the TLib and Intel's OpenCV libraries.

TLib provides the most important basic routines for image processing which are implemented as optimized low-level functions that are wrapped in C++ classes. The relative low number of mid-level functions and an easy data structure enable novice programmers to be proficient with this library after 1-2 hours of use [21]. Table 3.3 summarizes the properties of the two mentioned libraries. The main reason to select OpenCV for this work is its much greater

number of functions and options and integrated mathematical routines. It should be noted that during the work with OpenCV it turned out that many functions are very specialized towards certain applications or important variables of some methods were not accessible (e.g. the number of votes of a line for the Hough transform, see also section 3.7.2).

3.3 Image acquisition

The image processing is realized virtually completely in software. In order to process the images generated by the cameras it is necessary that the used personal computer can access them. The overview camera uses the USB 2.0 interface whereas the microscope camera uses the IEEE FireWire1394 interface to connect to the PC. In order to use the overview camera in the image processing application, high-level functions provided by the image processing library OpenCV are used. Depending which of these functions are used either Video for Windows (VfW) or DirectX routines are called by the library. The USB camera driver provides an interface which is compatible with VfW and DirectX (see also fig. 3.10). Although the DirectX functions should be preferred because they are better supported by Microsoft the VfW functionality was chosen. The main reasons are that the usage of the corresponding functions for VfW in OpenCV are easy to employ, its robustness and its easy access to the camera parameters (e.g. switching auto-gain control on/off). In addition, the OpenCV high-level functions for the DirectX did not work as reliable as the ones for VfW. In order to obtain good results with DirectX it is necessary to program the image acquisition on the DirectX-level. This requires programming in DirectShow which is part of DirectX which is a complex interface for handling video streams. Finally, the image acquisition of the overview camera is not crucial for the high-precision positioning and so the easy-to-use VfW functions were selected.

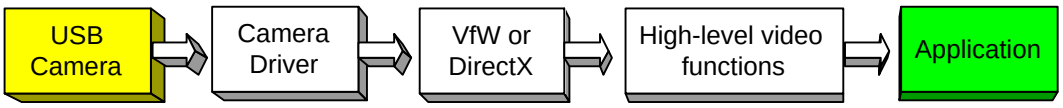

Figure 3.10: Image acquisition for the coarse camera (USB2.0) in the Windows environment.

The high-resolution camera uses the FireWire interface to connect to the PC. The FireWire specifications allow transfer rates up to 400 MBit/s which allow video rate (30Hz) transmission at maximum resolution of 1024×768 pixel. The same acquisition as for the USB camera is possible. However the default Windows driver did not work with the camera. A driver distributed by *The Imaging Source* was tested as trial version delivering good results for the DirectX functionality. However this driver is not free-ware. An alternative to access a FireWire camera is to use the driver provided for free by the Robotics Institute of the Carnegie Mellon University [19]. This driver uses the functions of the FireWire card driver and provides an API to access images and to camera controls (compare fig. 3.11). Including necessary conversions it takes about $85ms$ to grab and convert an image. This enables only a very low refresh rate of $11fps$ and is mainly due to the high resolution. Using the partial scanning option of the Sony camera, up to $80fps$ can be obtained for a small ROI. However, in order to change the video options the video stream has to be stopped and re-started.

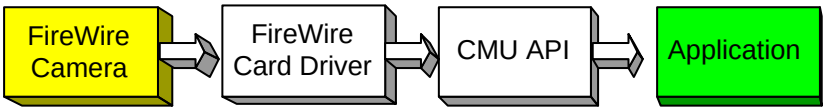

Figure 3.11: Image acquisition for the Sony FireWire camera.

3.4 Camera calibration

3.4.1 Introduction

Camera calibration is used to map image pixels to physical units and to align the image coordinate system with the field-of-view (FoV) coordinate system. Camera calibration is required to measure the repeatability of the manipulations and to control the gripper properly whose axes of motion are aligned with the FoV coordinate system.

Picture 3.12 depicts the definition of different coordinate systems. The image coordinate frame is attached to the axes in the image as defined in section 1.3. The real-world coordinate frame is aligned along the axes of the robot and is notated by the index w. The origin of this frame is defined by the zero position of the robot (*note:* this frame is not needed for the camera calibration, but is defined here for completeness). The field of view (FoV) frame is aligned with the world-coordinates frame, however the origin of this frame is defined by the upper left corner of the field of view of the objective for x_f and y_f and by the position of the focal plane for z_f.

For calibration a camera model is needed. This model is described by extrinsic and intrinsic parameters. Extrinsic parameters include the position and orientation of the imaging system and intrinsic parameters are properties of the optical system (e. g. focal length, pixel dimensions, etc.) [17].

Three camera models can be used to represent the mapping from a 3D world space to the 2D image space: Perspective projection, affine projection and scaled orthographic projection. Since in microscopy only one plane is considered and its distance to the camera is much greater than the relative depth of the imaged points, scaled orthographic projection is used [1].

In this work the numerical estimation method as proposed in [1] is used. The lateral components of the FoV coordinates x_f and y_f are mapped to the image coordinates accordingly to

$$\begin{pmatrix} \mathbf{x}_i \\ 1 \end{pmatrix} = \mathbf{T} \begin{pmatrix} x_f \\ y_f \\ 1 \end{pmatrix}, \tag{3.8}$$

where $\mathbf{T} \in \mathbb{R}^{3\times3}$ is the calibration matrix. $\mathbf{T}$ is estimated from a set of known points in FoV coordinates and the corresponding points in image coordinates using least-squares. Since $\mathbf{T}$ has 9 elements at least nine equations have to be used. Each point corresponds to three measurements so that at least three points with sufficient distance to each other have to be measured. $\mathbf{T}$ contains information about the angular displacement along the z_f axis, the scaling as well as the displacement in the $x-$ and the $y-$axis.

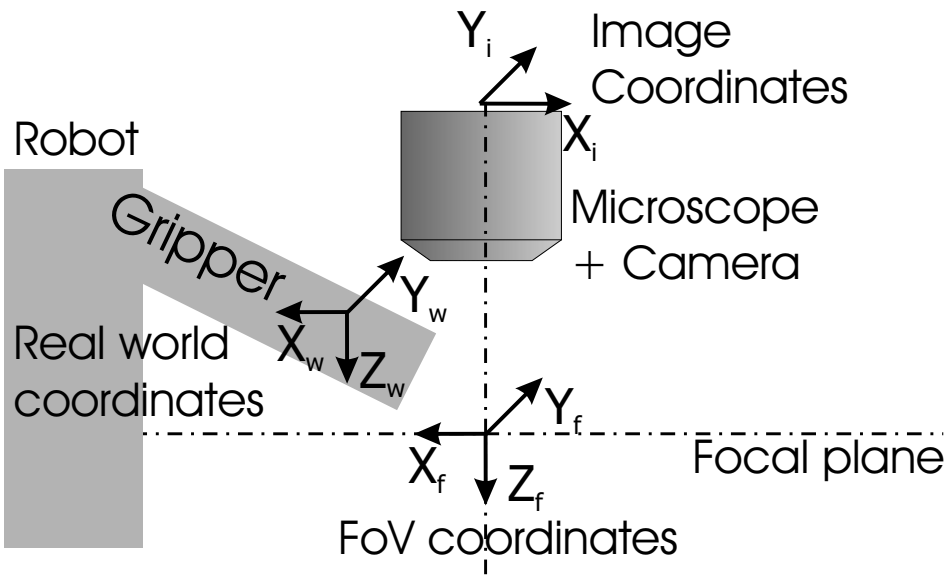

Figure 3.12: Definition of the image coordinate system, FoV coordinate system and the real world coordinates.

3.4.2 Choice of calibration pattern

The used camera calibration method requires a reference pattern whose geometry is known. An image of this calibration pattern is compared with the known geometry of the pattern. The author of [1] used one circle covering most of the field of view assuming that the image plane coordinates are aligned well enough with the FoV coordinates.

In [17] a set of 7×7 separated squares is employed. Grids are used in camera calibration as well. Comparing grid lines with squares, squares seem to be more robust to detect properly than lines, which might not have enough contrast or could be easier mixed up with environmental disturbances like dust or dirt.

Since the whole image plane should be mapped the calibration pattern should cover most of the field of view. Therefore for every magnification a different pattern is needed.

The calibration pattern needs high precision. Looking at the lateral resolution of the objective, the manufacturing precision should be at least as high. For lower manufacturing precision it is necessary to use more pixel for each square. It should be noted that only the pixels at the border of the square give information about its position. 50×50 pixel ($2 \times 50 + 2 \times 48 = 196$ pixel at the border of a square) seems to provide enough image points for a good detection of the position of the square. A calibration pattern for this setup is proposed in figure 3.13. The amount of pixels of the CCD and the square size results in 19×15 squares.

Another alternative is to use the robot itself for calibration. Since it is very desirable that the lateral coordinates of the image are aligned with the ones of the robot it makes sense to use the features of the gripper. Taking several images while moving the gripper in the field of view with known displacements results in a calibration pattern. The stored images are processed and the position of the gripper detected so that the required knowledge about the position

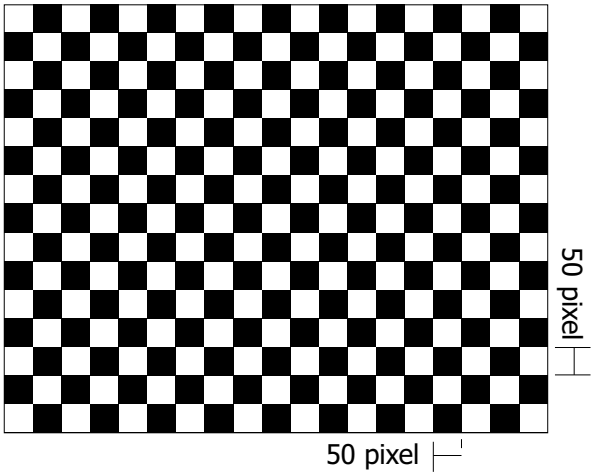

Figure 3.13: Proposed calibration pattern.

in both coordinate systems is known. The displacements in the field of view are not big and only relative displacements are needed for calibration. This means that the robots precision and the performance of the gripper detection limits the precision of the calibration. This technique has two big advantages: Firstly, no calibration pattern needs to be manufactured. Secondly, the calibration is along the axes of the robot and hence no alignment of a calibration pattern on the microscope table relative to the robot is needed.

3.4.3 Discussion of calibration results

The calibration of the microscope camera is done based on estimated gripper positions as described above. The correlation as described in section 3.7.2 is used for the detection. 7×7 calibration points distributed as a grid are taken. The grid covers about 90% of the FoV. Note that using the correlation for the gripper detection the grid cannot cover the whole FoV because the template must be always completely visible. Let $\mathbf{x}_f = [x_f, y_f]$, then the estimation error can be calculated for each calibration point by taking the difference between the known position and the estimated position in FoV coordinates $\hat{\mathbf{x}}_f$:

$$e = |\hat{\mathbf{x}}_f - \mathbf{x}_f| \tag{3.9}$$

Multiplying equation 3.8 with $\mathbf{T}^{-1}$ from left[2], $\hat{\mathbf{x}}_f$ can be written as

$$\hat{\mathbf{x}}_f = \mathbf{T}^{-1}\mathbf{x}_i. \tag{3.10}$$

Inserting equation 3.10 in equation 3.9, the error can be expressed as

$$e = |\mathbf{T}^{-1}\mathbf{x}_i - \mathbf{x}_f| \tag{3.11}$$

[2]The inverse of $\mathbf{T} = \begin{bmatrix} s_x & 0 & 0 \\ 0 & s_y & 0 \\ 0 & 0 & 1 \end{bmatrix} \mathbf{A}$ exists always since matrix $\mathbf{A}$ represents a homogenous 2D transformation which is multiplied by a diagonal matrix to obtain $\mathbf{T}$.

The calibration determined for the case with an overall magnification of 10× and the microscope looking from below the substrate to

$$\mathbf{T} = \begin{bmatrix} 2.1667 & -0.0771 & 651.8685 \\ 0.1021 & 2.1678 & 425.9150 \\ 0 & 0 & 1 \end{bmatrix} \frac{pixel}{\mu m}. \tag{3.12}$$

The diagonal terms are identical to two digits after the comma which is expected since the CCD pixels are squares. The off-diagonal elements are not completely zero indicating that there is a small angle between the FoV and the image coordinate systems.

The calibration errors are presented in figure 3.14. At each grid point the calibration error e is evaluated. The *maximum error* is $0.569\mu m$ and the *mean error* is $0.241\mu m$ with a standard deviation of $0.119\mu m$. Comparing with [17][3] and [1] who present for similar $\frac{\mu m}{pixel}$ of their vision systems a maximum error of $1.742\mu m$ and $4\mu m$, respectively. The calibration precision mainly limited by the the repeatability to detect the gripper (compare section 3.7.2) and the quality of the objective.

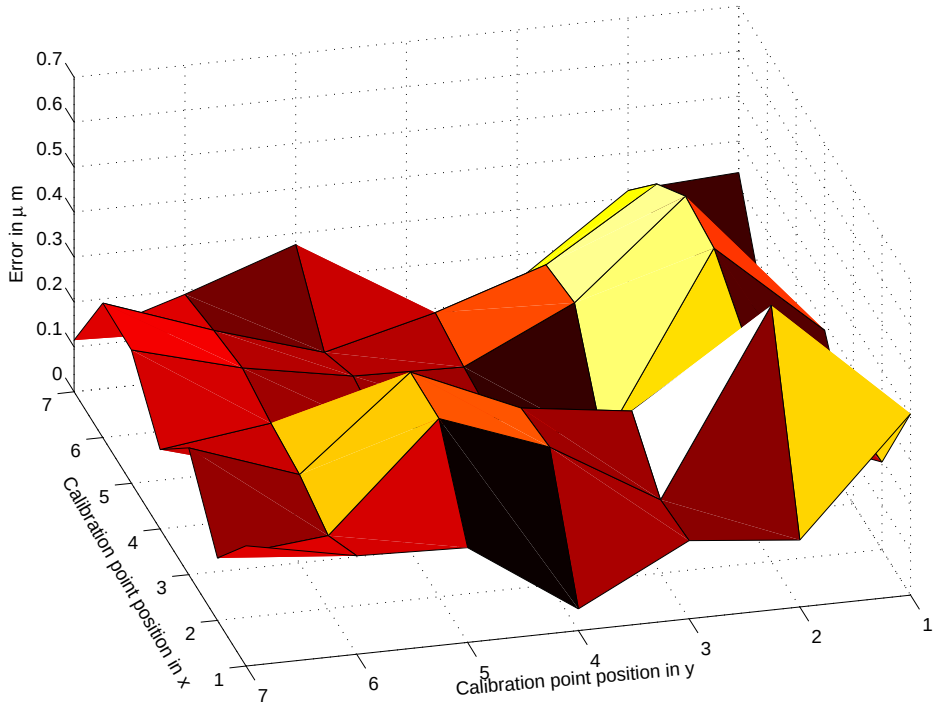

Figure 3.14: Calibration error e at each calibration point (x, y) in the grid.

3.5 Image properties

The captured images from the video stream are examined in this section. The properties of the images are discussed.

[3]This estimate includes also an error in the height component.

An image can be divided into background and foreground where the foreground is defined by the regions containing the objects of interest (*here:* gripper tip and objects to be manipulated). The background is made up of the image regions that are not foreground. There are several disturbances in the image:

- Noise from the CCD camera: This is often expressed using the signal-to-noise ratio in dB
$$\mathrm{SNR} = 20 \log_{10} \frac{max(I(\mathbf{x}_i)) - min(I(\mathbf{x}_i))}{\sigma_b}, \tag{3.13}$$
where σ_b is the standard deviation evaluated in a region of one image that contains only background [23]. The SNR was measured to be approximately $60dB$. Hence the noise within one grabbed image is not high. Furthermore the image contains no *salt&pepper*-noise which is characterized by single pixels that contain a completely wrong value. This happens for example if a CCD camera pixel is defect. Since the deviation of the intensity value of such pixels differs heavily from the average intensity of the background $\mathcal{E}\{I(\mathbf{x}_i)\}$ these pixels can be found when an image containing only the background is considered: In order to distinguish between the gaussian noise and the salt&pepper noise, it is assumed that the pixels that are not within a $\pm 3\sigma_b$ interval around $\mathcal{E}\{I(\mathbf{x}_i)\}$ are effects due to salt&pepper noise.

  The high SNR and no salt&pepper noise does not require special care of these effects. In general, Gaussian noise can be reduced by linear filters (e.g. Gaussian filter). Salt&pepper noise can be eliminated using a Median filter or a morphological opening (e.g. mask size 3×3).

- Besides the objects of interest, scratches in the substrate or dirt can be misinterpreted to be part of the foreground.

- Variations in lighting: Illumination of a scene can vary spatially and over time. Using backlighting, the illumination of the field of view is homogenous which can be seen in the histogram 3.25 indicated by the cluster of pixels around the intensity value of 150 representing the mean background value.

Since some object detection algorithms assume constant illumination it is important to examine this property. A scene was selected that contains only background and with 15 frames per second (fps) a video sequence containing $N_f = 500$ frames was considered. Since only the fluctuations of the light are subject of this measurement, the substrate was removed so that the results were not influenced by it. It should be noted, that the automatic gain control was switched off.

Figure 3.15 shows for the sequence of frames the average intensity value for each frame $\bar{I}_i = \mathcal{E}\{I(\mathbf{x}, t = t_i)\}, i = 0, .., N_f - 1$. If these average intensities are also averaged over time $\bar{I} = \mathcal{E}\{I(\mathbf{x}, t\}$ the mean intensity value of 165.8 is obtained. The standard deviation spatially within one frame i is $\sigma_i = \sqrt{\mathcal{V}\{I(\mathbf{x}, t = t_i)\}} \approx 4.7$ for all i. The standard deviation of $\bar{I}_i$ over time, respectively over all i is: $\sqrt{\mathcal{V}\{\bar{I}_i\}} = 1.8$. This means that the spatial inhomogeneities are 2.6 times greater than the temporal fluctuations. These temporal and spatial fluctuations in light are not very high. The effects of changes in illumination on the object and gripper detection will be discussed later in this chapter. Methods to minimize these fluctuations are presented below.

One standard approach to improve the spatial homogeneity is to store an image showing only the background. Dividing all images by the background improves the spatial homogeneity.

One approach to reduce the temporal variations in illumination is to average several images over time. However, this has the disadvantage that the window in time has to be chosen long enough to obtain constant illumination (about 5 frames). This is similar to increase the exposure time, respectively the shutter time of the camera. That means that the image intensity values are smoothed but also motion blur is introduced.

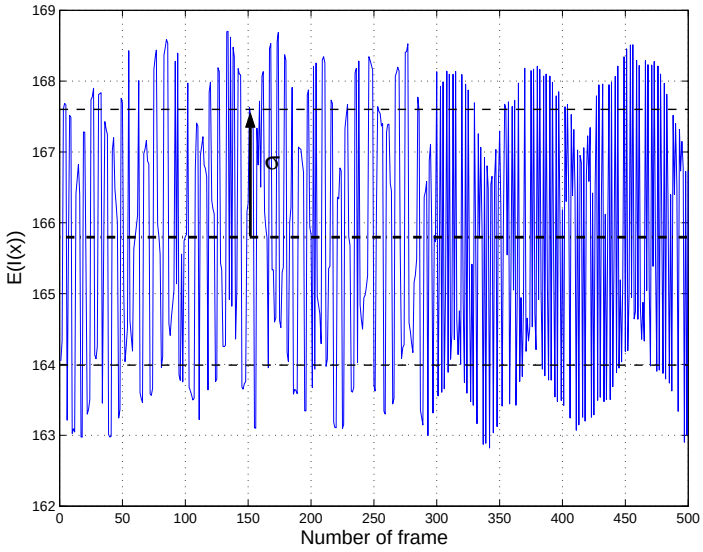

Figure 3.15: Average background intensity recorded over an interval of $N_f = 500$ frames. Average of the spatial average over time is 165.8 and the standard variation 1.8.

3.6 Autofocus

The principle of autofocussing is the maximization of the image sharpness. Moving the object plane relative to the focal plane the plane has the maximal sharpness if the two planes coincidence. Sharpness can be also interpreted as images with a high content of high spatial frequencies. In [3] different methods for autofocussing are mentioned and the Tenengrad criterion was selected because of its high robustness and functional accuracy. In this method first the image gradient magnitude

$$\|\nabla I(\mathbf{x})\| = \sqrt{(I_x(\mathbf{x})^2 + I_y(\mathbf{x})^2)} \tag{3.14}$$

is determined, where $I_x(\mathbf{x})$ and $I_y(\mathbf{x})$ are the spatial derivatives along the $x-$ and the $y-$axis. Summing up all gradients that are greater than a threshold T_a the Tenengrad criterion is defined as

$$C(z) = \sum_{\mathbf{x}\in\mathcal{I}} \|\nabla I(\mathbf{x})\|^2, \text{for} \|\nabla I(\mathbf{x})\| > T_a. \tag{3.15}$$

The authors in [3] propose to set T_a to 75% of the maximum gradient magnitude of the image taken in focus. Here a different approach was developed to determine T_a: An image is grabbed in focus (compare fig. 3.16) and the histogram of the image gradient magnitude (fig. 3.17) is considered when choosing the threshold. Figure 3.18 shows such a histogram for the surface of the glass substrate. The high gradient magnitudes belong to the imperfections in the glass surface (lines and little dots) and dirt on the lens[4] (circles) in the image. The very small magnitudes correspond to noise. The threshold is selected that this noise is filtered out and the threshold was set to $T_a = 25$. It should be noted that for processing reasons the magnitude gradients were scaled to a range from 0 to 255 so that it could be represented as a 8-bit bitmap.

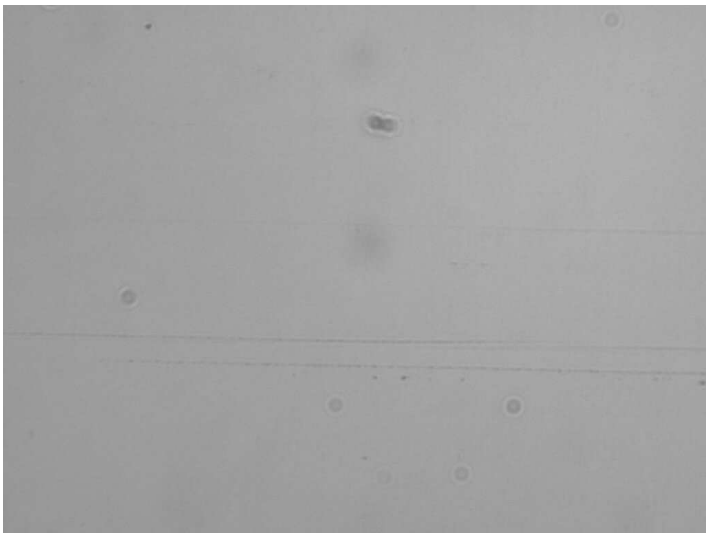

Figure 3.16: Image of the substrate in focus.

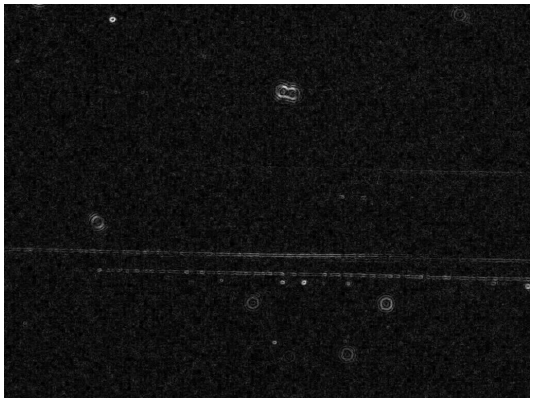

Figure 3.17: Normalized image gradient magnitude of substrate in focus.

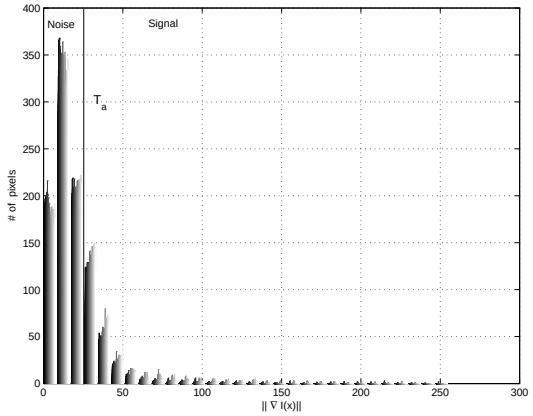

Figure 3.18: Histogram of the gradient magnitudes of image 3.17.

The automatic detection of the glass substrate surface was evaluated: The substrate was

[4]the dirt on the lens does not affect the autofocus procedure, because they do not change appearance in the image when the substrate position is changed.

mounted on a high-precision positioning stage which is actuated by a piezo-electric stick-slip drive and its position resolution is 100nm (sensor resolution).

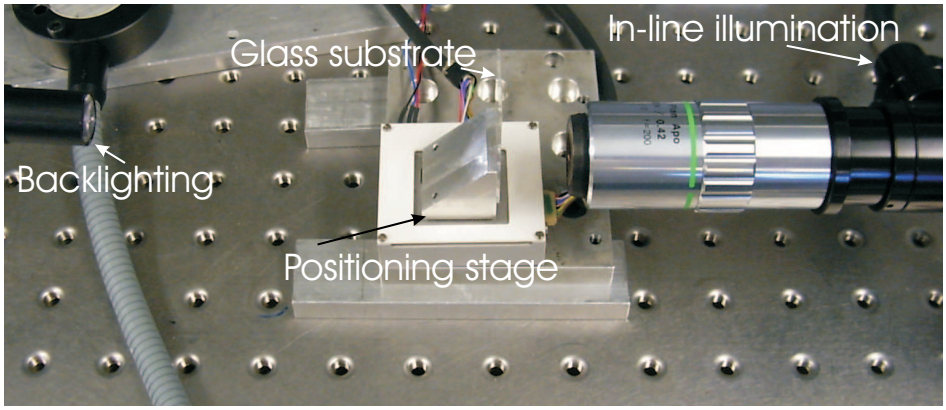

Figure 3.19: Substrate mounted on a high-precision stage.

First, the surface was brought into focus manually. The stage was then moved in a range of $\pm 30\mu m$ in steps of $1\mu m$ along the optical axis. At each step the Tenengrad criterion was calculated, respectively the contrast. Figure 3.20 presents the obtained experimental results. The measurements were repeated at three times. Each time starting with a different relative position to the focal plane, so that a shift in the three curves can be seen. In each experiment the maximum value was determined and an image was taken at the estimated substrate height which looked for all experiments as the one shown in figure 3.16. The shape of the measurements is similar for the three experiments. The absolute value are different because of changes in illumination between the experiments. Image 3.21 presents the results for the same experiment with the difference that the range was reduced to $\pm 3\mu m$ and the step size to $100nm$. The measurements contain noise which is due to fluctuations in the illumination. Looking at the broadest peak at $z = -1.5\mu m$ the repeatability of the detection of the position of the surface can be estimated as $\pm 0.5\mu m$.

The tilt of the substrate with respect to the optical axis can also be determined. This becomes important when the manipulator moves over a longer range (compare figure 3.22).

If the tilt is high, the security distance of the lower end of the end-effector would have to be increased. A solution to this problem is the detection of the substrate at different positions to describe the substrate plane Π relative to the focal plane. Since the substrate can be focussed in the FoV it is only necessary to consider the tilt angles η and ξ along the axes of the focal plane coordinate system (see figure 3.23).

At three points with know lateral coordinates $P_i = (x_{f_i}, y_{f_i}, z_{f_i})^T, i = 1, 2, 3$ the substrate height is detected. The points are chosen along the axes of the focal plane coordinate system, so that the angles can be calculated using a trigonometric relationship:

$$\begin{aligned} \xi &= \arctan(\tfrac{z_{f_1} - z_{f_2}}{x_{f_1} - x_{f_2}}) \\ \eta &= \arctan(\tfrac{z_{f_3} - z_{f_2}}{y_{f_3} - y_{f_2}}) \end{aligned} \tag{3.16}$$

Once these angles are known, the relative height between the substrate and the end-effector can be corrected when there is a relative lateral motion.

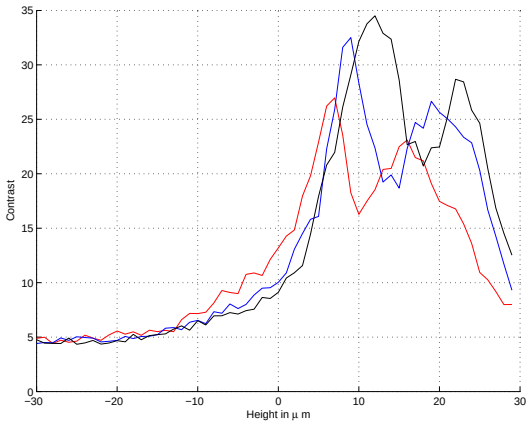

Figure 3.20: Contrast $C(z)$ over the positions of the substrate along the optical axis for three experiments (step-size $1\mu m$).

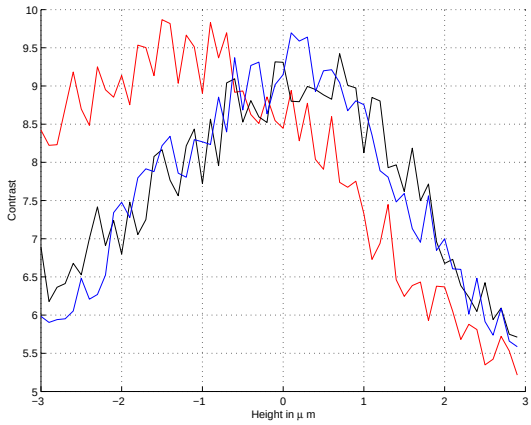

Figure 3.21: Contrast $C(z)$ over the positions of the substrate along the optical axis for three experiments (step-size $100nm$).

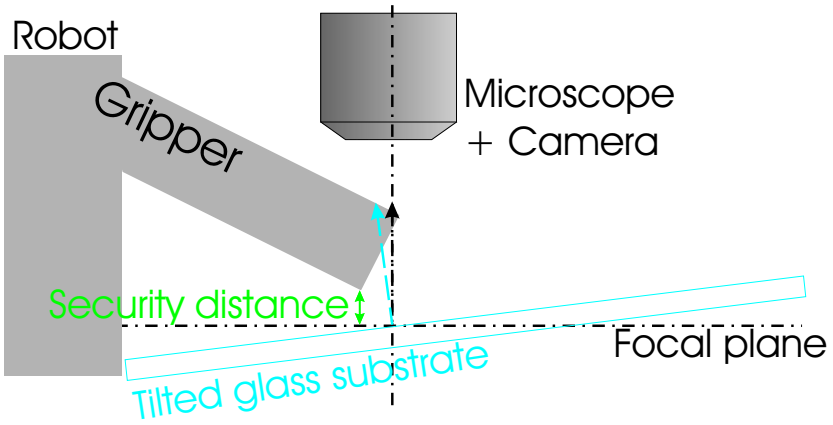

Figure 3.22: A tilted substrate with respect to the optical axis requires a greater security distance between the lower end of the end-effector and the focal plane.

Using the substrate height detection to determine the $z-$coordinate of the points has the drawback that it assumes that the end-effector is perfectly aligned with the focal plane. This can be improved when the tilt angles of the motion of the end-effector in the $(x, y)-$plane are determined as well. The height detection of the end-effector will be described later in section 4.1.

3.7 Object and end-effector detection

Gripping objects requires the knowledge of the position of the object and the gripper. The different properties of the objects and the gripper (e.g. *a-priori* information about shape) have been the reason to implement two different methods that are suitable for each case. A critical aspect in the location of objects and gripper is the scenario when the gripper is in

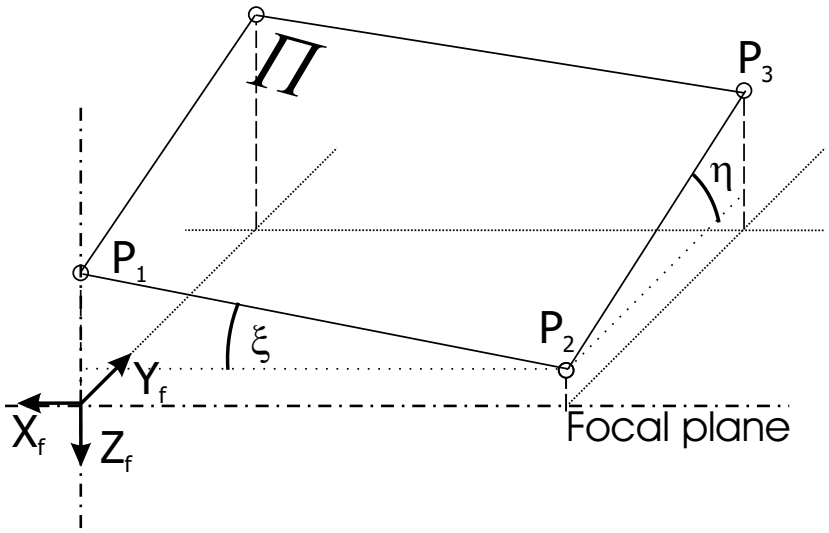

Figure 3.23: Plane Π tilted with respect to the focal plane.

contact with the object or only some pixel away from it. Without taking care of this situation the gripper as well as the object detection can be influenced heavily, even a completely wrong position estimate can occur. Based on these position estimates control inputs are sent to the robot.

3.7.1 Object detection

The objects to be manipulated have the following properties:

- Size: $5 \times 5 \times 5 - 50 \times 50 \times 50 \mu m^3$
- Lateral position $\{\mathbf{x}_O\}_i \in \mathbb{R}^2$ in the focal plane unknown
- Orientation in the plane $\{\alpha_O\}_i$ unknown
- Low rigidity: changes in contour can occur when a force is exerted onto the object

It is assumed that the object can be viewed at least one time sharp so that the object shape can be determined and that the object is never completely occluded by other objects.

After characterizing the object properties, it is important which information about the object has to be extracted from the image. This information is often also called *features*. Examples of features are corner points or the area of an object. In [27] it is shown that the center of gravity (CG) of an object $\{\mathbf{x}_{O,CG}\}_i \in \mathbb{R}^2$ is a good feature to consider because it is an unbiased estimate even if the object edges are wide (>5 pixels) when homogenous illumination is provided. It is also stated in [27] that also the CGs of blurry objects can be detected virtually unbiased. In order to manipulate an object it is important to determine at least the object height and width as well. These features are biased if the object has wide edges and if the intensity of illumination is varying or the threshold (see also following paragraph) is changed. The object size is also modified when the object is not in focus. The object width and height

are often approximated by a bounding rectangle. This is defined by the smallest rectangle that fits an object and its lines are aligned with the image coordinate axes.

There are several methods to obtain the desired features. A good overview of basic techniques can be found in [22]. An overview of methods applied in visual servoing is presented in [27] and [29] and methods used in machine vision can be found in [26]. Among these methods the contour retrieving algorithm described in [25] in combination with thresholding was selected. This algorithm needs usually only one pass through the image. For very rare cases more than one scan is required. The algorithm performs a line-by-line scanning and applies border following when a point is found that belongs to a new contour which is saved using chain code. The obtained contours are returned in a tree structure so that it is known which contours are surrounded by another. This algorithm was selected because only one pass through the image is expected and hence the computational effort is low. Furthermore, this function is already implemented in the OpenCV library. The input of this function is a binary image which is obtained through segmentation. Here thresholding is used. Alternatively, edge detection can be used or other techniques [22].

In an binary image, the foreground has to be separated from the background. One standard technique is thresholding the image intensity: The intensity values $I(\mathbf{x}_i)$ are mapped to zero if they are greater than a threshold T_b otherwise they are set to one. The threshold T_b can be determined automatically using Otsu's method that is explained in detail in [1]. Since the illumination conditions are stable (compare 3.5) it is sufficient to calculate the threshold only once when objects are selected. The thresholding together with the contour retrieving algorithm need 79ms to compute being applied to the whole image.

Figure 3.24 shows the microscope image of one scene containing some round objects, dirt and dust. The round objects are spheres with a diameter of $15\mu m$ and are the objects of interest. The result of thresholding this image is visualized in fig. 3.26. As it can be seen dust particles and traces of a dried fluid were interpreted as foreground as well. Morphological operations can be used to remove perturbations that are smaller than the objects of interest. In figure 3.27 it is shown how a morphological opening (mask: square of size=3×3) removes small disturbances and smoothes the contours of the objects. Details about morphological operators can be found in [22]. However to remove bigger objects the mask size would have to be increased resulting in a high computational effort. The computational time of this step is about $200ms$. Figure 3.28 and 3.29 depict the segmentation result overlaid into the original image. The white contours correspond to the estimated object boundary. The numbers label those objects that have an area of more than 100 pixel and are aligned with the CG of the object. Because this method of filtering out detected objects by their area the morphological opening is not needed since the SNR of the images is high enough for a good segmentation result. Figure 3.29 is a zoom of the segmentation result. The cluster of spheres is detected as one object since the spheres are touching each other. A few pixel above the number 5 the object is estimated at some places 2 pixels too big. The contour of the single sphere is determined correctly.

Figure 3.30 visualizes the results of a sensitivity analysis of the influence of changes in illumination on the contour detection and the calculation of the object CG.

The backlight illumination power was kept constant and the in-line illumination power I_i was modified. The optimal threshold T_b was calculated at 85% in-line illumination power and

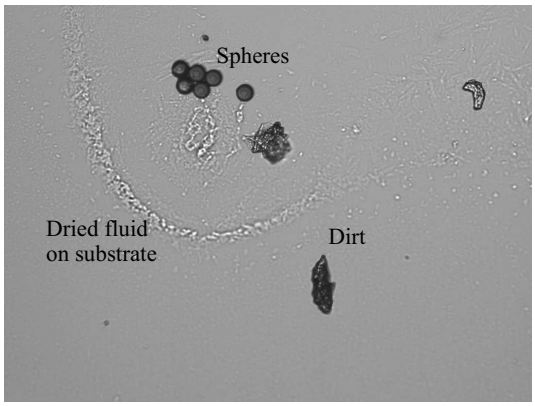

Figure 3.24: Original image.

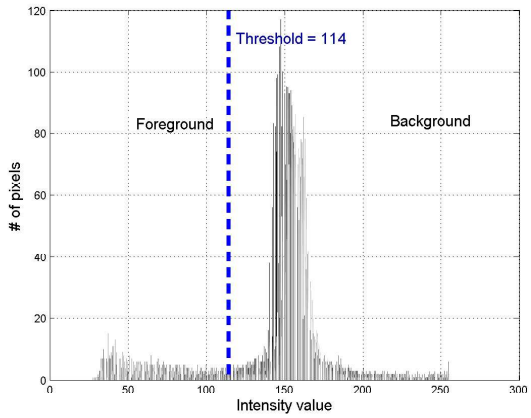

Figure 3.25: Image histogram.

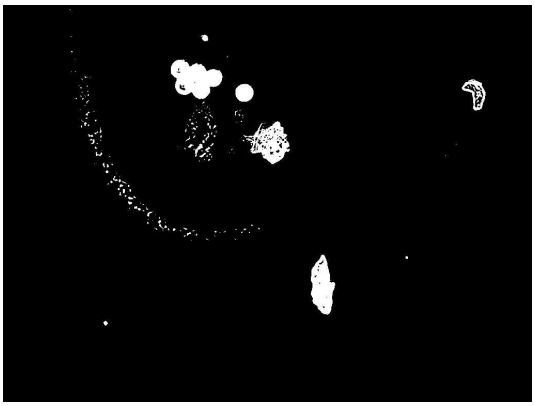

Figure 3.26: Image after thresholding.

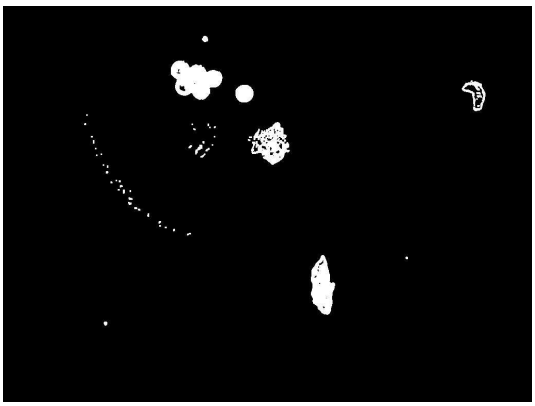

Figure 3.27: Image after morphological opening.

was not adapted for the changed illumination conditions. Reducing I_i to 50% a big region of the background is interpreted as object and the position estimate is completely wrong. The positions of the CG were determined for each illumination condition five times and were averaged. Table 3.4[5]

I_i in %	$\mathcal{E}\{x_O\}$ in μm	$\mathcal{E}\{y_O\}$ in μm	$\sqrt{\mathcal{V}\{x_O\}}$ in μm	$\sqrt{\mathcal{V}\{y_O\}}$ in μm
75	-353.1804	198.8384	0.2258	0.3362
85	-353.3880	199.4514	0.0772	0.0654
100	-353.5592	199.2160	0.0088	0.0540

Table 3.4: Results of sensitivity analysis regarding the effects of changes in I_i on the estimation of the object CG.

It can be seen that for decreased I_i the CG estimation is biased: Regarding in figure 3.30 the image for $I_i = 75\%$, it can be seen that the object was not well segmented and the disturbance

[5]The differences in the position estimates can barely seen in image 3.30 because the size of one pixel corresponds to $460nm \times 460nm$.

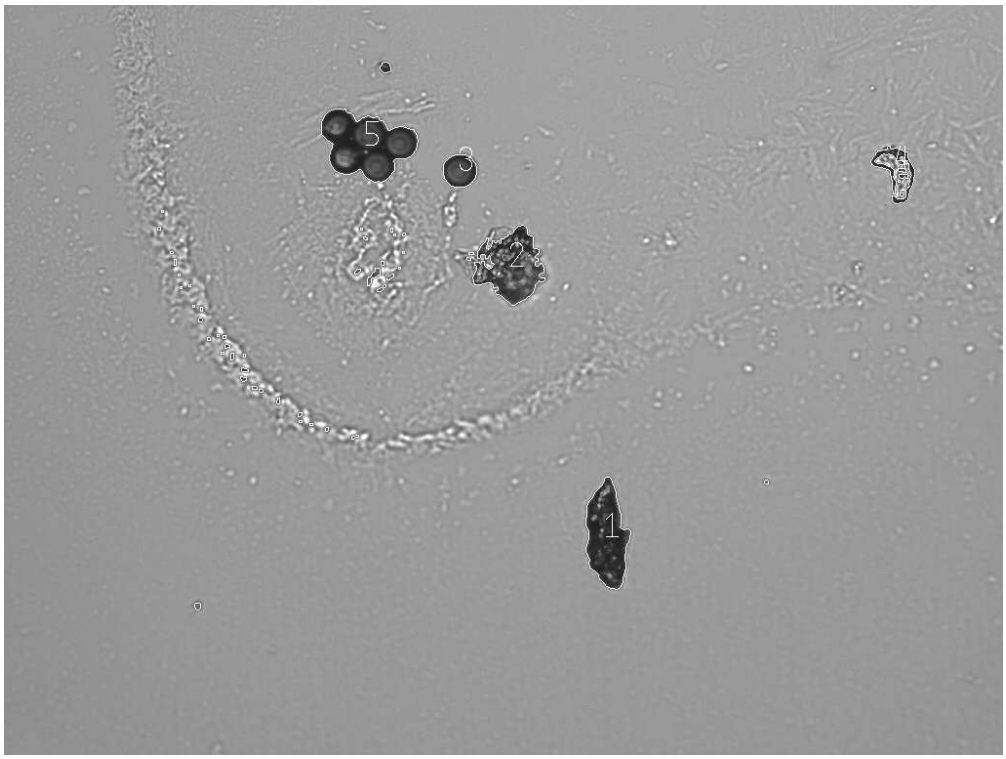

Figure 3.28: Labeled segmentation result.

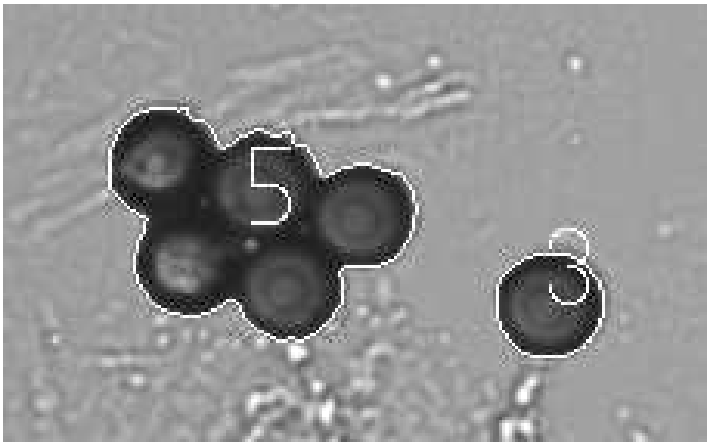

Figure 3.29: Zoomed view of segmented objects.

at the upper left boarder shifts the CG about $(208nm, 614nm)$. Since the disturbances around the object are close to the threshold, the standard deviation of the position estimates is approximately $(226nm, 336nm)$. For higher I_i the shift is about $(172nm, 224nm)$. But in this case the standard deviation decreases since the image contrast is higher.

At $I_i = 85\%$ and considering that at $z = 0$ the object is in focus, the object plane was moved $\pm 10\mu m$ relative to the focal plane. The results are presented in figure 3.31.

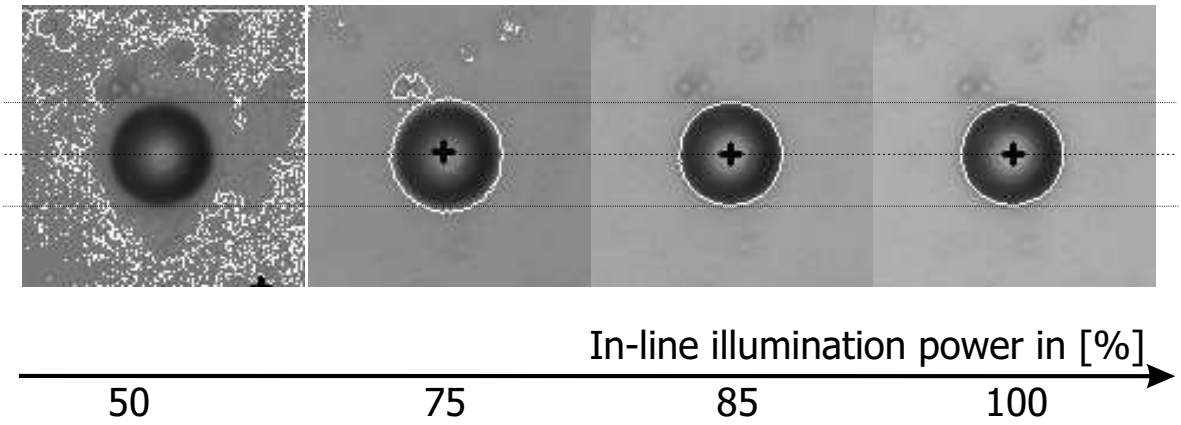

Figure 3.30: Influence of the in-line illumination power on the position and contour calculation. The optimal threshold was determined for a power level of 85% and was not adjusted for the changed illumination conditions.

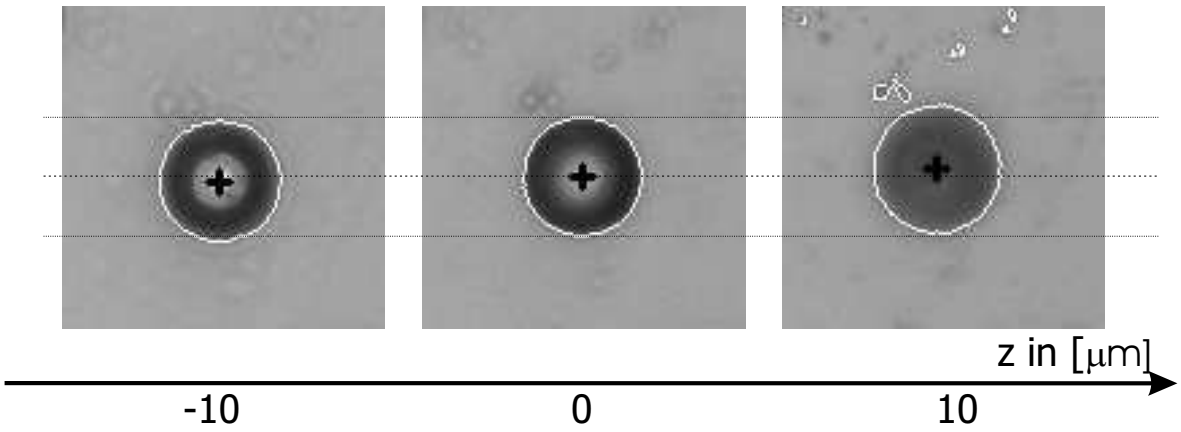

Figure 3.31: Influence of blur on the position and contour calculation. At $z = 0$ the object is in focus.

In all three cases the object was segmented properly in that sense that no other disturbances were interpreted to be part of the object. The object CG and the width of the bounding rectangle at each $z-$coordinate are averaged over several measurements and summarized in table 3.5.

z in μm	$\mathcal{E}\{x_O\}$ in μm	$\mathcal{E}\{y_O\}$ in μm	$\mathcal{E}\{$Bounding box width$\}$ in μm
-10	-353.4340	199.7784	21.2612
0	-353.3880	199.4514	21.0496
10	-352.9500	197.9656	22.4020

Table 3.5: Results of sensitivity analysis regarding the effects of blur on the estimation of the object CG.

A shift can in position detection especially in the $y - direction$ can be observed. This offset origins from the misalignment of the optical axis with the substrate motion along the z-

axis. The bounding rectangle grows about 1%$\hat{=}$0.21μm and 6%$\hat{=}$1.35μm for $z = -10\mu m$ and $z = 10\mu m$ respectively.

3.7.2 End-effector detection

In order to increase the manipulation precision, the relative distance in the FoV between the tip of the end-effector and the object of interest must be known very well. There is some previous knowledge how the zero position of the robot is related to the FoV. However thermal effects and un-modelled kinematics limit the accuracy to several μm. Therefore, the end-effector is detected locally in FoV coordinates so that the relative position of the gripper to the object can be determined based on this estimate. The lateral position of the end-effector in the FoV is denoted as $\{\mathbf{x}_E\}_i \in \mathbb{R}^2$ There are two types of end-effectors that are used in the current setup for manipulations: 1. a glass pipette for the manipulation of cells and 2. a two-jaw gripper.

Both end-effectors are mounted on the Delta robot which has 3 prismatic DOF so that the orientation of the end-effectors are constant. The geometries of the end-effectors are constant and can be measured with the microscope. Template matching algorithms are well known to give good results for such a detection problem. The three biggest drawbacks of template matching is the computational complexity: If the image has $N \times N$ pixel and the template $M \times M$ the computational complexity is $O(N^2M^2)$. The size of the template is an important parameter and determines the robustness of detection. Another disadvantage of template matching is the need to define a template for each end-effector in the focal plane. Finally, if one of the fingers of the two-jaw gripper is actuated to grasp an object, the tip is rotated. This will introduce an error into position estimate of the template match unless the templates are adapted if the match is below some threshold. However the template adaption process also introduces an error.

For improved performance, for the two-jaw gripper its geometry is employed: Instead of matching a whole region locally, the edges in the image are calculated and then the Hough transform (HT) is employed to detect lines and calculate their intersection points which represent the corner tips of the two-jaw gripper. The two methods are presented in the following, and their robustness against varying illumination and blur are examined.

Template matching to detect the pipette

Template matching compares a reference template pixel by pixel within an image at all possible positions. The comparison is based on an matching function. The position of the template that has the best match is then considered as position of the template. Figure 3.32 shows the selected template. The template size was chosen as a trade-off between computational cost and detection robustness. The template ends with the tip. Since the tip is imaged sharply it contains many features. Additionally, the tip has the form of a parabola. The likelihood that a similar shape will appear in the scene is low. Furthermore, robust position information in x- and y-direction is obtained because of the edges of the parabola. Another reason for this selection is that once an object is attached at the tip, the object would influence the detection heavily if it would interfere with the template frame.

For simplicity the index i indicating image coordinates is dropped in the following. $T(\mathbf{x}')$ denotes the intensity values of the template for all $\mathbf{x}' \in \mathcal{R}$. $\mathcal{R}$ is the set of pixels contained in the template (compare figure 3.33).

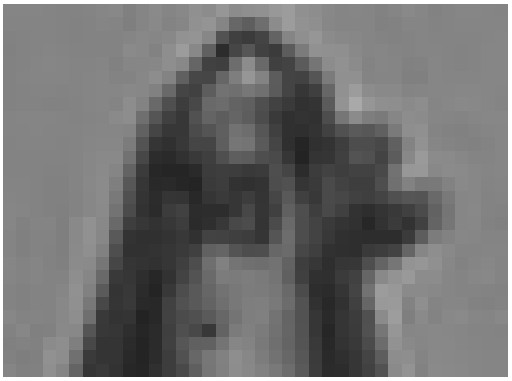

Figure 3.32: Template of the tip of a pipette (size: 38×24).

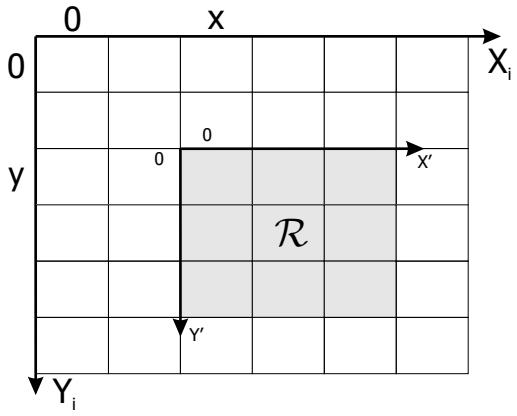

Figure 3.33: The region of a template of the size 3×3 is drawn into an image. Its position $\mathbf{x}$ is aligned with the upper left corner.

Three different matching functions were evaluated:

- *Normalized squared difference:*

$$S(\mathbf{x}) = \frac{\sum\limits_{\mathbf{x}' \in \mathcal{R}} (T(\mathbf{x}') - I(\mathbf{x}+\mathbf{x}'))^2}{\sqrt{\sum\limits_{\mathbf{x}' \in \mathcal{R}} T(\mathbf{x}')^2 \sum\limits_{\mathbf{x}' \in \mathcal{R}} I(\mathbf{x}+\mathbf{x}')^2}}. \tag{3.17}$$

- *Normalized cross-correlation:*

$$C(\mathbf{x}) = \frac{\sum\limits_{\mathbf{x}' \in \mathcal{R}} (T(\mathbf{x}') I(\mathbf{x}+\mathbf{x}'))}{\sqrt{\sum\limits_{\mathbf{x}' \in \mathcal{R}} T(\mathbf{x}')^2 \sum\limits_{\mathbf{x}' \in \mathcal{R}} I(\mathbf{x}+\mathbf{x}')^2}}. \tag{3.18}$$

- *Normalized correlation coefficient:*

$$R(\mathbf{x}) = \frac{\sum\limits_{\mathbf{x}' \in \mathcal{R}} (\tilde{T}(\mathbf{x}') \tilde{I}(\mathbf{x}+\mathbf{x}'))}{\sqrt{\sum\limits_{\mathbf{x}' \in \mathcal{R}} \tilde{T}(\mathbf{x}')^2 \sum\limits_{\mathbf{x}' \in \mathcal{R}} \tilde{I}(\mathbf{x}+\mathbf{x}')^2}}, \tag{3.19}$$

  where $\tilde{T}(\mathbf{x}') = T(\mathbf{x}') - \bar{T}$ and $\tilde{I}(\mathbf{x}+\mathbf{x}') = I(\mathbf{x}+\mathbf{x}') - \bar{I}$. $\bar{T}$ and $\bar{I}$ correspond to the average intensity level of the template and of the windowed image, respectively.

In the case of the *normalized squared difference*, the coordinate pair $\mathbf{x}$ that minimizes $S(\mathbf{x})$ represents the position of the template. For the correlation-based match functions, the position is given by the coordinate pair that maximizes $C(\mathbf{x})$ or $R(\mathbf{x})$.

Table 3.6 gives an overview of a set of experiments: The substrate was removed and the pipette was moved in the middle of the FoV. The illumination power was set to 75% and a template as shown in figure 3.32 was captured. The template has a size of 38×24 pixels. The template search in the whole image at full resolution takes $\approx 30s$. This time was reduced using a multi-scale approach: The image and the template are sampled down twice using the Gaussian pyramid decomposition[6]. This reduces the number of pixels in the image and the template by 16. Considering the complexity of template matching as described earlier the number of operations is reduced by 256. Certainly, the resolution of the result is reduced to ± 2 pixels. The coarse position of the template is then known and a region of interest (ROI) is defined in its proximity. Performing the template matching at full resolution in this ROI the resolution of ± 0.5 pixel is obtained. For the different matching functions the pipette was first detected at the same parameters at which the template was taken. The result for the normalized square difference is depicted in figure 3.34. The white box shows the detected position of the template in the image. The black cross marks the tip of the pipette[7]. Then the parameters illumination and blur were changed until the matching function failed to repeatable detect the tip of the pipette at the same position. These values are presented in table 3.6. Blur was considered in terms of vertical displacement to the focal plane in μm and the illumination sensitivity shows the range of illumination power in % of the maximum illumination power.

Matching function	**Illumination sensitivity in %**	**Displacement from focal plane in μm**	**Execution time in ms**
Normalized squared difference	60 - 100	± 100	175
Normalized cross-correlation	20 - 100	± 50	160
Normalized correlation coefficient	10 - 100	± 40	200

Table 3.6: Comparison of three different matching functions with respect to their sensitivity on changes in illumination, blur and their execution time.

Summarizing the results of table 3.6 it can be seen that the execution times are almost the same for the three different functions. It should be noted that the execution time is dominated by the initial search of the whole low-resolution image. The normalized square difference is very sensitive to changes in illumination considering that the template was taken at 75%. The correlation-based methods are very robust and were able to detect the tip of the pipette even for very few illumination. Considering the depth of focus of the $20\times$ objective of $1.6\mu m$, all functions give repeatable results till $\pm 40\mu m$. The normalized squared difference function even gives proper results for $\pm 100\mu m$.

This result can be further improved by interpolating the mask and the corresponding ROI in the image. The ROI is selected such that one pixel is added to each border of the detected position of the template. With this improvement a maximal error of ± 0.25 pixel is obtained

[6] details about this method can be found in [22] and [24]

[7] the offset from the upper left corner of the template to the tip of the pipette is measured off-line manually.

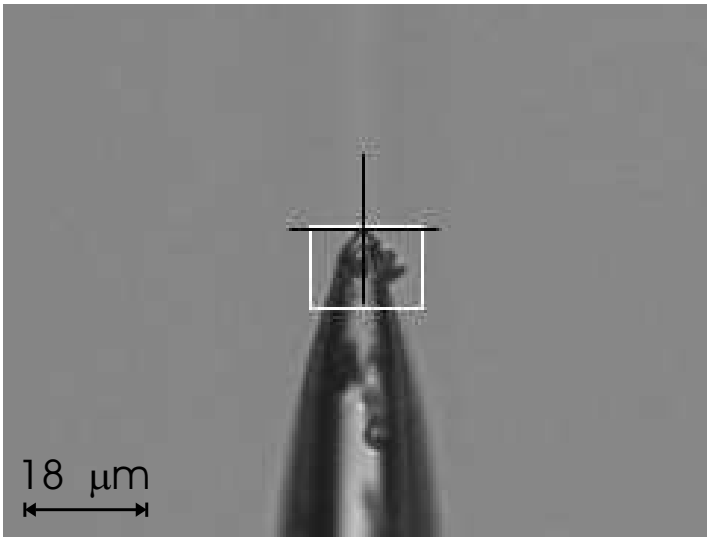

Figure 3.34: Result of template matching. Only the region around the tip is shown. The rest of the image has virtually the same background as seen in the zoomed representation.

which was verified in an experiment, where the end-effector was not moved and the position was measured 100 times. For this experiment the magnification of the optical system was reduced to 10× and the mounting alternative 2 (*from below*) was used, so that the maximum error in the experiment corresponds to $\pm 115nm$.

Line detection as position estimate of the two-jaw gripper

Line detection is used to detect the position of the two-jaw gripper. This involves the following steps:

- Sampling down the image twice
- Morphological closing
- Edge detection
- Standard Hough transformation (HT) to extract lines
- Assignment of lines to the gripper
- Calculation of intersections of the assign lines

The advantage of this method is that it uses only the geometry of the gripper: the angles γ_1 and γ_2 at the gripper tip and the opening distance of the gripper (compare image 3.35). The whole image is searched for lines. Since the gripper is much bigger than the objects of interest the longest lines are expected to belong to the gripper. In the case of the used two-jaw gripper only one horizontal line l_v and two vertical lines l_{h1} and l_{h2} are assigned to the gripper. Additionally, the lines can be expected to have certain orientations depending on the orientation of the gripper in the FoV and γ_1 and γ_2 . The lines that match these requirements the best are assigned to the gripper and the intersections of these lines are calculated to obtain

the position of the inner corners C_1 and C_2 of the jaws. The evaluation of the whole image leads to a evaluation of a big region which increases the robustness against disturbances. This means also that a grasped object affects the position detection of the gripper only few.

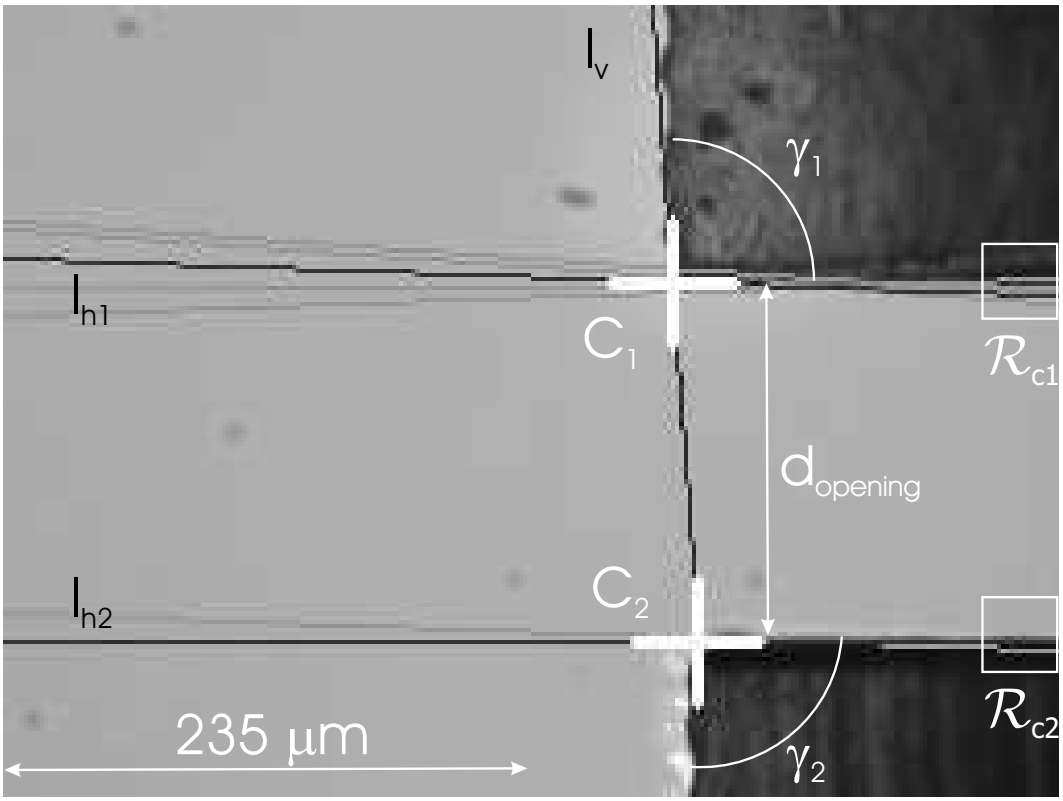

Figure 3.35: Two-jaw gripper entering the image from right and detected lines and intersections.

In image 3.35 it can be seen that the edges of the gripper have a roughness of several μm. Because of the dimensions of the gripper and the roughness of the edges, the magnification is reduced to $10\times$. Furthermore, firstly because of the roughness and secondly because of the computational effort of the line detection the captured image was sampled down twice using the Gaussian pyramid decomposition.

Some of the thin rills on the gripper surface might be interpreted as edges during the edge detection when the contrast on the surface is very high. In order to avoid this effect a morphological closing with mask size 5×5 is applied to the image after sampling down (compare image 3.36). Another option is to reduce the ratio of the in-line to the backlighting illumination power. This will increase the contrast of the gripper edge to the contrast of the surface details.

This image is input to a Canny edge detection[8]. Canny edge detection performs the following operations to produce a binary image as output in which the ones are indicating edges:

- Calculation of the image gradient magnitudes and orientations

[8]More details can be found in [25, p. 68ff]

- Non-maximum suppression
- *Hysteresis* thresholding of the gradient magnitudes. The strong edge between the background and the gripper gives raise to the selection of a high threshold. The lower and the higher limit was set to 150 and to 250 respectively.

Image 3.37 depicts the result of the Canny edge detection.

Figure 3.36: Two-jaw gripper after grey-level morphological opening.

Figure 3.37: Two-jaw gripper after Canny edge detection.

Only few disturbances can be recognizes in the image. The horizontal lines appear less rough than the one at the vertical edges at the tip. This is due to the mounting angle of the gripper towards the substrate $\beta = 20^o$ and the fact that the focus is on the tip. This results in blurrier edges the more one moves away from the tip so that the surface roughness is even out. Furthermore, the shadow at the lower jaw affects the line detection.

This image is the input to the Hough transform. The Hough transform describes a mapping from the image data space to a model space. In order to detect lines, the image is mapped to a model space which is spanned by two parameters required to describe a lines to which is also referred as parameter space. For numerical calculations the parameter space has to be discretized. The often used parameters slope and offset to describe a line are problematic because of the description of a vertical line which has slope infinity. Because of that the lines are parameterized by an angle ϕ and a distance d_{0l}:

$$x\cos(\phi) + y\sin(\phi) = d_{0l}, \tag{3.20}$$

where ϕ is the angle between the x_i-axis and the vector $\mathbf{n}_l \in \mathbb{R}^2$ that starts in the origin and is perpendicular to the line l and $d_{0l} = \|\mathbf{n}_l\|$ (see also figure 3.38). The discretization of the parameters was set to $\phi \in [0, \pi]$ with a resolution of 2^o and $d_{0l} \in [-\sqrt{w^2+h^2}, \sqrt{w^2+h^2}]$[9] with a resolution of 1 pixel.

The implementation of the HT in OpenCV checks all pixels of the binary input image. For each white pixel the algorithm checks on which lines the pixel lies. If the pixel lies on a line, it *votes* for the corresponding coordinate (ϕ, d_{0l}) in the parameter space. The votes for each line are accumulated. This is followed by a local maxima search of the votes over the

[9] w and h correspond to the image width and height, respectively.

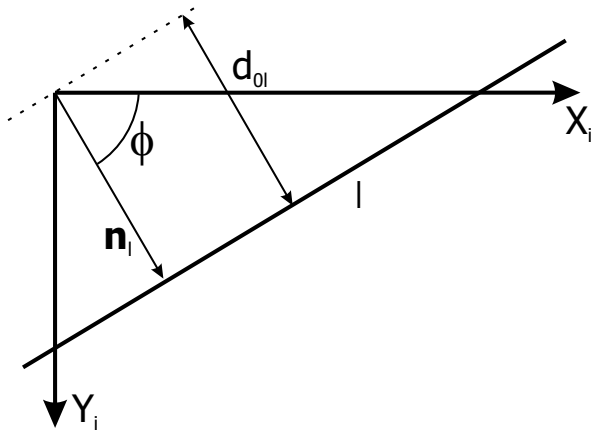

Figure 3.38: Alternative representation of a line.

whole parameter space. Only lines that exceed a threshold of votes are considered lines. This threshold was set to 20. This is a rather low value, however it is necessary: In the case when the gripper is only in a small part of the image the lines are shorter and hence these lines are defined by less pixels. The main drawback using a low threshold is that many lines (≈ 20, if the gripper in the middle of the FoV) are returned by the algorithm.

After applying the HT, the lines that correspond to l_v, l_{h1} and l_{h2} have to be found among the set of lines detected by the HT. In image 3.35 all lines returned by the HT are shown in grey color whereas the assigned lines are marked with black color. The steps to assign the lines are:

- Find two lines that have the maximum votes and whose angles fulfill the constraint (a) $|\phi - \phi_{expected}| < 5^o$. $\phi_{expected}$ is set to 0^o and to 90^o for the vertical line and the horizontal line, respectively. Since the parameter d_{0l} can be also negative, vertical lines can have also $\phi_{expected} = 180^o$. The vertical line is assigned to l_v. The horizontal line is initially considered as $line1$.
- The second horizontal $line2$ is found by considering the line with the maximum number of votes that fulfills constraint (a) with $\phi_{expected} = 90^o$ and the additional constraint that $|d_{0l,line1} - d_{0l,line2}| \in [d_{low}, d_{upper}]$[10]. The choice of d_{low} and d_{upper} is based on $d_{opening}$ and the minimum size of the objects of interest. The latter one gives an estimate how close the two lines l_{h1} and l_{h2} will approach and can be used to set d_{low}.
- Assuming that the stiff arm is always at the bottom and the actuated arm is at the top in the image, the two horizontal lines are assigned such that the line with the smaller absolute parameter $d_{0l,line}$ is assigned to l_{h1} and respectively the other line is assigned to l_{h2}.
- In order to check wether the horizontal lines are assigned properly, two ROI $\mathcal{R}_{Ci}, i = 1, 2$ are examined whose vertical position is centered by the position of the line and which is aligned with the border of the image where the gripper enters the FoV (compare figure

[10] Note that if the gripper enters the FoV from below and not from the side, the absolute value of $d_{0l,linei}, i = 1, 2$ has to be considered

3.35). The position of the ROIs was selected such that the likelihood that an object lies within them is minimized. The *sign* of the edge[11] within these ROIs is investigated. This tells if there is a transition from bright to dark or vice versa. In $\mathcal{R}_{C1}$ an edge with positive sign is expected and respectively a negative one for $\mathcal{R}_{C2}$. If the calculated values correspond to the expected ones, the lines are assumed to be assigned correctly.

- If there is a contradiction in the test above, a third horizontal line is searched that has a maximum number of votes and that fulfills constraint (a) for horizontal lines and the additional constraint

$$(|d_{0l,line1} - d_{0l,line3}| \in [d_{low}, d_{upper}] \wedge |d_{0l,line2} - d_{0l,line3}| \in [2d_{low}, 2d_{upper}]) \vee$$
$$(|d_{0l,line2} - d_{0l,line3}| \in [d_{low}, d_{upper}] \wedge |d_{0l,line1} - d_{0l,line3}| \in [2d_{low}, 2d_{upper}]).$$

  In words: the last constraint is restricting the candidates of possible lines to those that are in a distance d_e to each of the lines within a range Δ, however excluding themselves to be candidates (compare figure 3.39).

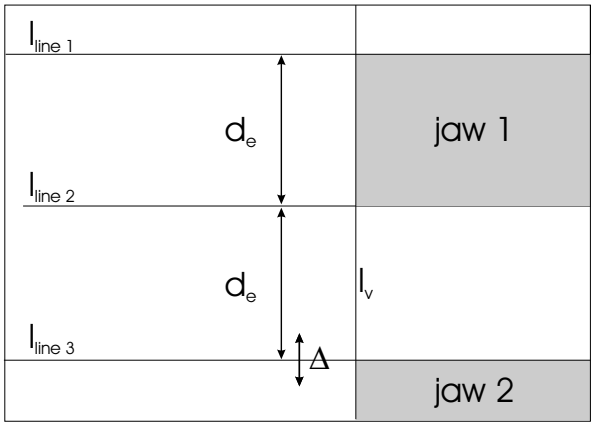

Figure 3.39: Visualization of the scenario when three horizontal lines being in a certain distance from each other can be found in an image.

  If a third line is found fulfilling the constraints the position corresponds to the situation as presented in figure 3.39: One of the fingers is completely in the image. Since the third line must be one of the horizontal lines one is looking for, the line that is closer to line 3 is selected to be the other horizontal line. Depending which line this is, the lines are assigned accordingly.

  If no third line is found, either the Canny edge detection did not find enough edges (due to blur or change in illumination) or, more likely, only one finger is in the FoV.

The corner points C_1 and C_2 can be found calculating the intersection points of the vertical line with the horizontal lines. The expression to calculate an intersection point $\mathbf{P} \in \mathbb{R}^2$ is:

$$\mathbf{P} = \mathbf{n}_{l,1} + \lambda_1 \mathbf{e}_1 \tag{3.21}$$

[11] The output of a sobel filtering in the direction of the y_i-axis is used

with

$$\mathbf{n}_{l,1} = \begin{bmatrix} \cos(\phi_1) \\ \sin(\phi_1) \end{bmatrix} d_{0l,1},$$

$$\mathbf{e}_1 = \begin{bmatrix} -\sin(\phi_1) \\ \cos(\phi_1) \end{bmatrix},$$

and

$$\lambda_1 = \frac{diag\{\cos(\phi_2), \sin(\phi_2)\}(\mathbf{n}_{l,2} - \mathbf{n}_{l,1})}{\sin(\phi_2 - \phi_1)},$$

where the introduced variables are defined in figure 3.40 and the derivation of this formula can be found in appendix A.

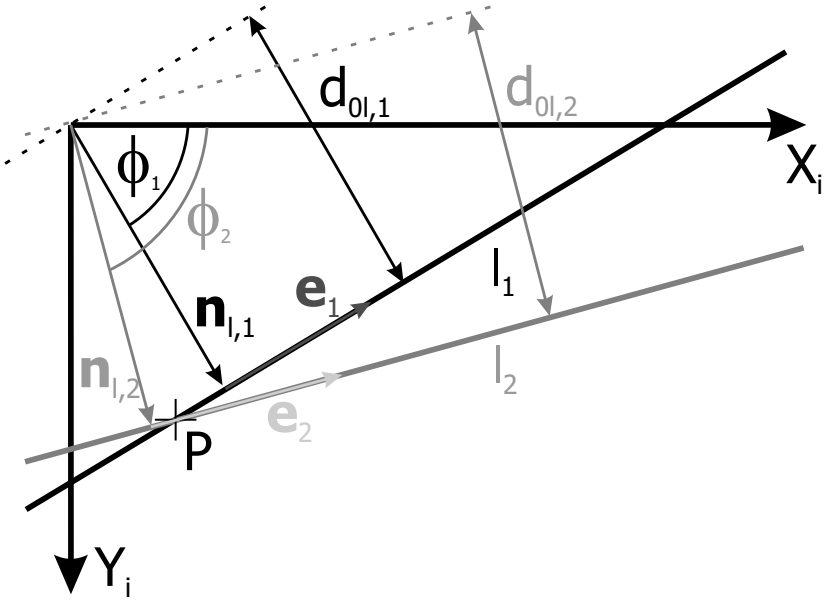

Figure 3.40: Intersection of two lines and related parameters.

A correlation is performed over a ROI in the proximity of C_1 and C_2. Since the estimate of C_1 and C_2 is robust to disturbances a small template and a small ROI aligned with these points can be used. The ROI must be of the size of the template and extent its size in theory by 2 pixels which is equivalent to the maximum error of the line detection itself. The discussion of this ROI size in the following section about experimental results gives a size of the ROI of 10 pixel.

From the detected corner point an geometric offset can be defined which can be used to align the desired point on the gripper with the object.

Experimental results

The line detection was investigated towards it robustness to changes in illumination and blur. A series of experiments was performed using a power of the backlighting of 75% and constant

in-line illumination power of $\approx 50\%$. In the following discussions only the results of the line detection itself is discussed not considering a local template matching of the corners. Figure 3.35 presents the results for this situation when the upper gripper finger tip is in focus. In figure 3.41 the results of the line detection with different offsets of the gripper tip from the focal plane $z = 0\mu m$ are presented. For $z = -60\mu m$ and $z = 80\mu m$ the edge contrast of the gripper becomes too low and not the complete contour of the gripper is obtained leading to a failure in the line detection. At $z = 60\mu m$ l_v obtained all of its votes by the upper gripper jaw. The rough surface of the tip and the introduced blur results in an edge which is not parallel to the vertical edge determined for the lower gripper jaw. This behavior could be improved detecting a second vertical line and adding e.g. the constraint to check wether the lines are assigned correctly using a check box at the expected gripper corners C_1 and C_2. The position of C_1 at $z_1 = -40\mu m$ is calculated to be $(148.4\mu m, 65.3\mu m)$ and at $z_2 = 40\mu m$ the position is estimated to be at $(148.1\mu m, 60.7\mu m)$. Moving from z_1 to z_2 the position estimate along the x_i-axis shifts by $300nm$ whereas the shift along the $y-$axis is $4.6\mu m$. This shift results mainly from a misalignment of the optical axis with the $z-$axis of the used piezo positioning stage to generate the offsets (compare figure 4.8).

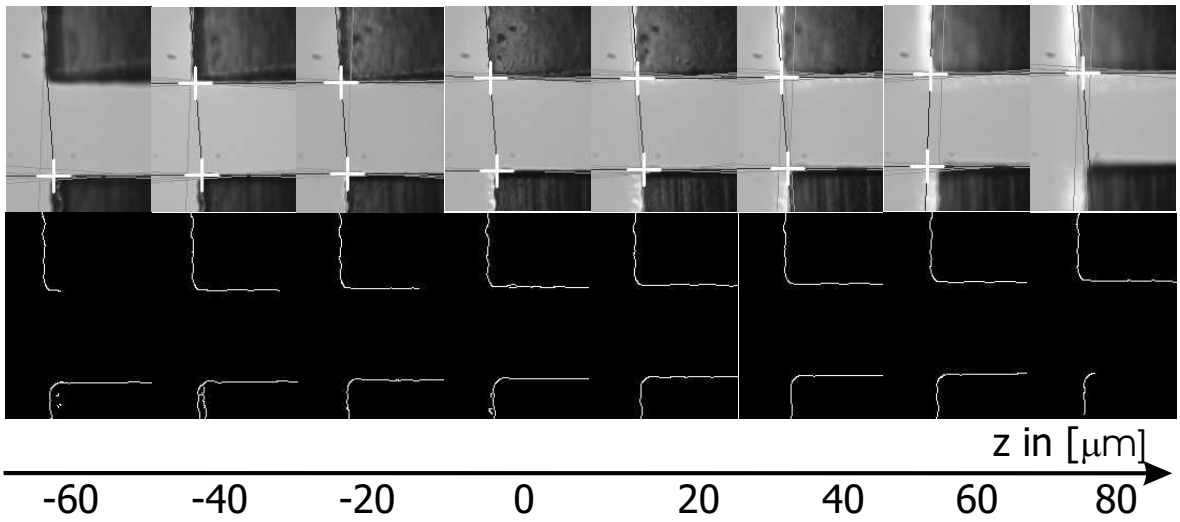

Figure 3.41: Results of the line detection in the first line for different offsets of the gripper tip from the focal plane $z = 0\mu m$. In the second line are the corresponding results of the Canny edge detection. Only the right half of the whole captured image is presented.

Figure 3.42 shows the results of a sensitivity analysis of changes in illumination. The power of the backlighting was varied over a range from $25 - 100\%$ and its effects on the line detection examined. For 25% illumination power I_p the contrast of the edges is not sufficient to exceed the threshold defined for the Canny edge detector. The threshold was chosen to work well at 75% I_p. At $50\% I_p$ the upper edge of the gripper is not completely detected. Instead the edges of surface details influence the determination of the vertical line. An adaptive threshold for the edge detection could improve the results. The threshold could be lowered until all three lines can be assigned. The problem is the necessary assumption that the gripper is in the FoV and all three lines could be detected. For higher I_p values the position estimate stays constant. In general, using a rather high ratio of backlighting to in-line illumination power will give robust results for the case when the gripper is viewed from above (compare figure

3.8).

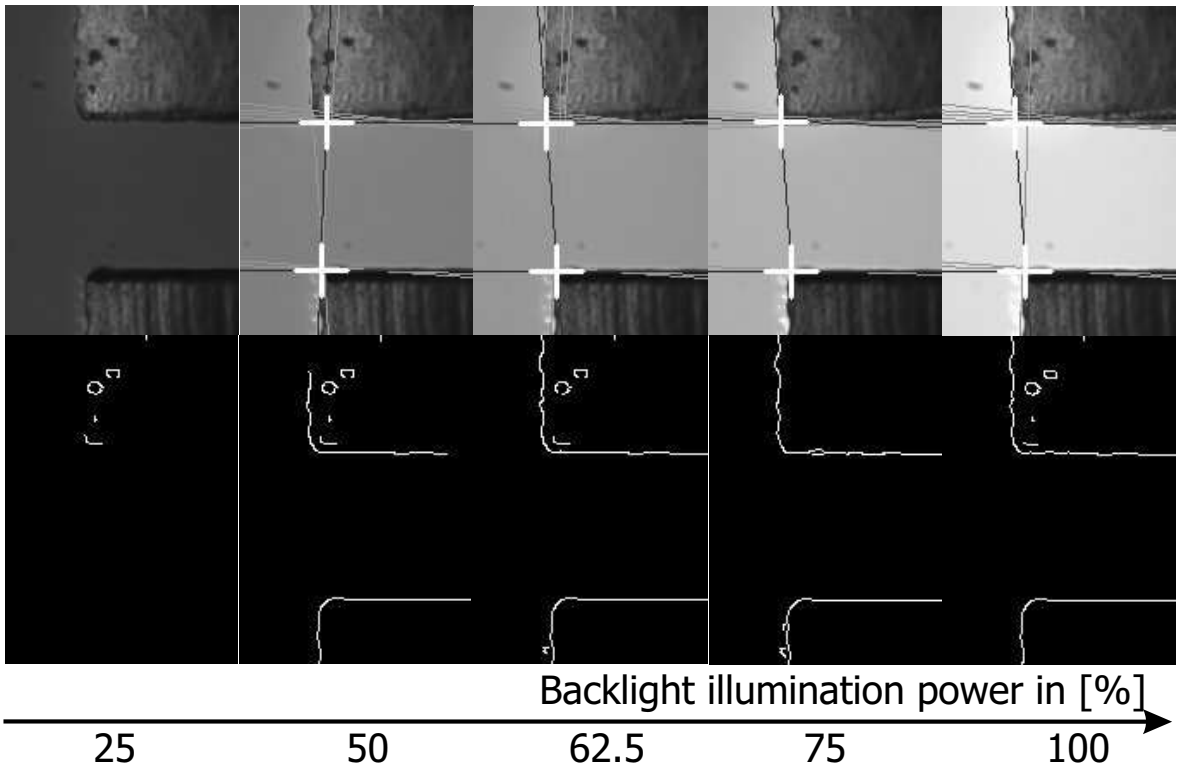

Figure 3.42: Results of the line detection in the first line for different power levels of the backlighting. In the second line are the corresponding results of the Canny edge detection. Only the right half of the whole captured image is presented.

Considering the case when the gripper is observed from below, again the edges of the contour of the gripper are considered.

Note that due to the laser machining process of the gripper its edges are not only rough but are also modified in vertical direction (compare fig. 3.43). If the backlighting is not well-aligned shadow effects can be observed that occlude the object. In fact, since the shadow represents the contour of the gripper, the line detection is affected. This means that although it looks as if the gripper touches an object it does not, because it is just the gripper shadow. Note that this effect disturbs the gripper detection in general (e.g. also when template matching is used).

This can also lead to a biased estimate of the $y-$coordinate of the object CG as discussed later in section 3.7.4 since the object is pushed into the shadow when closing the gripper. In order to reduce shadow effects and to obtain a better estimate of the *real* gripper position, the shadow needs to be distinguished from the actual gripper: An approach to tackle this problem is to perform a search for the texture of the gripper surface. Whereas the shadow appears completely black, the surface, that is high-lighted by the in-line illumination, has a certain texture containing high spatial frequencies. Another option is certainly to switch of the backlighting only using in-line illumination. Reducing the thickness or increasing the object size reduces the shadow effect as well.

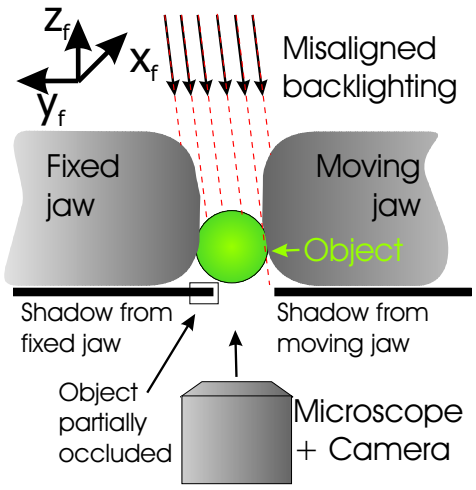

Figure 3.43: Sketch of the gripper tip and shadow effects for misaligned illumination when looking with the microscope *from below*.

Table 3.7 summarizes the results of the sensitivity analysis and gives the execution time of the line detection.

Illumination sensitivity in %	**Displacement from focal plane in** μm	**Execution time in** ms
62.5 - 100	$-40..60$	300

Table 3.7: Results of the line detection with respect to its sensitivity on changes in illumination, blur and its execution time.

Some other effects were observed while using the line detection with the specific presented gripper:

- Due to machining and assembly precision of the gripper the assumption of one vertical line l_v is not completely fulfilled. The parameters of l_v are set only by the gripper jaw that is more in the FoV. The intersection of l_v with the other jaw still gives an useful estimate. In order to still be able to obtain a constant estimate when the gripper is moving along the y_i-axis, the corner position estimates are refined using template matching in a sufficiently big ROI. This ROI was determined experimentally to be in the proximity of ± 10 pixels around the intersection point for this specific gripper. It is expected that using a gripper that fulfills the assumption of parallel jaw tips, a much smaller ROI can be used or even the template matching is not needed.
- When the gripper jaws approach each other the ROIs $\mathcal{R}_{Ci}$ have to be decreased in size since both edges of the gripper could appear in these ROIs. However reducing the ROI size is reducing the certainty with which can be decided wether the line assignment is correct.
- Manipulating very small objects, the assignment of the lines becomes difficult: If d_{low} is selected very small (e.g. 10 pixel), two lines that belong to the same gripper edge might be interpreted as the two proper horizontal lines. In the case in which one is not sure

if the lines are assigned properly, the position estimates of the previous detection are used and are refined by the local template matching. This solution has the disadvantage that the ROI in which the corners are searched must be chosen bigger (e.g. 30 pixel) leading to an increased execution time and making the results heavily dependent of the outcome of the local template search. An alternative approach to solve this problem is tracking of the parameters of the horizontal lines how it will be described in detail in section 3.7.3.

The theoretic resolution of only the line detection can be derived from the fact that the image is sampled down twice. Considering that the magnification was set to 10× and considering the size of the pixel on the chip $0.46\mu m$ a maximum error of $\pm 0.92\mu m$ can be calculated. Using the local template matching without interpolation the precision can be given as ± 0.5pixel$\hat{=}\pm$ $0.23\mu m$.

3.7.3 Tracking

After an initial detection of the object of interest and the end-effector it is not necessary to evaluate the whole image for following time steps. Once the position of the gripper $\{\mathbf{x}_E\}_i$ and the object $\{\mathbf{x}_O\}_i$ are known, this information can be used at the next time step to define a ROI in which the new search is performed. Additionally, the robot displacement is known which can be added to the position estimate at $t = k-1$. Figure 3.44 visualizes this concept, where $\hat{\mathbf{x}}_k^+$ is a position estimate taken at $t = k$, $\{\Delta\mathbf{x}_{k-1}\}_i$ the robot motion commanded at $t = k-1$, $\hat{\mathbf{x}}_k^-$ the prior position estimate at $t = k$.

In the following, the tracking of

- an object,
- a template
- and lines

is discussed.

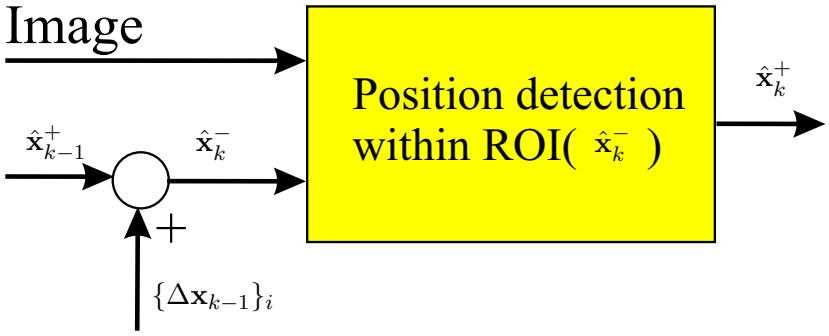

Figure 3.44: Signal flow chart for tracking, respectively estimation of the position of the ROI.

Object tracking

The definition of a ROI is especially of help for the tracking of the objects: The same routine as described in section 3.7.1 is used to find an object. In the object tracking the user is not asked to choose an object. At the initial object selection its moments, respectively e.g. its area, are saved. When several objects are detected within a ROI, the one that matches the area the best is chosen. The size of the ROI is aligned by the bounding rectangle of the previous detection and inflated by 20 pixel in all directions. The size of the ROI after inflation is dependent on the expected motion of the object. The displacement $\{\Delta\mathbf{x}_{k-1}\}_i$ is added to the position of the ROI if it is gripped.

Template tracking

Using the template matching a definition of a ROI saves a lot of computational time. The ROI is aligned with $\hat{\mathbf{x}}_k^-$ and the size is determined by the template size and is inflated by 15 pixel in all directions. The correlation is only applied to this ROI. The initial low-resolution search is not performed. After the detection the correlation value is considered. If the value is below a certain threshold it is assumed that the template does not match and a search in the whole image is initiated accordingly to 3.7.2. One can calculate easily that for a template which is reasonably small one should avoid this case. That is why the ROI has to be chosen sufficiently big. To determine the size, the certainty of the prior position estimates $\hat{\mathbf{x}}_k^-$ can be used. The threshold depends on the size and the specific features of the template.

Line tracking

The assignment of lines as it is proposed in section 3.7.2 has difficulties when the gripper is almost closed, hence the opening distance is changed a lot. To overcome this problem it is easier to track the parameters of the horizontal lines l_{h1} and l_{h2}. The tracking is initialized by the line detection as described when the gripper state is open. The rule for assigning the lines is now based on the line parameters of the previous step:

For each horizontal line l_{hi} find the line that has the maximum votes and whose parameters ϕ and d_{0l} fulfill the constraints[12].

$$\begin{aligned} |\phi - \phi_i| &< 2^o \\ |d_{0l} - d_{0l,i}| &< 5. \end{aligned} \tag{3.22}$$

In order to use this for tracking of the position of the fixed and the moving gripper jaw the robot motion and the gripper opening has to be forwarded. This avoids the need for many tracking updates or big search regions in the line parameter space. Since the parameters of the lines are not in a cartesian coordinate system. The equations to feed-forward the robot motion $\{\Delta\mathbf{x}_{k-1}\}_i$ to obtain a prior line parameter estimate are:

$$\begin{aligned} d_{0l,i} &= \frac{\mathbf{n}_{l,i}^T}{\|\mathbf{n}_{l,i}\|} \frac{\{\Delta\mathbf{x}_{k-1}\}_i}{4} + d_{0l,i} \\ \phi_i &= \phi_i. \end{aligned} \tag{3.23}$$

[12]Note that the discretization for the angles is now set to 1^o

$\{\Delta\mathbf{x}_{k-1}\}_i$ only affects the parameter $d_{0l,i}$ in the direction of $\mathbf{n}_{l,i}$. The division by 4 is coming from the fact that the lines are tracked in images that are twice sampled down. Certainly, motion along the line does not change the line parameters. Since the gripper orientation stays constant ϕ_i is constant as well.

The parameters for the moving gripper jaw are updated considering the change in gripper opening distance $\Delta d_{opening}$:

$$\begin{aligned} d_{0l,1} &= \Delta d_{opening} + d_{0l,1} \\ \phi_1 &= \tfrac{\Delta d_{opening}}{d_{rn}} + \phi_1, \end{aligned} \tag{3.24}$$

where the small-angle approximation was used and d_{rn} is the distance from the rotational center of the gripper jaw to the point described by $\mathbf{n}_{l,1}$ in coordinates that correspond those of d_{0l}. This distance can be approximated by the distance from the rotational center to the end of the gripper which is known from CAD data.

3.7.4 Combined gripper-object detection

So far the object and the gripper detection were discussed separately. However the gripper obviously interacts with the object during the manipulation task. In fact, this poses a difficult problem since the object disturbs the detection of the gripper and vice versa. In particular, the object detection becomes more difficult because of its detection method: The image on the left in figure 3.45 shows an image of the gripper while it grips an object. The contour of the object and the gripper are connected and would be detected as one big contour (compare also figure 3.29). In order to avoid this, the following procedure is used:

- Detect the position of the gripper using line detection.
- Use the gripper position estimates and the orientation of the lines to fade out the gripper. Since the object detection classifies dark intensities as foreground, a thick white line is superimposed on the gripper tip (compare image in the middle of fig. 3.45).
- Detect the object as described in section 3.7.3.

Figure 3.45: Combined gripper-object detection while the object is gripped.

In figure 3.46 several examples for the combined gripper-object detection are presented. Spheres as presented earlier are gripped.

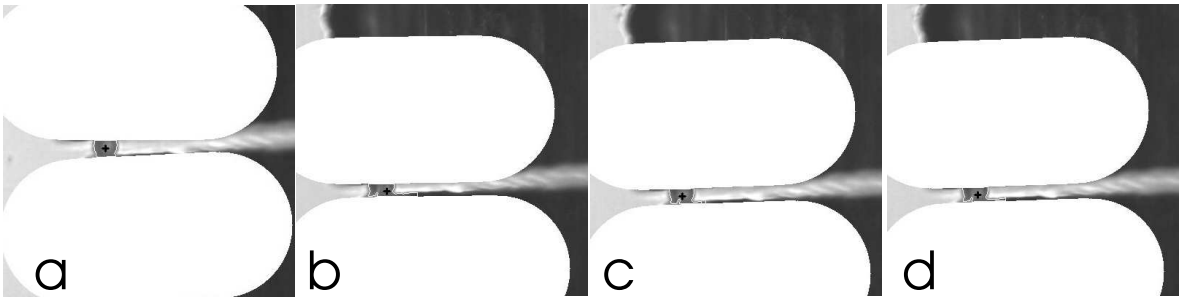

Figure 3.46: Examples of several combined gripper-object detection results.

In these image several problems of the combined detection can be seen: The gripper is not completely faded out so that these reminders are still counted towards the object contour (compare images (b)-(d)). This effect can be reduced by employing a morphological opening that removes the connection of the object to the gripper edge. Using a morphological opening here, it is very important that the object is in focus rather than the gripper, because blur effects might introduce brighter areas in the object center. These effects are no problem if the whole object is viewed, but if the object border is cut off, this part is not treated to be part of the object and hence the overall object contour is changed and the CG is biased. In images (b)-(d) the reminder of the lower gripper edge is too much, so that the object-gripper connection could not be removed leading to a bias of the object CG estimate. In image (a) the object detection is better: Because of the symmetric object geometry the CG is only slightly biased in x-direction, however more biased in y-direction because of the asymmetric occlusions. Note that for these four examples the upper line was not properly detected because a different line assignment strategy was employed. The above described effect is however similar.

Since the ROI for the object tracking is based on the previous bounding box size it is possible that the ROI increases in size heavily. This can happen when the reminder of the gripper is counted towards the object resulting in an increased size of the bounding rectangle in x-direction. This problem can be solved in restricting the possible size of the bounding rectangle. Here, the maximum size was chosen to be 1.3 of the object bounding rect size obtained during the object selection.

If the gripper is almost completely closed, there is the case in which the white lines touch each other. This is certainly the limit of this approach because the object cannot be detected anymore.

Note that the bias of the estimated object CG is reduced when the ratio of the object size to the gripper surface roughness is increased. Furthermore, the detection of objects with higher stiffness is easier, because of their small deformations.

The quantitative performance of this combined detection is discussed along the experimental system evaluation in chapter 5.

Chapter 4

End-effector calibration and control

For the manipulation of micron-sized objects either a two-jaw microgripper or a pipette is mounted onto the high-precision robot Δ^3 *version 2* (compare section 2.1). The accuracy of the Δ^3 *version 2* without calibration is about $10\mu m$. In order to improve the positioning results, the robot manipulates objects below the microscope. The relative position (x_f, y_f) between the origin of the FoV and the position of the robot zero position is known with a certainty of the robot accuracy, unless the microscope was moved. That means that it is possible to move the end-effector into or out of the FoV without prior detection. Once the gripper is in the FoV the robot can be controlled using the microscope vision feedback. A reliable gripper detection is only possible if the end-effector is at a certain z_f- position. Usually $z_f = 0$ is required and corresponds to the focal plane. In this chapter the

- height calibration of the end-effector relative to the focal plane/substrate and
- the detailed alignment procedure of the gripper and the object of interest

are presented.

4.1 End-effector height calibration

The positioning of the end-effector along the optical axis needs special attention because it cannot be determined directly from the microscope images. The manipulations are done in the focal plane so that the relationship between the z-position of the robot and the focal plane has to be determined.

The most used method to detect this coordinate is contrast-based which was already presented in section 3.6. Another method used by the authors of [6] is to push the ground with the end-effector and detect if the ground is moving. The end-effector is mounted with modelling clay to a robot so that it is questionable which repeatability can be achieved. The authors of [10] detect and correlate edges at the tip of their end-effector.

Three methods were investigated in this work to obtain height information about the end-effector, respectively its calibration.

- Image contrast

- Correlation
- Employing optical effects

The three methods will be presented in the following and applied to the two different manipulators. The methods based on optical effects are only used for the two-jaw gripper. The examinations are done $\pm 30 \mu m$ around the focal plane, since there is enough a-priori knowledge to move the end-effector from its zero position into this range.

4.1.1 Contrast

As already discussed in section 3.6 the image sharpness is maximized. Moving the tip of the end-effector into the focal plane, the sharpness at the tip of the end-effector will be maximal. The autofocus method described in 3.6 uses the whole image and detects only a plane well. Since the end-effector is tilted towards the optical axis only a small part of the end-effector can is in focus. Therefore a ROI $\mathcal{R}_a$ is defined at the tip of the end-effector. Figure 4.1 depicts the used $\mathcal{R}_a$ for the tip of a pipette. In order to find the ROI the correlation as described in section 3.7.2 is used. Only in this region the contrast is calculated.

In order to determine the threshold T_a for the end-effectors the histogram of the gradient magnitudes is used. Figure 4.2 shows for example the histogram for the case of the two-jaw gripper. The high gradient magnitudes belong to the edge of the gripper. The very small magnitudes correspond to noise. The threshold is selected that this noise is filtered out and the threshold was set to $T_a = 25$.

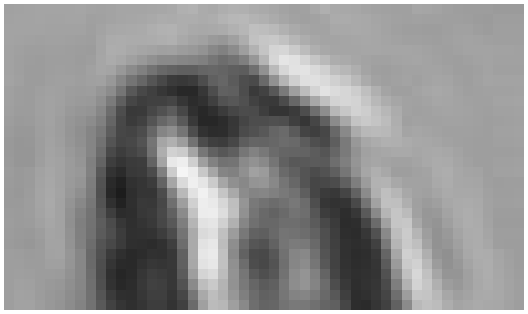

Figure 4.1: ROI $\mathcal{R}_a$ at the tip of a pipette (size: 51×30) used for the contrast-based autofocus.

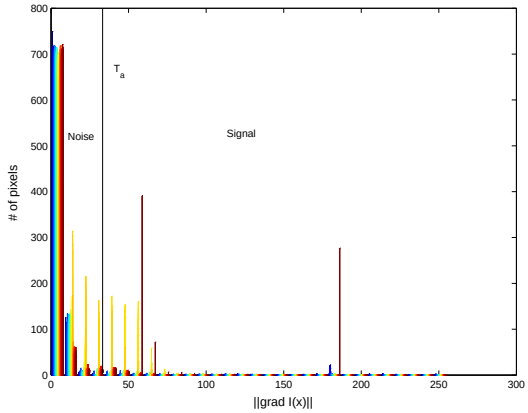

Figure 4.2: Histogram of the gradient magnitudes of the ROI $\mathcal{R}_a$ that contains the tip of the two-jaw gripper in focus.

The main advantage of the contrast-based methods is the fact that no precise knowledge about the end-effector geometry is required. Still a template of the tip is needed however it must not be a defined exactly in focal plane, because it is only needed to determine the position of $\mathcal{R}_a$. A drawback of this method is the sensitivity to changes in illumination. For example a high illumination produces higher gradients at edges and hence a higher contrast.

4.1.2 Correlation

Normalized correlation as earlier described in equation 3.18 for the lateral gripper position detection was also employed to obtain height information about the gripper: A template is taken when the tip of the end-effector is in focus. Correlating the template with the image at different heights, the height with the maximum correlation value corresponds to the situation in which the tip of the end-effector is in focus.

The main difficulty and the limiting factor is the definition of a template that corresponds to the tip of the end-effector when it is in the focal plane. There are two ways to obtain a template: One approach is to move the tip of the end-effector manually into focus and define a template. Another approach is to find the template by micro-pushing (compare figure 4.3):

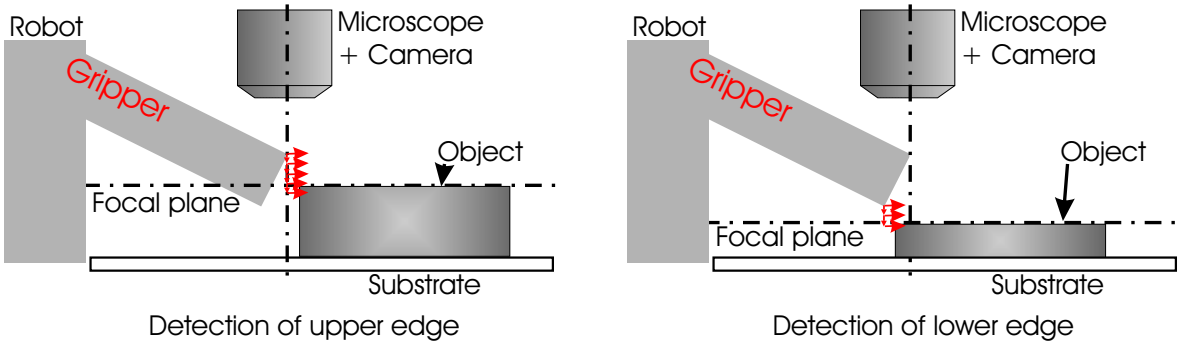

Figure 4.3: Process to bring either the upper or the lower edge of the two-jaw gripper into focus.

An object with a surface, which is parallel to the focal plane[1], is brought into focus. The end-effector is driven closely above the focal plane and positioned next to the object. The end-effector is then moved into the direction of the object to push it. Certainly if it is above the object, the object does not move. Step-wise the position of the robot is lowered by $100nm$ and moved back and forward ($500nm$) until the object is pushed. Then the whole substrate on which the object lies is lowered so that only the end-effector stays in the image and the template can be defined. Depending how this motion is performed either the upper or the lower edge of the gripper can be detected (see also figure 4.3). The pipette has no edges in that sense but its tip can be brought into focus using the same technique.

The drawback of this approach is certainly that the definition of the template has to be done for every single end-effector. Furthermore, this procedure is limited by the repeatability of the autofocus to bring the objects into the focal plane. Another limitation are attracting forces between the end-effector and the object. So that the object might be pulled by those forces and an object motion can be observed although the gripper is still above it. The template can be defined considering the constraints discussed in section 3.7.2.

One advantage of this method is that it works for arbitrary shaped end-effectors (assuming there are enough features at the tip). This includes also that it works for most angles β. Another advantage is the robustness against changes in illumination as they were already

[1] Objects with surfaces parallel to the focal plane can be brought into focus very robust (compare 3.6 and [1]

discussed earlier in section 3.7.2. The possibility to define a template of the upper edge while the lower edge is at the level of the focal plane makes it possible to move the gripper close to the substrate without crashing into it.

Note that if *tracking of the end-effector height* is required the correlation can be used with any other technique to improve results: Usually autofocus techniques cannot determine in which direction the end-effector moved out of focus. The scanning into the wrong direction is time-consuming. Assuming the correct height is known, templates can be taken of the tip several μm below and above the focal plane. Since these images look quite different a correlation with both of these templates can tell wether the end-effector is below or above the focal plane so that scanning into the direction towards the focal plane is guaranteed.

4.1.3 Optical effects

Under different illumination conditions optical effects are employed. Four different setups are considered. All of them have in common that the substrate is involved and that rather the optical flow is used than single images at different positions. The basic idea is to approach the substrate step-wise and to detect if something in the image changes. This is a similar approach as shown in [6], however the authors pushed the whole substrate and detected its motion. In this setup the substrate is fixed and has a much higher stiffness than the gripper. Once the gripper touches the substrate it will bend. This effect is however so small that it can hardly be observed in the microscope images when the gripper is moved several μm into the substrate. Furthermore, it is difficult to distinguish if a lateral movement of the gripper tip corresponds to a bending due to contact with the substrate or if it corresponds to a misalignment of the z-axis of the manipulator to the optical axis. Instead, illumination effects (diffraction and shadows), that produce a significant change in the image when the gripper height is changed only minimal, are exploited. That means that there is a sufficient optical flow when the height of the gripper is changing and that there will be a very low optical flow when the height of the gripper does not change. This means that few optical flow corresponds to the scenario where the end-effector touches the substrate.

In order to determine the changes in an image over time a very easy approach is to take an image at t_0, move the end-effector and take another image at t_1. Subtracting the second image from the first and calculating the L_1 norm N of the subtraction can be written as

$$N = \|I(\mathbf{x}, t_1) - I(\mathbf{x}, t_0)\|_{L_1} = \sum_{\mathbf{x} \in \mathcal{I}} |I(\mathbf{x}, t_1) - I(\mathbf{x}, t_0)|. \tag{4.1}$$

Comparing the sum of changes in an image N with an threshold N_{th} can then be used to detect if the end-effector is moving or already touching the substrate.

Figure 4.4 depicts the principle of the first setup: A glass substrate in combination with backlighting is employed. The focal plane is aligned with the substrate. This means that the upper edge of the gripper will appear blurry even when it touches the substrate with its lower edge. This is due to the thickness of the gripper of $50\mu m$, the tilting angle β and the depth of focus for the $20\times$ objective of $1.6\mu m$. Furthermore, diffraction patterns are observed which move when the gripper height is changed. The idea is to detect the change in the images.

Once the gripper touches the substrate the diffraction pattern do not change anymore, since the gripper height stays constant.

Setup 2 is presented in figure 4.5. Instead of a glass substrate a metal substrate is employed. This makes the use of backlighting impossible. In order to minimize the direct reflections of the light into the microscope and to maximize the reflections at the upper edge of the gripper, the directed light has only a small angle towards the focal plane. Again the dependence of the diffraction patterns on the gripper height is used.

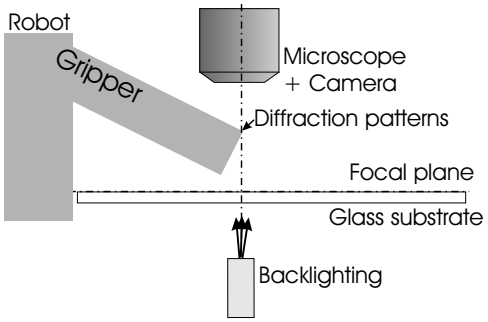

Figure 4.4: Calibration of the gripper height: Setup 1.

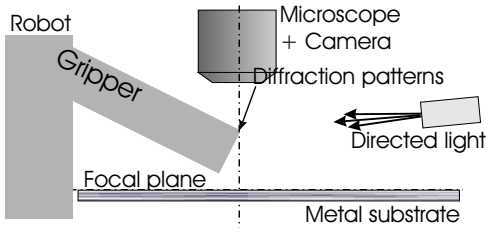

Figure 4.5: Calibration of the gripper height: Setup 2.

The third setup is shown in figure 4.6. A metal substrate is used and the illumination is above the gripper and points towards the gripper tip. Diffraction patterns and shadow effects can be observed that can be used to calibrate the gripper height.

Figure 4.7 presents the fourth setup: A metal substrate is used and the light passes between the substrate and the gripper, almost aligned with the focal plane. The gripper throws a shadow on the metal substrate whose lateral position is very sensitive to the height of the gripper. The idea to calibrate the height is to approach the substrate until the shadow does not move anymore.

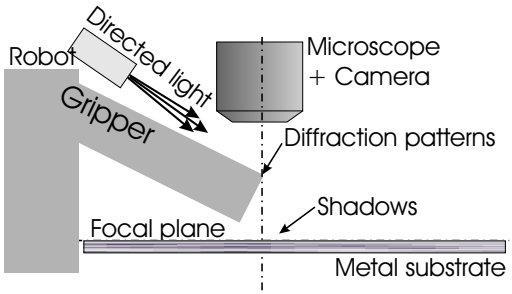

Figure 4.6: Calibration of the gripper height: Setup 3.

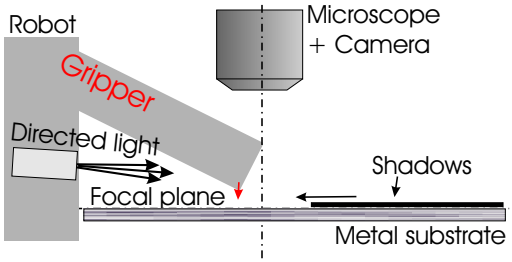

Figure 4.7: Calibration of the gripper height: Setup 4. Small motions of the gripper result in large motion of its shadow.

These methods appear interesting since the resolution is expected to be high (especially for setup 4) and the position of the lower edge of the gripper is detected directly. A drawback is introduced by the use of the optical flow: Optical flow techniques are sensitive to changes

in illumination. If there is a change in the intensity of a background pixel from one frame to another this will be interpreted as optical flow.

4.1.4 Experimental evaluation

Autofocus and correlation

Experiments were performed to evaluate the functionality of the presented concepts: Over a range of $\pm 30\mu m$ with a step size of $1\mu m$ close to the focal plane, the autofocus and the correlation methods were tested using a similar hardware setup as described in section 3.6. Figure 4.8 shows the setup for the height calibration.

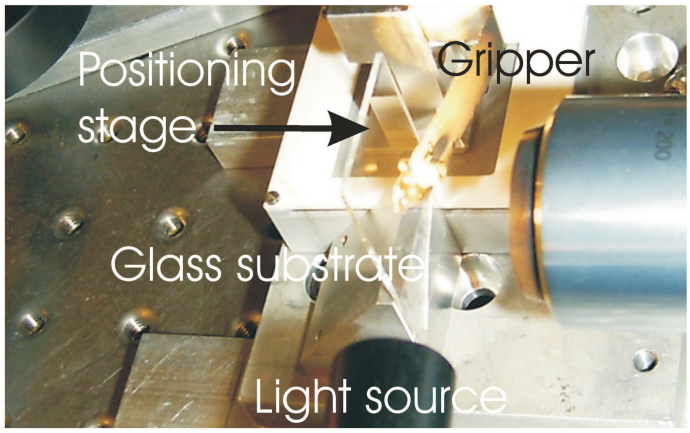

Figure 4.8: Gripper mounted on a high-precision stage. The substrate is fixed to the mounting plate of the stage.

In order to discuss qualitatively the different methods, each experiment was performed three times at different positions of the focal plane so that correspondingly the obtained sharpness over height curves should be shifted. The similarity of the shape of these profiles indicates the repeatability of the method. Another aspect is the curvature of the sharpness peak. A thin peak corresponds to a high repeatability. Most important however is the fact that the detected maximum in sharpness really corresponds to the situation in which the tip of the end-effector is in focus. In order to compare the contrast-based and the correlation method, both criteria were evaluated parallel during the experiments. The templates for the end-effectors were obtained manually. A comparison between the images, which were obtained at the position with maximum correlation, and the image, which were taken when defining the template, showed that the two images were almost identical. This makes it possible to evaluate the contrast-based function relative to the correlation results.

An important parameter is the angle β which is defined between the orientation of the end-effector and the focal plane (compare figure 2.3).

The experiments were performed for several angles β for a small ($1\mu m$ diameter at the tip) and a big ($5\mu m$ diameter) pipette and for the two-jaw gripper. The result for the big pipette are presented in the figures 4.9, 4.10 and 4.11.

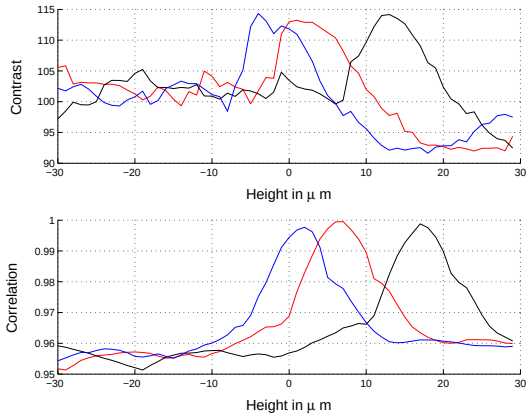

Figure 4.9: Contrast and correlation calculated at different height positions of the pipette ($\beta = 0^o$).

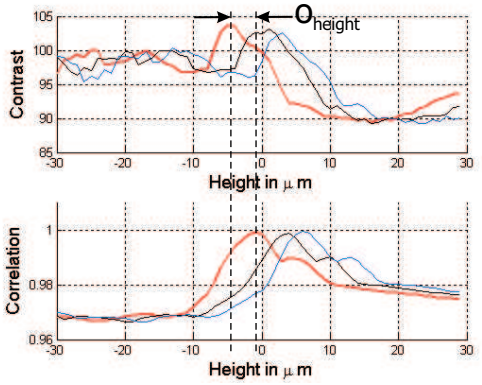

Figure 4.10: Contrast and correlation calculated at different height positions of the pipette ($\beta = 10^o$).

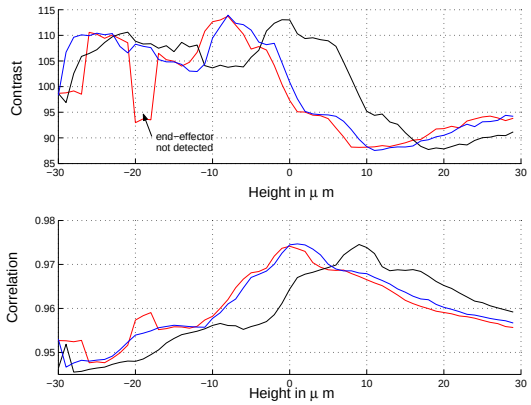

Figure 4.11: Contrast and correlation calculated at different height positions of the pipette ($\beta = 30^o$).

It can be seen that for $\beta = 0^o$, the maxima of the contrast function is broader than for the other β. Additionally, it can be seen that the shape of the peak changes. This is due to the transparent property of the pipette: Depending on the β the pipette refracts the light in a different way. In figure 4.12 the big pipette is imaged for different cases: Image (a) shows the pipette in which the tip is brought into focus manually. In image (b)-(d) the height with maximum contrast within the ROI are presented. With an increased angle β it can be seen that a very bright spot almost at the end of the tip has an increased intensity.

Due to these refractions (in particular the bright spot) there are heights at which the pipette tip focussed some of the backlight resulting in a high contrast in the images. This generates a high contrast in the image so that this height is interpreted to be the tip of the pipette. A constant displacement of the height of the pipette is the consequence. In order to obtain the *true* height of the tip of the pipette, an height offset o_{height} is needed. Such an offset is indicated in figure 4.10. This offset is based on the correlation and hence requires the definition of a reference template or the use of some other calibration method.

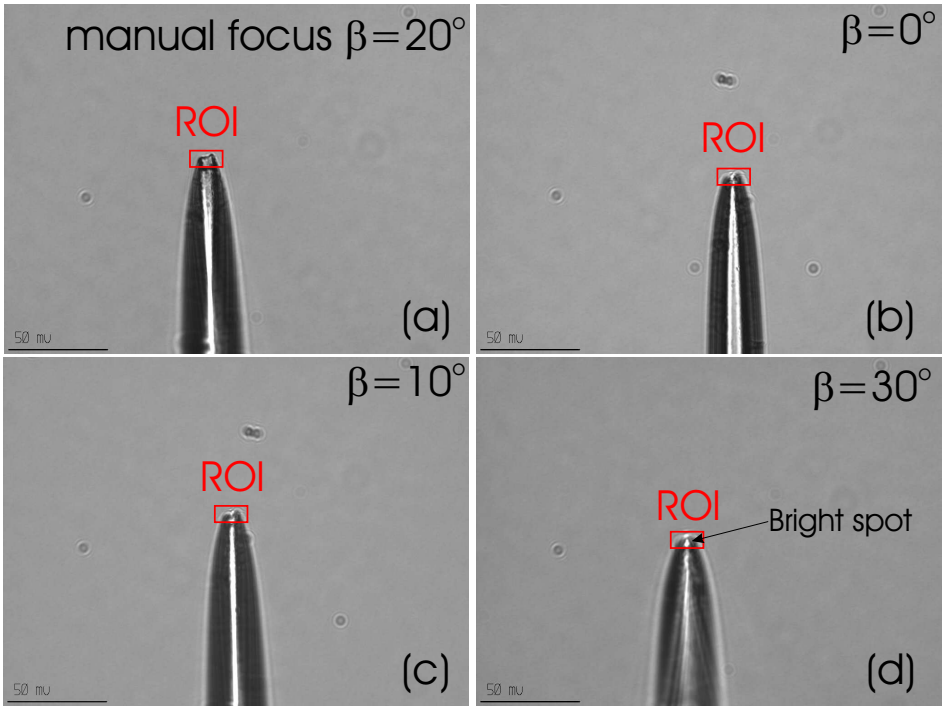

Figure 4.12: Images of the big pipette: (a) Pipette is manually driven into focus (b) Pipette in focus at the height determined by the autofocus for $\beta = 0^o$ (c) for $\beta = 10^o$ (d) for $\beta = 30^o$.

Figure 4.13 presents the results obtained for the small pipette ($\beta \approx 30^o$): The maxima of the contrast and of the correlation fall almost together. The contrast-based method shows noise but the peaks are smaller than the one of the correlation method. On the right side of figure 4.13 the tip is shown at the height z_0 as determined by the autofocus. The images below and above the focal plane show images of the tip out of focus (offset from the focal plane $\pm 15\mu m$).

The results obtained for the two-jaw gripper are depicted in figure 4.14 for $z \in [-30\mu m, 30\mu m]$ and in figure 4.15 for $z \in [-3\mu m, 3\mu m]$. For each run the contrast over height curve shows similar maxima. There is some noise in the curve which is evoked by varying illumination. Since this calibration process is not time-critical, instead of using one image at each height position, several images at one height can be averaged to reduce noise. Image 4.16 shows the gripper at the height at maximum contrast in the ROI. Along the edge of the gripper surface details of the gripper can be seen clearly. Subjectively, the tip of the gripper is in focus. Considering the roughness of the end-effector, a repeatability of $\pm 1\mu m$ seems sufficient for the present setup. A possible improvement of this result is to fit a parabola to the peak of the height-contrast curve and to determine the maximum of the parabola.

With the microscope mounting alternative 2 (*from below*) a repeatability test was performed and 20 times within 5 minutes the gripper position was detected. The step-size was set to $250nm$ and the range was set to $\pm 3\mu m$ around the initially detected proper height: the standard deviation of the results of the contrast-based function was $0.52\mu m$ and the one of the correlation-based method was $0.24\mu m$. Note that for the correlation-based method the template was defined briefly before the test. The influence of changes of the end-effector (for

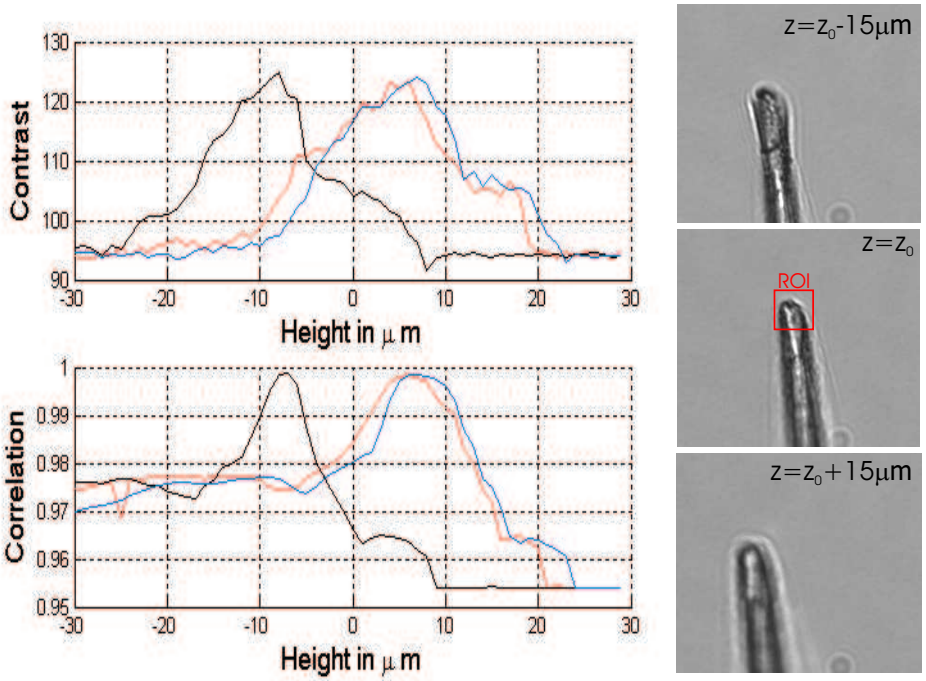

Figure 4.13: Contrast and correlation calculated at different height positions of the pipette ($\beta \approx 30^o$). On the right side the images of the pipette are shown at different heights: The pipette is in focus in the ROI at $z = z_0$.

example dirt, scratches) were not examined.

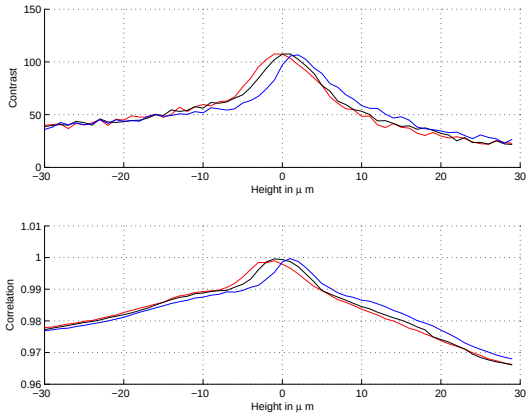

Figure 4.14: Contrast and correlation calculated at different height positions of the gripper ($\beta = 20^o$).

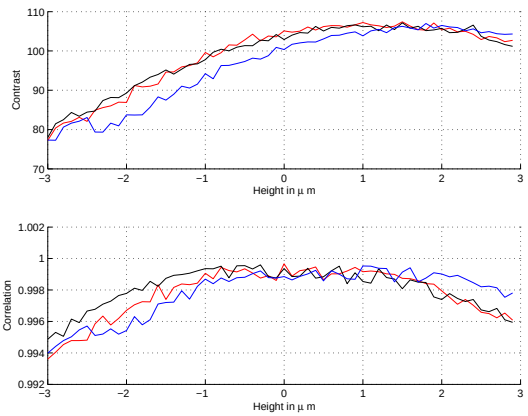

Figure 4.15: Contrast and correlation calculated at different height positions of the gripper ($\beta = 20^o$) for a step-size of $100nm$.

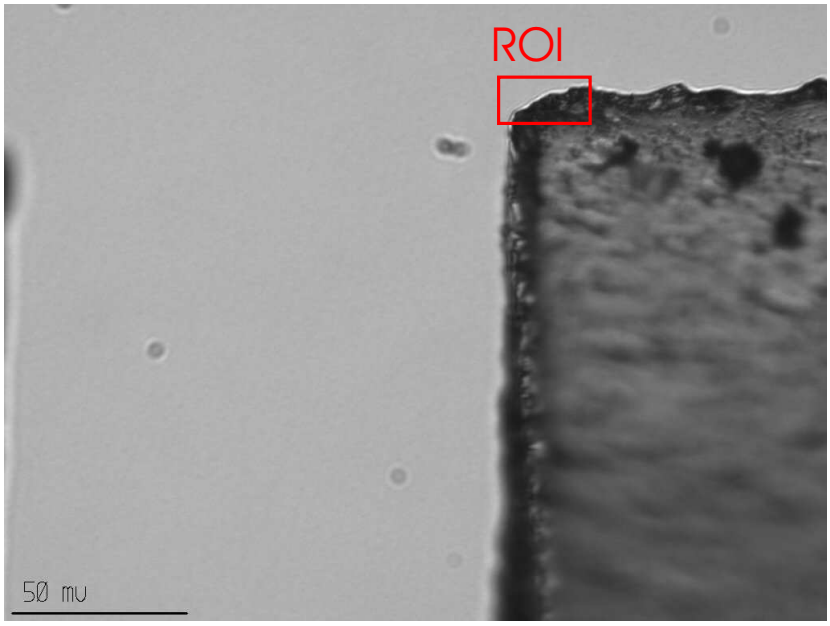

Figure 4.16: Image of one finger of the gripper at the height determined by the contrast-based method.

Optical effects

The experimental results of the four different setups exploiting optical effects are described in this section. The first three setups are discussed only qualitatively and for setup four quantitative results are presented.

In figure 4.17 the results for setup 1 are illustrated: At different distances z relative to the substrate images were taken. The optical flow from the previous image to the following image is indicated qualitatively with arrows. It can be seen that the diffraction pattern approaches the edge of the gripper when coming closer to the substrate. The change from image to image does not seem very high so that no good position resolution can be expected using equation 4.1. In order to obtain better results the fact that the direction of motion is known can be exploited. Alternatively, the diffraction pattern could be tracked and when no significant change in position is detected the gripper touches the substrate. The sensitivity could be also increased when the angle β is increased, since more diffraction patterns can be observed far away from the tip.

The results for setup 2 are depicted in figure 4.18. Approaching the substrate there is an optical flow into the direction of the y-axis of the image. This is mainly due to shift of the bright reflective region of the image that is moving when approaching the substrate. The magnitude of the optical flow is however not very high. When the gripper touches the surface there can be an optical flow observed against the direction of the y-axis. This could be used to improve the height detection resolution.

One images that shows the gripper in contact with the substrate obtained with setup 3 can be seen in figure 4.19.

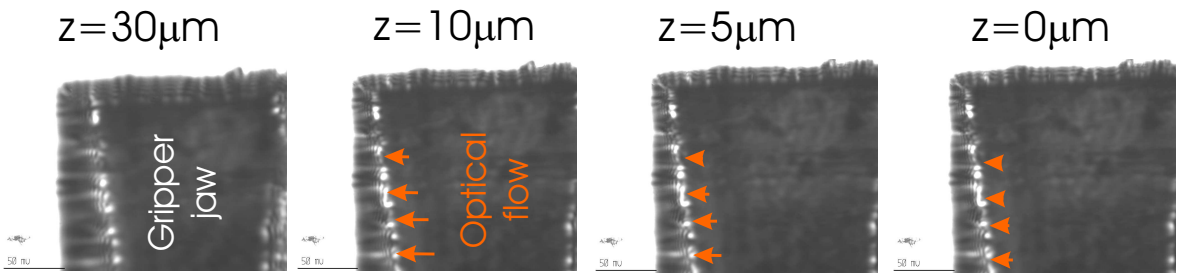

Figure 4.17: Results for setup 1 in combination with the gripper. The images show the gripper at different heights relative to the substrate. The optical flow is indicated with arrows.

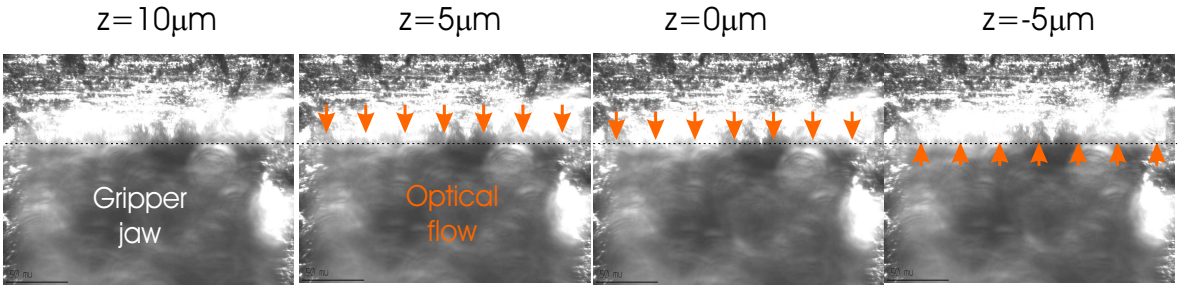

Figure 4.18: Results for setup 2 in combination with the gripper. The images show the gripper at different heights relative to the substrate. The optical flow is indicated with arrows.

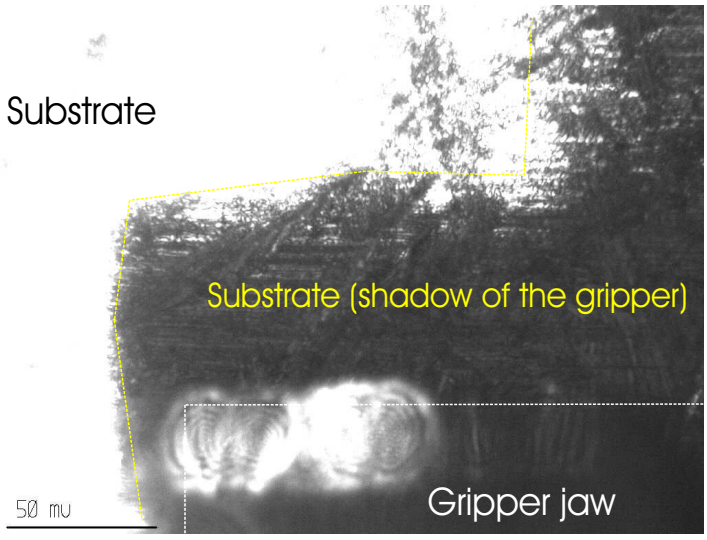

Figure 4.19: Image of the gripper touching the substrate for setup 3 in combination with the gripper.

The gripper (indicated by the lines in the lower part of the image) is touching the substrate and it throws shadow onto the substrate (yellow line). At the edge of the gripper several diffraction patterns can be observed. As the sequence of images in figure 4.20 illustrates, the diffraction patterns shrink when coming close to the substrate. Once the substrate is touched

and the gripper is pushed onto the substrate the radius of the diffraction pattern virtually does not change but moves against the direction of the y-axis. Again, the overall changes during the steps are not high. An alternative detection principle could be to track the radius of the diffraction pattern and detect at which height the radius stays constant and starts moving laterally.

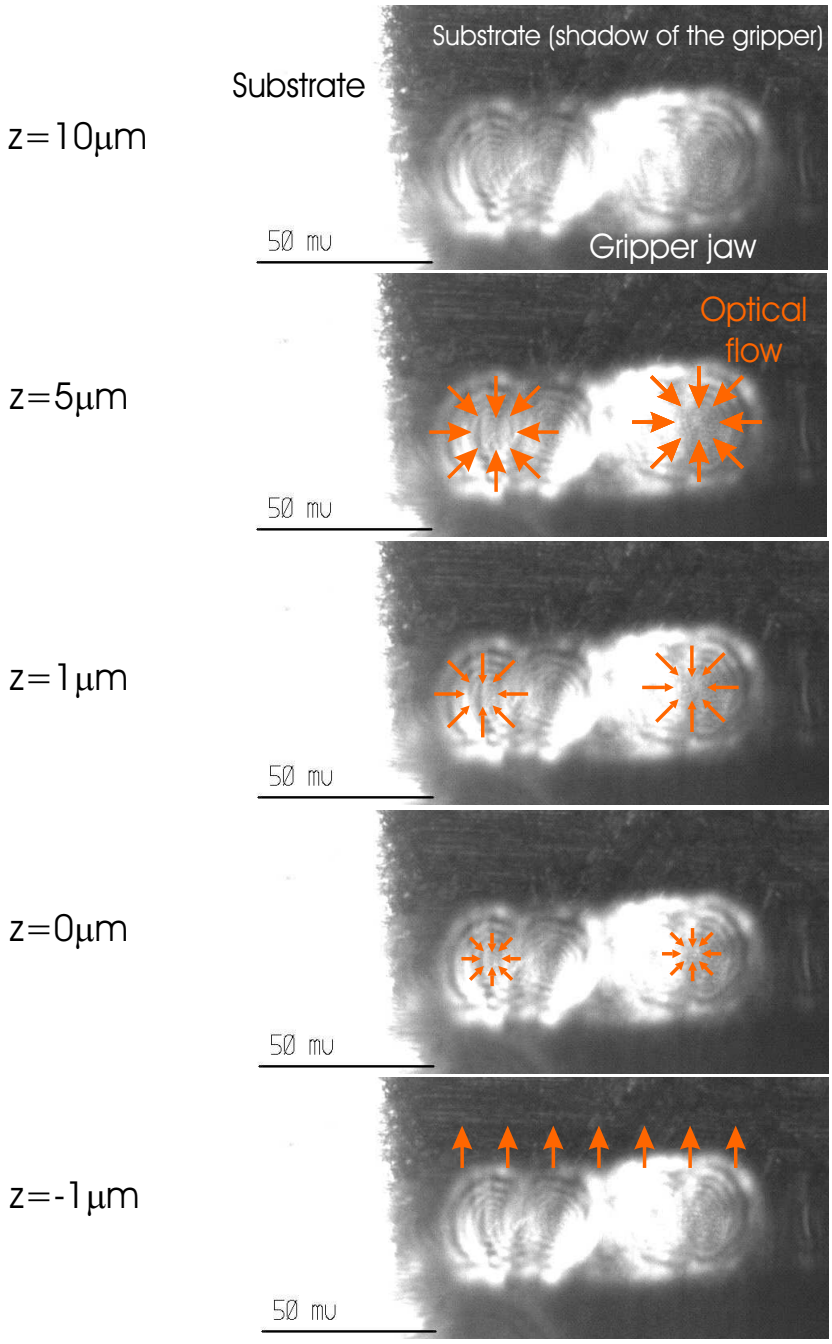

Figure 4.20: Results for setup 3 in combination with the gripper. The images show the gripper at different heights relative to the substrate. The optical flow is indicated with arrows. Only the region of interest is shown.

The results of the fourth setup are shown in 4.21: The gripper throws a big shadow that is moving when approaching the substrate. When the substrate is touched the shadow stops moving, however a small motion of the edge can be observed when pushing into the substrate.

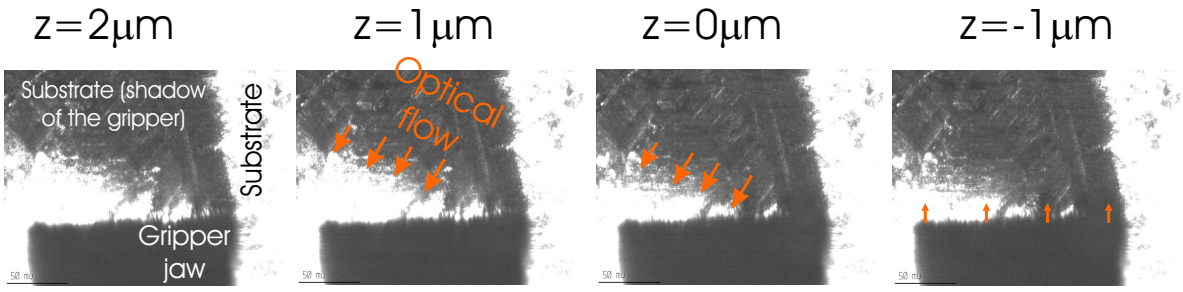

Figure 4.21: Results for setup 4 in combination with the gripper. The images show the gripper at different heights relative to the substrate. The optical flow is indicated with arrows.

Figure 4.22 depicts the result of a calibration based on equation 4.1. The gripper was positioned within several μm away from the substrate and was approached in steps of $1\mu m$. At each step an image was taken and N was calculated and compared with the threshold $N_{th} = 100000$. N_{th} was selected such that it is about 50% greater than the N when the height is not changed between the capture of the two consecutive images. In order to improve results, a morphological opening (mask size 5×5) was performed to reduce noise before calculating N. For example the detected height in the experiment marked by the thick line is at $z = 8\mu m$, because $N(z = 9\mu m) < N_{th}$. During the four experiments depicted by the different curves the relative distance between the substrate and the gripper changed $\pm 1\mu m$. When the gripper pushes into the surface the stiffness of the mounting (substrate and gripper are taped to their support) is not sufficient so that there is a small displacement. The repeatability can be given then as the step size as $\pm 1\mu m$.

The results could be certainly improved when really the optical flow in the image is calculated. Another improvement could be achieved only considering the optical flow of the region where the shadow changes. Additionally, assuming the relative position of light source and substrate does not change, the expansion direction of the shadow is known. Projecting the optical flow onto this direction should eliminate even more disturbances.

Note that the methods presented in the setups 2-4 can be also used in combination with a transparent substrate: Onto a part of a transparent substrate a layer of opaque material could be deposited with a known height. This layer can be used for the height calibration and the manipulations are done in the regions without depositions so that backlighting can be used.

Recognizing the change in the diffraction patterns can be certainly also exploited with the pipettes. However, the surface needs to be approached slowly, because of the high fragility of the pipette tips. When a pipette is coated with an opaque material the technique with the shadow can be applied, too.

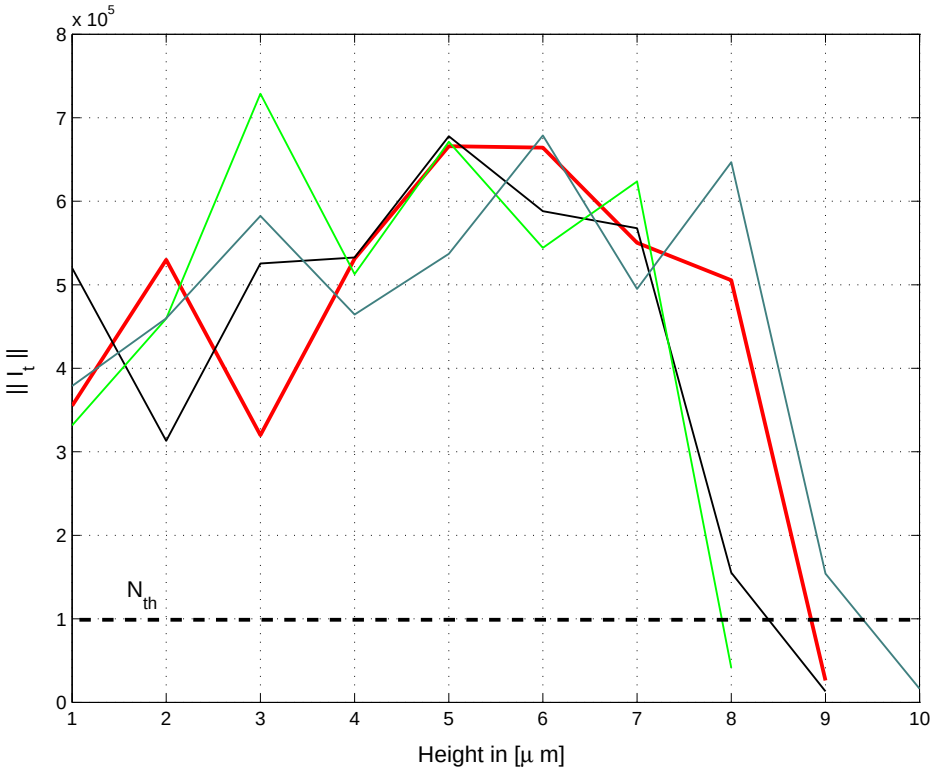

Figure 4.22: Results for setup 4 using equation 4.1 to detect the substrate height.

4.1.5 Summary

Several methods to detect, respectively to calibrate the height position of an end-effector were developed. The well-known autofocus technique, correlation with a reference template and concepts using optical flow were discussed. One advantage all methods have in common is that *no external sensor* is required to obtain the height information. A completely different approach uses a special calibration tool tip (for example a touching probe) that can detect the substrate position with a precision in the nanometer range. In this case the tool needs to be changed or the distance between the calibration tool tip and the actual end-effector must be known very well and requires certainly much more calibration effort. The presented methods have the advantage that the height of the actual end-effector is measured directly. Table 4.1 summarizes the advantages and disadvantages of the presented methods.

Method	**Advantages**	**Disadvantages**	**Repeatability in $\pm\mu m$**
Auto-focus	• Few knowledge about geometry required	• Diffraction patterns and other optical effects are interpreted as high contrast • Problems detecting tilted surfaces • Sensitive to changes in illumination	$\approx 3/\approx 0.5/0.5$
Correlation	• Robustly repeatable • Robust against changes in light • Works for most angles β	• Requires the definition of a *precise* template for each end-effector	$\approx 1/\approx 1/0.24$
Optical flow	• Position of lower edge is detected directly needed for collision avoidance	• Sensitive to changes in illumination	-/-/<1

Table 4.1: Comparison of the presented methods to detect the height of the employed end-effectors. The repeatability is given for the big pipette/small pipette/gripper.

Depending which height calibration is needed the following methods should be applied:

- For the *small pipette* and the *upper edge of the gripper* the contrast-based method provides a high repeatability of $\pm 0.5\mu m$. Especially, few knowledge is required about the end-effector's tip.
- The *lower edge position of the gripper* cannot be determined by the contrast-based function since it is occluded the upper edge. The optical flow method of setup 4 gives a repeatability of $\pm 1\mu m$ to detect the lower edge. Note that the lower edge can be viewed if the *microscope is mounted horizontally* and viewing from below the substrate (compare figure 3.8).
- For the *big pipette* correlation showed the best results with an estimated repeatability of $\approx \pm 1\mu m$.

4.2 Relative position control

The relative position control is used to overcome the necessity of a completely calibrated system. All manipulations take place in the FoV of the microscope. The fact that the lateral

position of the FoV is given by the accuracy of the robot[2] and hence the *global position accuracy* is not important, since the goal is to achieve a high *relative position accuracy* only within the FoV.

This paragraph gives a detailed description how the end-effector is aligned with the object of interest (compare the points *Start object and gripper tracking* and *Move gripper towards object* in the complete manipulation sequence in figure 2.6). At this moment in the manipulation sequence the gripper and the object of interest are both in the FoV and their positions are estimated.

- *Determine relative initial error in focal plane coordinates:* The relative position error $\mathbf{e} \in \mathbb{R}^2$ can be formulated as

$$\{\hat{\mathbf{e}}\}_i = \{\hat{\mathbf{x}}_O\}_i - \{\hat{\mathbf{x}}_E\}_i, \tag{4.2}$$

  where $\{\hat{\mathbf{x}}_O\}_i$ and $\{\hat{\mathbf{x}}_E\}_i$ are estimates of the definitions of the gripper and the object coordinates as defined in section 3.7.

  The inverse of the calibration matrix is used to determine the error $\{\hat{\mathbf{e}}\}_f = \mathbf{T}^{-1}\{\hat{\mathbf{e}}\}_i$ in focal plane coordinates (compare equation 3.10).

- *Move to approach point:* Based on $\{\hat{\mathbf{e}}\}_f$ and the bounding box of the object a relative motion $\Delta\mathbf{x}_f$ to an approach point P_a can be calculated:

$$\Delta\mathbf{x}_f = \begin{bmatrix} \{\hat{e}_x\}_f \\ \{\hat{b}_{lower}\}_f + d_{approach} - \{\hat{y}_E\}_f \end{bmatrix}, \tag{4.3}$$

  where $\{\hat{b}_{lower}\}_f$ is the position of the lower boundary of the object bounding rectangle and $d_{approach}$ is an initial distance to the border of the object.

  The latter parameter is used as a security distance: if the manipulator travels a long trajectory (several $100\mu m$) the uncertainty in the camera calibration introduces an error so that a direct alignment is only reasonable with a perfect calibration matrix. $d_{approach}$ was set to $5\mu m$. In order to avoid collisions with objects, the gripper is lifted $50\mu m$ above the focal plane before moving to P_a. The motion commands are given using the robot programming interface of the Δ^3 enabling the generation of robot trajectories. The calculated desired robot motions are sent to the robot as linear trajectories with the desired position at its end.

- *Align with object:* The positions of the object and the gripper are calculated repeatedly and at each time step k the robot will be moved accordingly to

$$\Delta\mathbf{x}_f = a \begin{bmatrix} \{\hat{e}_x\}_f \\ \{\hat{b}_{lower}\}_f - \{\hat{y}_E\}_f \end{bmatrix}, \tag{4.4}$$

  where $a \in (0, 1]$ parameterizes the control error[3]. This will be repeated until

$$\|\Delta\mathbf{x}_f\| < \varepsilon_{alignment}, \tag{4.5}$$

[2] Assuming the relative position of the zero position of the robot and the microscope does not change.
[3] a is rather a parameter than a control gain, since it merely adjusts the length of a linear trajectory.

where $\varepsilon_{alignment}$ gives the desired precision of the alignment between the lower object boundary and the gripper.

Convergence considerations: Assuming an underlying stable cartesian robot controller and considering $\Delta\mathbf{x}_f^k$ as a time-dependent function on k, one can write for iteration $k+1$:

$$\Delta\mathbf{x}_f^{k+1} = \Delta\mathbf{x}_f^k - a\Delta\mathbf{x}_f^k = (1-a)\Delta\mathbf{x}_f^k \tag{4.6}$$

The sequence $\{\Delta\mathbf{x}_f^k\}_1^\infty$ is *strictly monotonic decreasing* and is *bounded* by the upper limit $\Delta\mathbf{x}_f^0$ and will reach the lower bound 0 for $k \to \infty$ for $a \in (0, 1]$.

Chapter 5

Experimental system evaluation

The manipulation system was tested to manipulate a polystyrene ball with a diameter of $15\mu m$ that is moved to a user specified destination in FoV coordinates. The experiment was performed at a room temperature of 23^oC and a humidity of 35%. The manipulation sequence is similar to the sequence as described in section 2.2 and the differences and the reasons of the modifications will be explained. During one manipulation several images were captured which are depicted in figure 5.1:

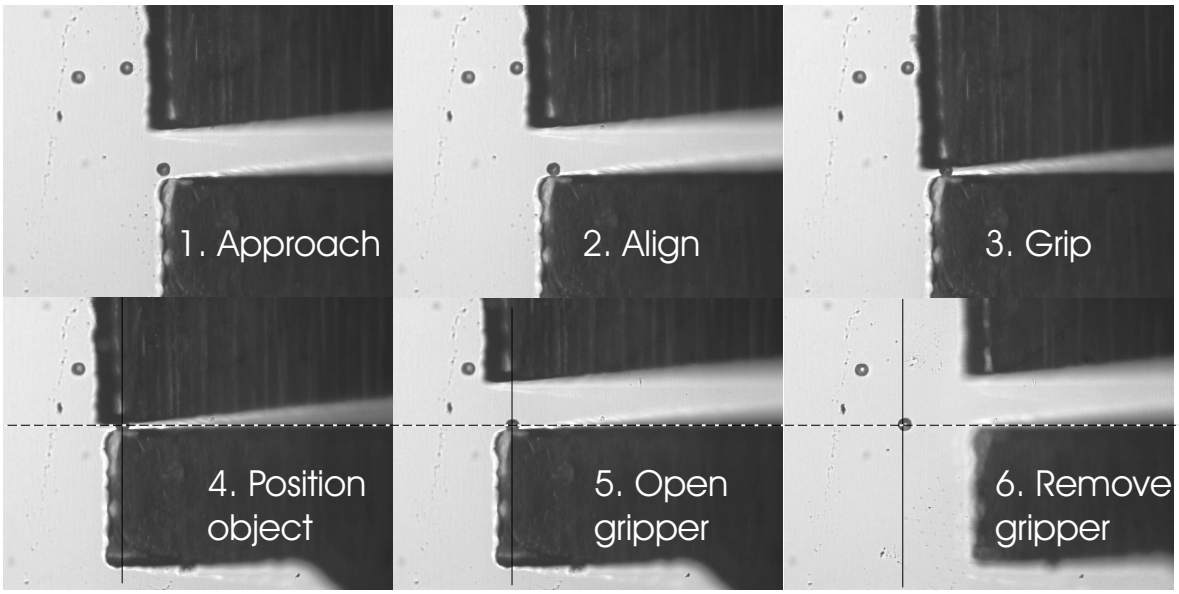

Figure 5.1: Microscope images captured at key moments during the manipulation experiment of a polystyrene ball.

- *Image 1*: After calibrating the height of the end-effector and the substrate, both are moved in the focal plane. The ROI of the autofocus was applied to the jaw whose distance to the focal plane is greater. This means that the other jaw will be pushed into the substrate and hence the height misalignment between the gripper jaws is *corrected*. After the user selected the polystyrene ball and the gripper is detected, the relative

position is used to direct the gripper to the approach position.

- *Image 2*: The lower gripper jaw is aligned with the object. A position tolerance of $2\mu m$ is selected, so that the gripper is aligned in average after 3 iterations.
- *Image 3*: The gripper is closed manually. It can be seen that the ball is squeezed slightly and is also partially occluded by the shadows of the gripper. Also the object moved during the gripper closing.
- *Image 4*: The user can select now the destination coordinates. In this sequence the desired object position $\{\mathbf{x}_{O,CG,des}\}_f$ was set to $[150\mu m, 150\mu m]^T$. After this selection the object was positioned using the error $\mathbf{e}_O$ between $\{\mathbf{x}_{O,CG,des}\}_f$ and the object position estimate $\{\hat{\mathbf{x}}_{O,CG}\}_f$. This means that the object position is controlled directly and object motion within the gripper can be compensated. The object is aligned until $\|\mathbf{e}_O\| < 300nm$. Note that the gripper was neither lifted nor lowered when moving from the pick-up position to the desired position, because of the height asymmetry of the gripper jaws. Lifting the gripper from the substrate would result in a re-orientation of the height of the gripper jaws leading to a possible heavy squeezing of the object.
- *Image 5*: After the positioning the gripper is opened. The ball stays at the lower gripper. This behavior was observed in more than 95% of the experiments.
- *Image 6*: The manipulator was moved $60\mu m$ above the substrate and moved $50\mu m$ in the $x-$direction. The ball stuck to the gripper at the end of the manipulation sequence although the object sticks rather to the substrate than to the gripper. This can be explained that during the gripping the object was lifted within the gripper so that the object is not touching the substrate anymore and the gravity forces are not sufficient to overcome the surface forces between gripper and object. The gripper was tele-operated back to the release position. With tweezers the substrate was hit carefully to introduce vibrations with small amplitudes. This intervention was sufficient to re-establish the object-substrate contact and the gripper could be removed without any further influence on the object position. The object position was determined again and the object positioning error was measured to $\mathbf{e}_O = [1.2\mu m, 3.3\mu m]^T$, $\|\mathbf{e}_O\| = 3.5\mu m$. This error is higher than the expected error stated in the following. The main reason is the *forced object* release.

Repeatability considerations

Considering only one manipulation the system performance cannot be evaluated. 30 experiments with spheres were conducted. Some of the spheres were manipulated multiple times. In only one case an object stuck to the actuated gripper jaw (3.3%). In 9 cases (30%) the object could not be released directly after opening the gripper. In 4 cases (13%) the object was released after tipping onto the substrate. During all these experiments the object positioning strategy was modified, so that the discussion of the positioning measurements are based on the last 13 experiments. The results are presented in table 5.1 and 5.2.

In table 5.1 it can be seen that the positioning errors in x-direction are roughly 50% of the one in the y-direction. This is due to:

$\mathcal{E}\{\mathbf{e}_O^T\}$ in μm	$\sqrt{\mathcal{V}\{\mathbf{e}_O^T\}}$ in μm	$\min\{\mathbf{e}_O^T\}$ in μm	$\max\{\mathbf{e}_O^T\}$ in μm
[0.59, 1.69]	[0.59, 1.35]	[0.1, 0.2]	[2.1, 4.3]

Table 5.1: Results of the object positioning error $\mathbf{e}_O$ in μm. The minimal and maximal errors separated by axis is given for the estimate in one direction and not for the best estimated point.

- The difficult position estimation of the object while it is gripped. The occlusions bias the position estimate especially in the y-direction. Because of the object symmetry the x-coordinate is less biased.
- The object is deformed rather in the y-direction leading to a different object CG while the object is gripped.
- When removing the gripper from the object it moves. This motion is stronger in y-direction.

In table 5.2 the distance between the actual position and the desired one is presented. Considering object and gripper size an average positioning error of $1.88\mu m$ seems very good. The maximum error of $4.79\mu m$ is quite far away from the average. The high error comes from the fact that in this manipulation the object was not released immediately and it was *tipped* onto the substrate.

$\mathcal{E}\{\|\mathbf{e}_O\|\}$ in μm	$\sqrt{\mathcal{V}\{\|\mathbf{e}_O\|\}}$ in μm	$\min\{\|\mathbf{e}_O\|\}$ in μm	$\max\{\|\mathbf{e}_O\|\}$ in μm
1.88	1.35	0.45	4.79

Table 5.2: Results of the object positioning error $\|\mathbf{e}_O\|$ in μm.

A comparison with other systems from the literature is difficult since usually inserting or gluing tasks are examined rather than pick&place operations. Furthermore, the considered object size is much greater than the manipulated polystyrene spheres with a diameter of $\approx 15\mu m$. In terms of merely positioning the gripper this system can be compared. Using template matching as described in section 3.7.2 the robot can be positioned with a precision of at least $\pm\frac{1}{4}pixel \hat{=} \pm 115nm$ (using $10\times$ magnification). In combination with the certainty of the calibration matrix only few positioning updates are needed to obtain this resolution. This is about the resolution which is obtained by the authors of [7]. For an objective with a $NA = 0.42$ and $0.175\frac{pixel}{\mu m}$ the authors of [11] give a positioning repeatability of $0.17\mu m$.

Chapter 6

Conclusions and future work

6.1 Conclusions

The topic the optimal use of vision for micromanipulation is very general and covers several subproblems. Many different aspects that influence the overall performance of a micromanipulation system were discussed:

A manipulation sequence providing a semi-automatic manipulation procedure was defined. Instead of picking up an object the substrate is lowered such that the behavior of the object during this phase can be observed in focus. An object-oriented software design was developed meeting the specifications of modularity, re-useability and maintainability. A graphical user interface was developed visualizing the process data.

At the beginning of chapter 3 the vision system itself was examined: The critical factors to the vision system performance were discussed. The most important variables are the magnification and the numerical aperture of the objective, the illumination and the way the microscope is mounted. If a transparent substrate is used and the magnifications are not higher than 20×, the microscope should be mounted horizontally observing the substrate from below. Otherwise, the microscope has to be mounted above the substrate. Because of its rich amount of features and implementations Intel's OpenCV was selected to facilitate the implementation of the computer vision. The integration of an USB and a FireWire camera using video for windows and the CMU FireWire camera driver was shown briefly. An examination of the captured microscope images showed a high SNR of 60dB and no salt&pepper noise. The changes in illumination were measured to have a standard deviation of 1.8 measured over several frames that is about factor 0.38 of the spatial change of illumination within one frame. A contrast-based autofocussing algorithm using the Tenengrad criterion was used to obtain height information about the substrate position. An estimated repeatability of $\approx 0.5\mu m$ was obtained using only the imperfections in the substrate surface to detect its relative height to the focal plane. Methods to locate objects and end-effectors within the focal plane were presented: Therefore the properties of the objects and the end-effectors were clearly discussed. The main difference is that the end-effector geometry and orientation is well know, which is not the case for the objects. Based on this characterization three different methods were applied. The methods are compared in table 6.1 and 6.2.

	Advantages	**Disadvantage**	**Application**
Segmentation and labeling	• Arbitrary object geometry • Repeatability = $\pm 65nm$ for $10\times$ magnification • Execution time $200ms$ (full frame), several ms for a ROI	• Requires user input or a classifier to choose a segment • High image contrast needed • Items that touch the object of interest are interpreted to belong to it unless no further processing is done	Object detection and tracking
Correlation	• Maximum error = $\pm 115nm$ for $10\times$ magnification • High robustness against changes in illumination • Robust against blur • Arbitrary constant shape	• The upper left template corner is detected so that a offset within the template has to be defined • Computational effort is $O(MNM_tN_t)$, with M_t and N_t the dimension of the template • Sensitive to scaling and rotations • Disturbances in the template region can influence the detection	Precise end-effector detection

Table 6.1: Comparison of the three different used methods.

In chapter 4 three different approaches to calibrate the relative height of the end-effector to the focal plane were shown and experimental results were presented. Depending on the type of end-effector and also how the microscope is mounted, different calibration methods should be used (compare table 4.1 and the list below). A robot trajectory-based relative positioning law is used. First an approach position is passed so that the object is not touched before the actual approach to the object. Moving only a fraction of the relative error to align the end-effector with the object several positioning updates are calculated.

The complete system was tested in an experiment manipulating a polystyrene ball with a diameter of $15\mu m$. Considering the misalignment of the two jaws of $7\mu m$, the thickness of the gripper $50\mu m$ and its rough border, it is a good result being able to perform a pick&place operation of such an object. The inelasticity of the sphere lead to plastic deformations of the object. Objects were released at a user specified position in 13 experiments with an average error of $1.88\mu m$ and a maximum error of $4.79\mu m$, where the error in y-direction was twice as high as the one in x-direction. The main disturbances to this direct object-based approach

	Advantages	Disadvantage	Application
Line detection	• Only geometric information is used • Robust position estimate because of the evaluation of the whole image with reasonable computational effort • Allows rotations within a specified range • medium robustness against blur and changes in illumination	• Maximum error = $\pm 0.92 \mu m$ without using the refinement with the correlation	Robust detection of a gripper

Table 6.2: Comparison of the three different used methods.

are the motion of the object when the gripper is removed from it, the biased object position estimate during the gripping phase due to occlusions and the object deformation. The effects of the deformation can be reduced using an integrated sensor measuring the gripping force. Occlusions can be reduced decreasing the thickness and surface roughness of the gripper.

6.2 Future work

This work gives the base for further examinations.

- The developed system should be tested with other end-effectors especially two-jaw grippers.
- The method *setup 4* using shadow effects to detect the lower edge of the gripper is very interesting and should be investigated more in detail.
- The improvement of the robustness of the detection of an object while it is gripped can increase the positioning repeatability.
- Pattern classification of the objects can be used to full-automate the manipulation.
- Non-spherical objects should be manipulated to examine the properties of an end-effector to align objects in terms of orientation.
- Control gripper opening based on the position estimate obtained by the line detection
- Examine how the object placement repeatability changes when 20x magnification is used.
- Partial scanning of the microscope camera can be used to record the gripping and release process at high frame rates. In a post-processing step position information can

be obtained to characterize the two processes.

Appendix A

Derivation of the calculation of the intersection point of two lines represented in polar coordinates

In this section the calculation of the intersection point $\mathbf{P} \in \mathbb{R}^2$ of two lines represented in polar coordinates (ϕ, d_{0l}) is derived. The used symbols are defined in figure A.1.

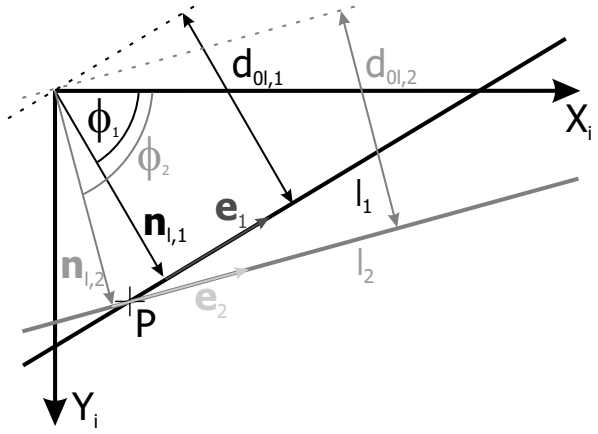

Figure A.1: Intersection of two lines and related parameters.

Using trigonometric relationships the vectors $\mathbf{n}_{l,i}$ and $\mathbf{e}_i$ can be related to the description in polar coordinates as

$$\mathbf{n}_{l,1} = \begin{bmatrix} \cos(\phi_i) \\ \sin(\phi_i) \end{bmatrix} d_{0l,i},$$

$$\mathbf{e}_i = \begin{bmatrix} -\sin(\phi_i) \\ \cos(\phi_i) \end{bmatrix}, \tag{A.1}$$

In order to calculate the intersection of the lines l_1 and l_2, they are at first represented as

$$l_i : \mathbf{x} = \mathbf{n}_{l,i} + \lambda_i \mathbf{e}_i, \lambda_i \in \mathbb{R}, \tag{A.2}$$

where $\mathbf{n}_{l,i}$ represents a point on l_i and $\mathbf{e}_i$ the direction vector of the line being parameterized by λ_i.

When the coordinates $\mathbf{x}$ of both lines are the same, respectively $l_1 = l_2$. Using equation A.2

$$\mathbf{n}_{l,1} + \lambda_1 \mathbf{e}_1 = \mathbf{n}_{l,2} + \lambda_2 \mathbf{e}_2 \tag{A.3}$$

can be formulated. Inserting A.1 and collecting the terms containing λ_i on one side

$$\lambda_1 \begin{bmatrix} -\sin(\phi_1) \\ \cos(\phi_1) \end{bmatrix} + \lambda_2 \begin{bmatrix} \sin(\phi_2) \\ -\cos(\phi_2) \end{bmatrix} = \mathbf{n}_{l,2} - \mathbf{n}_{l,1} \tag{A.4}$$

is obtained. This can be rewritten as

$$\underbrace{\begin{bmatrix} -\sin(\phi_1) & \sin(\phi_2) \\ \cos(\phi_1) & -\cos(\phi_2) \end{bmatrix}}_{\mathbf{A}} \begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix} = \mathbf{n}_{l,2} - \mathbf{n}_{l,1}. \tag{A.5}$$

The determinant of $\mathbf{A}$ can be calculated and written down using a trigonometric simplification as

$$\det(\mathbf{A}) = \sin(\phi_1 - \phi_2).$$

It follows by inspection that the matrix A is singular for $\phi_1 \text{modulo}(\pi) = \phi_2 \text{modulo}(\pi)$, which is the case when $l_1 \| l_2$. Assuming this case is excluded, $\mathbf{A}^{-1}$ exists and can be multiplied from left to equation A.5. This solves for

$$\begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix} = \frac{1}{\sin(\phi_2 - \phi_1)} \begin{bmatrix} \cos(\phi_2) & \sin(\phi_2) \\ \cos(\phi_1) & \sin(\phi_1) \end{bmatrix} (\mathbf{n}_{l,2} - \mathbf{n}_{l,1}). \tag{A.6}$$

For the calculation of $\mathbf{P}$ only one parameter is required. If only λ_1 is considered the following expression can be formulated:

$$\lambda_1 = \frac{diag\{\cos(\phi_2), \sin(\phi_2)\}(\mathbf{n}_{l,2} - \mathbf{n}_{l,1})}{\sin(\phi_2 - \phi_1)},$$

Inserting λ_1 into l_1

$$\mathbf{P} = \mathbf{n}_{l,1} + \lambda_1 \mathbf{e}_1 \tag{A.7}$$

is obtained.

Bibliography

Micromanipulation:

[1] Silvia Allegro:*Automatic Microassembly by Means of visually guided micromanipulation*; Dissertation, EPF Lausanne, No. 1804, 1998

[2] Jean-Philippe Bacher: *Conception de robots de très haute précision à articulations flexibles: Interaction dynamique-commande*; Dissertation, EPF Lausanne, No. 2907, 2003

[3] A. Buerkle, F. Schmoeckel, M. Kiefer, B. P. Amavasai, F. Caparrelli, A. N. Selvan, J. R. Travis: *Vision-based closed-loop control of mobile microrobots for micro handling tasks*, Proceedings of SPIE Vol. 4568: Microrobotics and Microassembly III, Boston, USA, October 2001.

[4] S. Fatikow, J. Seyfried, S. Fahlbusch, A. Buerkle, and F. Schmoeckel: *A Flexible Microrobot-Based Microassembly Station*; Journal of Intelligent and Robotic Systems 27: pp. 135169, 2000

[5] J. T.Feddema; R. W.Simon: *Visual servoing and CAD-driven microassembly*; IEEE Robotics & Automation Magazine, Volume: 5 , Issue: 4 , pp. 18-24, Dec. 1998

[6] A. Georgiev, P. K. Allen: *Visually-guided Protein Crystal Manipulation Using Micromachined Silicon Tools*; To appear at IROS'04, Sendai, Japan, 2004

[7] I. Pappas, A. Codourey: *Visual Control of a Microrobot Operating under a Microscope*; IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, Osaka, Japan, pp. 993-1000, 1996

[8] D. Popa, H. Stephanou: *Micro and Meso Scale Robotic Assembly*; to appear in the special issue of SME Journal of Manufacturing Systems, 2004.

[9] S. Ralis, B. Vikramaditya, and B. Nelson: *Micropositioning of a Weakly Calibrated Microassembly System Using Coarse-to-Fine Visual Servoing Strategies*; IEEE Transactions on Electronics Packaging Manufacturing, Vol. 23, No. 2, April 2000

[10] T. Trüper, A. Kortschack, M. Jähnisch, H. Hülsen, S. Fatikow: *Transporting and Sorting of Cells with Microrobots*; Proc. Micro.tec, Munich, Oct. 2003

[11] Yu Zhou, B. Nelson, and B. Vikramaditya:*Integrating Optical Force Sensing with Visual Servoing for Microassembly*; Journal of Intelligent and Robotic Systems 28: pp. 259-276, 2000

Microscopy:

[12] M. Abramowitz, M. Davidson: *Anatomy of the Microscope - Light Sources*; www.olympusmicro.com/primer/anatomy/sources.html

[13] K. Spring, M. Parry-Hill, M. Davidson: *Microscope optical systems*; http://www.microscopyu.com/articles/optics/opticshome.html

[14] K. Spring, M. Davidson: *Image Brightness*; www.microscopyu.com/articles/formulas/formulasimagebrightness.html

[15] K. Spring, K. Tchourioukanov, M. Davidson: *Airy Patterns and the Rayleigh Criterion*; www.olympusmicro.com/primer/java/imageformation/rayleighdisks/index.html

[16] Ian T. Young: *Quantitative Microscopy*; Engineering in Medicine and Biology, Vol.15, No. 1, pp. 59-66, 1996

[17] Yu Zhou, B. Nelson:*Calibration of a parametric Model of an optical Microscope*; Optical Engineering, Vol. 38, No. 12, pp. 1989-1995, Dec. 1999

[18] www.edmundoptics.com/TechSupport/

Image Processing:

[19] *CMU 1394 Digital Camera Driver*; http://www-2.cs.cmu.edu/~iwan/1394/

[20] D. Forsyth, J. Ponce: *Computer Vision - A Modern Approach*; Prentice Hall, Upper Sadle River, NJ, USA, 2003

[21] S. Grange, T. Fong, C. Baur: *TLIB: a Real-time Computer Vision Library for HCI*; Proc. VIIth Digital Image Computing Techniques and Applications, Sydney, Australia, pp. 1017-1026, Dec. 2003

[22] Bernd Jähne: *Digitale Bildverarbeitung*, 5., überarb. und erw. Auflage, Springer, Berlin, 2002

[23] Jaap van de Loosdrecht: *Class slides for Introduction to Computer Vision* http://www.engineering.tech.nhl.nl/engineering/personeel/loosdrec/-vision_course/acquis.pdf

[24] *OpenCV Yahoo forum*; http://groups.yahoo.com/group/OpenCV/

[25] *Open source computer vision library - Reference manual*; http://www.sourceforge.net/projects/opencvlibrary

[26] W. E. Snyder, Hairong Qi: *Machine Vision*, Cambridge University Press, Cambridge, 2004

Visual Servoing:

[27] Peter I. Corke:*Visual Control of Robots : High-performance Visual Servoing*; Research Studies Press Ltd., Taunton, Somerset, England, 1996

[28] Koichi Hashimoto: *Visual Servoing - Real-Time Control of Robot Manipulators Based on Visual Sensory Feedback*; World Scientific Publishing, Singapore, New Jersey, London, Hong Kong, 1993

[29] S. Hutchinson, G. Hager, P. Corke: *A Tutorial on Visual Servo Control*; IEEE Trans. on Robotics and Automation, Vol. 12 No. 5, pp. 651-670, 1996.