CARL VON OSSIETZKY universität OLDENBURG

Fakultät II – Informatik, Wirtschafts- und Rechtswissenschaften
Department für Informatik

Automatic Image Analysis in Micro- and Nanorobotic Environments

Dissertation zur Erlangung des Grades eines
Doktors der Ingenieurwissenschaften (Dr.-Ing.)

von

Dipl.-Ing. Tim Wortmann

Gutachter:

Prof. Dr.-Ing. habil Sergej Fatikow
Prof. Pasi Kallio, D.Tech.

Tag der Disputation: 30. Mai 2012

Übersicht

Die Fähigkeit zur Handhabung, Montage und Inspektion von mikro- und nanoskaligen Objekten und Strukturen schafft derzeit die Voraussetzung für die Entwicklung einer Vielzahl von neuen Produkten, Verfahren sowie Werkstoffen. Dabei ist die Automatisierung dieser Vorgänge in der jüngeren Vergangenheit weit vorangeschritten. Die Auswertung von Bilddaten hat sich hierbei als geeignete Methode zur automatischen Positionsbestimmung von Werkzeugen und Werkstücken erwiesen. Die meisten bekannten Verfahren müssen manuell mit Position, Art und Ausrichtung von Objekten initialisiert werden, um deren Bewegung kontinuierlich verfolgen zu können. Dadurch ist der maximal erreichbare Automatisierungsgrad begrenzt. Im Rahmen der vorliegenden Arbeit wird diese Einschränkung durch die Einführung von zwei neuen Verfahren zur Bildauswertung überwunden.

Das erste Verfahren beschäftigt sich mit dem Problem, in einer komplexen Bildszene einzelne Objekte aufzufinden und zu klassifizieren. In diesem Bereich kommen oft so genannte Template Matching Verfahren zum Einsatz, deren Fähigkeit zur Generalisierung jedoch stark begrenzt ist. Stattdessen nutzt das vorgeschlagene Verfahren Methoden der statistischen Mustererkennung und lernt optimale Entscheidungsregeln. Die Strategie zur Merkmalsauswahl stellt sicher, dass die problemspezifischen Charakteristika berücksichtigt werden. Die Klassifizierung basiert auf Support Vector Machines. Anhand von drei verschiedenen Anwendungsszenarien und Bildmodalitäten wird das Verfahren validiert. Bei der ersten Anwendung werden Kohlenstoffnanoröhren unter Nutzung des Rasterelektronenmikroskops auf der Oberfläche eines Siliziumwafers erkannt. Die zweite Anwendung ist die Erkennung von biologischen Zellen in einem mikrofluidischen Kanal, wobei das Lichtmikroskop Anwendung findet. Zuletzt werden magnetische Partikel unter Verwendung der Magnetresonanztomographie detektiert.

Das zweite Verfahren widmet sich dem Problem, den örtlichen Zusammenhang zwischen mehreren Aufnahmen zu rekonstruieren. Die Aufnahmen wurden von heterogenen Bildsensoren erzeugt. Im Allgemeinen können zwei Strategien verfolgt werden. Bei der merkmalsbasierten Strategie wird in allen Aufnahmen nach lokalen Strukturen gesucht, zwischen denen dann eine Korrespondenz ermittelt wird. Bei der flächenbasierten Strategie wird ein Ähnlichkeitsmaß zwischen den Aufnahmen optimiert. Beide Ansätze unterscheiden sich hin-

sichtlich Konvergenzeigenschaften, maximaler Genauigkeit und Rechenaufwand. Das neue Verfahren kombiniert die Vorzüge beider Ansätze, indem es zunächst eine merkmalsbasierte Ausrichtung vornimmt und das Ergebnis dann der flächenbasierten Strategie folgend verbessert. Dadurch wird die Konvergenz gegen die bestmögliche Ausrichtung bei gleichzeitig geringer Verarbeitungszeit sichergestellt. Eine große Zahl an Merkmalsdetektoren zur Erkennung lokaler Bildstrukturen wurde integriert. Je nach vorherrschendem Bildinhalt wird mit einer geeigneten Teilmenge an Detektoren gearbeitet. Eine weitere wichtige Eigenschaft ist die Fähigkeit, ein variables Maß an Vorwissen im Abgleichverfahren zu berücksichtigen. Das neue Verfahren wurde anhand der Bilddatenfusion zwischen Rasterkraftmikroskop und Rasterelektronenmikroskop validiert. Eine vollautomatisierte Bilddatenfusion wurde erfolgreich an einer großen Auswahl von Probenoberflächen demonstriert.

Abstract

The ability to handle, assemble and inspect micro- and nanoscale objects and structures currently enables the development of a multitude of new products, procedures and materials. Automation of these tasks made significant progress in the recent past. Especially the automatic interpretation of image data turned out to be a very useful form of sensory feedback. Most methods known in this field deal with the problem of continuous pose estimation of micro- and nanoscale tools and workpieces. A common drawback of these methods is the need for a manual initialization step, which prevents fully automated assembly or inspection processes. In this work, two methods for automatic image analysis are introduced which will help to increase the level of automation.

The first method faces the problem of locating and classifying objects in a complex image scene. Template matching algorithms are a popular solution to this task, but they suffer from a poor generalization capability. Instead, the proposed method makes use of statistical pattern recognition by learning optimal classification rules from a set of training samples. The feature selection and training strategy makes sure, the specific problem's characteristics are taken into account. Classification is based on a support vector machine classifier. The method is validated using three different application scenarios and imaging modalities. First, carbon nanotubes are localized on the surface of a silicon wafer, inspected by the scanning electron microscope. Next, biological cells in a micro fluidic channel are inspected for mechanical damage using the optical microscope. The last application is the detection of magnetic particles using magnetic resonance imaging. In all scenarios, the proposed method has been applied successfully.

The second method is designed for the spatial alignment of images acquired by multiple heterogeneous sensors. Generally, two strategies can be followed. One strategy is the detection of local image structures and the identification of corresponding structure (feature-based approach). The other strategy optimizes a similarity measure between images (area-based approach). Both approaches are different in terms of accuracy, convergence properties and execution time. The new method combines the benefits of both approaches by performing a feature-based alignment first, followed by an area-based refinement step. This assures convergence towards the best alignment and a low processing time. A large number of feature detection methods have been integrated. Depending on the

prevailing image contents, the method works with a suitable subset of detectors. Another important feature is the incorporation of a variable level of prior knowledge. For the validation of the new procedure, it has been applied to the task of fusing scans acquired by the atomic force microscope and scanning electron microscope. Fully automatic image fusion has been demonstrated successfully on a multitude of sample surfaces.

Acknowledgements

I would like to thank all the people who supported me and helped making this work a success. The results presented in this thesis have been achieved at the Division Microrobotics and Control Engineering (AMiR) of the University of Oldenburg. Especially, I would like to thank the head of the group and supervisor of this thesis, Prof. Dr.-Ing. Sergej Fatikow, for all the support and the confidence he had in my work. I am deeply grateful for the excellent working conditions and high degree of freedom I had over the last years. Also, I appreciated the close collaboration with so many international partners. I would like to thank Prof. Pasi Kallio, D.Tech., for reviewing my dissertation.

Furthermore, I would like to thank all my colleagues at AMiR for the friendly atmosphere and all the interesting discussions. In particular I would like to thank Christian Dahmen for the very good teamwork, many great business trips and also for proofreading this thesis. I would like to thank my student assistant Christian Geldmann for his help with MRI sequence development and data analysis. Also, I would like to thank Dr. med. Alexander Kluge for all the experiments carried out at Pius Hospital Oldenburg.

I would like to thank my parents Karin and Joachim for their support throughout my life and for making my studies in Harburg and Graz possible. Also I would like to thank my sister Wiebke for her support and advice over all the years. Last but not least I would like to thank my girlfriend Anabel for her support and for constantly motivating me. You strongly helped making this work a success.

Table of Contents

List of Figures

List of Tables

1 Introduction

Micro- and nanotechnology play a key role in many areas of research and industry. Structures and objects with dimensions in the micro- and nanometer range exhibit unique electrical, mechanical and optical properties. These properties are exploited already in a multitude of novel products and applications such as displays or sensors for mobile phones. Nevertheless, the production technology is improving continuously, making applications of micro- and nanotechnology better and more available. Robotic manipulations enable characterization, assembly, handling and structuring of micro- and nanoscale components [1, 24]. Therefore, micro- and nanorobotics provide important tools for basic research but also for optimizing existing and establishing new production technologies.

Common application scenarios of microrobotics include the inspection and manipulation of biological cells and also the assembly of microelectromechanical systems. Performing these tasks imposes difficulties originating from the scaling effects of the governing forces such as surface forces or weight. On the nanoscale, managing the scaling effects becomes even more important. Possible applications for nanorobots are bending experiments for testing the mechanical properties of nanoscale objects. Also, nanoelectromechanical systems can be assembled by nanorobots.

Many setups for micro- and nanorobotic manipulations have been presented, including commercially available nanomanipulators and positioning stages. A more flexible solution is the construction of mobile platforms for micro- and nanorobotic tasks. The components of a micro- or nanorobotic system typically include a set of sensors and actuators and also a control system. Frequently used types of end-effectors are tips for indentation and grippers for pick-and-place handling of objects. Sensors for multiple quantities are available, exploiting electrical, optical or mechanical principles of measurement. One of the most important forms of sensory feedback is obtained by image analysis. Image material provided by microscopes is not only valuable for visual inspection of micro- and nanoscale objects and structures. It can also be used for the automatic extraction of object positions and indirect measurements such as optical force measurement.

In the context of micro- and nanorobotics, most prior work on computer vision focuses on the problems of object tracking and depth estimation. Object tracking refers to the task of continuously following the movement of an object such as

a microgripper or a nanowire. Together with methods for depth estimation, the introduction of object tracking helped to tremendously increase the level of automation in micro- and nanorobotic tasks. Two of the most advanced methods which have been successfully applied are the active contours algorithm and rigid-body tracking based on a geometric model. Both require a manual initialization step and in the latter case also a model of the object geometry. This limits the possible level of automation which can be achieved, and also the applicability to targets with high variations in visual appearance.

A special application of micro- and nanorobotic systems is the preparation and execution of material inspections. An example is the identification of targets and preparation of lamellae for an examination using the transmission electron microscope. However, the characteristics of imaging modalities with the capability of image acquisition at the micro- or nanoscale are very different. These include features such as the maximal magnification or also the type of contrast. The different nature of the imaging modalities motivates their side-by-side use. By using principles of micro- and nanorobotics, fully automated multimodal surface inspections seem possible. This might include sample preparation, the selection of a region of interest, setting of the imaging parameters and presentation of a fused view of all collected data. While most of these task can be carried out with state-of-the-art components of micro- and nanorobotic systems, there is a lack of a suitable strategy for fusion of the collected data.

1.1 Objective

The goal of this thesis is to further increase the level of automation in micro- and nanorobotics by providing new methods for computer vision. Specifically, two problems must be solved:

- Localization and classification of micro- and nanoscale objects from image scenes,
- Data fusion between heterogeneous sensors with imaging capabilities at the micro- and nanoscale.

Both procedures will supersede manual image interpretation steps such as labeling objects or landmark points which are common in state-of-the-art processes. A requirement is a high level of integration with existing control architectures, enabling fully automated manipulation and inspection procedures at the micro- and nanoscale.

1.2 Thesis outline

This thesis is structured into eight chapters. In the following chapters, the author contributes to the field of image registration and object classification at the micro- and nanoscale. Chapter 2 describes application scenarios and state-of-the-art methods used for image-based object detection in micro- and nanorobotics. The problem of multimodal image fusion and the most important fields of application are explained in Chapter 3. Also, state-of-the-art methods for image registration are categorized. In Chapter 4, a system for micro- and nanoscale object classification is developed. It is based on statistical pattern recognition and designed for being integrated into fully automated micro- and nanorobotic setups. Chapter 5 introduces a new strategy for multimodal image registration. It combines the benefits of area-based and feature-based registration schemes. The system for micro- and nanoscale object classification is validated in Chapter 6. Three experimental setups with different imaging modalities are used. The tasks carried out are the localization of carbon nanotubes on the surface of a silicon wafer, a quality check for biological cells and the detection of magnetic particles using magnetic resonance imaging. Chapter 7 validates the new strategy for multimodal image registration by registering scans obtained from the atomic force microscope and scanning electron microscope. The thesis is summarized and an outlook is provided in Chapter 8.

2 Image-based object detection in micro- and nanorobotics

Transferring computer vision methods to micro- and nanorobotic applications imposes a multitude of difficulties. Those are arising from the characteristics of sensors with imaging capabilities at the micro- and nanoscale but also from the high variation in appearance of the target objects. Nevertheless, image analysis provides a valuable form of feedback for process automation. This chapter provides an overview on methods, which are used for image-based object detection in micro- and nanorobotics.

2.1 Problem statement

In the recent past the interpretation of image data turned out to be one of the most important forms of sensory feedback in micro- and nanorobotic tasks. A well-established application scenario is the analysis of image scenes showing a number of usually movable objects. The most relevant types of objects are tools and workpieces. Typical tasks to be carried out are the inspection, manipulation or assembly of objects. A common difficulty is that image scenes can also contain other contents such as acquisition-related artifacts or contaminations of the setup. Although this configuration shows some similarities with macroscale robotic tasks, the characteristics of the imaging sensors used and also the behavior of the objects is very different in micro- and nanorobotics.

The problem targeted here is the detection of micro- and nanoscale objects in an image scene. Object detection is the process of localizing objects belonging to a given class of objects [15]. It is closely related to a number of other tasks. Object recognition refers to the localization of a specific object instance. On the other hand the success of automatic image analysis on the micro- and nanoscale strongly relies on the formulation of prior knowledge about the setup. In cases dealing with unique objects such as a microgripper, the distinction between the different tasks might vanish. Pattern recognition is the process of recognizing patterns by analyzing object features. Pattern recognition can be used for object detection. Object tracking is the continuous position determination of a specific

object instance. Generally, the object tracking procedure needs to be initialized with the initial object position and sometimes also orientation.

The following section describes a collection of state-of-the-art methods currently in use in the field of micro and nanorobotics. Some applications have a strong focus on object tracking and thus need an initialization. However, in most cases the same methods can also be used for object detection by reformulating the search space or the optimization criterion. For example, correlation-based techniques are currently in use for both, object detection and object tracking.

2.2 State-of-the-art object detection methods in micro- and nanorobotics

2.2.1 Marker-based position determination

The problem of determining object locations can be ill-posed, depending on the viewing perspective and amount of visible object detail. A possible solution is to mark objects with a unique label, which can be recognized easily in an image scene. This approach is very popular in macroscale robotics. However, attaching markers to micro- and nanoscale objects is much more difficult and not possible for all applications. On the other hand, it strongly simplifies the localization procedure.

An example of active markers has been shown using light-emitting diodes (LEDs) [18]. The LEDs are positioned at the bottom of a mobile robotic platform. As the LEDs are the only source of illumination in this setup, the camera image acquired from the bottom view shows nothing else but bright spots at the location of the LEDs. By computing the center of gravity of these bright spots, the location of the LEDs and therefore the position and orientation of the mobile platform can be concluded.

In [61], a system for automated microassembly is presented. Localization of the microobjects is carried out using a three-dimensional optical vision sensor, based on a miniature camera. The microobjects have been created in a lithographic process. Therefore it was possible to attach special circular positioning marks already in the manufacturing process. For improved visibility, the positioning marks have been created using a fluorescent substance. If appropriate illumination is chosen, the camera image will show an almost perfect segmentation of the positioning marks. Object localization is now based on the known object geometry and location of the positioning marks. Besides the very high speed of computation, this approach also hardly suffers from unwanted reflections. The authors report a high accuracy and repeatability of the optical vision sensor.

A system for fast object tracking in the scanning electron microscope (SEM) is presented in [53]. In contrast to [83], the system is not working with full images but works with line scans along arbitrarily shaped patterns. Avoiding the time-expensive full image acquisition in the SEM, the system works with update rates in the kHz range. For successful object tracking, it needs to be initialized with the object location. In order to achieve the very high update rate and arbitrarily-shaped scan patterns, custom-built hardware has been used. Several marker geometries have been tested. The markers have been manufactured using a focused ion beam (FIB).

2.2.2 Position detection from known scene arrangement

In automated microassembly tasks there can be situations where objects need to be localized in an image scene of which the general arrangement is known a priori. This can be the case if workpieces and tools have been moved to the field of view by means of a non-vision-based position control or even in open-loop mode. The constellation can also be known from prior assembly steps. In some cases the general arrangement of the image scene may be formulated by a set of rules. A set of rules might look as follows:

- There are 4 separated objects in the scene.
- Three of them are tools (T1, T2, T3), one of them is a workpiece (W1).
- T1 is located above all other objects.
- T3 is located lower than all other objects.
- W1 is the leftmost object and is connected to the left image border.
- T2 is the rightmost object and is connected to the right image border.

From such a set of rules, individual objects can be identified in the image scene. The recognition of the gripping arms of a microgripper has been successfully demonstrated [117]. Initially, objects are segmented using Canny edge detection [13]. The objects are identified by computing their centers of gravity and by using the prior knowledge about the scene arrangement. In a subsequent assembly step, it is known from the context that the microobject will randomly adhere to one of the gripping arms. This is detected automatically by observing a reduction in the number of contours and by comparing their centers of gravity.

2.2.3 Correlation-based approaches

When assembly tasks on the micro- and nanoscale are performed, available tools such as grippers or probes are often reused multiple times or tools of similar shape are used. Additionally, the workpieces used in an assembly setup will often also show similarities. Therefore, in some application scenarios explicit search patterns are available for the detection of objects from an image scene. The direct way of measuring the image similarity is to compute the normalized cross-correlation. For a given input image $I(x, y)$ and a search pattern $P(x, y)$, the correlation coefficient matrix $C(x, y)$ can be computed from

$$\mathbf{C} = \mathcal{F}^{-1}\left[\mathcal{F}\left[\mathbf{I}\right] \cdot \mathcal{F}\left[\mathbf{P}\right]^{*}\right], \tag{2.1}$$

where $\mathcal{F}$ denotes the Fourier transform and * the conjugate complex. The normalized cross-correlation matrix $C_N(x, y)$ is derived as follows:

$$\mathbf{C_N} = \frac{\mathbf{C}}{\sqrt{\sum\limits_{x,y} I(x, y) \cdot \sum\limits_{x,y} P(x, y)}}. \tag{2.2}$$

Figures 2.1 and 2.2 depict how cross-correlation can be utilized for object detection. The input image scene can be seen in Figure 2.1. It shows an electrothermally actuated microgripper approaching a silicon wafer. The scene was captured using the SEM. Figure 2.2 shows the cross-correlation coefficients and the search pattern, which shows the gripper jaws. A sharp tip can be observed in the correlation matrix at the actual position of the jaws.

The technique has been applied to SEM images with a focus on processing speed [98]. Initially, the authors localize the object in a full image scene. This can be speeded up by downsampling of image scene and search pattern. Once the object is found, it's position can be continuously followed by performing correlation only in the local neighborhood of the last known object position. This neighborhood is referred to as region of interest (ROI). A big benefit of using ROIs in SEM imaging is the speedup not only in image processing but also in image acquisition. Unlike other image sensors, the SEM is capable of scanning small selected areas of an image scene. The acquisition time is reduced proportional to the scanning area. Using these extensions of the direct cross-correlation method the authors demonstrate real-time object tracking. The high robustness of the correlation method to image noise is identified as the crucial factor for the success in SEM image processing.

Correlation-based object detection in the direct form as shown in Equations 2.1 and 2.2 is sensitive to a number of effects. Any difference in scaling or rotation

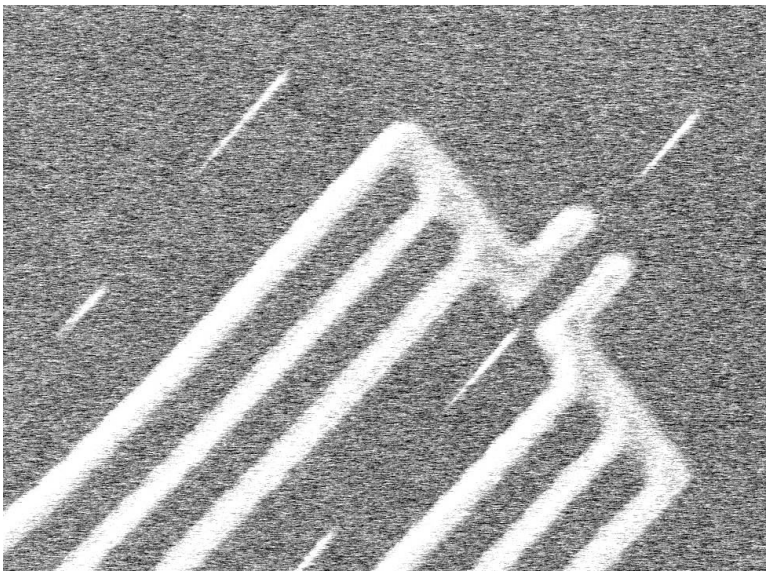


Figure 2.1: Electro-thermally actuated microgripper approaching a silicon wafer with carbon nanotubes.

between input image and search pattern will dramatically decrease the correlation coefficient at the target position. Even small differences can reduce the correlation coefficient below the noise level, making successful object detection impossible. Also, the correlation coefficient will suffer from variations in image signal intensity. The authors in [98] target the problem of object rotation by generating multiple rotated search patterns. In object tracking mode, not only limited object displacements but also a limited object rotation can be assumed between consecutive images. By limiting the search space to small rotations, the method preserves it's real-time capability.

Extending the orientation estimation procedure from object tracking to a full-frame object detection increases the computational burden by orders of magnitude. Another effect is that due to the large search space, additional peaks in the correlation coefficient will accidently show up and complicate the identification of the true object position. Therefore, correlation-based object detection can be regarded as a tradeoff between detection performance, size of the search space and computation time. Methods for the design of optimal detectors with respect to a variety of criteria can be found in [63]. The idea of making correlation-based detectors robust to rotation can be extended to cover three-dimensional rotation, scaling and also multiple types of objects in one detector. These extensions are generally paid for by a loss in detector performance.

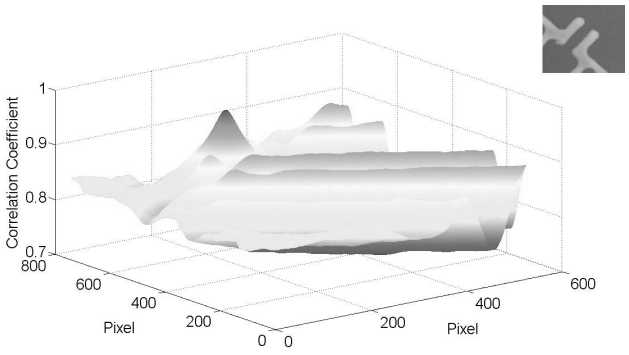


Figure 2.2: Result of cross-correlation between a new image scene and a search pattern (upper-right corner). The search pattern was obtained by temporal averaging of images showing the gripper in relaxed state. A sharp tip can be seen at the actual jaw position.

2.2.4 Object localization based on model fitting

Depending on the application scenario, the geometry of micro- and nanoscale tools and workpieces can be described in the form of a mathematical model. This is especially the case if the object has been created by a manufacturing method which itself relies on computer-assisted design (CAD) modeling. An example are microgrippers which have been produced using a lithographic process. As an alternative, the model can be extracted from multiple views of an object by means of a contour description. If the object model is not known from the manufacturing process and cannot be extracted, there is still the option to describe it manually using basic geometric shapes. The model can then be used for the localization of objects in an image scene.

Microscale object tracking in the SEM has been demonstrated using the active contour algorithm [99]. An active contour (also referred to as snake) is a parameterized contour of an object. The contour is modified iteratively with the aim of minimizing an assigned energy E^*_{snake}. The optimization can take place in a single image or also in a sequence of images. In the original formulation given in [60], the energy of a parameterized snake $v(s) = (x(s), y(s))$ is obtained from

$$E^*_{snake} = \int_0^1 E_{snake}(v(s))\,ds = \int_0^1 E_{int}(v(s)) + E_{image}(v(s)) + E_{con}(v(s))\,ds. \tag{2.3}$$

The behavior of the snake is determined by the choice of the energy terms E_{int}, E_{image} and E_{con}. These terms need to be adapted for special applications and imaging sensors. The internal energy E_{int} is derived only from the shape of the snake but not from the image contents. E_{int} usually penalizes a high number of windings in the contour and preserves smoothness. E_{image} is based on the image contents in the surrounding of the snake. One popular choice is to favor contours which align to image edges. E_{con} depends on external constraint forces such as user interaction.

By changing the energy terms during an application, the algorithm can switch between different modes of operation. For example, the active contour can be initialized with a basic shape (e.g. a rectangle) for iterative adaption towards the target object shape. This shape is locked when switching to object tracking mode. In tracking mode, only a combination of rotation, scale and translation is allowed to update to active contour. This behavior can be implemented by making other changes prohibitively expensive in terms of E_{int}. The described method has been applied to microscale object tracking inside the SEM using a special formulation of E_{image} which is adapted to the noise distribution of the SEM [97].

Generally, active contours can also be utilized for object detection by initializing $v(s)$ with the targeted shape. Changes in object appearance can be handled by extracting contours from a set of training images [16]. This set of training shapes is fused into a representative shape which can be used for object detection. The idea covers only the aspect of contour shape but has also been extended to image contents inside the contour such as texture.

Another approach to microscale object tracking in the SEM is based on predefined three-dimensional CAD models [62]. Initialized with an estimate of the three-dimensional object pose, the algorithm computes a two-dimensional projection image which simulates the actual SEM view. The projection view contains only the visible edges of the CAD model, and ideally model edges and object edges from the SEM image should align. With any movement of the microobject, model and image edges will not match any longer. Assuming only small movements are possible between consecutive frames, the algorithm detects object edges in the local neighborhood of the model edges. With the new edge positions the object pose can be updated. This procedure is applied iteratively for continuous object pose estimation.

2.3 Limitations of the state of the art

In summary, it can be said that the methods described helped increasing the level of automation in micro- and nanorobotics substantially. The marker-based solutions are reliable and fast and should be used wherever permitted by the application. Exploiting knowledge about the scene arrangement is also a good option, if available. Correlation-based methods can be used efficiently for object detection tasks in cases where sufficient constraints concerning object orientation and scale can be made. The probably most flexible method is the model fitting approach.

However, many micro- and nanorobotic procedures still contain multiple user interaction steps. A main reason is that the image processing routines described so far suffer from two types of limitations. The first group is limited to special application cases, inherent to the functional principle. This group includes marker-based detection and the exploitation of knowledge about the scene arrangement. The second group is limited by a weak capability of generalization. This applies to correlation-based and model fitting techniques. Both are mainly good for finding objects of known geometry. All attempts to make these methods robust to a variety of geometric transformations mainly aim at compensating for acquisition-related effects such as viewpoint or illumination. Unlike in most macroscale robotic tasks, micro- and nanoscale tools and especially workpieces can exhibit a strongly variable outer appearance.

Both, correlation-based and model fitting techniques can be extended with generalized search patterns created out of multiple training images. The problem is that the more general a search pattern becomes, the more the detector tends to respond to arbitrary image patterns. Another problem is the definition of decision rules: each detector must not only detect the presence but also the absence of objects of interest. These rules must be specified explicitly by means of minimum correlation coefficients or a minimum similarity measure for model fitting. The limitations of state-of-the-art micro- and nanoscale object detection methods become most obvious in applications where objects of interest exhibit a high intra-class variability not only in shape but also in other object properties such as texture or color.

3 Multimodal image fusion

For the inspection of micro- and nanoscale structures, multiple imaging modalities are available. Each is providing special imaging capabilities but also imposing restrictions. Micro- and nanorobotics can help to combine the benefits of multiple imaging modalities by assisting sample preparation, identification of regions of interest and also by supporting data fusion. This chapter gives an overview of the state of the art in multimodal image fusion. It presents common methods used in material inspection and other important application fields of multimodal image fusion.

3.1 Problem statement

Image fusion is a special case of sensor data fusion. In contrast to many other applications of data fusion, the sensors deliver multi-dimensional data sets. The motivation for image data fusion is to create a spatially aligned view of an image scene which is *better* in a certain sense. In comparison to the separate images, the fused view allows a deeper understanding of the image contents. A popular example is image stitching [110], where multiple overlapping views of a large-area scene are fused to an overview. Image stitching is a form of unimodal image fusion, as only a single imaging modality is incorporated. In multimodal image fusion, heterogeneous image sensors with different imaging characteristics are used. Multimodal image fusion is in most cases motivated by complementary image information obtained from the different sensors. The spatially aligned view simplifies the combined interpretation of the sensor data.

An image fusion procedure typically requires at least two processing steps which are spatial alignment and visualization of the image data. The process of spatially aligning images is also referred to as image registration [42]. Because registration is by far the most complex task of image fusion, both expressions are often used interchangeably. In some cases, multimodal image acquisition can be carried out in a pre-registered setup. An example is the use of multispectral sensors in a camera, where both sensors use the same optical system simultaneously. In all other cases, registration must be carried out using additional position sensors or directly, by evaluating the image contents.

According to [42], the result of an image registration is a point-to-point correspondence between two images of a scene. The point-to-point correspondence is formulated in a transformation function, which is of the form $T(x, y)$ for two-dimensional image data. It has to be noted that image registration is not a commutative operation. One of the images is named the reference or base image and it is kept unchanged. The other image is referred to as target image. The task of image registration is to transform point coordinates from the target image to the coordinate system of the base image. $T(x, y)$ can most efficiently be described with the help of a parameterized transformation model. In probably the most simple case this would be a shift of the image coordinates (x, y) by a translation vector $\mathbf{t}$:

$$T(x, y) = \begin{pmatrix} x \\ y \end{pmatrix} + \mathbf{t}. \tag{3.1}$$

The transformation model should reflect the type of geometric differences which can be expected between the images.

3.2 Application fields of multimodal image fusion

With the increasing number of imaging techniques it has become more and more difficult for a human observer to have an overview of multi-dimensional data, showing complementary but also redundant information. Therefore, the demand for multimodal image fusion has grown rapidly over many fields of application. A number of recent developments is summarized in [118]. In the following, three of the most important application fields of multimodal image fusion are sketched.

3.2.1 Medical image processing

Medical diagnosis holds a huge repertoire of imaging modalities with strongly different imaging capabilities. In [92], the distinction between primarily morphological modalities and primarily functional modalities has been proposed. Primarily morphological modalities highlight mainly anatomical structure. Examples include magnetic resonance imaging (MRI), computed tomography (CT) or conventional X-ray. Primarily functional modalities such as functional magnetic resonance imaging (fMRI) or positron emission tomography (PET) are capable of highlighting functionally active areas, for example in the brain. A popular application of medical image fusion is the combination of primary morphological and primary functional modalities, allowing the precise localization of (mal-)functional areas in relation to the surrounding tissue. Another common

usage of medical image fusion is the combination of primarily morphological modalities with complementary imaging characteristics. An example is the fusion of MRI (good soft-tissue contrast) and CT (good contrast on bones and calcifications).

For some widely used combinations of modalities such as PET and CT, integrated imaging equipment is commercially available. This brings the benefit of working in a pre-registered coordinate system. If this is not possible, the use of markers such as screws with good visibility in both imaging modalities is a customary. The latter method is usually referred to as extrinsic registration. In contrast, intrinsic registration relies purely on the image data and is by far the most challenging task. A common difficulty in medical image fusion is the registration of elastic tissue, which requires a non-rigid transformation model [45]. It has been pointed out that many of these registration problems are ill-posed and require proper measures of regularization [33].

3.2.2 Remote sensing

The interpretation of remote sensing data is essential in various areas of applications including cartography, detection of land usage or emergency management. In most cases, imaging sensors are mounted on aircrafts or earth observation satellites. Multispectral scanners are able to highlight different surface properties [7, 52]. Other principles of measurement such as interferometric synthetic aperture radar (InSAR) or laser altimetry provide a good contrast on surface topography. Most remote sensing data are referenced with information about position and viewing direction of the sensor as provided by the navigation system of the aircraft or satellite. However, due to the high travelling speed and the comparably small field of view this information usually does not allow pixel-accurate registration. An appropriate procedure for image registration is needed to further improve the transformation parameters [42]. Although the nature of the problem suggests using rigid transformation models, sometimes a correction of sensor-specific distortions is integrated into the transformation model. Marker-based registration is not a relevant topic in remote sensing applications.

3.2.3 Material inspection

Some main motivations for material inspection in industry are defect detection and measurement of material quality during or after a production process. It is also closely related to material analysis performed in research. Image-based material inspection includes methods for surface investigations, projection images

and also tomography. Although most image-based testing methods are non-destructive, sometimes objects are sliced and surface inspections are performed on the cut faces. Some popular modalities in material inspection are industrial CT, ultrasound testing and optical coherence tomography (OCT). An example of combined usage of ultrasound testing, thermography and X-ray inspection has been reported [38]. For image-based surface investigations a high number of techniques for microscopy is available, covering many different forms of contrast.

Multimodal microscopy, also referred to as correlative microscopy, is in most cases carried out during separate investigations [90]. However, some techniques of microscopy allow different forms of image contrast which can be integrated in a single device. For example SEMs nearly always integrate electron detectors for secondary electrons (SE) and back-scattered electrons (BSE). Both types of scans can be obtained in the same frame of reference. If this is not possible, usually time-consuming re-locating of the area of interest is needed after changing the microscope. Recently there have been attempts to automate this workflow by introducing sample holders equipped with dedicated markers [39]. The sample holders can be used in the optical microscope first. When moving it to the SEM, the field of view position with respect to the markers is saved. The SEM stage control recognized the markers and relocates the field of view with an accuracy in the micron range. Exact registration of optical microscope and SEM scan is performed interactively.

3.3 Basic strategies for image registration

In the previous sections, image fusion tasks have been categorized according to the type of input data and type of alignment. Types of input data include uni-/multimodal and two-/higher-dimensional. Pre-calibrated and marker-based alignment have been discussed as well as image content-based registration. Image content-based registration is the most complex task but also the only option in many application scenarios.

A multitude of approaches to the problem of image content-based registration have been reported but no universally best solution could be identified. The methods can be categorized into feature-based and area-based methods [120]. The basic working principles are depicted in Figure 3.1. Feature-based methods try to estimate the transformation model parameters by a set of corresponding image features. The area-based approach starts with an estimate of the transformation model parameters and iteratively improves them by maximizing a similarity measure between the images.

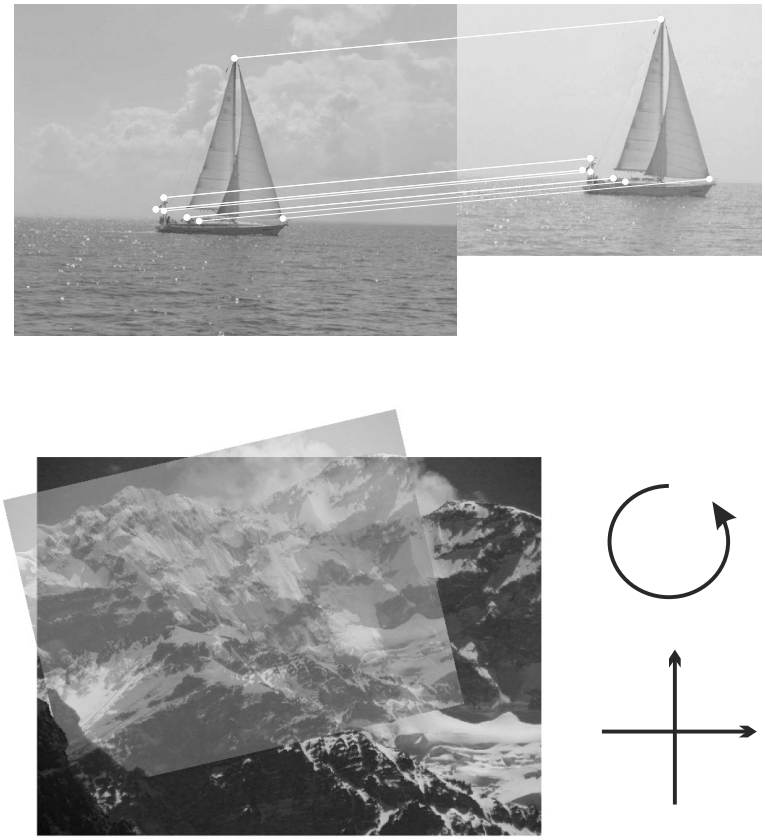


Figure 3.1: Basic types of image registration strategies. The upper image shows corresponding pairs of features which are used to derive transformation model parameters (feature-based registration). In the lower image, two views of an image scene are registered by iteratively modifying the transformation model parameters and evaluating a similarity measure (area-based registration).

3.3.1 Feature-based methods

Feature-based image registration has some similarities with registration based on markers or manually selected landmark points. In all cases, the transformation parameters are estimated by fitting the transformation model into a set of corresponding features. Unlike in marker-based registration or manual landmark labeling, theses features must be identified automatically from the image contents. Features which are reproduced in both, base- and target image, can potentially be used for feature-based registration. These features can represent physical structure but should not be caused by modality-specific imaging artifacts

or reflections. Feature points should be available in a sufficient number, allowing the computation of the model parameters and the compensation of measurement uncertainty. Besides the reproducibility and quantity requirement, features must be distinct and allow the determination of feature correspondence.

Many methods for the detection of suitable image features for feature-based registration have been reported [42]. Naturally, image contents vary strongly between different applications. In contrast to the object detection methods described in Chapter 2, the goal of feature detection is not the localization of compact objects or functionally linked structures. Instead, a reliable source of anchor points for geometric registration is needed. Feature detectors can be specialized on given applications or constructed for general usage.

If prior knowledge about the scene contents is available, shape-based detection methods as they have been described in Section 2.2 can also be used in the context of feature detection. It has to be noted that this approach is strongly tailored to a specific application and requires well-defined geometries of the structures. Feature detection by shape-matching is further described in [78]. The idea of shape detection can be generalized by detecting basic geometric shapes such as circles, lines or line intersections. Image registration has also been demonstrated using arbitrarily-shaped regions with closed boundaries [43]. A reason why feature detectors mostly focus on shapes is that shapes are reproduced much more reliably under a change in viewpoint, illumination or modality than for instance color or texture features.

One of the most prominent feature detectors is the combined corner and edge detector also known as *Harris corner detector* [48]. For a two-dimensional intensity image $\mathbf{I}$, initially a structure tensor $\mathbf{M}$ is computed:

$$\mathbf{M} = \begin{bmatrix} \left(\frac{\partial \mathbf{I}}{\partial x}\right)^2 * \mathbf{w} & \frac{\partial \mathbf{I}}{\partial x}\frac{\partial \mathbf{I}}{\partial y} * \mathbf{w} \\ \frac{\partial \mathbf{I}}{\partial x}\frac{\partial \mathbf{I}}{\partial y} * \mathbf{w} & \left(\frac{\partial \mathbf{I}}{\partial y}\right)^2 * \mathbf{w} \end{bmatrix}. \tag{3.2}$$

$\mathbf{w}$ denotes an appropriate smoothing kernel such as the Gaussian function. $*$ is the convolution operator. With a tunable parameter k, the corner response $\mathbf{R}$ is obtained from:

$$\mathbf{R} = \det(\mathbf{M}) - k \cdot \operatorname{trace}(\mathbf{M}). \tag{3.3}$$

In the surrounding of corners, positive values of $\mathbf{R}$ are obtained. Negative values indicate edge regions and small values are found in flat regions. Points around edge regions are usually unstable feature points, as their position along the edge is not well defined. Therefore, corner points are generally the preferred feature points.

Another source of features are line segments. In [87], a modified variant of the snake algorithm (see Section 2.2) has been used to detect straight line segments. The authors use a formulation of E_{int} which favors straight lines. Another frequently used method for straight line detection is the so called *Hough transform* [36]. Each point in Hough space corresponds to a line object in the input image, defined by angle and offset. Peaks in Hough space are used to identify dominant line objects. A disadvantage of the Hough transform is its high computational cost.

Depending on the type of detected features, a suitable method for establishing feature correspondence must be selected [42]. Direct point-matching methods rely purely on feature point coordinates and typically try to exploit scene coherence. Similar methods are available for line matching. Better results can be obtained, if additional information about the features is incorporated into the matching procedure. If regions are used as features, invariant shape descriptors such as the Fourier descriptors provide a way to measure feature similarity. Fourier descriptors interpret the contour of a region as a complex signal. The magnitude of the coefficients obtained from the Fourier transform of this signal serve as region descriptor and are invariant to rotations of the region.

The elimination of incorrect feature matches can be carried out using robust model fitting algorithms such as the Random Sample Consensus (RANSAC) [34]. RANSAC checks an initial set of matched features for consistency with the transformation model. In each iteration, a subset is randomly chosen from the initial set of feature matches. The subset has the minimal size necessary for computing the transformation model parameters. All remaining feature pairs are then checked for geometric consistency with the model parameters derived from the subset. A feature pair is referred to as *inlier*, if it fits with the transformation model within the limits of a given error threshold. The inliers form a geometrically consistent set, the consensus set. Finally, the transformation parameters are computed by linear regression from the consensus set with highest cardinality over all iterations. RANSAC extensions such as m-estimator sample consensus (MSAC) or maximum likelihood sample consensus (MLESAC) define a loss function also for inliers or fit a combined error model to inliers and outliers.

3.3.2 Area-based methods

Area-based image registration is also known as intensity-based registration or direct method. The basic idea behind area-based registration is to measure and optimize the image similarity directly over the image area or a subimage. Starting from an initial estimate of the transformation model parameters, the target image is registered to the base image. The quality of the registration is then

evaluated using an appropriate similarity measure. By iteratively modifying the transformation model parameters, the similarity measure is optimized. Besides the transformation model, two components are required in order to perform area-based registration: a similarity measure and an optimization strategy.

A large number of similarity measures has been put to the test [46]. They differ in the assumptions made about the image signal. Most methods neglect the aspect of color and work on single-channel intensity images. The direct way of comparing image intensities is to compute the normalized cross-correlation which has been already introduced in the context of object detection (see Section 2.2). Correlation-based registration can also be carried out in Fourier domain [72]. If the image signal in target- and base image shows a sufficient degree of similarity, the correlation coefficient will show a maximum for the best image alignment. A further simplification is to compute the sum of squared differences between pixel intensities. This method is limited to unimodal applications with highly similar values of signal intensity.

Multimodal registration requires a similarity measure which tolerates different intensity levels in the single imaging modalities. *Correlation ratio* is an extended variant of cross-correlation which is explicitly suitable for multimodal registration [86]. Correlation ratio does not require similarity between intensity levels but assumes a functional relationship between intensity levels in each modality. An alternative similarity measure for multimodal registration is *mutual information (MI)*. MI is a measure of statistical dependency between two data sources [69]. The underlying assumption is that if target- and base image are well-aligned, it is likely that there is a high co-occurrence between corresponding pixel values. This means that image parts with a given intensity in the target image are probably mapped to a limited but not necessarily equal intensity range in the base image. MI for an image pair $I_1(x, y)$ and $I_2(x, y)$ can be computed from

$$MI(I_1, I_2) = \sum_a \sum_b P_{I_1 I_2}(a, b) \log \frac{P_{I_1 I_2}(a, b)}{P_{I_1}(a)\, P_{I_2}(b)}, \tag{3.4}$$

where the probabilities P_{I1} and P_{I2} can be estimated from the image intensity histograms and the joint probability P_{I1I2} from the co-occurrence histogram. At the best alignment, the value of MI is assumed to have a maximum.

Area-based image registration can be regarded as an optimization problem of which generally no closed-form solution can be found. Common iterative optimization schemes include variants of the Gauß-Newton algorithm, which requires an estimate of the local gradients of the target function. The downhill-simplex method [64] works without the explicit computation of gradients. Generally, the optimization problem can be ill-conditioned. This problem increases with the

degrees of freedom of the transformation model. To assure convergence towards the optimal solution, proper measures of regularization are needed.

A further improvement of area-based image registration has been reported in [94]. The author targets the task of surface reconstruction from scans obtained by using the atomic force microscope (AFM) and SEM, which is a multimodal registration problem. Instead of directly registering the image data obtained from AFM and SEM, a further step of preprocessing is introduced. The AFM image is converted into a simulated SEM image by modeling parts of the image acquisition process. As a result, the images to be registered look more alike, making successful registration more likely.

3.4 The coarse-to-fine approach

Multimodal image registration based purely on image contents is a computationally expensive procedure. In medical image registration where three-dimensional scans are registered, processing times in the range of multiple hours are not unusual. A common difficulty especially in area-based registration is the high amount of data to be processed. Given the case that all options for parallel processing are exhausted, data reduction is another means of speeding up the registration procedure. On the other hand, in contrast to parallel processing, data reduction will probably have a negative effect on the registration accuracy.

One of the simplest registration schemes which exploits data reduction works with a single reduced copy of target and base image [88]. The data reduction has been performed by block averaging, which is a form of subsampling. The authors suggest usage of the subsampled copies first. If registration leads to sufficiently high similarity values, the procedure terminates. Otherwise, registration is carried out using the original copies of target- and base image. This idea can be extended by applying the subsampling step multiple times. The resulting structure is a multigrid representation of the images [55], also called image pyramid. Usually, subsampling procedures with spectrally more favorable properties than block averaging are applied. Two common examples are the Gaussian pyramid and the Laplacian pyramid.

The resulting registration scheme is an implementation of the coarse-to-fine registration strategy [42]. Figure 3.2 shows a multigrid representation of a pair of multispectral aerial photographs. Coarse-to-fine registration can be performed by consecutively registering each level of the pyramid and thereby iteratively improving the transformation model parameters. The procedure ideally transfers both from coarse to fine: image resolution and transformation model parameters. Although most coarse-to-fine registration schemes work with multiscale

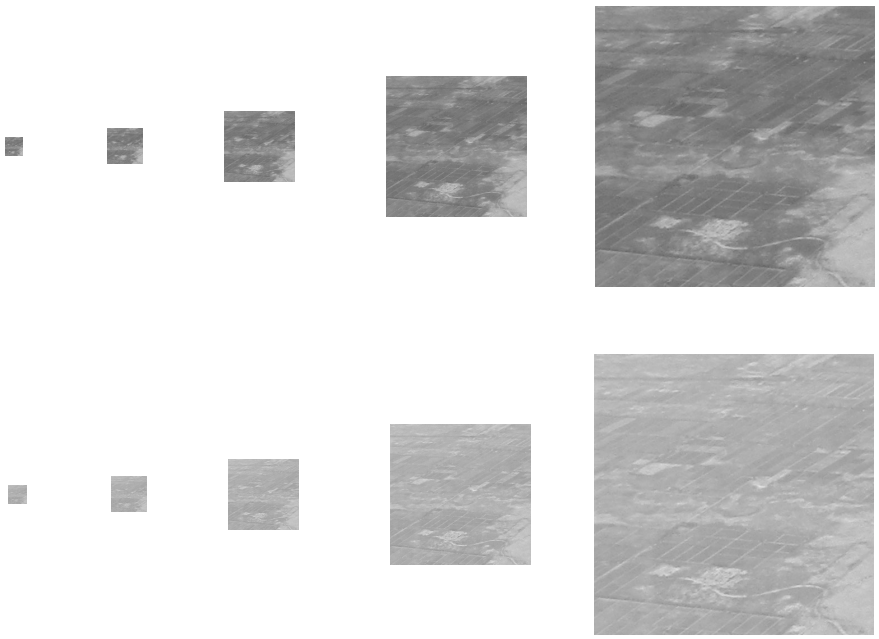


Figure 3.2: The traditional coarse-to-fine approach. Registration is carried out in a series of subsampled copies of target- and base image. Multiple subsampled copies of a pair of multispectral aerial photographs are shown.

representations, the idea of determining a coarse estimate the registration and then improving it can be exploited in different ways. For instance, an initial registration obtained by feature-matching and edge alignment has been improved by optimizing feature correspondences [51]. Improving a manual registration result by a simulation-based fine registration step [94] is another implementation of the coarse-to-fine idea.

3.5 Limitations of the state of the art

It has been shown throughout the chapter that a wide variety of registration methods is well-established and used in daily practice in a multitude of applications. The most difficult applications are those containing multimodal registration based purely on image contents. The methods which have been mentioned differ significantly under multiple aspects of performance. It has to be noted that the number of available methods is comparably high and a lot of application-specific improvements have been proposed. Nevertheless, the different categories

of algorithms can be characterized with respect to computational complexity, convergence properties and maximal precision.

Due to high computational complexity of measuring image similarity over an area, area-based registration schemes are generally require more execution time than feature-based methods. Depending on the transformation model, feature-based methods can work with a small number of feature correspondences which is beneficial under the aspect of processing speed. The number of features can usually be controlled with the help of a detector threshold. Once, feature correspondence is known, registration can be carried out in a single step, avoiding the time-consuming iterative optimization procedure common in area-based registration. Feature-based methods generally do not require an initialization but can work directly in an unconstrained search space of transformation model parameters. Area-based methods need an initial estimate of the parameters and usually, the radius of convergence is small. Generally, they tend to have a problem with managing larger differences in scale.

On the other hand, area-based registration schemes measure image similarity directly. If an appropriate similarity measure is chosen, the perceived similarity is maximized and optimal results are obtained. Feature-based registration can suffer from incorrect feature correspondences. Even if model fitting procedures such as RANSAC are incorporated, there is no guarantee that systematic errors in feature localization will be compensated. Therefore, generally the most accurate results are obtained by area-based registration procedures.

In combination with multiscale representations and the coarse-to-fine approach, area-based registration can be significantly speeded up. Anyways, the convergence properties do not benefit from this modification but are more likely to degrade. Both registration strategies can be made more alike by either working with a large number of features or by performing area-based registration on a large number of sub-patches. In these cases, the strict separation between the two categories of methods will vanish. However, instead of combining the benefits this approach merely combines the drawbacks of both registration strategies.

The state of the art lacks a universally applicable method for multimodal image registration which overcomes the conceptual limitations of feature- and area-based registration.

4 Development of an SPR-based system for micro- and nanoscale object classification

Performing automated manipulation or assembly tasks at the micro- and nanoscale requires awareness of workpiece and tool locations and state. Visual feedback has been proven to be one of the most important sources of such information. This chapter presents a system for micro- and nanoscale object classification, which extends the capabilities of the state-of-the-art procedures mentioned in Chapter 2. It incorporates statistical pattern recognition for determining the class membership of micro- and nanoscale objects.

4.1 Object classification in automated micro- and nanorobotic tasks

The state-of-the-art methods for micro- and nanoscale object detection described in Chapter 2 suffer from two types of limitations. They are limited to a special arrangement of the setup and they show a low generalization capability. Another problem is the manual setting of decision thresholds. The proposed procedure overcomes these limitations by incorporating methods of statistical pattern recognition (SPR) for the classification of individual objects. This brings the benefit that instead of explicitly specifying decision rules, a machine learning procedure performs this task. In contrast to other methods, the proposed system can be trained for a large range of problem settings. Thereby, object inspection is not limited to the object shape, but can cover all object properties which can be extracted from the image material or measured by other means.

In the targeted application scenario, the proposed system is integrated into an automated environment for micro- and nanorobotic tasks. For being universally applicable, the system is required to be remotely configurable by the high-level controller. Depending on the actual subtask carried out, components of the system can be selected automatically. The system is required to handle a large

range of possible micro- and nanoscale objects. Those will be imaged with highly different image sensors, providing two- or (pseudo-)three-dimensional image data.

Another requirement to the proposed system is the capability of real-time processing of the image data during execution of an assembly or inspection task. In this context, the capability of real-time processing is defined to be the ability to process the incoming amount of image data continuously. Naturally, this rate is varying strongly between different applications.

4.2 Modular system for object classification at the micro- and nanoscale

The proposed system includes four processing steps, which implement SPR for the classification of micro and nanoscale objects. In order to assure versatile applicability, the functionality of each processing step can be exchanged modularly. The processing starts with acquisition and preprocessing of the image data. This step includes reformatting of the image data, selection of a region of interest and optional conversions of the color space. Next, objects are segmented from the image background and a list of connected objects is created. The following processing steps are carried out separately for each object. A set of object features relevant for object classification is extracted for each object. Finally, the object class is determined by the classifier. The system output is a list of object locations and their class memberships. In the following, the four processing steps will be described in more detail.

4.2.1 Acquisition and preprocessing

Imaging sensors used in micro- and nanorobotics have different characteristics than those used in other fields of machine vision, for instance industrial product inspection. The acquisition and preprocessing step capsules all tasks necessary to create a uniform representation of the image data obtained from different sensors. Also noise and other sensor-specific sources of image degradation are handled in this step. Some characteristics of four image sensors which are frequently used in micro- and nanorobotics are summarized in Table 4.1.

In microrobotics, a widely used source of image data is the optical microscope (OM), equipped with a digital image sensor. The resolving capacity is limited to some hundreds of nanometers. Also the depth of field is traded off with the field of view size and is generally a limiting factor. On the other hand, the acquisition speed is high and is only limited by the capabilities of the digital image sensor. Depending on th actual type of image sensor used, the optical microscope is able

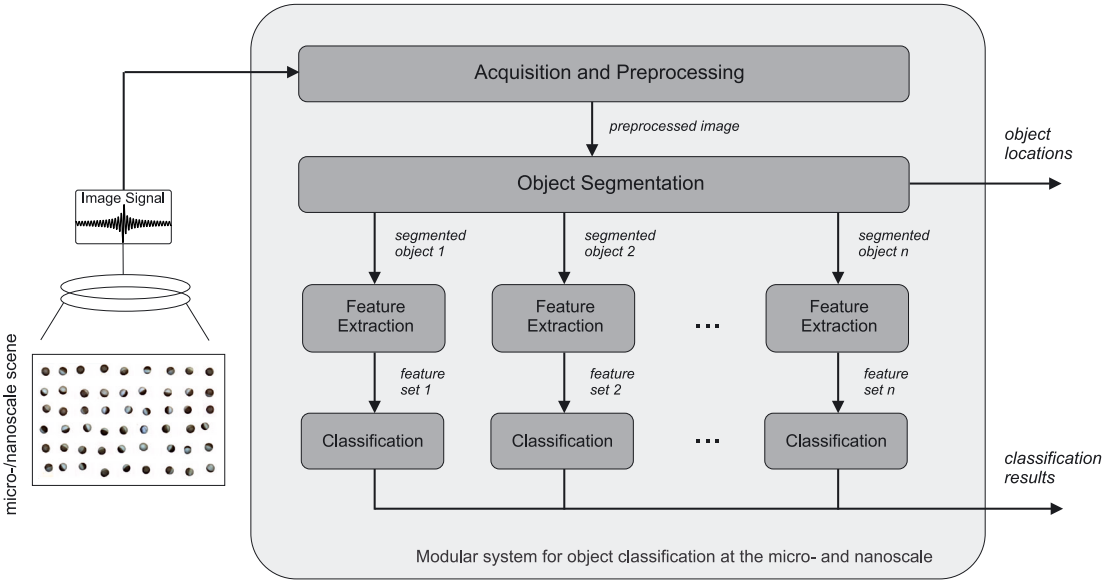


Figure 4.1: Overview of the modular system for object classification. The first processing block performs image acquisition and image preprocessing. In the image segmentation step, candidate objects are located. For each object, a set of features is extracted and serves as input for object classification.

to acquire color images. This is in contrast to the other imaging techniques, which acquire intensity images.

The SEM forms an electron beam, which is used to scan the sample surface [84]. An electron gun equipped with a tungsten filament or a field emission gun is used as an electron source. The electron beam is accelerated by an electric field and focused by condenser lenses. Another set of lenses performs the beam deflection in a raster-like fashion. From the electrons or photons emitted by the sample, a signal can be measured that is used in order to form an image. In contrast to the optical microscope, the SEM has a higher resolving capacity and also a higher depth of field.

The AFM is a special form of the scanning probe microscope (SPM). It probes the sample surface with the tip of a cantilever [8]. From the deflection of a laser beam pointing at the cantilever, the force between tip an surface is concluded. An alternative approach to measurement of the cantilever bending incorporates piezoresistive elements into the cantilever. Generally, three modes of AFM operation can be distinguished, based on the tip-sample interaction: *contact mode*, *intermittent contact mode* and *non-contact mode*. These modes differ in envi-

	OM	SEM	AFM	MRI
dim. of image data	2D	2D	pseudo 3D	2D, 3D
type of image data	intensity/color	intensity	intensity	intensity
typical FOV width	0.5 - 5 mm	20 - 200 µm	10 - 50 µm	5 - 30 cm
typical T_{acq}	50ms	1s	5 - 10 min	1s - 10min

Table 4.1: Image sensors with imaging capabilities at the micro- and nanoscale

ronmental requirements, signal quality and the risk of causing damage to the tip or sample. The resulting AFM scan is a pseudo-three-dimensional view of the scanned surface. It can be stored in the form of an intensity image.

MRI exploits the effect of magnetic resonance with the help of a spatial coding mechanism. In contrast to the other imaging techniques which have been mentioned, MRI is able to acquire true three-dimensional image data. MRI can be applied in order to guide medical interventions performed by micro- and nanorobotic devices. Clinical MRI scanners have a resolving capacity in the range of hundreds of microns. However, with the help of contrast agents, also smaller structures can be localized.

The proposed system for micro- and nanoscale object classification is not limited to these four image sensors. By modifying the acquisition and preprocessing routine, the system can be prepared to work with additional sensors with imaging capabilities on the micro- and nanoscale. These might include a transmission electron microscope (TEM), equipped with a digital image sensor. Also, additional forms of SPM can be supported.

A goal of preprocessing the raw image data is to remove sensor-specific artifacts and simultaneously preserve all information relevant for the classification task. Preprocessing also supports subsequent image segmentation. It has to be noted that in contrast to image enhancement routines, the preprocessing incorporated here does not aim at producing detailed and realistic images for visual inspection. Instead, the preprocessing is carried out in order to support a concrete segmentation and classification task. In contrast to enhanced micrographs usually used in manual image interpretation, the preprocessed images will look comparably poor in image detail.

The probably most important preprocessing task is the removal of image noise. An effective method for noise reduction is frame averaging. However, in time-critical applications and also in scenes captured in motion, frame averaging is not an option. An alternative method which can be applied in a single frame is filtering using a Gaussian low-pass filter. This method is applicable, if high-frequency components of object shapes are irrelevant for object classification.

Another preprocessing step needed for AFM scans is fitting of a common ground level.

4.2.2 Segmentation

Once, the preprocessed image is available, candidate objects can be segmented from the image background. The result of the image segmentation step is a set of mask images of similar size as the preprocessed image. The masks indicate the position of all segmented objects in the image scene. Together with the preprocessed image, the mask of each segmented object is delivered separately as input to the following processing steps.

Image segmentation is carried out by exploiting specific properties of the micro- or nanorobotic setup. If possible, the aspect of image segmentation should be taken into account already when arranging the setup. In many cases, a favorable arrangement is easier to create than to solve a difficult segmentation problem resulting from an unfavorable arrangement. An example of this behavior is found in setups, where background subtraction can be considered. Background subtraction is probably the most simple form of image segmentation. It requires an image background which can be modeled by a simple statistical model and which is not effected by foreground objects. The perfect background for background subtraction is constant over time and is not changed by shadows or other effects caused by foreground objects. If such a behavior can be created by rearranging the setup, this should be the preferred solution rather than composing a sophisticated segmentation algorithm.

In application cases where the image background is not constant over time, image thresholding techniques can be considered [95]. A requirement is that objects of interest can be differentiated from the image background by their intensity values. The concept can also be transferred to color image segmentation. The simplest way of performing image thresholding is to define a threshold value T_{th} and to assign a constant value to it. The segmentation result is obtained by performing the $\leq T_{th}$ or $\geq T_{th}$ operation on each image pixel. On the other hand, for most sensors with imaging capabilities on the micro- and nanoscale, a stationary level of pixel intensities cannot be guaranteed. Therefore, adaptive thresholding techniques are a better choice.

Adaptive thresholding techniques analyze the histogram of image intensities and determine an individual threshold value T_{th} in each image frame. A multitude of different strategies for determining the threshold value are in use [95]. Some of the most popular methods are based on clustering [79] or entropy of the foreground an background signal source [59]. These methods allow a fast computation of a global threshold value.

An adaptive procedure widely used in medical image processing is the expectation maximization (EM) algorithm [77]. The EM algorithm requires a statistical model and aims at iterative estimation of the model parameters. A Gaussian mixture model (GMM) is a common choice of such a model [36]. The model assumes that each segmentation zone in the image is a random process with a Gaussian characteristic. In contrast to the thresholding techniques mentioned before, the number of segmentation zones is not limited to two. Image pixels generated by each source must be described by a feature vector, which can include image intensity or also color information. For each random process, mean value and standard deviation are the model parameters. Also a-priory probabilities for a pixel originating from each source must be specified. In each iteration, two steps are carried out:

- The expectation step computes the a-posteriori probabilities of each pixel belonging to each process. For obtaining an intermediate result, each pixel can be assigned to the process with highest a-posteriori probability.
- In the maximization step, the parameter vector is updated by pretending, the results from the expectation step were new measurement data.

If image segmentation cannot be carried out based on image intensities, object edges can be exploited in order to detect the outer object boundaries. For example, the Canny edge detector [13] can be utilized for object segmentation. However, this approach requires a low number of edges in the image background. Also, additional processing will be needed in order to handle edges inside object regions.

In many applications, none of the segmentation methods which have been described will directly lead to a satisfying segmentation result. A common problem are regions inside an object region, which have been incorrectly classified as image background. This problem can be overcome by applying region filling. Initially, connected components are identified in the segmented image. Region filling is then performed separately for each connected component, by including all regions which are completely enclosed by the connected component. The resulting regions indicate the candidate object location and can be described in the form of a mask image or by the outer enclosing contour.

4.2.3 Feature Extraction

The goal of the feature extraction step is to compute a feature vector $\mathbf{v}$ for each candidate object. The feature vector contains object features which are relevant for object classification. Good object features are invariant to the viewing perspective and other acquisition-related settings. On the other hand, in micro- and nanorobotic setups the acquisition conditions can normally be controlled to a certain degree. Therefore, the level of invariance needed must be chosen according to the actual application. A minimum requirement for most setups is feature invariance to translation and rotation. Naturally, the success of the SPR-based approach strongly depends on the selection of meaningful features and an appropriate method for feature reduction, if needed [44]. The selection of features should show a high discriminative power with a high inter-class variance and a low intra-class variance of the feature values. Some broad categories of object features are geometric-, densitometric-, texture- and color features.

Geometric features are obtained by analyzing object shape. Simple geometric features such as area, perimeter, or largest diameter are variant to scale. Nevertheless they are useful for feature description for two reasons. They serve as building blocks for composing invariant features. Also, in applications with a fixed setup arrangement, these quantities can be meaningful features and can be used directly. A scale-invariant feature composed out of two basic shape features is object roundness R as described in [49]:

$$R_{2D} = \frac{4\pi \cdot Area}{Perimeter^2}. \tag{4.1}$$

A similar measure can be derived for three-dimensional objects:

$$R_{3D} = \frac{6 \cdot \sqrt{\pi} \cdot Volume}{SurfaceArea^{1.5}}. \tag{4.2}$$

Other geometric features can be obtained by directly processing the enclosing contour of the object. BE_{2D} is a scalar value expressing the energy that is stored in an object's contour [49]. The 2D bending energy is calculated by integrating squared curvature values along the object contour. κ is defined as the derivative of contour direction θ_c with respect to arc length s:

$$\kappa = \frac{d\theta_c}{ds} \qquad BE_{2D} = \int_{contour} \kappa^2 ds. \tag{4.3}$$

Normalizing BE_{2D} by the contour length yields a translation, rotation and scaling invariant descriptor of object shape.

The length and shape of the enclosing object contour can be sensitive to variations in the segmentation and preprocessing procedure. High-frequency components of the contour may be filtered out by the noise removal filter. A measure of object straightness which is insensitive to these variations can be derived from the geometrical distribution of the object points $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n]$. It requires computation of the scatter matrix $\mathbf{S_c}$ which is an estimate of the covariance matrix:

$$\mathbf{S_c} = \sum_{i=1}^{n} (\mathbf{x}_i - \overline{\mathbf{x}}) (\mathbf{x}_i - \overline{\mathbf{x}})' \quad \textit{where} \quad \overline{\mathbf{x}} = \frac{1}{n} \sum_{j=1}^{n} \mathbf{x}_j \ . \tag{4.4}$$

Since $\mathbf{S_c}$ is positive semidefinite, two non-negative eigenvalues $\lambda_{1,2}$ and corresponding eigenvectors can be computed. The absolute values of $\lambda_{1,2}$ depend on scale and object elongation. A normalized score indicating object straightness is obtained from

$$\mathrm{PC_E} = \frac{2 \cdot \lambda_1}{\lambda_1 + \lambda_2} - 1 \ . \tag{4.5}$$

Not only object shape but also object surface properties can be relevant to object classification. Densitometric features are extracted from the statistical distribution of the intensity values found on the object surface. They include statistical moments and also entropy measures of the intensity histogram. Texture features additionally take into account the spatial distribution of the intensity patterns. A common method generates surface texture features from a two-dimensional histogram of co-occurring intensity values [47].

If the image sensor provides color information which is also relevant for the classification task, color features can be included in the feature vector. Color features are especially important in biological cell classification. Color images are most likely available in RGB color space, which is a three-channel image with spectral components for red, green and blue color components. These color values change with illumination and many other effects and are unlikely to reproduce. Color features with better reproducibility can be obtained by transforming the RGB values into another color space, which partly separates color properties from illumination intensity. One of the simplest forms of such a color space transforms RGB value pairs into hue, saturation, value (HSV) color space [41]. The hue value is a characteristic property of a color and shows much better reproducibility than the RGB values.

4.2.4 Classification

The feature vector $\mathbf{v}$ obtained for each object is the basis for determining the object class membership ω. This assignment is performed by the classifier, which is a mapping between feature space and object category space. Training is the process of adapting the classifier to the actual problem characteristics. The proposed system for micro- and nanoscale object classification requires training to be carried out as a preparation step. In an automated assembly or inspection operation, appropriately trained classifiers are selected by the control system in order to carry out the specific classification task.

A multitude of methods are known for the construction of classifiers [25, 93]. Besides others, two of the most popular classifiers are artificial neural networks (ANN) and support vector machines (SVM). For a number of reasons, SVMs have been selected as the exclusive classifier in the proposed system. SVMs provide a high level of performance while having a minimum number of open design parameters [12, 100, 106]. This makes SVMs easily adoptable to most classification problems found in automated assembly or inspection tasks on the micro- and nanoscale. SVMs automatically find a balance between two concurring requirements: the capability to abstract from the training data (generalization) and to classify the training data correctly (reclassification). Another benefit of SVMs is the convex nature of the optimization problem, leading to a unique solution.

Training an SVM for a classicication task requires a set of prelabeled data points. For instance, SPM tips can be categorized manually as *suitable* or *defect*. The most basic form of SVMs assumes two object categories which are linearly separable in feature space. Linear refers to the existence of a hyperplane, separating the training data points of both object classes. If such a hyperplane exists, it is most likely not unique. During the training procedure, the SVM searches for the hyperplane surrounded by the largest margin free of training points. This hyperplane is unique and serves as the decision boundary. The non-linearly separable case is depicted in Figure 4.2 a). The training point lying on the border of the margin are referred to as support vectors. Another benefit of SVMs is that adding training points which do not touch the margin does not affect the training result.

Linear SVMs can be extended to work on non-linearly separable data by relaxing the constraints on the margin. The optimization problem for finding the optimal hyperplane is extended by penalty terms for data points, falling into or beyond the margin. The non-linearly separable case is depicted in Figure 4.2 b). Training SVMs includes a training parameter C, which allows balancing training error and margin width. For linear SVMs, C is the only parameter to be selected before training.

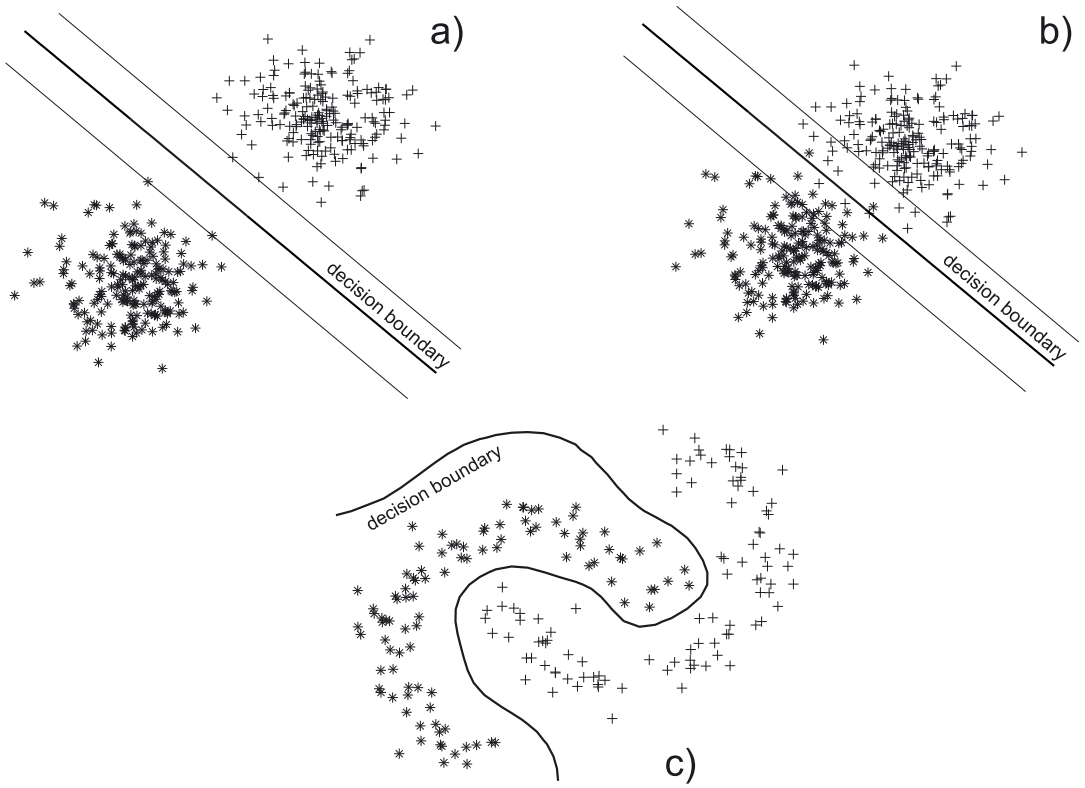


Figure 4.2: Working principle of the SVM in a two-dimensional feature space. The linear SVM places the decision boundary between two classes by maximizing the size of a margin which is free of samples (a). The two lines parallel to the decision boundary indicate this margin. Soft-margin SVMs allow samples inside the margin and even beyond the decision boundary (b). These contribute to the optimization problem with a penalty term. By applying the kernel trick, SVMs can be extended for nonlinear decision boundaries (c).

Many classification problems cannot be solved satisfactorily by means of a linear decision boundary. SVMs solve this problem with the help of the so-called kernel trick. A reformulation of the optimization problem causes the data points to occur only in the form of dot products. It can be shown that exchanging these dot products with a kernel function meeting some requirements is equivalent to transforming the data points into a higher-dimensional space, where linear SVMs are applied. The huge benefit of this procedure is that the transform does not need to be specified or executed explicitly. Practically, only a limited number of kernel functions are used, including polynomials and radial

basis functions (RBF). The kernel function usually depends on a second design parameter. Figure 4.2 c) depicts a decision boundary for a non-linearly separable data set, obtained by training an SVM with an RBF kernel function.

In summary, SVMs automatically generate a classifier of appropriate complexity and depend on only few open design parameters. Practically, these parameters can be selected by training the SVM repeatedly and varying C and the kernel parameter by a power sequence. These properties led to the selection of SVMs as the exclusive classifier used in the proposed system. Although SVMs are originally formulated for classification problems with only two object categories, they can easily be combined for multi-class problems [85].

4.3 System Integration

The proposed system has been designed for being integrated into an automated setup for robotic tasks at the micro- and nanoscale. Typical tasks carried out in such a setup are inspection, manipulation or assembly of micro- and nanoscale objects and structures. Process automation is performed by a control system. The control system collects all available sensor data and computes signals for controlling all actuators. Thereby, the proposed system for micro- and nanoscale object classification is used in the function of a sensor, providing location and object class information. Other typical sensors include position sensors and force sensors. Common actuators include microgrippers and measurement probes.

A simplified view of the proposed system integrated into an automated setup is depicted in Figure 4.3. The proposed system has two types of inputs. It receives image signals from a global visual sensor, observing the sample holder and the manipulation instruments. It also receives configuration commands from the control system. This allows switching of the classification task, depending on the actual stage of the automated procedure. In a usual application case, the automation procedure is implemented as a sequence of tasks including tool calibration, target localization, assembly or manipulation and quality control. The proposed system for object classification can be incorporated at multiple of these steps, especially target localization and quality control. All four processing steps performed by the proposed system are configured by the control system according to the requirements of the actual subtask carried out.

The proposed system can be configured manually with the help of a graphical user interface (GUI). This type of usage is needed mainly for testing a configuration and tuning the algorithm parameters. Also, the data for training the SVM can be generated in manual mode. The four processing steps the proposed system is composed of are implemented in the form of processing nodes. These nodes

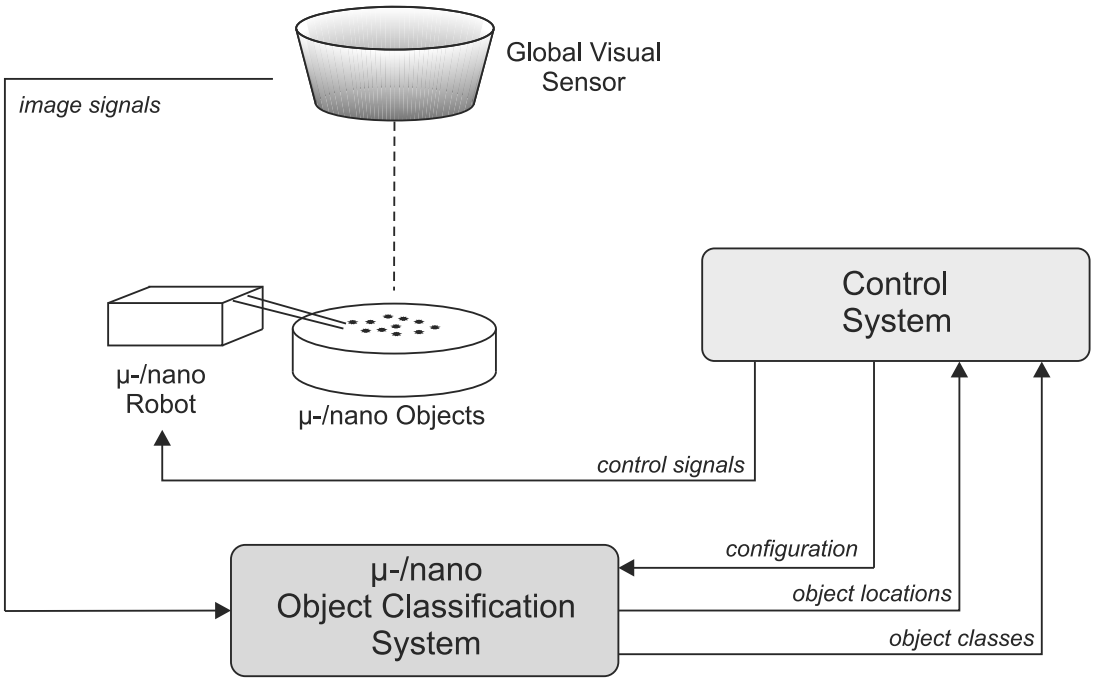


Figure 4.3: Integration of the SPR-based system for micro- and nanoscale object classification into an automated setup. A global imaging sensor provides images of the micro- and nanoscale objects and structures. The classification system performs object classification and transmits classification results and object locations to the control system. The control system configures the classification system and controls all operations performed by the micro- or nanorobot.

are combined in the form of a data flow graph. This method brings a number of advantages. The building blocks can be composed and configured graphically from a set of pre-implemented and pre-tested algorithms. The configuration can be restored conveniently by the control system by saving and loading the data flow graph. Also, the abstraction of processing nodes simplifies parallel processing. Figure 4.4 shows the GUI used for manual composition and configuration of the processing steps. The implementation of the processing graph is described in more detail in references [18] and [114].

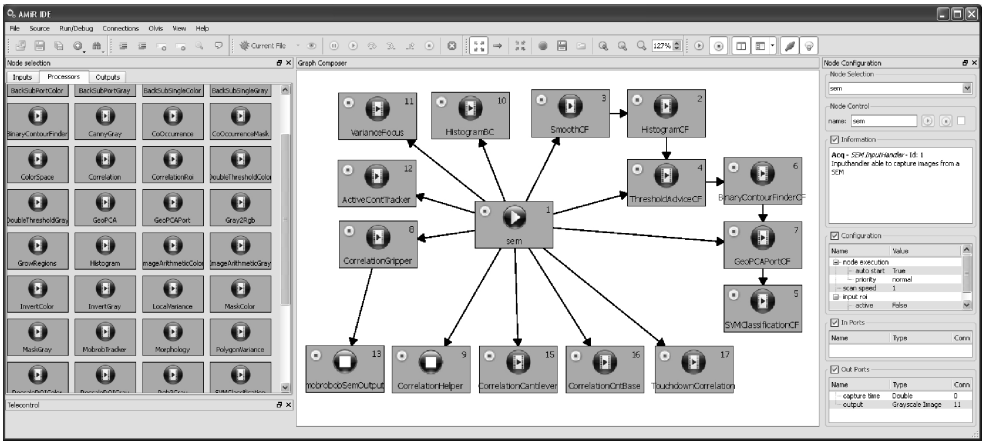


Figure 4.4: Screenshot of the GUI used to compose the components of the SPR-based system for micro- and nanoscale object classification. The system configuration is stored in the form of a data flow graph. These graphs are restored dynamically by the control system.

5 Development of a new multimodal image registration strategy

In this chapter, a new coarse-to-fine image registration strategy for multimodal contents is presented. The proposed method initially performs a feature-based coarse registration and refines the result in an area-based registration step, thereby combining the benefits of both approaches. Main novelties are the usage of scale-invariant local features for multimodal registration, the combination of multiple feature detectors and regularization in the feature-based registration step.

5.1 Multistage procedure for image registration

5.1.1 Overview of the proposed registration scheme

In Chapter 3, the distinction between feature- and area-based registration has been introduced. Both approaches differ under multiple aspects. While feature-based registration schemes tend to show better convergence and shorter execution times, area-based registration usually provides the most accurate results. A combination of both approaches can be achieved by area-based registration of small sub-patches. However, this procedure strongly increases the degrees of freedom and will probably decrease the overall performance. Instead, the proposed registration procedure combines the benefits of feature-based and area-based registration by executing them subsequently. It results in a fast and accurate registration scheme with good global convergence.

An overview of the proposed registration scheme is depicted in Figure 5.1. It assumes two imaging modalities, which provide overlapping image scenes. In the initial step, features are extracted from both image scenes. For each feature, a location and a feature description must be obtained. Next, correspondence between the two sets of features is established, based on the feature descriptors. The initial feature correspondence is checked for consistency in a refinement

step. This is also the most likely point, where failure of the registration can be detected. If the image scenes do not overlap, no consistent subset of feature correspondences with adequate size will be found in this step. From the consistent subset, initial transformation parameters are derived. These parameters are improved in the area-based registration step, which is likely to converge now. At last, the image data is fused and displayed, based on the final transformation parameters.

By an appropriate choice of feature detectors, descriptors and a similarity measure, the aspect of multimodality is covered. The detectors must produce comparable responses on structures imaged by both imaging modalities. Also, the feature descriptors must show enough similarity to allow matching between both modalities. In the proposed registration scheme, a comparably young class of feature detectors and descriptors is used for multimodal registration. These methods are known as local image features [103]. Although local features are mostly known in the context of image retrieval and unimodal registration, surprisingly good results were obtained during initial multimodal registration experiments. This was the motivation for further pursuing local features as a component in the proposed registration scheme. For the area-based fine registration step, the MI criterion is a promising candidate for handling the aspect of multimodality. Nevertheless, also other criteria have been tested.

The proposed registration scheme can be regarded as an implementation of the coarse-to-fine registration strategy. In the traditional coarse-to-fine approach, the transformation parameters are improved from the coarse to the fine level of a multiscale representation. In contrast, the proposed registration scheme improves the transformation parameters by changing to another registration strategy. This does not imply that the proposed scheme cannot handle differences in scale or does not benefit from a multiscale representation. The opposite is the case. Local features extensively make use of a scale-invariant feature representation. In the following it will be shown how the proposed registration scheme enables fully automatic registration while making only few assumptions about the image contents.

5.1.2 Registration using local image features

Local features are image contents which can be distinguished from surrounding image contents. The difference can be in any measurable image property. Local features emerged in the context of image retrieval, where the main task is to identify somehow similar images from a database. A concurring form of image features are global features such as image histograms or other statistics obtained over the whole image area. Global features are usually expressed by a descriptor

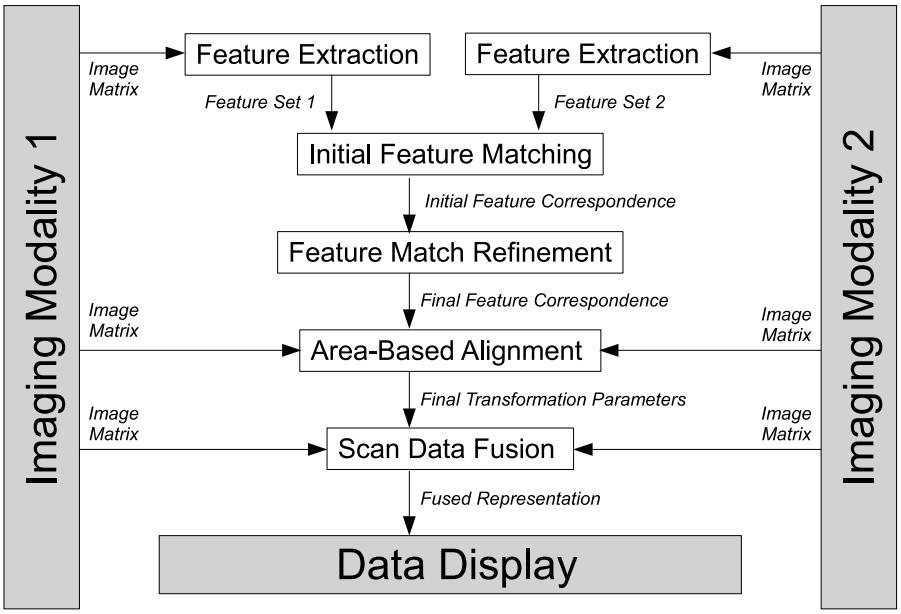


Figure 5.1: Overview of proposed registration scheme. Initially, local features are extracted from both imaging modalities. The features are matched against each other and the match is checked for consistency in a refinement step. Based on the feature correspondence, the transformation model parameters are computed. In the following step, area-based fine alignment is performed. The final fused view is generated and then displayed.

vector. In contrast, each local feature is defined by a 2-tuple, consisting of a region A and a descriptor D. Similar to global features, local features can be grouped and be used to retrieve similar images. The problem of image retrieval can also be reformulated in order to solve object detection tasks [116]. In these contexts, images are described by distributions obtained from the descriptors of all local features found in the image scenes. The local feature location is of minor importance.

On the other hand, the algorithms used for the computation of A and D are powerful tools in the context of feature-based image registration. For example, local features originally developed for image retrieval have been successfully used for photograph stitching [11]. In contrast to application-specific feature detectors (streets, houses in remote sensing applications), local features generally do not provide an interpretation of the physical structure that they have been produced

by. Instead, they should meet a number of ideal properties which have been stated in the literature [103] and will be summarized in the following.

Feature detection should be *repeatable*, which means that features are repeatedly detected under changes in viewpoint, illumination or other imaging conditions. Local features should show *distinctive* image contents, which enable feature matching. The detected features should be *local* and not be effected by distant image contents. Also, the *quantity* should be sufficient for the targeted application. Feature location and extent should be detected *accurately*. The computation should be *efficient*. If any severe deformations are expected these should be modeled by mathematical transformations, so that the detected features become *invariant* to these deformations. Features should also be *robust* to small deformations which are not modeled.

Most algorithms used for the computation of A and D have been developed in the context of image retrieval and content-based object recognition [73]. The vast majority of images used in this context are camera images showing for instance landscape photography or portraits. Therefore, the properties stated in the previous paragraph were optimized with respect to the prevailing forms of geometric and photometric transformations found between corresponding pairs of camera images. Those are mainly:

- *Changes in scale or rotation*, for instance due to varying object distances and orientations,
- *Changes in viewing direction or viewpoint*, typically found in photography and mobile robot vision,
- *Changes in illumination*, in type and/or direction,
- *Compression artifacts*, as most images stored in databases are compressed and
- *Image blur*, as different objects may be focused in multiple images of the same scene.

The development of algorithms for local feature extraction makes the implicit assumption, that image acquisition of the same scene under identical viewpoint, illumination, etc., will result in identical images, optionally corrupted by additive noise. This assumption is clearly violated in the case of multimodal image registration. Even for identical image scenes, the change in imaging modality itself can already be regarded as a geometric and photometric transformation. It will be shown, that multimodal registration based on local features can yield

good results anyways. Apparently, many of the geometric and photometric transformations the algorithms are optimized for are also relevant to the problem of multimodal image registration.

5.1.3 Local feature detectors

The detection of local features is the process of determining a list of regions A_i from a given image. From the properties required of these regions, repeatability is probably the most important. In the context of multimodal image registration, repeatability means that if a region is detected in one modality, a corresponding region is also detected in the other modality. On the other hand, repeatability is concurring with the requirement of distinctiveness. A trivial detector which returns a complete list of image positions and region shapes will show a perfect repeatability but little use for the task of image registration. Especially homogeneous regions will not show any distinctive information which can be used for feature matching.

Detectors based on 1st and 2nd order derivatives

In order to meet the desirable properties, the associated region A of a local feature should somehow differ from the surrounding image area. A way of directly measuring changes in image intensity is to compute local derivatives of the image intensity $I(x, y)$. For example, the Harris corner detector which has been already introduced in Section 3.3 makes use of partial derivatives. As the image intensity is a discrete function, the derivatives are approximated by local differences. In Figure 5.2, local derivatives are shown using the example of a tin-coated surface, imaged by the SEM. Derivatives are generally signed and therefore absolute values are used for visualization. 1st order derivatives highlight image edges and highest responses can be seen around the tin spheres. 2nd order derivatives produce double peaks along edges. The lowest response can be seen in the dark areas which are cavities in the carbon surface.

Besides the Harris detector, two other important detectors are exploiting local derivatives directly: The *Determinant of Hessian (DoH)* and the *Laplacian of Gaussian (LoG)* detectors. Both detectors work purely on 2nd order derivatives and can be derived from the Hessian matrix $\mathbf{H}$:

$$\mathbf{H} = \begin{bmatrix} \frac{\partial^2 \mathbf{I}}{\partial x^2} * \mathbf{w} & \frac{\partial^2 \mathbf{I}}{\partial x \partial y} * \mathbf{w} \\ \frac{\partial^2 \mathbf{I}}{\partial x \partial y} * \mathbf{w} & \frac{\partial^2 \mathbf{I}}{\partial y^2} * \mathbf{w} \end{bmatrix}. \tag{5.1}$$

Again, $\mathbf{w}$ can be a Gaussian smoothing kernel and $*$ is the convolution operator. The DoH detector response is obtained by computing the determinant of $\mathbf{H}$.

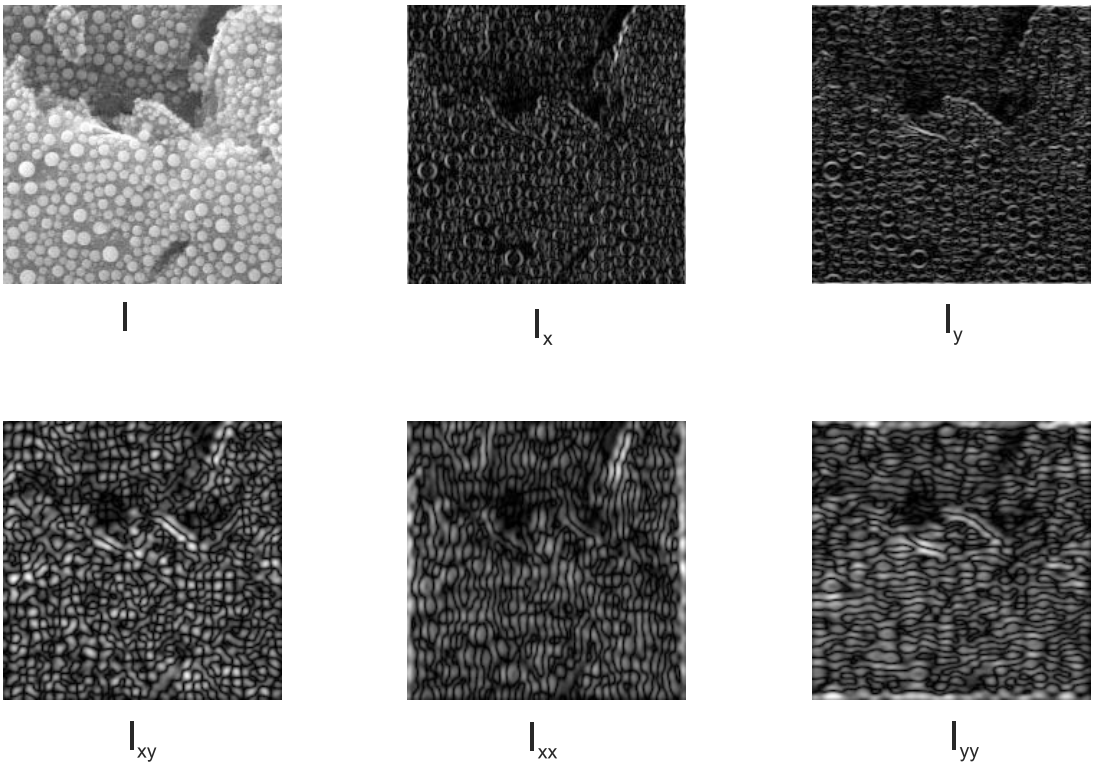
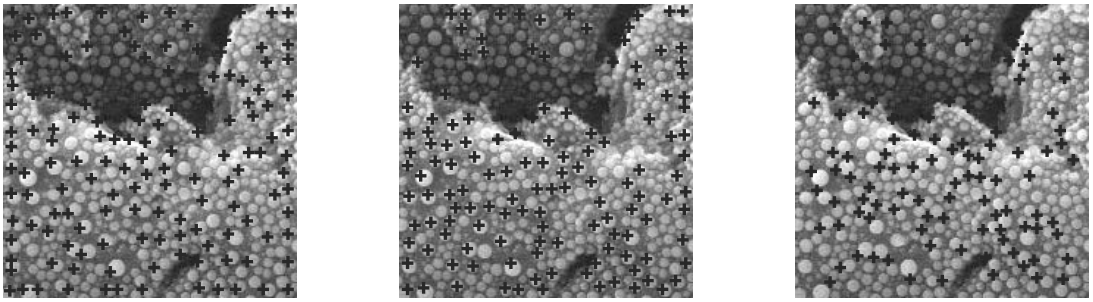


Figure 5.2: SEM scan of a tin-coated carbon surface. The field of view width is 2.84 µm. Besides the original scan I, a selection of discrete derivatives absolute values are shown.

The LoG response is given by the trace of **H**. Figure 5.3 shows some peaks in the detector responses of the native Harris detector (tuning parameter $k = 0.04$), the DoH and the LoG detector. The Gaussian parameter is $\sigma = 2$ in all cases. It can be seen that the Harris detector responds mainly to corner-like structures between the tin spheres. The DoH mainly responds to the spheres. Most of the LoG responses are found along the edges of bigger tin spheres.

All detectors fulfill the locality requirement because each peak in the detector responses is caused by a locally limited region of support. Due to the use of the Gaussian smoothing function, the borders of the region of support are unsharp and depend on the discrete approximations actually used. An estimate of the size of this region can be obtained by assuming the following approximation of the local derivatives in each direction:



a) b) c)

Figure 5.3: Peaks in detector responses of native Harris (a), Determinant of Hessian (b) and Laplacian of Gaussian (c) detector. The image scene is identical with Figure 5.2.

$$\mathbf{D}_x = \begin{bmatrix} 0 & 0 & 0 \\ -1 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}, \quad \mathbf{D}_y = \begin{bmatrix} 0 & -1 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}. \tag{5.2}$$

With a 7×7 approximation of the Gaussian kernel, the region of support will cover an area of 9×9 pixels. On the other hand, the contribution to the detector response of pixels at the corners of this region is minimal. A better approximation of the region of support is the incircle of this square neighborhood.

When the image is rescaled, the region of support will cover a bigger or smaller physical area. Compared to the area covered in the original image, the new region of support will probably cover a different amount of image structure. As a result, the detector response of all three detectors is scale-variant. It can be seen very well in Figure 5.3 (b), that the DoH preferably detects tin spheres of medium size. Although smaller or bigger spheres appear as a rescaled version of the medium-sized spheres, they produce lower detector responses. Figure 5.4 shows the scale-variance of the Harris detector response for a given point in the image scene. As the image scene is rescaled, the detector response reaches two peaks.

The two peaks of the detector response shown in Figure 5.4 are actually peaks not only over image area but in scale-space. Regardless of the actual scale in which an image scene is presented, the regions of support associated with these peaks can be repeatedly detected. It has been shown that the Harris and DoH detectors deliver responses which are spatially most stable [74]. On the other hand, the LoG responses are most stable over scale. As a result, the *Harris-Laplace* and *Hessian-Laplace* detectors have been proposed [73, 74]. The idea is to work with peaks obtained by the Harris and DoH detectors in a coarsely

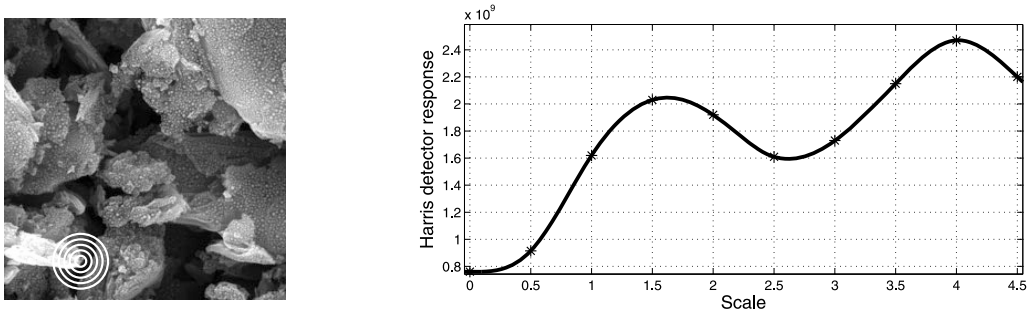


Figure 5.4: A feature point has been detected using the Harris detector (left). Cavities in a tin-coated carbon surface are shown. The field of view width is 14.45 µm. The diagram shows the detector response for the marked position in relation to scale. Scale=0 corresponds to the original image resolution which was 1024×1024 pixels. The detector response shows two peaks.

sampled scale-space. The scale-detection is then refined by interpolating the peaks of the LoG detector.

Maximally stable extremal regions (MSER)

Thresholding an image scene is a standard method in image segmentation. Each image pixel is classified as being of higher or smaller (or equal) intensity than a given threshold value. By switching between the $<$ and $>$ operation, connected regions which are darker or brighter than the surrounding image area can be segmented. Based on this principle, a local feature detector has been constructed [70]. Instead of using a global threshold value for all segmented regions, the *maximally stable extremal regions (MSER)* detector varies the threshold value for each segmented region and both thresholding variants ($<$ and $>$). By modifying the threshold value, the region size will likely either increase or decrease. For meeting the repeatability and robustness requirement, the MSER detector is looking for threshold values which lead to a maximally stable region size. The resulting list of regions can be repeatedly detected under a number of geometric and photometric transformations. Generally, the detector output can be an arbitrarily-shaped connected image region. However, this shape is unlikely to reproduce exactly under image transformations and especially in the multimodal application case. For this reason and better comparability with the other detectors, an ellipse is fitted into each output region and used as the final detector output.

Intensity-based regions (IBR)

There is little hope that image intensities will be reproduced under photometric transformations. On the other hand, the position of local extrema in image intensity can be stable under several transformations. *Intensity-based regions (IBR)* is a local feature detector which is based on this assumption [104]. The IBR detector starts with a list of local extrema of image intensity in scale-space. From each extremum, a set of radial straight lines is sampled and sudden changes in image intensity are detected. The area enclosed by connecting these points form a region, which is stable under changes in scale and other transformations. The working principle is depicted in Figure 5.5. Again, the region can be approximated by fitting an ellipse to the boundary points.

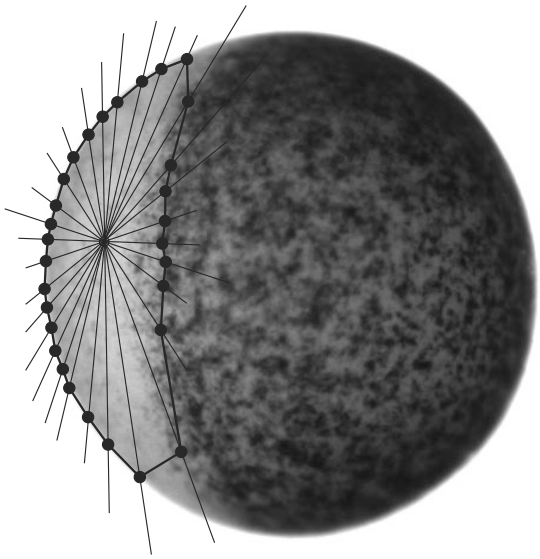


Figure 5.5: Working principle of the IBR detector. Starting from an intensity extremum (center point), straight lines are sampled and sudden changes in image intensity are detected. The image shows an oocyte of Xenopus Laevis, imaged using an optical microscope. The diameter is approximately 1mm.

Salient region detector

A requirement to local feature detectors is the distinctiveness or informativeness of the detected image region. The *Salient region detector* is based on directly measuring the informativeness in the form of an entropy measure over a candidate

region [58]. Therefore, a grid of circular candidate regions is placed over the complete image area. By varying the region radius, the detection is carried out at multiple scales. The entropy measure is based on the probability density function (PDF) of image intensities in the enclosed area. For this reason, peaks in entropy are widely invariant to many geometric transformations. A problem of the entropy measure is the bad localization of extrema over scale. This problem is solved by adding a second criterion, which is the derivative of the PDF of a candidate point with respect to scale. Both criteria are combined for computing a saliency score. This score serves as the detection threshold value.

Scale-invariant feature transform (SIFT) feature detector

The *Scale-invariant feature transform (SIFT)* uses a very efficient implementation of the LoG detector [68]. Instead of creating a multiscale representation of the input image first and then computing the LoG detector response for each point in scale-space, the SIFT detector combines both steps. The SIFT detector uses the Gaussian smoothing kernel with the scale-space parameter σ. For an input image $I(x, y)$, the scale-space is defined by the function

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y), \tag{5.3}$$

with the Gaussian smoothing kernel

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-\left(x^2+y^2\right)/2\sigma^2}. \tag{5.4}$$

The LoG detector response can be approximated by the *Difference of Gaussian (DoG)* detector:

$$D(x, y, \sigma) = \Big(G(x, y, k\sigma) - G(x, y, \sigma)\Big) * I(x, y) = L(x, y, k\sigma) - L(x, y, \sigma) \ . \tag{5.5}$$

k is a constant factor separating two close-by scales. Practically, only a limited number of scales is explicitly computed and maxima in $D(x, y, \sigma)$ are interpolated. Because the LoG as well as the DoG give strong responses along image edges, The SIFT feature detector computes the local Hessian matrix for candidate feature points. The Hessian helps to eliminate edge responses and needs to be computed only for a limited number of candidate points.

Speeded-up robust feature (SURF) feature detector

Similar to the LoG detector, the DoH detector can also be simplified and integrated with the scale-space construction. The speeded-up robust feature (SURF)

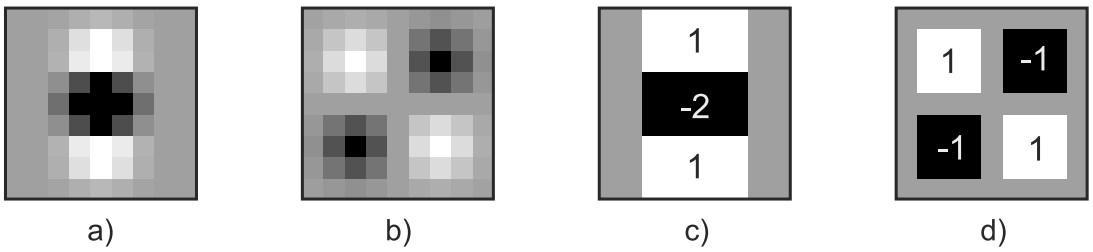


Figure 5.6: Construction of the box filters used in SURF feature detection. The second-order Gaussian derivatives (discretized, cropped) $\frac{\partial^2 G(x,y,\sigma)}{\partial y^2}$ (a) and $\frac{\partial^2 G(x,y,\sigma)}{\partial x \partial y}$ (b) are approximated by the box filters D_{yy} (c) and D_{xy} (d). All filters are signed. Zero value is indicated by grey color. The smallest box filter has a size of 9 ×9 pixels.

[4] feature detector follows this approach, which results in a very efficient implementation. An important concept exploited by the SURF detector is the usage of integral images. For each position, an integral image stores the sum of all image pixels with lower or equal image indices. Once the integral image is stored, integrals over rectangular areas of arbitrary size can be computed with a constant effort of two subtractions and one addition operation. The integral image is reused in the construction of the scale space, the approximation of the DoH and also later in the computation of the feature descriptor.

The SURF detector approximates the DoH detector with the help of so-called box filters. Figure 5.6 depicts two of the Gaussian derivatives and their box filter approximation. The rectangular areas the box-filters are composed of are evaluated efficiently using the integral image. Also, the coefficients (-2, -1, 0, 1) are computationally friendly. The scale-space is not constructed explicitly in the form of an image pyramid but instead implemented implicitly by varying the box filter size. Thereby, the smallest box size of 9×9 pixels corresponds to a scale of $\sigma = 1.2$. Detector responses can be estimated continuously for any scale by interpolating the detector responses of box filters with corresponding neighboring scales.

Random sampling

With respect to the desirable properties of a local feature such as repeatability and distinctiveness, randomly selecting points in scale-space promises the lowest probability of success. On the other hand, randomly selected points may cover areas which other detector do not respond to. Also, the number of points and range of detected scales may be controlled very easily. However, in most cases

randomly selected points are used only for performance comparison with other detectors. Figure 5.7 shows a number of detected regions in an SEM scan, in which clusters of melamine spheres can be seen. For IBR and MSER, the ellipse-fitting approach has been used. It can be seen clearly, that the detectors respond to different formations of melamine spheres. Also the detected region sizes are very different.

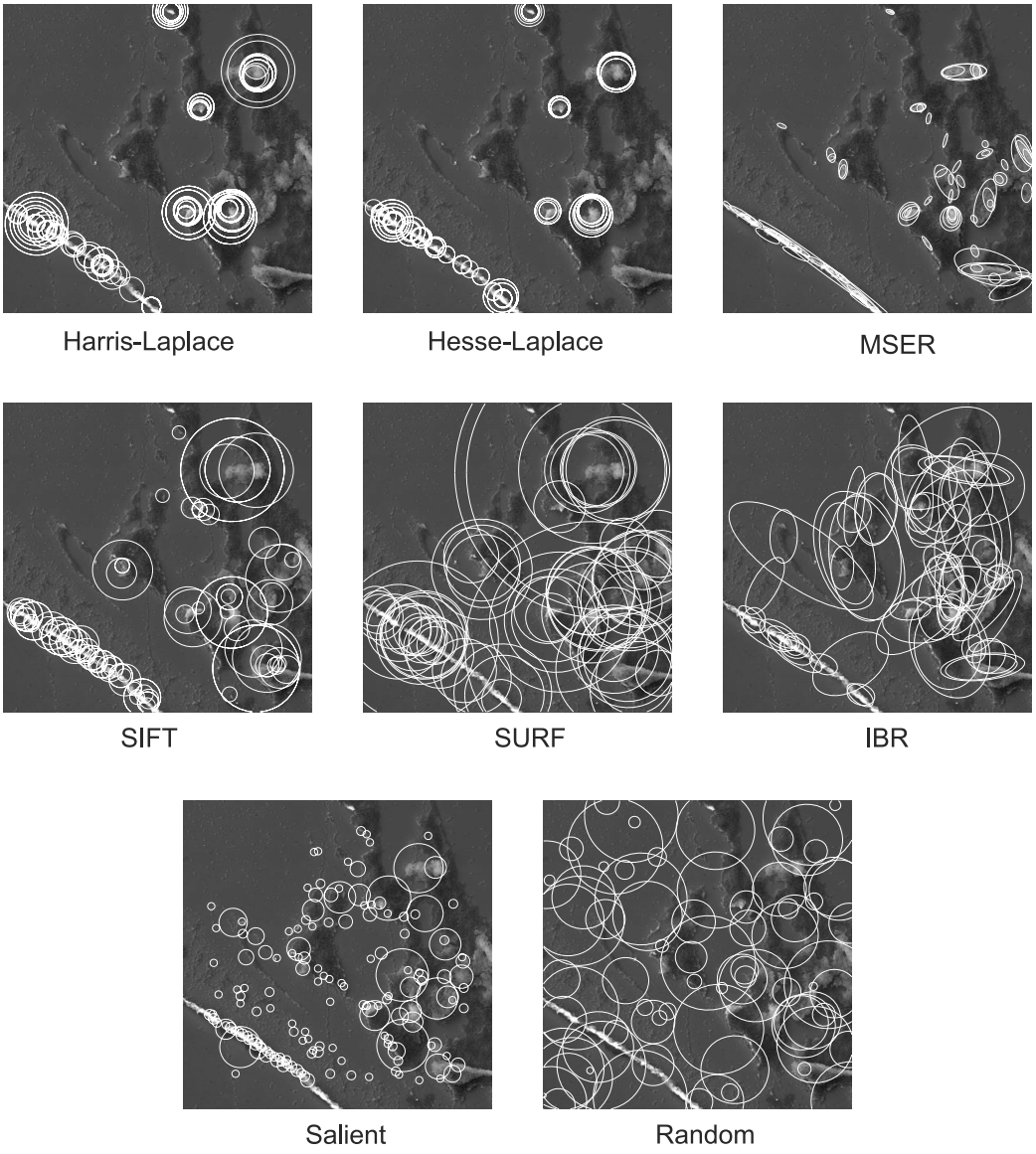


Figure 5.7: Comparison of all detector responses on an SEM micrograph, showing melamine spheres. The field of view width is 127.1 µm.

5.1.4 Local feature descriptors

From each detected region a descriptor needs to be extracted, which can be used to establish a feature correspondence. Similar to the region detector, the ideal descriptor should reproduce under a number of transformations and also in different imaging modalities. On the other hand, the descriptor must contain enough information, allowing to distinguish between features which have a high degree of similarity. The ideal descriptor will extract information relevant for the identification of the feature and discard modality-specific imaging artifacts, noise and other irrelevant information. Naturally, a clear separation is impossible and all descriptors implement a trade-off between these concurring requirements. Additionally, the descriptor must be robust to positioning inaccuracies of the feature detectors. Corresponding regions in both imaging modalities will rarely cover exactly the same physical area but always be biased in position and/or scale. A good descriptor will show small variations to this sort of inaccuracies.

In order to allow a simple measurement of similarity, the descriptor should produce an output in the form of a vector of scalar values. Many descriptors are rotation-variant. They can be computed invariant to rotation by aligning them with a *dominant orientation* of the region. The SIFT and SURF feature description method each include a procedure for determining dominant feature orientations. For all descriptors without such a procedure, the SIFT method has been applied. In order to compensate for differences in intensity levels, image intensities inside each region are normalized.

Moment invariants

Geometric moments can be used to describe the shape and intensity distribution of local features. However, they are variant to geometric and photometric transformations. Originally proposed for color images [105], moment invariants are an invariant feature descriptor which can be used in the context of feature-based registration. For an intensity image $I(x, y)$ and a region A, the centered shape- and intensity moments are obtained from

$$MS_{Apq} = \int\int_A x^p y^q dxdy \quad and \quad MI_{Apq} = \int\int_A I(x, y)\, x^p y^q dxdy. \tag{5.6}$$

The most simple invariant is given by MI_{00}/MS_{00}. More complex invariants include higher-order moments. Instead of directly computing the intensity moments on the intensity pattern of the image, derivatives of $I(x, y)$ can be used instead [76].

Steerable filters

The local derivatives of an image patch with respect to the image coordinates are variant with rotation and have no use for feature description. It has been demonstrated, that orthogonal filters can serve as a basis for composing arbitrarily oriented filters. This procedure is referred to as *steerable filters* [37]. Two examples of Gaussian derivatives are shown in Figure 5.6 a) and b). By steering Gaussian derivatives in the direction of the dominant feature orientation, rotation-invariant derivatives are obtained. Computing all derivatives up to fourth order yields a descriptor vector with 14 entries.

Complex filters

Similar to the steerable filters method, complex filters construct a filter bank which delivers the entries for the descriptor vector [91]. The filters use complex convolution kernels derived from the following formula:

$$K_{mn}(x,y) = (x+iy)^m (x-iy)^n G(x,y). \tag{5.7}$$

Thereby, $G(x,y)$ is the Gaussian smoothing kernel. Rotation by an angle of θ has the same effect as multiplying the filter response by the complex number $e^{i(m-n)\theta}$. This changes the phase, but not the magnitude of the filter response. The absolute value of the filter responses for different combinations of m and n form the rotation-invariant descriptor. Using all combinations with $m+n <= 6$ and $m > n$ leads to 16 descriptor entries.

Spin images

Intensity histograms discard all information about feature geometry and therefore are of limited use for feature description. However, multi-dimensional histograms can combine information about intensity and location. *Spin images* follow this approach by creating a two-dimensional histogram over intensity and distance from the center point [65]. The histogram is vectorized and forms a rotation-invariant descriptor. The descriptor length can be varied by modifying the number of histogram bins. Figure 5.8 shows three examples of normalized image patches and corresponding spin images.

Scale-invariant feature transform (SIFT) feature descriptor

The SIFT algorithm includes a feature descriptor which is equipped with a procedure for determining the dominant orientation of a feature point [68]. Initially,

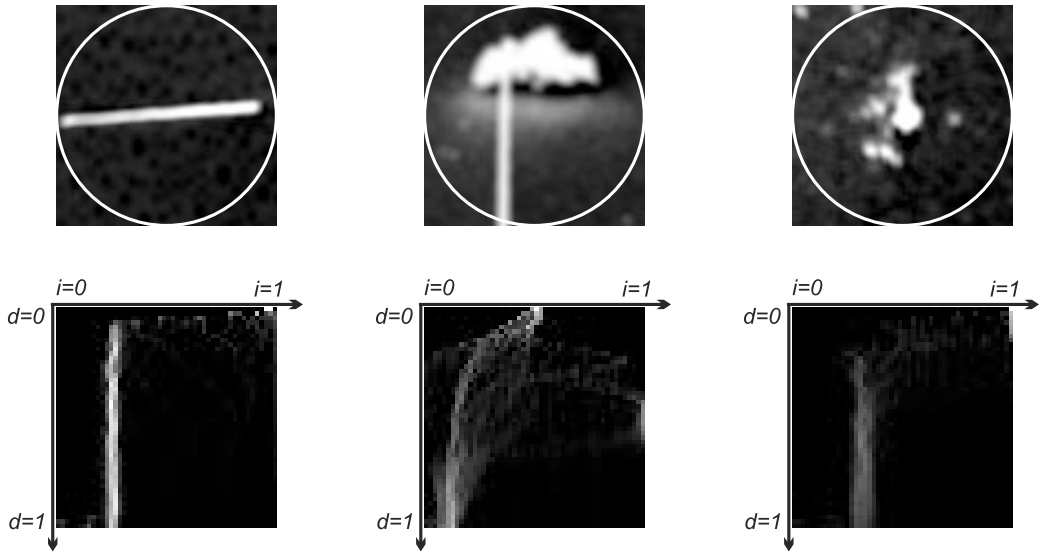


Figure 5.8: Feature description using spin images. The upper row shows three image patches, where the circle indicates the normalized distance *d=1* from the center point of the patch. The lower row shows the extracted spin images.

the smoothed image copy $L(x, y)$ which best fits with the detected scale is selected from the multiscale representation (Equation 5.3). Next, gradient magnitudes and orientations are computed in a neighborhood of the feature point:

$$m(x, y) = \sqrt{(L(x+1, y) - L(x-1, y))^2 + (L(x, y+1) - L(x, y-1))^2} \quad (5.8)$$

$$\theta(x, y) = \tan^{-1}\left(\frac{L(x, y+1) - L(x, y-1)}{L(x+1, y) - L(x-1, y)}\right) \quad (5.9)$$

Additionally, the gradient magnitudes are weighted by a Gaussian kernel, centered at the feature point location. The weighted gradient magnitudes are collected in a 36-bin histogram, covering all 360 degrees of possible orientations. Figure 5.9 a) shows an example of an orientation histogram. The peak in the histogram corresponds to the dominant feature orientation. However, if another orientation exceeds 80% of the peak strength, a further descriptor is assigned to the feature point.

The construction of the SIFT descriptor is depicted in Figure 5.9 b) and c). Gradient magnitudes and orientations for a grid of locations are computed using Equations 5.8 and 5.9. Again, the magnitudes are weighted by a Gaussian,

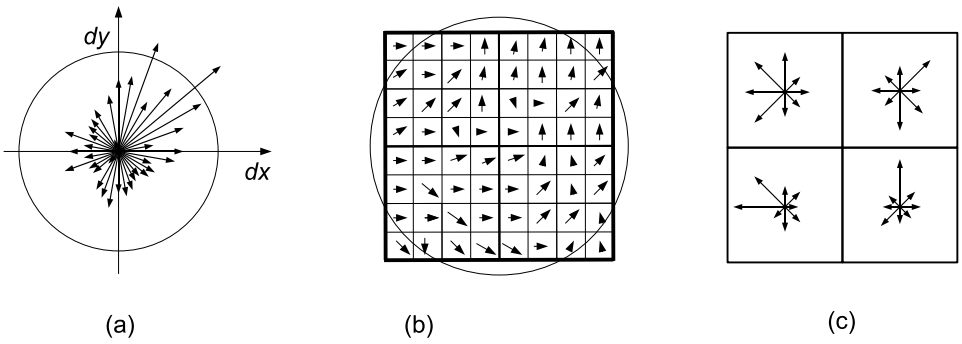


Figure 5.9: Generation of the SIFT descriptor. Initially, local gradient magnitudes and orientations in the neighborhood of a feature point are weighted with a Gaussian window centered at the feature position. From the results an orientation histogram (a) is computed. For all orientations with a magnitude of >80% of the highest peak (indicated by the circle), a descriptor is assigned. Therefore, a grid (b) is placed around the feature point, and local gradient magnitudes and orientations with respect to the assigned direction are computed and weighted by a Gaussian window. From the gradients in each sector of the grid, an orientation histogram is computed (c).

centered at the feature point location. The grid is divided into a number of sectors and the weighted gradients are grouped into orientation histograms for each sector. The original SIFT descriptor uses a grid of 4×4 sectors and 8-bin orientation histograms, resulting in a descriptor vector with 128 components. The vector is normalized in order to compensate for changes in illumination.

Gradient location and orientation histogram (GLOH)

The *Gradient location and orientation histogram (GLOH)* is a modified version of the SIFT descriptor [76]. Instead of using the rectangular grid of 4×4 sectors in which the orientation histograms are computed, a log-polar grid of 17 sectors is introduced. With an increased number of orientation histogram bins, the descriptor length becomes 272. Using principle component analysis, the descriptor length is reduced to 128.

Shape context

Shape context is a descriptor with similarities to SIFT and GLOH [6]. However, instead of gradient locations and orientations, shape contexts use edge points obtained by an appropriate detector such as the Canny edge detector [13]. The original version computes a two-dimensional histogram of log-polar edge point locations. This histogram is referred to as the shape context. An extended version computes one shape context for each edge orientation (horizontal, vertical, two diagonals) [76].

Speeded-up robust feature (SURF) descriptor

The descriptor included in the *Speeded-up robust feature (SURF)* algorithm includes a separate procedure for determining the dominant feature orientation [4]. Similar to the detector, the feature orientation estimation and also the descriptor generation rely on the evaluation of integral images. Both steps compute two-dimensional Haar-wavelet responses, which can be evaluated at constant time with the help of integral images. Figure 5.10 a) depicts the SURF method for determining the dominant feature orientation. In the neighborhood of a feature point, the horizontal and vertical Haar-wavelet responses are computed and summarized in a two dimensional coordinate system. A sliding window of size $\pi/3$ is rotated around the center of the coordinate system. For each position of the window, the vector sum of all wavelet responses inside the window is computed. The orientation corresponding to the highest vector sum found is the dominant orientation of the feature point.

Once the feature orientation has been determined, wavelet responses are computed in a rectangular location grid which is oriented along the dominant orientation (see Figure 5.10 b)). Instead of computing orientation histograms, the SURF descriptor groups the wavelet responses in a sector of the grid by computing local sums. Again, the responses are weighted by a Gaussian, centered at the feature point location. For each sector, the signed sum and sum of absolute values is computed in horizontal and vertical direction. With a grid of 4×4 sectors, a descriptor length of 64 is obtained. The descriptor is normalized to make it invariant to changes in contrast. An extended version of the descriptor separates positive and negative components of each sum and thereby doubles the descriptor length.

Cross correlation

The direct way of expressing an image patch is to sample a scaled and rotated copy of it and to vectorize the obtained pixel grid. If cross-correlation is used to

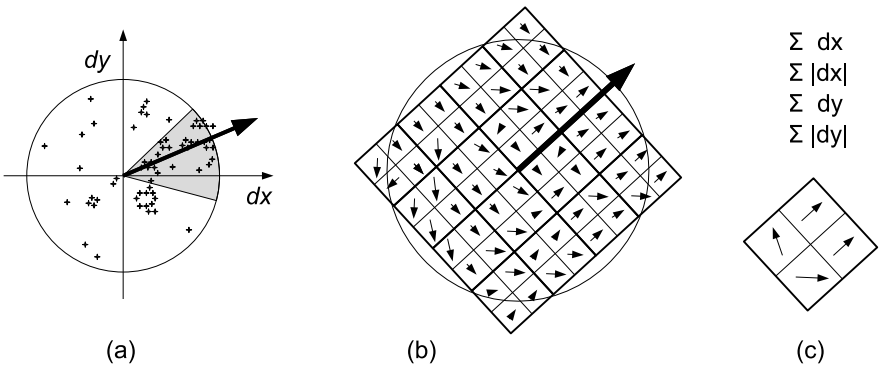


Figure 5.10: The SURF feature descriptor. Wavelet responses for the x any y direction are computed in the neighborhood of a feature point and weighted with a Gaussian window centered at the feature position. The feature orientation is assigned with the help of a sliding window (a). A grid of wavelet responses relative to the assigned feature orientation (b) is then used to compute a descriptor entry. The descriptor entry for each block (c) is built of the sum and the sum of absolute values of the local wavelet responses. Finally, the SURF descriptor is composed out of all block entries.

measure the vector similarity, the method becomes equivalent to the correlation-based object detection methods mentioned in Chapter 2. Differences in the intensity levels of the imaging modalities can be compensated for by normalization. The density of the sampling grid trades off the level of feature detail and descriptor robustness.

5.1.5 Feature matching

By establishing the feature correspondence between the two feature sets, the spatial transformation parameters can be estimated. Generally, feature correspondence can be determined by a nearest neighbor search using the invariant descriptor vectors. However, no algorithm is known for solving the exact nearest neighbor problem in high-dimensional space which is faster than exhaustive search. The Best-Bin-First algorithm [5] is a common approximation of the nearest neighbor search, which gives a speedup by about 2 orders of magnitude. Also, unstable feature matches where the nearest neighbor distance is too close to the second nearest neighbor distance can be excluded. The approximation brings the risk of not finding the exact nearest neighbor. On the other hand, even for true

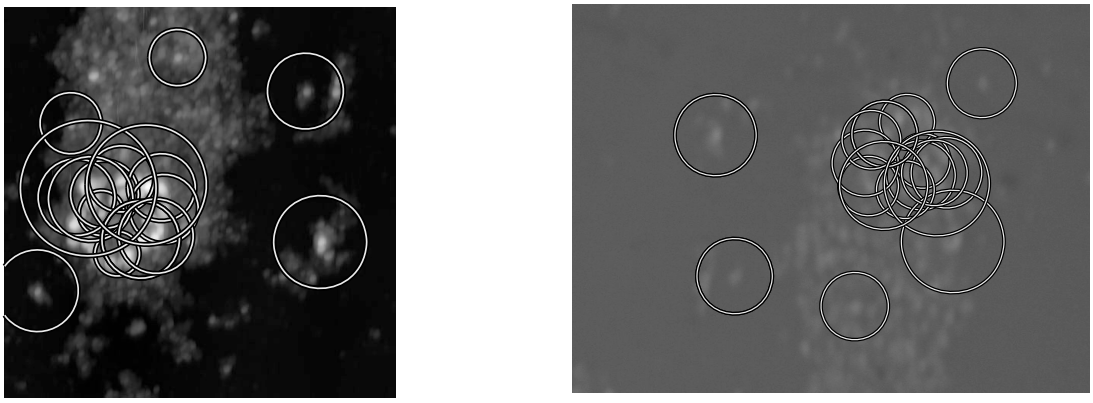


Figure 5.11: Successfully matched corresponding regions in an AFM scan (left) and SEM scan (right) of a nanocluster sample. SIFT detector and descriptor have been used. The circle diameter indicates the scale on which the region has been detected. The field of view width is 10µm for the AFM scan and 13.3µm for the SEM scan.

nearest neighbors there is no guarantee for feature correspondence, especially if ambiguous image structure is present in the scene. A number of successfully matches regions on a nanoscale sample can be seen in Figure 5.11.

5.1.6 Area-based improvement

For a number of reasons, the transformation parameters can still be suboptimal. This can happen, if feature points are rare or cover only a small part of the image scenes. Also, the computed location of distinct image features can be slightly different between the two imaging modalities. For instance, the reproduction of edges and corners can be affected by shadowing, depending on the sensor arrangement. For further improvement of the registration result, area-based fine registration is applied.

As the similarity measure to be optimized, mutual information MI as introduced in Section 3.3.2 has been considered. However, a bias of the MI measure has been reported regarding the proportion of image foreground and background [101]. The authors propose an improved measure called normalized mutual information (NMI), which is used instead of MI. Equation 3.4 can be reformulated by computing the image entropies $H(I_1)$ and $H(I_2)$ and the joint entropy $H(I_1, I_2)$:

$$H\left(I_1\right) = -\sum_a P_{I_1}\left(a\right) \log P_{I_1}\left(a\right), \tag{5.10}$$

$$H\left(I_2\right) = -\sum_b P_{I_2}\left(b\right) \log P_{I_2}\left(b\right), \tag{5.11}$$

$$H\left(I_1, I_2\right) = -\sum_a \sum_b P_{I_1 I_2}\left(a, b\right) \log P_{I_1 I_2}\left(a, b\right). \tag{5.12}$$

MI is then defined by:

$$MI\left(I_1, I_2\right) = H\left(I_1\right) + H\left(I_2\right) - H\left(I_1, I_2\right). \tag{5.13}$$

A drawback of MI is that the absolute entropy values change with the proportion of foreground and background. NMI on the other hand is largely invariant to this effect:

$$NMI\left(I_1, I_2\right) = \frac{H\left(I_1\right) + H\left(I_2\right)}{H\left(I_1, I_2\right)} \tag{5.14}$$

For optimization, the downhill-simplex method is used [64]. It is a direct search method for multidimensional unconstrained minimization of scalar-valued functions. The downhill-simplex method works on function values only and does not require any derivative information. Each processing step of the iterative optimization maintains a simplex, which is a geometric figure in n dimensions, where n is the number of model parameters. The simplex is the convex hull of $n + 1$ vertices. The downhill-simplex method is as follows:

1. Select the starting point as the transformation model parameters obtained by the feature-based registration.
2. Choose $n + 1$ mutated versions of starting point.
3. Compute all function values.
4. Identify best and worst value.
5. Quit if the best value satisfies the termination criterion, else replace worst point according to rules and continue with step 3.

Replacing the worst point first requires reflection of the point at the center of the simplex.

- If this point beats all others then expand simplex and reflect further.

- If this point brings just an improvement accept it and continue.
- If this point brings a degradation shrink simplex and start again.

Typical termination criteria are a targeted function value or a limit in the variation of the function value over the last iterations.

5.1.7 Visualization

Once, the spatial interrelationship between the scans of both imaging modalities is known, a fused image can be computed. Three methods have been discussed in the context of AFM and SEM image fusion [109]: color space fusion, multiresolution fusion and surface rendering. Color space fusion is one of the simplest forms of image fusion, in which the available data sets are copied into the different color channels of a color image. A drawback is that small changes in color are hardly visible. Multiresolution fusion performs fusion in all levels of a scale-space representation and then reconstructs an assembled image out of it. This procedure can cause a loss of image detail but is a good solution if a monochromatic planar fused scan is desired. The preferred method for (pseudo-) three-dimensional imaging modalities is surface rendering.

5.2 Multiple detectors for feature extraction

5.2.1 Motivation for combining detectors

It has been discussed already that the feature detectors differ in their capability of reproducing similar detection results in different imaging modalities. This is due to different imaging characteristics of the single modalities. On the other hand, image contents will also vary between different applications. The feature detectors which have been presented respond to corners, blob-like structures or arbitrarily-shaped regions. Naturally, most real-world scenes will show not only a single category of these features. Also, the distinction between these categories is rather fuzzy. Figure 5.12 shows that additionally to preferring different image structures, the feature detectors also prefer regions of different sizes.

The question arises, which feature detector to choose. Unlike application-specific detectors such as street detectors in aerial photography, the detectors considered here focus on basic structures found in most types of image material. Some requirements such as repeatability and distinctiveness of the detected regions have been mentioned. The performance of the detectors with

respect to these criteria varies with the image contents. For a collection of natural photographs, highest repeatability and robustness has been reported for the Harris-Laplace detector [103]. However, for image contents with mostly blob-like structures, other detectors will be superior.

For a given set of images with known spatial orientation, the best detector can be identified. If the targeted application and image contents can be narrowed down sufficiently, this detector can be chosen to be *the best* and be used exclusively in the future. A similar approach can be applied for determining *the best* feature descriptor. On the other hand, the targeted image contents can be numerous, appear simultaneously in a single image scene and also change over time. A good example of this behavior can be found in microscopy, where different samples and the high range of magnifications produce strongly variant image contents.

Often, the requirement is not to find *the best* detector for a given application scenario but *a good* detector for a large number of known and unknown image contents. In this case, it can be argued why detection should be limited to a single detector. The expression of the detection result in the form of a set of ellipses makes the fusion of detection results very simple. The combined detector response of two detectors D_1 and D_2 is obtained by forming the union set:

$$D_{1,2} = D_1 \bigcup D_2 \tag{5.15}$$

It has to be noted that this procedure is not applicable to descriptors, as similarity cannot be measured between different description schemes. Once, the combined detector response is known it can be treated like a single detector response: a descriptor is computed for each region in the set and used for matching. In the following, strategies for the selection of detector combinations will be presented.

5.2.2 Strategies for selecting detector combinations

Two application scenarios for the selection of a combination of detectors for the automated registration scheme can be distinguished:

- Sample image pairs of a targeted application are given and future images to be registered are assumed to show a degree of similarity in the type of image contents presented.
- No prior knowledge of image contents is available.

In the first scenario, the available sample image data is used to select a combination of detectors, which is optimal with respect to a set of quality criteria.

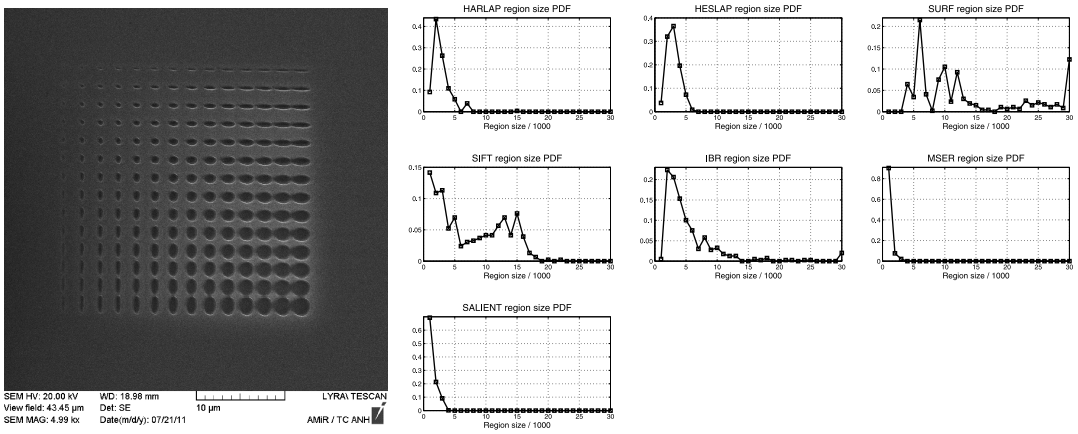


Figure 5.12: SEM scan of FIB-milled structures in elliptical shape (left). The diagrams show the probability density functions (PDFs) of region sizes obtained by the different detectors.

In the second scenario, an optimal combination of detectors is selected automatically for the given pair of images.

Application-specific detector combinations

In the context of feature detection and matching, the ratio between finally resulting correct and total matches is often used as the only performance indicator. This has the benefit of being well defined, easy to compute, and also helps to estimate the chance of success of model-fitting tools such as RANSAC. On the other hand, it does not provide a measure on how many existing correspondences are rejected incorrectly and are finally unavailable for the transformation model parameter estimation. This can be pointed up with a simple example: A detector failing to reproduce responses under minimal variations of the imaging conditions and a strongly transformation-variant descriptor (e.g. binary equality check of the surrounding image patch). This detector/descriptor combination will find a small number of feature correspondences, but probably most of them will be correct matches. Although this method obtains a high correct ratio, it is useless for most applications.

This problem can be overcome by establishing a ground truth feature correspondence for a given pair of images, based on a ground truth spatial transformation. By projecting the interest points of the target image to the coordinate system of the base image, corresponding points can be identified. A simple way of testing for correspondence is to define a fixed-level distance threshold: projected

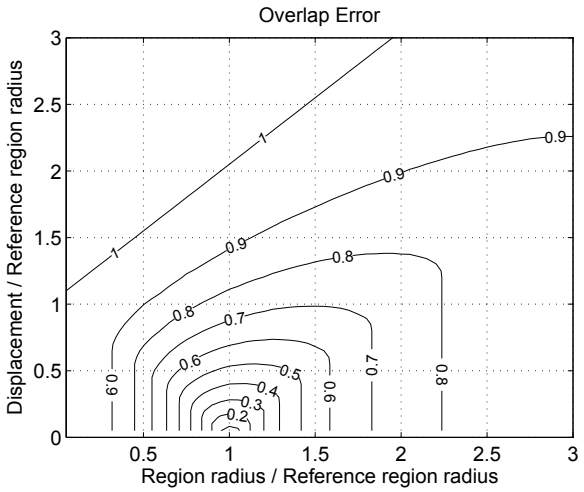


Figure 5.13: Overlap error between two regions mapped onto each other as a function of difference in region size (proportional to scale) and location. For overlap errors $< 50\%$ region correspondence is assumed.

points with a lower distance are defined to be corresponding points. In fact, this method has been used for the performance evaluation of early feature extraction algorithms. However, it is not invariant to scale. A scale-invariant definition of feature correspondence should allow a variable displacement between projected features. Additionally, projected feature pairs with large differences in scale should not be defined as corresponding features, because successful matching cannot be expected in this case.

A method integrating all these needs is based on the idea of not establishing an interest point correspondence but instead a correspondence of regions [76]. The region size of detector window and descriptor window is proportional to the feature scale. Figure 5.11 shows successfully matched corresponding regions in an AFM and SEM scan detected on a nanocluster sample using the SIFT detector and descriptor. The region size has been selected to be the circumcircle of the descriptor window. Region correspondence can be defined with the help of the overlap error ϵ_S for two regions $\mathbf{A}$ and $\mathbf{B}$:

$$\epsilon_S = 1 - \frac{\mathbf{A} \cap T(\mathbf{B})}{\mathbf{A} \cup T(\mathbf{B})}. \tag{5.16}$$

The formula is based on the ratio between intersection and union of the regions and returns an overlap error of zero for identical regions. $T(\mathbf{x})$ is the ground truth

transformation function. According to [76], region correspondence is assumed for overlap errors $< 50\%$. In this case, strong descriptors are still able to detect feature correspondence. The overlap error ϵ_S between two regions mapped onto each other as a function of difference in region size (proportional to scale) and location can be seen in Figure 5.13.

A measure of detector performance is obtained by computing the repeatability score for an image pair. It is the ratio between the number of feature correspondences and the smaller number of features detected in one of the scans. Only the image areas present in both scans are regarded here. A good detector will reproduce responses in both modalities and therefore yield high repeatability scores. Although the repeatability score provides a realistic view on the expected ability to reproduce results in a different modality, for detector comparison it has a drawback. Using the above definition of ϵ_S, detectors returning large regions are privileged. To avoid this behavior, the detected regions can be rescaled so that each base region is normalized to the same size and the size ratio between corresponding regions remains untouched.

The repeatability criterion allows ranking of a set of detectors by their mean repeatability for a set of sample images. A simple strategy for selecting a suitable combination of feature detectors is to combine the two or more detectors which obtained the highest ranking position. However, this procedure does not guarantee complementary behavior of the set of detectors. For instance, a slightly modified variant of the detector with the highest rank will be ranked on position two. The union set of both detector responses will contain a lot of double entries. This detector combination does not bring a benefit over the single detectors, neither in terms of repeatability nor in the ability to generalize and being applicable to a broader range of image contents.

The problem of detector similarity can be overcome by measuring detector similarity using the repeatability criterion. This idea is depicted in Figure 5.14. The illustration is limited to two detectors but the principle can be extended to an arbitrary number of detectors. The detector response of detector D_1 applied on an image scene acquired by modality M_1 is denoted M_1D_1. The ground truth transformation functions are used to compute inter-modal repeatabilities. Inter-modal repeatability is depicted in horizontal direction and high values of $Rep(M_1D_1, M_2D_1)$ and $Rep(M_1D_2, M_2D_2)$ are a desirable property. The inter-detector repeatabilities $Rep(M_1D_1, M_1D_2)$ and $Rep(M_2D_1, M_2D_2)$ measure similarity between the detectors and therefore small values are favorable if a combination of D_1 and D_2 is considered.

Unfortunately, repeatability is a necessary but not a sufficient requirement to a detector or combination of detectors. A reason is the bias towards large numbers of features. A good detector obtains a high repeatability, regardless of

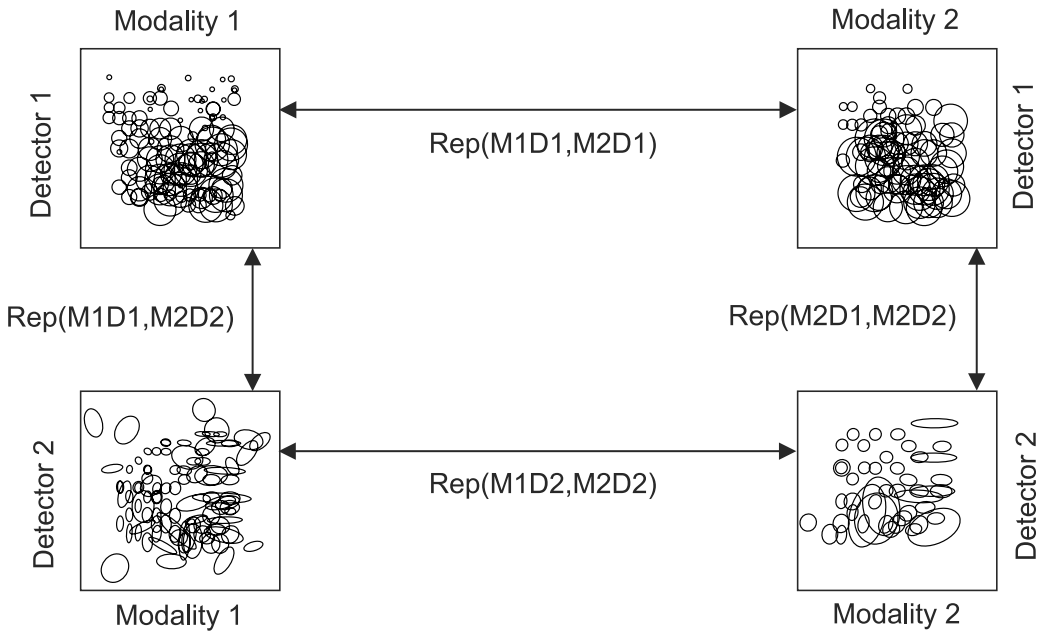


Figure 5.14: Repeatability between modalities and detectors. Repeatability of a feature detector applied using multiple imaging modalities is a requirement for the success of the feature-based registration stage. Repeatability can also be computed between the responses of two different detectors, applied to a single image. The result is a similarity measure between detectors which can be used to select complementary rather than similar detectors.

the number of features. For high numbers of features, the detector responses of a bad detector are likely to repeat just because of the limited image area. This effect is demonstrated in Figure 5.15. The SIFT detector shows a high repeatability which remains almost constant when the detection threshold is lowered and more features are detected. A detector which is based on randomly selecting points and scales starts with a very low repeatability. As the number of randomly selected regions is increased, the repeatability also increases. It has to be noted that feature matching becomes more difficult and also time-consuming for higher numbers of features. Including the number of features in a quality measure is misleading, as the number is highly variable with image contents. Practically, the number of features should be in the range of some tens to some thousands of features per megapixel and all detectors which do not meet this requirement can be excluded.

A second reason why repeatability is a necessary but not a sufficient require-

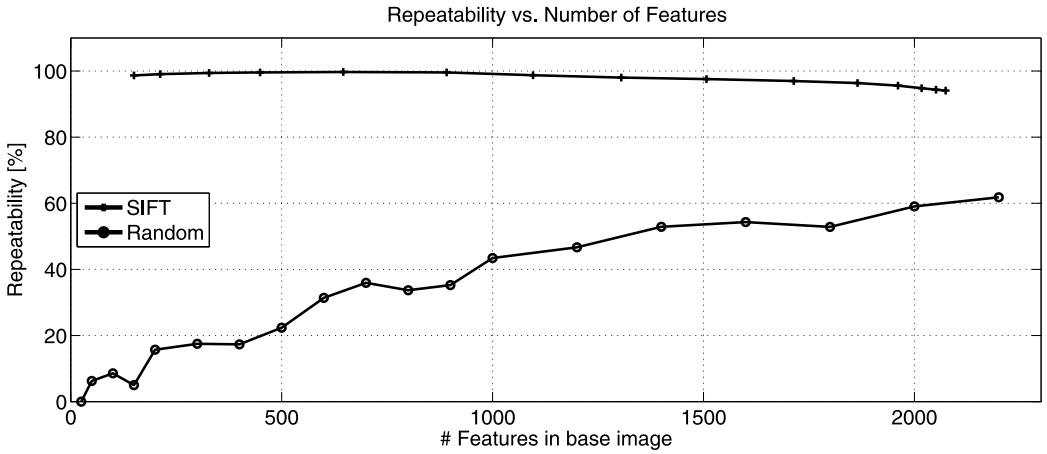


Figure 5.15: Bias of the repeatability with respect to the number of features. A good detector shows a high repeatability for any number of features. As the number increases, all detectors will obtain higher repeatability values - even a random detector. The values have been computed on the image scene shown in Figure 5.11

ment is the problem of informativeness of the detected regions. The image contents shown in the detected region must be sufficient in order to allow feature matching based solely on the descriptor extracted from it. If the region does not contain enough salient contents which are reproduced in both imaging modalities, matching becomes infeasible. Practically, it is difficult to measure the informativeness of a detected region, because informative image structure cannot be distinguished from modality-specific imaging artifacts. Therefore, the best way of assuring feature informativeness is to perform feature matching experiments. This requires a proven and tested feature descriptor to be specified already. A simple performance measure is the *matching score* [76], which is the ratio between the number of correct matches and the smaller number of features detected in one of the images. Detectors with low informativeness obtain low matching scores.

Although the repeatability criterion is biased in multiple ways it can still be considered as the best criterion available, if a number of necessary precautions are taken. The strategy for the selection of a combination of detectors is summarized in the following. It is assumed that a number of image pairs with known ground truth transformation are available. The procedure is described for combining two detectors but can be extended to higher numbers.

1. Compute all detector responses and descriptors on all image material available.
2. Remove all detector responses which do not meet the required minimal number of features.
3. Compute all inter-modal repeatabilities $Rep(M_1D_i, M_2D_i)$ for all image pairs and rank detectors in decreasing order of the mean value.
4. Select a subset of detectors which obtain the highest ranks.
5. Form all 2-combinations of the detectors in the subset.
6. Compute all inter-detector repeatabilities $Rep(M_1D_l, M_1D_m)$ and $Rep(M_2D_l, M_2D_m)$ and rank detector combination in increasing order of the mean value.
7. Select a subset of detector combinations which obtain the highest ranks
8. Compute the matching scores for all detector combinations in the subset.
9. Select the detector combination with the highest matching score.

The resulting combination of detectors produces a high number of finally resulting correct matches in the sample image material. At the same time, a too high degree of specialization on the image contents included in the training material is avoided by assuring dissimilarity of the detectors. It has to be noted that in general it cannot be expected that combining detectors will lead to a performance gain in terms of mean repeatability. Rather, the combined detector becomes applicable to a broader range of image contents. The performance gain is in terms of minimal repeatability.

Automatic profile selection

The method for selecting a combination of detectors based on inter-modal and inter-detector repeatability requires an amount of sample image material. In contrast, there are application scenarios where the type of image contents cannot be narrowed down sufficiently or where no sample image material is available in advance. Also, the selection procedure and especially the computation of the repeatability is a time-consuming task. However, the proposed registration scheme can also be applied to a new application without prior search for on optimal combination of detectors. It requires pre-computed ranking results from a generic image data set. If no multimodal data set is available, the ranking

results reported in Mikolajczyk et al. [76] can be re-computed for combinations of detectors. The resulting combinations can be regarded as profiles, which are chosen in dependency of the actual image contents. In the application case, the procedure is as follows:

1. Specify minimum number of consistent matches n_{min} and maximal number of iterations i_{max}.
2. Start with detector combination of highest rank.
3. Compute detectors and descriptors, perform matching.
4. Apply RANSAC to compute cardinality of consensus set.
5. Terminate if number of consistent matches exceeds n_{min} or i_{max}. Else repeat with next lower-ranked combination of detectors.
6. Proceed with area-based registration step.

This run-time selection procedure requires the choice of a descriptor in advance. However, the loop (Steps 2 to 5) can also be extended to optimize over several descriptors. On the other hand, the execution time will suffer severely.

5.3 New feature matching strategies

5.3.1 Integration of prior knowledge

In many application scenarios, the transformation model parameters do not need to be optimized in an infinite search space. Often, some prior knowledge is available about the range of each model parameter. For instance, the initialization vector an area-based registration scheme is started with implies, that the final registration result will be found in a limited neighborhood of the initialization vector. In other applications, the difference in scale is known to a certain accuracy. The integration of this knowledge into the iterative optimization procedure in area-based registration is a form of regularization. It means that the solution is driven into a plausible direction. The cost function to be optimized can be extended by a regularization term:

$$C_R(\mathbf{x}) = C(\mathbf{x}) + R(\mathbf{x}) \tag{5.17}$$

A possible choice of $R(\mathbf{x})$ is an inverted Gaussian kernel centered at the initialization vector and giving penalty to large displacements. In other applications,

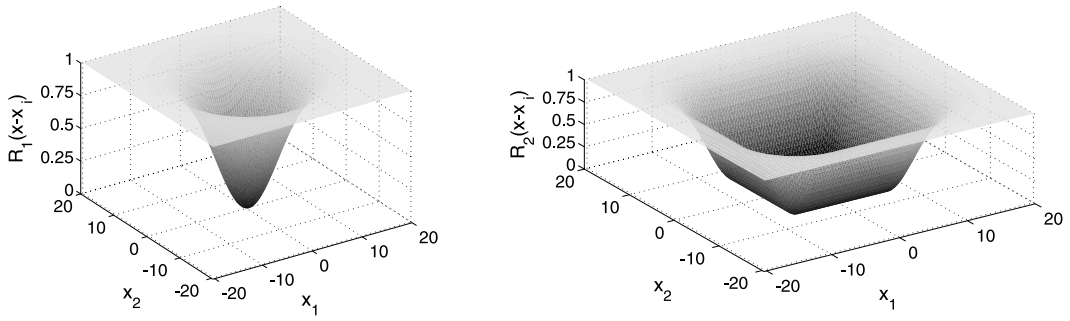


Figure 5.16: Two penalty functions for steering the iterative optimization procedure in area-based registration towards a plausible direction. $R_1(\mathbf{x})$ is an inverted Gaussian kernel centered at the initialization vector $\mathbf{x_i}$. Depending on the kernel parameters, $R_1(\mathbf{x})$ penalizes large displacements from the initial transformation parameters. $R_2(\mathbf{x})$ is an expanded variant with a constant zero-level region centered around $\mathbf{x_i}$. This region is the plausible subspace in which the solution is expected. Leaving this region is penalized. For simplicity, only two model parameters are displayed, although the principle can easily be transferred to higher-dimensional space.

the search space is limited by a soft border, driving the solution back towards a sound subspace. These approaches are depicted in Figure 5.16. In most applications, extreme differences in scale are impossible due to the digital nature of the image contents: extremely subsampled structures do not show enough image detail for successful registration. Also, the translation vectors are limited by the image geometry.

The proposed registration scheme starts with a feature-based registration step. This prohibits the direct use of a regularization term as described in Equation 5.17. Nevertheless, prior information about the transformation parameters can be integrated into the feature-based registration step.

5.3.2 Geometric consistency

The proposed registration scheme establishes feature correspondence based solely on the descriptor vector of each local feature. This creates a conflict between the requirement of the detector to be robust under the aspect of multimodality but at the same time highly distinctive and preventing accidental feature matches. As a result, the set of feature correspondences obtained after the initial feature

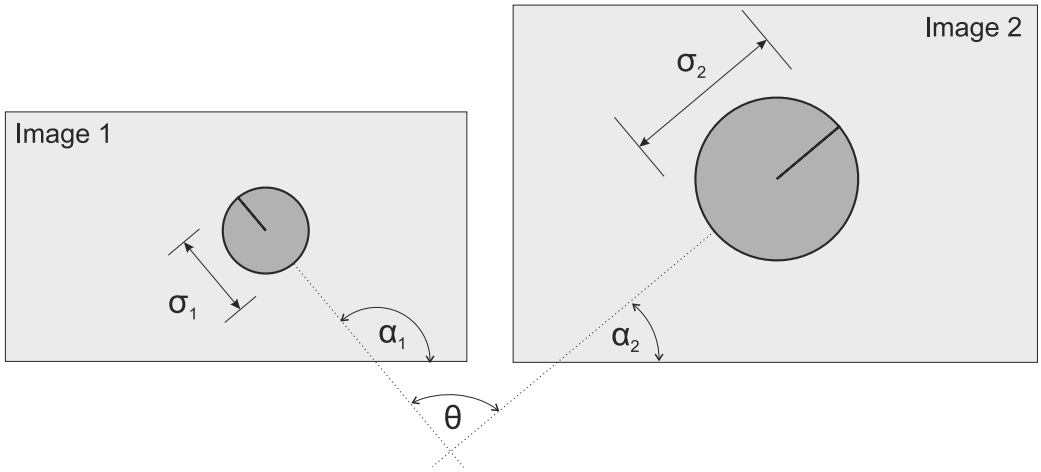


Figure 5.17: Scale and Orientation of a local feature. Two regions have been detected in different images. Region scale σ and orientation α with respect to the image grid are different. For a geometrically consistent set of feature correspondences, the scale ratios and differences in orientation can be expected to cluster around the true transformation parameters.

matching will contain a non-negligible percentage of incorrect entries. With the help of model-fitting tools such as RANSAC, a geometrically consistent subset of this correspondence of features can be determined. However, for very high percentages of outliers, RANSAC and derivatives of this algorithm are not likely to succeed. The percentage of outliers for which failure must be expected depends on the algorithm parameters and the level of measurement noise. For the targeted application it can be assumed to be $> 80\%$.

By incorporating prior knowledge about the transformation model parameters, the probability of success for the model fitting algorithm can be increased. The idea exploited in the following requires a transformation model which can be globally approximated by a combination of rotation, scaling and translation. It is applied as a postmatching step after the initial feature matching but before determination of the largest geometrically consistent subset. By reducing the percentage of outliers, the success rate of the complete registration scheme can be increased. The proposed postmatching step implies a limitation of the search space and under this aspect is comparable to the penalty functions $R(\mathbf{x})$. In contrast, it uses a hard- rather than a soft border.

In the application case, the detected scale σ and orientation α with respect to the originating image grid are known for each detected feature. For each pair

of matched features, these properties can be compared. Figure 5.17 depicts two detected regions and a geometric interpretation of scale and orientation. With the detected scales $\sigma_{1,2}$ and orientations $\alpha_{1,2}$, two quantities can be derived:

$$s = \frac{\sigma_1}{\sigma_2}, \tag{5.18}$$

$$\theta = (\alpha_1 - \alpha_2) \mod 2\pi \tag{5.19}$$

A perfect combination of feature detector and descriptor would produce perfectly localized and oriented regions and no incorrect feature correspondences. With such a combination of detector and descriptor, a single pair of corresponding features would be sufficient for image registration. If image 1 is defined to be the base image and image 2 is the reference image, then s is the scale factor and θ the rotation required in order to perform registration. However, in real applications all measurements of scale and orientation will be inaccurate. Due to the locality of the detectors, this cannot be avoided. The geometric consistency in single pairs of features is also referred to as weak geometric consistency. It has been exploited in the context of image retrieval from databases for re-ranking of query results [54].

Without the postmatching step, scale and orientation of a feature are only determined in order to compute the invariant feature descriptor. The geometric consistency is assured by model-fitting tools such as RANSAC, based purely on the feature locations. Nevertheless, feature scale and orientation contain valuable information. This will be demonstrated with the help of an example. Figure 5.18 shows a multimodal pair of images which are rich of distinctive structures. Feature detection and descriptor computation both have been carried out using the SIFT algorithm. The initial set of feature matches contained 1422 correspondences of which only 117 are correct. Figure 5.19 shows histograms of the computed values of s and θ. The histograms show clusters around the true transformation parameters which are a rescaling factor of $s = 1.1451$ and a rotation by $\theta = 88.0442°$. On the other hand, many correspondences deviate strongly from these values. Especially the histogram of θ shows a second cluster in reverse direction. This is mostly caused by ambiguity of the dominant feature orientation in symmetric image structures. For this reason, a previously reported exploitation of differences in scale for SIFT feature matching in video scenes cannot be applied [3]. It relies on the presence of unique peaks in the matching histograms and resamples the target image accordingly.

If a range of plausible values for s or θ can be specified, correspondences falling out of this range can be eliminated from the list. A way of specifying this range is to start from an estimate $\hat{s}$ or $\hat{\theta}$ of the transformation parameters and to define

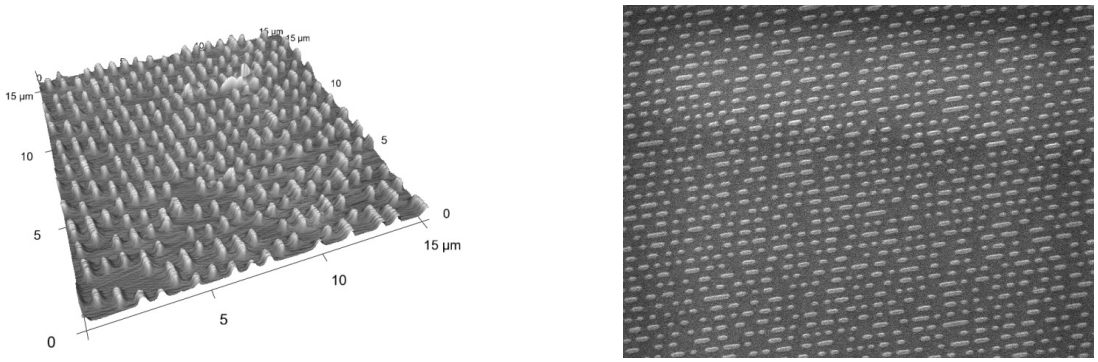


Figure 5.18: Surface of a DVD imaged by AFM (left) and SEM (right). The AFM scan is shown as a rendered surface.

a margin around these estimates. For example a window with a width of $\pi/8$ centered around $\hat{\theta}$ can be chosen. The window for the scaling factor is best chosen relative to the amount of $\hat{s}$, e.g. $\hat{s} \pm 25\%$. The window width chosen will also depend on the expected accuracy of the estimates $\hat{s}$ and $\hat{\theta}$. If the estimates are believed to be inaccurate a large window width may be selected. According to the low level of prior knowledge included in this case, the matching procedure will benefit only moderately from this step. On the other hand, if scaling and rotation are known precisely from the imaging setup and the task of registration reduces to the problem of translation, narrow windows can be selected.

It has to be noted, that the proposed postmatching procedure is not capable of increasing the total number of correct feature correspondences, because no new correspondences are established. The goal of the procedure is the removal of incorrect correspondences. This brings the risk of also removing true feature correspondences, mainly due to two reasons. On the one hand, the detection of scale and orientation is inaccurate but there are cases where features can be matched successfully, even with scale and orientation detected inaccurately. An example are isolated corner-like structures appearing almost identical over scale. The second reason is the inaccuracy of the estimates $\hat{s}$ or $\hat{\theta}$ in combination with a narrow window width in the postmatching step.

The postmatching procedure has been applied to the initial 1422 feature correspondences extracted from the image pair shown in Figure 5.18 with accurate estimates $\hat{s}$ or $\hat{\theta}$ and the window parameters chosen as mentioned above. If the postmatching steps for scale and rotation are carried out separately, 1114 correspondences are inliers with respect to scale and 431 are inliers with respect to rotation. Executing both steps subsequently, leaves 395 inliers. These include 109 of the 117 correct matches or in other words, 8 true correspondences have

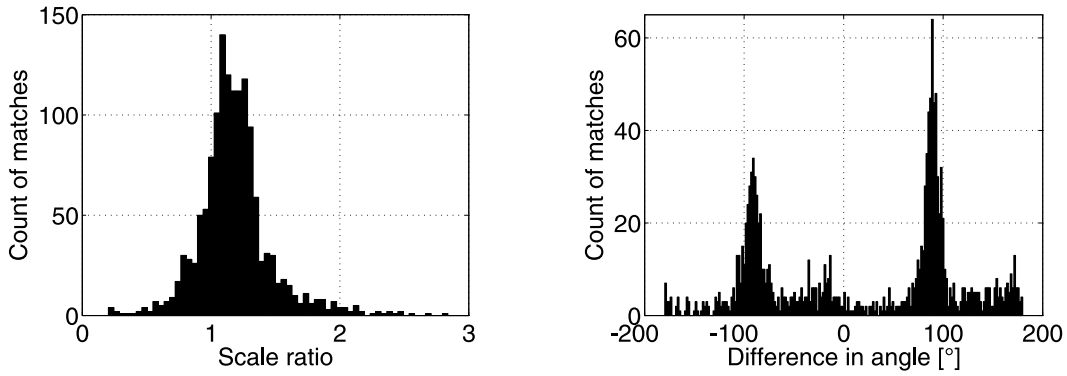


Figure 5.19: Histogram of scale ratios and differences in orientation for matched features of the image pair shown in Figure 5.18. The SIFT algorithm has been used for both, feature detection and description. With the AFM scan as base image and the SEM scan selected as reference image, the histograms show clusters around the true transformation parameters. These are a rescaling factor of $s = 1.1451$ and a rotation by $\theta = 88.0442°$.

been removed incorrectly. On the other hand, the correct ratio in the correspondence set has been improved tremendously, which was the intended effect.

6 Experimental validation of the system for object classification

The proposed system for micro- and nanoscale object classification (Chapter 4) has been validated in three application scenarios. These differ in the type and scale of objects and also in the imaging modalities used. The applications originate from three EU-funded projects. Classification of nanoscale objects in SEM scans has been performed in the context of the project *Micro-nano system for automatic handling of nano-objects* (NANOHAND). In the context of *Hybrid ultra-precision manufacturing process based on positional- and self-assembly for complex micro-products* (HYDROMEL), quality monitoring and defect detection of Xenopus Laevis oocytes using an optical microscope has been implemented. Localization of magnetic particles using the MRI was part of the project *Nano-actuators and nano-sensors for medical applications* (NANOMA). This chapter describes the characteristics of all three tasks and shows how the proposed system can be applied successfully.

6.1 Recognition of carbon nanotubes (SEM)

For automatic manipulation, testing or assembly of carbon nanotubes (CNTs), a nanorobotic system needs to be aware of the locations of individual CNTs. Due to imperfect cleanroom conditions and also complex assembly setups, CNTs will be not the only type of structures found on a substrate. For this reason, the proposed system has been applied in order to identify isolated CNTs, which can be handled automatically in subsequent processing steps. The results have been published before [27, 115].

6.1.1 Automated handling of carbon nanotubes

CNTs are one of the most promising materials in nanotechnologic applications. Their most interesting nanoelectric properties are the ballistic (scattering-free) and spin-conserving transport of electrons and their ability to show metallic as well as semiconductive behavior. Also, CNTs can handle a current density

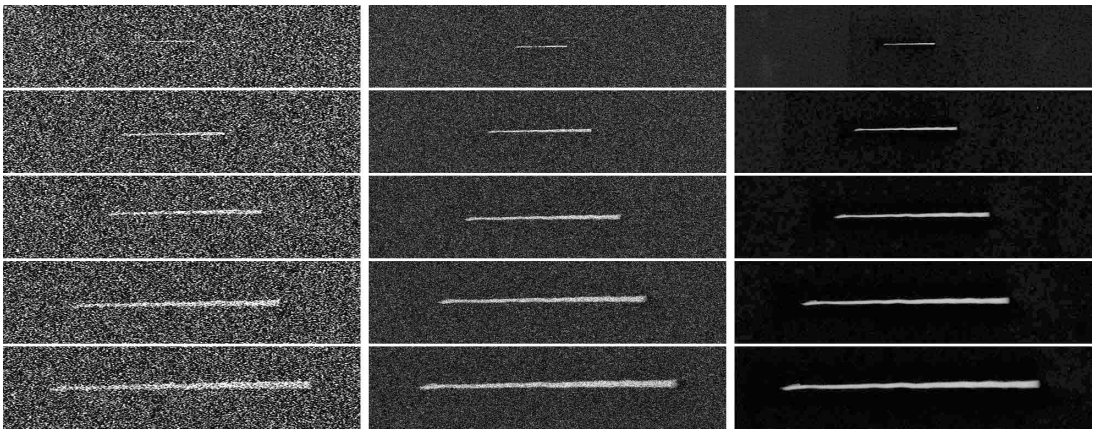


Figure 6.1: Sequence of SEM images of a multi-walled CNT with increasing level of magnification ranging from 800× (top) up to 4000× (bottom). The full-frame image acquisition time is 90ms/640ms/40.3s (left to right). The length of the CNT is 12.75µm.

1000 times higher than copper. This makes CNTs the perfect candidate for novel interconnects in the fabrication of integrated circuits [80]. Furthermore, nanoelectric components such as the CNT field effect transistor have been built and show superior characteristics compared to silicon-based transistors [2, 50]. Besides their nanoelectric properties, CNTs are also an auspicious material in nanomechanic applications.

Before these properties can be exploited in the large-scale production of nanodevices, reliable methods for automated handling and assembly of CNTs are needed. These differ significantly from the techniques used in conventional robotics, because the behavior of nanoscale objects such as CNTs is considerably less predictable than that of macroscale objects. Visual feedback from SEM scans turned out to be the most important type of sensory feedback for this task.

Today, CNTs are commercially available and three different production methods are well-established: arc discharge, laser ablation and chemical vapor deposition (CVD). In the following, the focus is on multi-walled CNTs grown by plasma enhanced CVD with a typical length of 10 - 20 µm and diameter of 150 - 350 nm. Nevertheless, the results can be easily transferred to other types of CNTs. Electrothermal microgrippers have shown to be a useful tool for mechanical handling of individual CNTs. Latest approaches demonstrated pick-and-place handling of single CNTs following different goals including their electrical and mechanical characterization and the assembly of sensory devices [89].

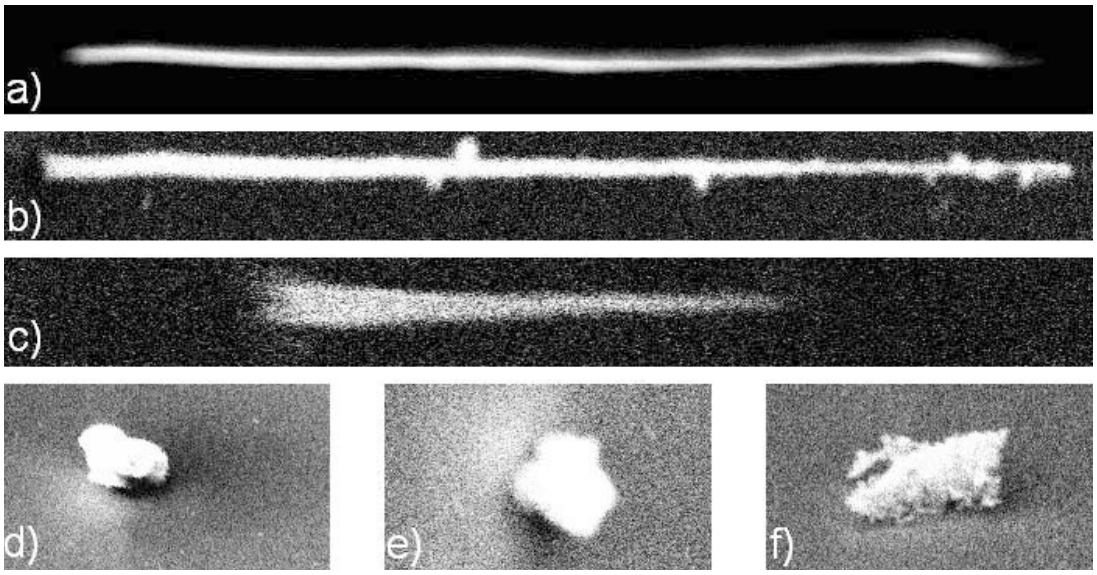


Figure 6.2: CNT rising out of the focus plane (a), material attached to CNT (b), conical shaped CNT (c), debris (d-f)

The main portion of noise in SEM images is caused by the secondary electron detector and is usually reduced using temporal averaging. This is why real-time processing of SEM images is always a tradeoff between image acquisition time and image quality. Figure 6.1 depicts this relationship showing multiple SEM images of an isolated CNT grown by CVD onto a silicon wafer. An acceptable acquisition time for full-frame CNT search is <1s. Another source of image degradation is grey level fluctuation which arises due to electrostatic charge and due to changes in the alignment of target, electron beam and detector.

There are multiple requirements to a CNT detection algorithm for the search of CNTs that are located at the surface of a silicon wafer. The algorithm must be fast enough for automation procedures and therefore handle a high level of image noise. It must detect CNTs at a wide range of different magnifications, different lengths and in any orientation. Depending on the fabrication process, multi-walled CNTs may be conically shaped. Evaporation of the electrothermal gripper may also lead to the deposition of material on the CNT surface. Although the SEM is considered to be an image sensor with high depth of field, a CNT may partly rise out of the focus plane. The algorithm must provide orientation and endpoints of CNTs and finally reliably reject any non-CNT particles that cannot be avoided if working under low cleanroom standards. Some of these requirements are illustrated in Figure 6.2.

Combined with a scanning procedure, the proposed system is able to provide a complete map of the wafer surface including CNTs and non-CNT objects. This information is intended to be used for the initialization of tracking algorithms that are needed for automated CNT handling. Additionally, it may be used as an error recovery sequence.

6.1.2 Application of the new system

Depending on the imaging conditions, CNT fabrication process and manipulation history, only few assumption can be made regarding the appearance of CNTs in SEM images. One assumption that holds in many cases is the idea that the outline of a CNT can be approximated by a straight line. Since other objects or structures found on the substrate tend to arrange and shape in a chaotic or spherical way, this property will be exploited for CNT detection. Three commonly used methods for straight line detection were considered.

The active contour algorithm can be reformulated for straight line detection [87]. It uses an inner contour energy depending on line straightness and an outer contour energy derived from the image gradient perpendicular to the contour. This method is not suitable for the targeted problem since the image gradient is highly sensitive to noise and because thick lines result in double hits. Another algorithm for straight line detection is the Hough transform. Each point in Hough space corresponds to a line object in the input image, defined by angle and offset. Peaks in Hough space are used to identify dominant line objects. A disadvantage of the Hough transform is its high computational cost.

The principle component analysis (PCA) is a tool well-known in multivariate statistics. In the context of pattern recognition it is mainly used for dimension reduction of the input feature space. However, applied to the geometrical distribution of an object, it may also be utilized for straight line detection [67]. An ideal straight line in two-dimensional space has a principle component, which can be calculated from the eigenvectors and eigenvalues of its scatter matrix. This leads to the elongation measure PC_E, which has been introduced in Equation 4.5. In the following, the components selected for the application of the proposed system will be explained in detail.

Noise reduction SEM image noise is assumed to be Poisson-distributed due to the small number of secondary electrons measured by the detector. The median filter is expected to provide good results in this case because it is capable of edge preservation and shot noise removal. In this application, median filtering is critical because it removes thin lines from the image, e.g. CNTs that are the scope of this experiment. Therefore, Gaussian low-pass filtering is used in

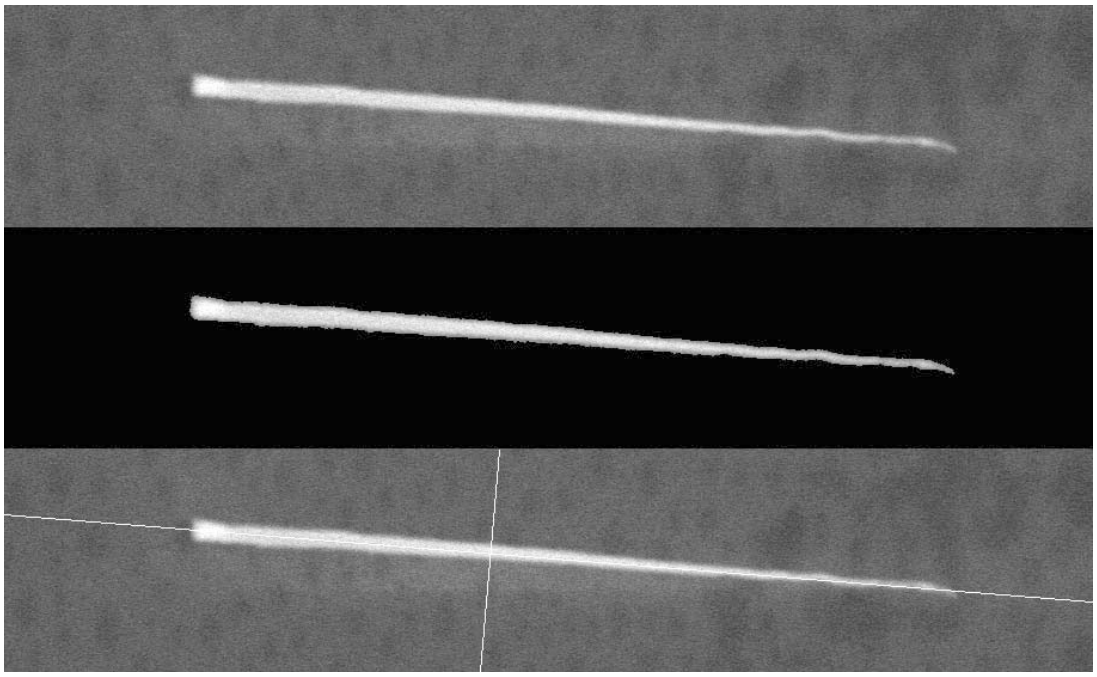


Figure 6.3: Masked CNT image after segmentation (middle), detected principal and secondary component of CNT's geometrical distribution (lower). The original scan can be seen in the upper image.

addition to temporal averaging. Gaussian filtering also removes high-frequency image detail, which is irrelevant to the problem of CNT detection.

Image segmentation It has to be noted that the CNT detection presented here does not rely on a perfect object segmentation from the background. Locally adaptive thresholding techniques turned out to be sensitive to slight variations in substrate brightness. Consequently, a global thresholding strategy is chosen. By experiment, three methods which derive a threshold value from the image grey level histogram [95] have been compared: Gaussian mixture modeling (GMM), within-class variance minimization and class-entropy maximization. All methods successfully separate CNT and debris objects from the substrate at a wide range of brightness. However, in image scenes showing only the substrate, the GMM and variance-based approaches fail to grade the whole image as background. For this reason, class-entropy maximization has been selected for image segmentation.

Individual objects are identified in the binary segmented image as 8-connected pixel-clouds. Object holes are filled afterwards. As this problem has a unique solution and is solved by any naive approach in acceptable time the actual implementation is irrelevant. An exterior contour retrieval algorithm with additional region filling is used. For m objects, m object point sets $\mathbf{X}^1 \dots \mathbf{X}^m$ are received.

Feature extraction PCA is applied to every point set $\mathbf{X}^1 \dots \mathbf{X}^m$ individually and the principle component energy scores $\mathrm{PC}_\mathrm{E}^1 \dots \mathrm{PC}_\mathrm{E}^m$ are obtained. A score of $\mathrm{PC}_\mathrm{E} = 1$ corresponds to an ideal straight line whereas if the score is $\mathrm{PC}_\mathrm{E} = 0$, no dominant direction is found. This would be the case if the object is a circular disc. Single point objects are excluded because the corresponding scatter matrix $\mathbf{S_c}$ is singular. The eigenvectors $\mathbf{v}_{1,2}$ of each object form a rotation matrix, which is used to transform each set of object points $\mathbf{X}^1 \dots \mathbf{X}^m$ into its own PCA space. The transformed coordinates are denoted $\hat{\mathbf{X}}^1 \dots \hat{\mathbf{X}}^m$.

Endpoint retrieval From each set of transformed object points $\hat{\mathbf{X}}$, two peak positions are calculated:

$$\hat{\mathbf{p}}_1 = \left[\min_{i \in 1 \dots n} \left\{ \hat{\mathbf{x}}_i \cdot \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right\}, 0 \right]^T \tag{6.1}$$

$$\hat{\mathbf{p}}_2 = \left[\max_{i \in 1 \dots n} \left\{ \hat{\mathbf{x}}_i \cdot \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right\}, 0 \right]^T . \tag{6.2}$$

By transforming $\hat{\mathbf{p}}_{1,2}$ back to image coordinates, an estimate of the CNT endpoints $\mathbf{p}_{1,2}$ is obtained. Figure 6.3 depicts the working principle of PCA applied to the geometrical distribution of a CNT image.

Object classification The principle component energy based score PC_E is a meaningful indicator for CNT objects. However, if an object is formed by only a few points the result becomes arbitrary. For this reason, the object size also needs to be taken into account. A rough measure of object size is computed by using the relative projection area n, an appropriate scaling factor k_{P_A} and magnification scale M:

$$\mathrm{P}_A = k_{P_A} \cdot \sqrt{n}/M. \tag{6.3}$$

The decision boundary is learned automatically by training a SVM based on PC_E and P_A.

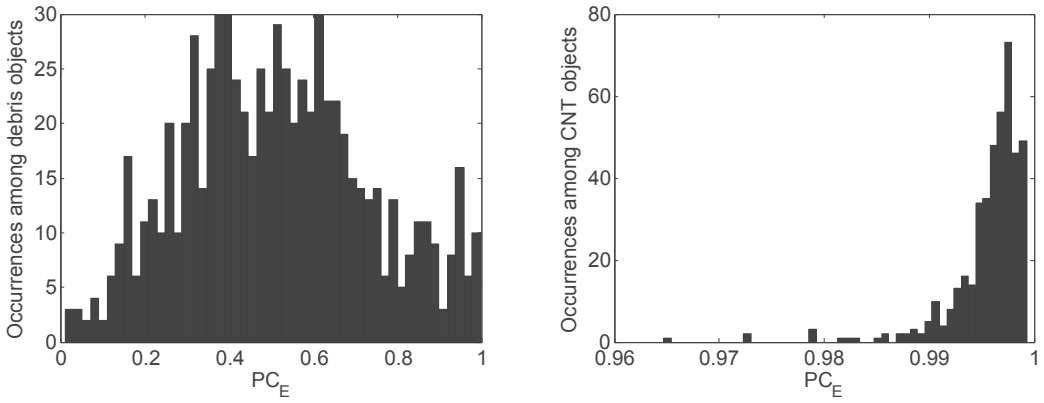


Figure 6.4: Statistics of principle component energy score $\mathrm{PC_E}$ for debris and CNT objects. Values of CNT objects cluster close to $\mathrm{PC_E} = 1$ while debris objects fill the whole range.

6.1.3 Results of carbon nanotube detection

The majority of experiments have been carried out using a LEO 1450 by Carl Zeiss. For comparison, additional tests have been made using a Quanta 600 by FEI. Both SEMs are equipped with additional image acquisition hardware by Point Electronic. The silicon wafer is movable by piezoelectric actuators. Initially, the influence of imaging conditions on the outcome of the segmentation procedure has been investigated. Both SEMs come with an automated brightness and contrast adjustment which aims at best usage of the dynamic range. The effect of variations in brightness has been studied. The segmentation procedure was found largely insensitive to variations in brightness, as long as the object contrast remains sufficient.

In Chapter 4, the importance of feature invariance to rotation, scaling and translation has been pointed out. These properties have been evaluated using a representative CNT object. For this purpose a sequence of 60 images was taken. The orientation was changed using electron beam rotation. Magnification was varied from 800× to 4000×. As expected, the features are almost invariant to translation and scale within the limits of numerical calculations and noise ($\sigma^2\left(\mathrm{PC_E}\right) = 4.79 \cdot 10^{-8}$, $\sigma^2\left(\mathrm{P_A}\right) = 6.12 \cdot 10^{-4}$). Changing the object orientation relative to the scanning direction effects contrast and shadows introducing a higher level of variation which is still uncritical for object classification ($\sigma^2\left(\mathrm{PC_E}\right) = 5.61 \cdot 10^{-7}$, $\sigma^2\left(\mathrm{P_A}\right) = 6.70 \cdot 10^{-3}$).

A collection of 300 image scenes was used to generate statistics of possible

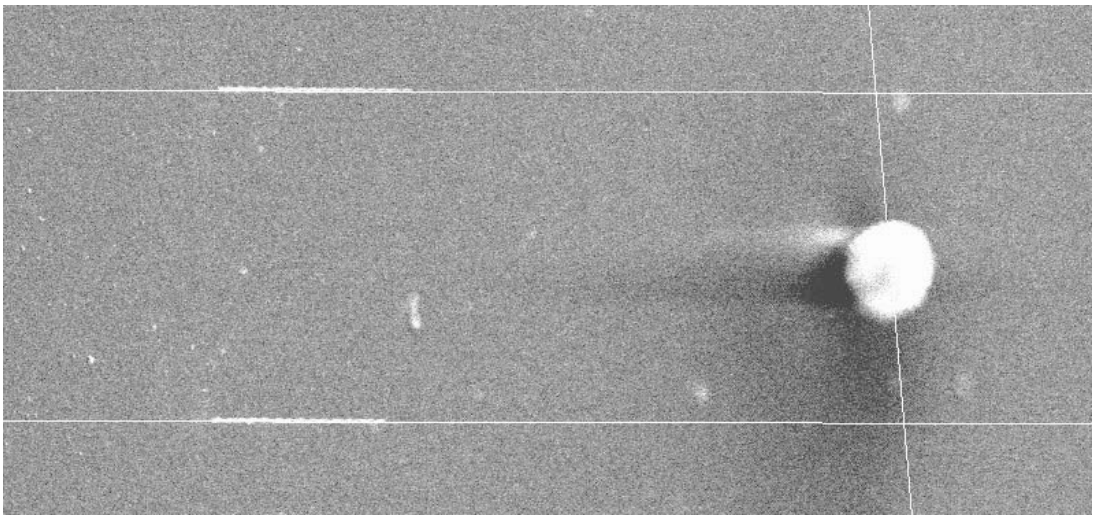


Figure 6.5: Image scene assessed by the proposed algorithm. Principle component directions are indicated. The two CNTs obtain a PC_E of 0.9988 and 0.9987, whereas the debris object obtains a PC_E of 0.1457.

object shapes. Some of them can be seen from Figures 6.5 and 6.6. 1177 objects have been identified manually, which included 432 CNTs. Small parts broken from CNTs were marked as debris, because they are useless for further processing. Figure 6.4 shows the histogram of PC_E for CNT and debris objects. All CNTs obtain high values of PC_E. This also includes conically shaped CNTs and those with material attached to them. Small debris objects come in arbitrary shape, which occasionally results in high values of PC_E.

The algorithm has been fully integrated into a pre-existing distributed control architecture for automated nanohandling of CNTs. All actuators as well as the SEM control are available via a common interface. This allows testing the detection algorithm in its dedicated application environment without any manual interaction. An automation script has been written which scans the wafer surface through the SEMs field of view. Only if CNTs are detected, the object endpoint coordinates are transmitted to the controller. The controller uses its knowledge of the wafer position to translate the image coordinates to world coordinates.

For testing, the wafer used in training has been exchanged in order to provide a new set of objects. The scan was performed along multiple straight lines which are 500 µm in length. Throughout multiple test sequences, no CNT was detected incorrectly (false positive). However, a few CNTs were ignored which were attached to or occluded by large debris objects. Also, the algorithm is not

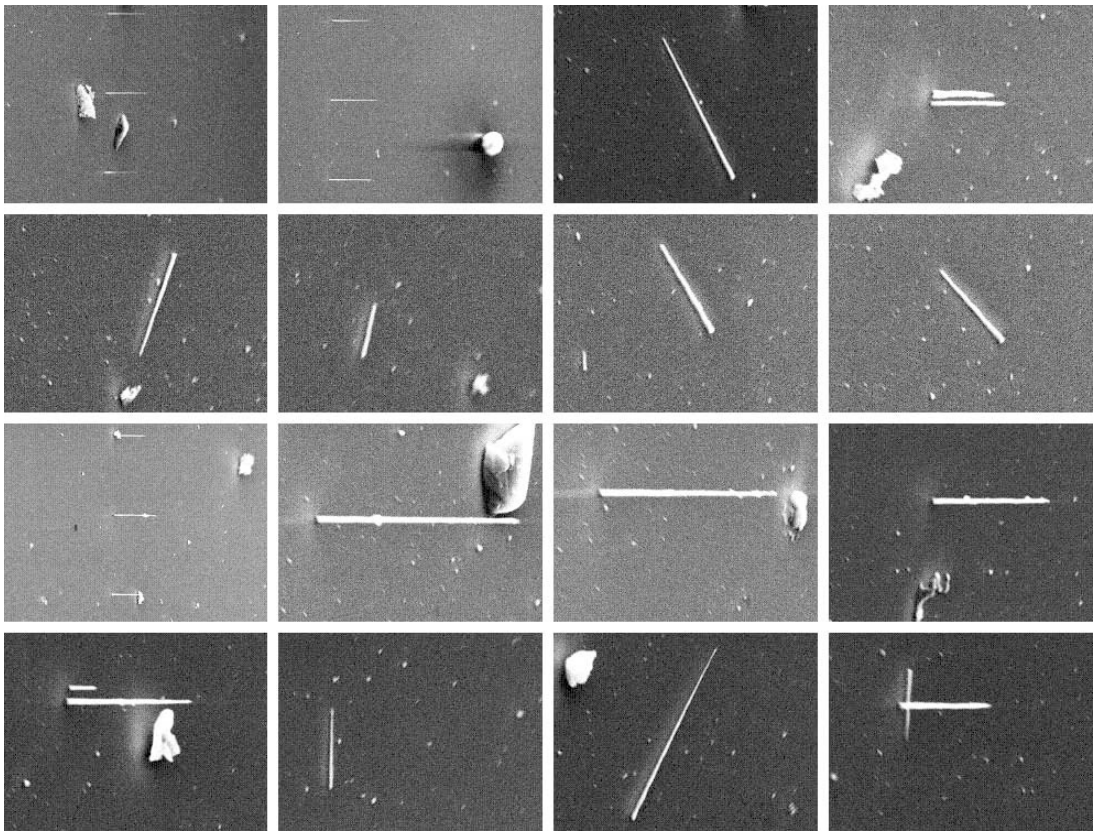


Figure 6.6: Selection of image scenes collected during the experiments at magnifications ranging from 800× to 4000×. A total number of 1177 objects were identified in the complete set and used for training: 432 CNTs and 745 debris objects.

able to separate CNTs crossing each other (Figure 6.6, lower right scene). This is not contrary to the aim of this procedure which is the detection of isolated CNTs on the substrate that can be revisited for automated nanohandling.

The accuracy of the CNT endpoint coordinates could not be evaluated due to the lack of a measurement method for comparison. By manually inspecting the collected data, most endpoints are found at the correct position. Some misplacements occur if the CNT tip is out of focus. In that case, the detected tip position is shifted towards the CNT center. The system shows a data-dependent processing time which was below the image acquisition time for all real-world experiments. It has to be noted that the magnifications mentioned in this section are calibrated for a 17" display with a resolution of 1280×1024 pixels.

6.2 Defect detection for biological cells (optical microscope)

Automated microinjection of biological cells can potentially speed-up drug discovery or production by orders of magnitude. The success rate of this procedure can be significantly increased by adding an automated quality check prior to the automated microinjection. In this context, the proposed system has been applied in order to detect defects such as mechanical damage in the input cell material. Defect cells can be sorted out automatically and will be removed from further processing. The results have been published before [18, 31, 114].

6.2.1 Automated injection of fluids into biological cells

Xenopus Laevis are very popular in research because they proliferate rapidly and their oocytes and embryos are very robust to experimental manipulation. Fully automated microinjection would simplify large scale studies and provide an important tool for research. As the oocytes are manually separated from the frog, their quality cannot be guaranteed. Defects include oocyte death, inadequate oocyte separation and external damage. The proposed system is applied in a preparation step for later micro injection of fluids into the oocytes.

Automatic characterization of biological cells is successfully performed in flow cytometry since around 40 years. Traditional flow cytometry is based on the measurement of laser light scatter and can monitor several thousands of cells per minute, e.g. in a blood stream. In contrast, image cytometry measures cell properties from image data [108]. It enables quantification of complex staining patterns but comes with a throughput far below laser flow cytometry. Some commercially available devices combine multispectral laser flow cytometry with additional microscope cameras [9]. However, these systems are not designed for single cell characterization and not suitable for large cells such as Xenopus Laevis oocytes with a diameter of approximately 1mm. Instead, the proposed system is applied for cell classification.

6.2.2 Application of the new system

An experimental setup has been built for oocyte monitoring, which can be seen in Figure 6.7. The oocyte suspension flows through a glass tube where it is in the scope of the macro lens system (reproduction scale 0.75). A FireWire camera with a 1/4" CCD is used for image acquisition. The proposed system performs image segmentation, feature retrieval and the actual object classifica-

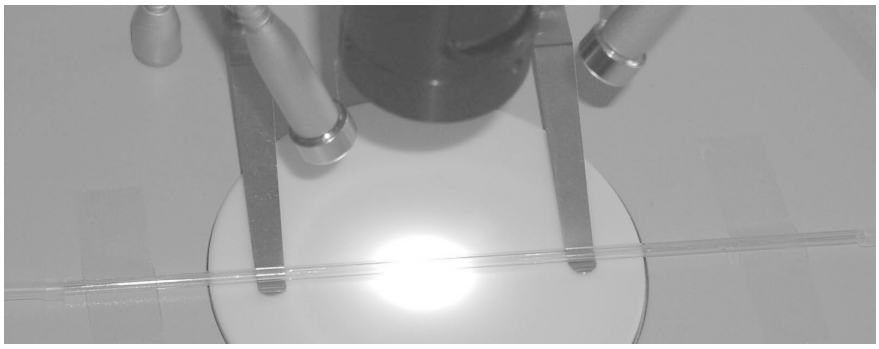


Figure 6.7: Experimental setup: the oocyte stream flows through a glass tube, which is in the scope of the macro lens

tion. Finally, the classification result is transmitted to the sensor server and may cause the controller to remove defect oocytes or particles from the stream using a downstream pump. In order to reduce the effect of motion blur, the exposure time has been set to 3ms at a frame rate of 60fps. The stream speed is 30mm/s in average and the field of view width is 4.74mm. The objects are in scope for up to 10 frames. Two white LEDs have been used for illumination.

Segmentation In this well defined environment, image segmentation can be performed by background subtraction in combination with fixed-level thresholding. Assuming that there are no changes in setup arrangement or illumination, a background image (or a time averaged series of n images) may be captured for calibration. Two aspects have to be considered:

- The image background is of white color. Illumination and exposure time are chosen so that background pixels are in saturation. This yields an almost noise-free background in the difference image and hence produces sharp contours after fixed-level thresholding.

- Shadows might appear when objects pass the field of view. Illumination may be arranged so that shadows only show up around the glass tube walls. These parts may be excluded from thresholding by initially setting the image region of interest to the inner part of the tube.

Figure 6.8 illustrates the process of image segmentation. The captured image (top row) is subtracted from the background image and the result (middle row) is thresholded, yielding a well-segmented image. By applying an additional thresh-

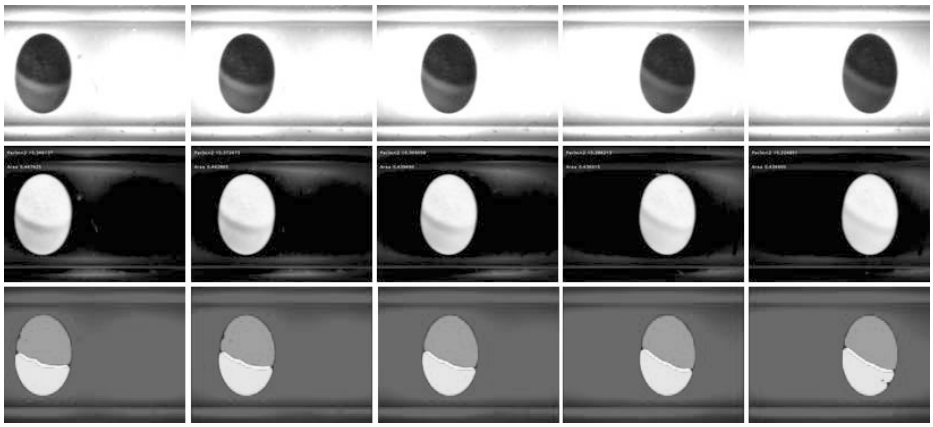


Figure 6.8: Passage of a single oocyte: original scene (upper row), difference image (middle row) and labeled poles (lower row). The field of view width is 4.74mm for each image.

olding procedure in the detected cell area, the dark and bright cell poles can be differentiated (lower row).

Feature extraction A variety of damaged oocytes and fragments can be seen in Figure 6.9. For the separation of viable oocytes (cp. Figure 6.8) from damaged oocytes or particles, two shape features have been selected:

1. Object size obtained by the number of object pixels. As the distance between object and camera is fixed by the pipe, no scale-invariant descriptor is needed.

2. Object roundness R_{2D} as introduced in Equation 4.1

These two measurement results may be rescaled in order to influence their impact on the classifier design. Object size has been normalized to half the ROI size. The influence of R_{2D} is outweighted by a factor of 18. This choice has shown best performance in experiments. For large objects, R_{2D} is a reliable feature for detecting defects. In contrast, the roundness of a small particle of only a few pixels is arbitrary and not suitable for proper discrimination. Therefore the object size feature is needed additionally.

Classifier training A nonlinear SVM classifier was trained using 130 representative samples, where 95 of them were negative. The regularization parameter has been set to $C = 0.5$. The RBF kernel function

$$K\left(\mathbf{x_i}, \mathbf{x_j}\right) = e^{-||\mathbf{x_i}-\mathbf{x_j}||^2} \tag{6.4}$$

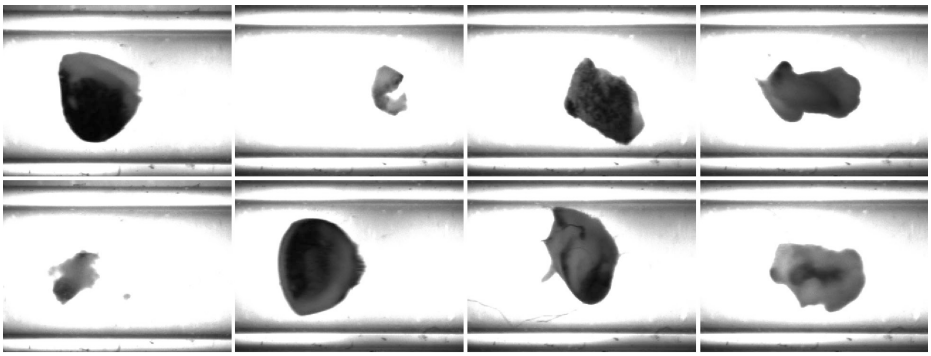


Figure 6.9: Oocytes with external damage and particles of blasted oocytes. The field of view width is 4.74mm for each image.

has been chosen. Training has been performed using SVM^{light} software [56]. A list of $N_s = 56$ support vectors s_i with associated Lagrange multipliers α_i and the hyperplane threshold b were obtained, which together serve as input to the classification module.

6.2.3 Results of cell defect detection

The classifier output can be seen in Figure 6.10. An output ≥ 0 signals positive samples (viable oocytes), whereas an output < 0 corresponds to objects that have to be removed. As intended, objects of deficient size or improper roundness are rejected. The classifier is far less tolerant to deviations in roundness than to little object dimensions. It has to be noted that perfect roundness corresponds to a value of $R_{2D} = 1$ which will never be reached by spherical oocytes, because the image is distorted by the glass tube surface.

A test series with 10 oocytes and particles has been performed multiple times. The classification results matched user judgment in all cases. Nevertheless the current setup brings some limitations:

- For a clear assignment of classification results, only single objects are expected to be in scope.

- Only defects which can be seen from the shape are detected. Cell death results in discoloration and cannot be detected.

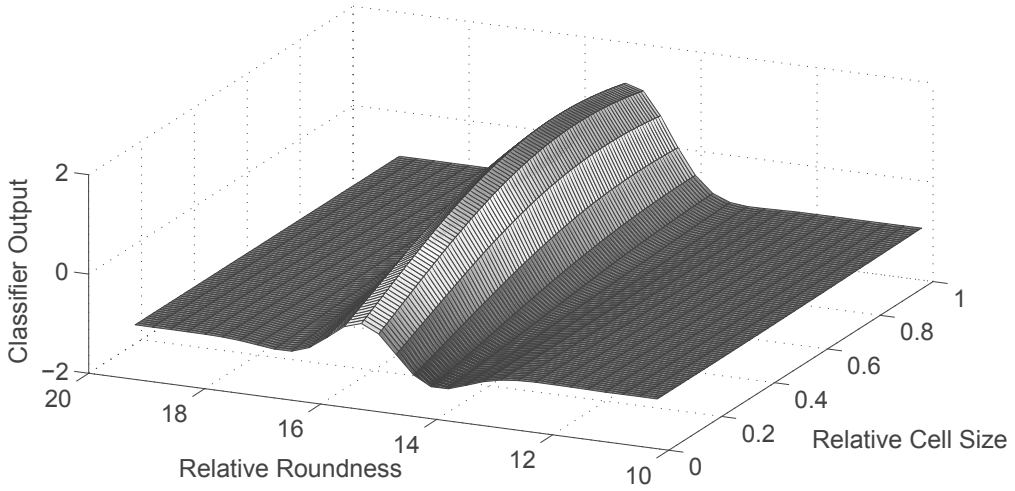


Figure 6.10: Classifier output to parts of the input space (features have been rescaled). The classifier rejects objects (output < 0) which are too small or deviate too much from optimal roundness.

6.3 Localization of magnetic particles (MRI)

The gradient fields of a clinical MRI scanner have the capability of inducing a force effect on magnetic particles. This property can potentially be exploited for applications such as targeted drug delivery or catheter navigation. With the help of a dedicated pulse sequence, the MRI can perform imaging and force effect in a time-interleaved way. The proposed system has been applied for the detection of magnetic particles in MRI scans. Combined with a suitable routine for object tracking and a controller, closed-loop position control of magnetic objects is enabled. Results have been partly published before with a focus on force effect [20], the overall setup [17, 19, 21] and navigation and imaging [35, 112, 113].

6.3.1 Navigation of magnetic particles using MRI

MRI is based on the effect of magnetic resonance (MR). If a strong external magnetic field B_0 is applied to a specimen, the nuclear magnetic moments (spins) align with the external field. Specially-shaped radio frequency (RF) pulses are capable of changing the orientation of the spins (excitation). The magnetization then starts a precession movement with a frequency ω_L, which depends on the

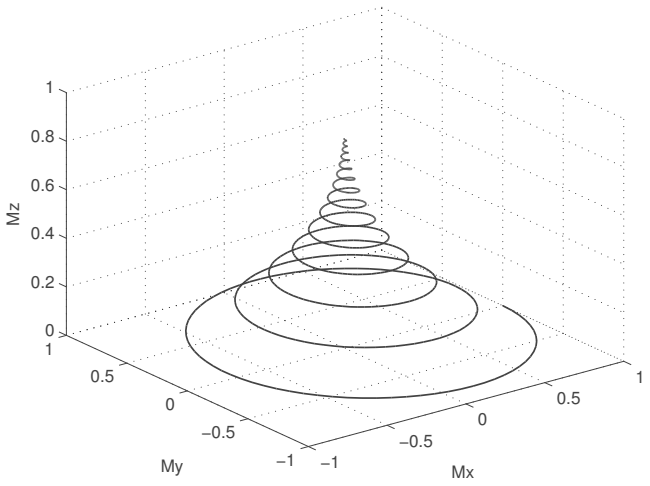


Figure 6.11: Relaxation trajectory of a nuclear magnetic moment (spin), obtained by evaluating Equation 6.5. Initially, the spin has been flipped into the xy-plane by applying a 90° excitation pulse. The B_0 field is aligned with the z-direction.

strength of B_0. After some time, the spin realigns with B_0. This process is referred to as relaxation and is governed by two timing constants $T_{1,2}$, which are material constants of the surrounding material. From a macroscopic view, the relaxation can be described by the Bloch equation [26], which has been evaluated in Figure 6.11:

$$\vec{M}(t) = \begin{pmatrix} e^{-t/T_2} & 0 & 0 \\ 0 & e^{-t/T_2} & 0 \\ 0 & 0 & e^{-t/T_1} \end{pmatrix} \cdot \begin{pmatrix} B & A & 0 \\ -A & B & 0 \\ 0 & 0 & 1 \end{pmatrix} \cdot M(t_0) + \begin{pmatrix} 0 \\ 0 \\ M_0 \cdot C \end{pmatrix}, \qquad (6.5)$$

with $A = \sin(\omega_L \cdot t)$, $B = \cos(\omega_L \cdot t)$ and $C = 1 - e^{-t/T_1}$.

Most imaging modes are based on visualizing local differences in T_1 and T_2. While B_0 is spatially and temporally constant, MRI scanners allow setting three gradient fields G_X, G_Y and G_Z dynamically. As the gradient fields are capable of locally changing ω_L, they are required for spatial coding. Also, they are capable of aligning spins and causing an echo, which is a measurable RF signal. All excitation, signal acquisition and gradient switching events are listed in a so-called pulse sequence. Depending on the method of creating an echo, spin-echo and gradient-echo sequences can be distinguished.

Besides the importance of G_X, G_Y and G_Z for the purpose of image acquisition, they also cause a force effect on magnetic objects. In common imaging

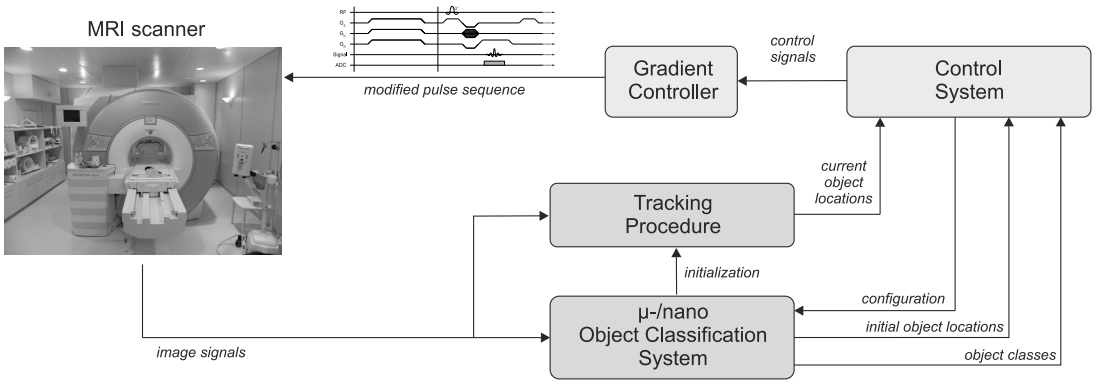


Figure 6.12: Application of the SPR-based system for micro- and nanoscale object classification for magnetic particle navigation using MRI. The MRI scanner serves as both, imaging modality and actuator. In addition to the object classification system, a tracking procedure is incorporated.

modes, this effect is minimal due to the frequent changes of gradient directions. On the other hand, the force effect can be utilized by switching the gradient fields in a controlled way. This method can potentially be exploited in applications such as steering magnetic drug carriers or actuating magnetic catheter tips. Figure 6.12 depicts, how the proposed system can be applied in such a scenario. Initially, one or multiple magnetic target objects are localized by the proposed system. The initial positions are transferred to a tracking procedure, which follows the movement of the objects in subsequent scans. Based on the actual object positions and the path-planning, the control system computes a propulsion strength and direction. The gradient controller interleaves propulsion gradients with the imaging sequence.

An alternative method for object localization has been reported, which does not reconstruct complete scans but is based on one-dimensional signal projections [32]. The method shown in Figure 6.12 is generally slower but on the other hand delivers updated image material of surrounding tissue, which can be used for path-planning. The detection of magnetic objects is based on the exploitation of magnetic susceptibility artifacts, which will be explained in the following.

6.3.2 Susceptibility artifacts

MRI scanners rely on the assumption of a set of well-defined magnetic fields which are the base field B_0 and gradient fields G_X, G_Y and G_Z for each direction. Much effort is spent during the scanner's construction to ensure, that those fields match

their theoretical shape as perfectly as possible. Therefore, a uniform magnetic susceptibility χ is needed inside the field of view. This requirement is violated to a small extent by the human body, which shows tissue and gas-filled cavities with different χ. Insertion of magnetic material into the field of view of a MRI scanner causes a tremendous impact on a number of image formation principles, because χ will vary by many orders of magnitude. The resulting image artifacts are called susceptibility artifacts. Nevertheless, a thorough knowledge about their occurrence and scaling laws may be exploited for the localization of magnetic capsules or nanoparticles inside the MRI. The size and shape of susceptibility artifacts strongly varies with the actual imaging sequence used.

There are two dominant effects caused by strong local susceptibility differences. These are spatial misregistration and intravoxel dephasing. Spatial misregistration results from the spatial coding principle of most MRI sequences. They rely on uniform gradient fields G_X, G_Y and G_Z to enable slice selection, frequency- and phase encoding as well as signal acquisition. Distortion of the gradient fields caused by susceptibility differences leads to incorrect frequency and phase of the spins surrounding the magnetic object. As a result, magnetic resonance signals are cancelled or misregistered in the image reconstruction [22]. The second effect is intravoxel dephasing, which is caused by the local gradient field around a magnetic object. Spins inside a specific voxel are usually assumed to behave uniformly and align at echo time. An additional dephasing dampens the echo signal and can lead to a complete loss of signal. This effect can be partly corrected in spin-echo sequences, due to the refocusing 180° RF pulse, which is characteristic to all spin-echo sequences. Gradient-echo sequences lack the 180° refocusing pulse and can only compensate for the systematic dephasing caused by G_X, G_Y and G_Z. Therefore, the effect of intravoxel dephasing can be considered to be severe in gradient-echo sequences.

Most studies about susceptibility artifacts focus on their avoidance in clinical examinations. A typical source of susceptibility artifacts are medical implants, for instance dental casting alloys [96] or intervertebral spacers [29]. Artifacts caused by three different metallic screws have been studied with respect to object size and orientation [66]. Ferromagnetic steel screws produced bigger artifacts as compared to paramagnetic titanium screws, a more severe signal cancellation was observed in gradient-echo sequences. The screws have been oriented along the direction of B_0. A comparable setup but with perpendicular screw orientation was presented in [71]. The artifacts show a larger extension along the B_0 direction. This is a result of the varying magnetic field shapes obtained for anisotropic objects with different orientations against an external magnetic field. Again, the artifact size varies with the material used. In medical examinations, susceptibility artifact occurrence can be reduced by modifying the patient ori-

entation [82]. This phenomenon may be exploited to localize objects with large aspect ratios such as needles [23].

For imaging of magnetic capsules, low aspect ratios may be assumed and therefore orientation against the B_0 field will be of minor importance. Instead, imaging sequence specific parameters are expected to play a key role. From the observations reported in clinical examinations, four parameters are supposed to have the strongest influence on artifact occurrence: *echo type*, due to the presence or absence of the 180° refocusing pulse, *echo time*, which is critical for the amount of dephasing, *voxel size* in frequency encoding direction and *flip angle*, which is the angle between the B_0 field and spins after excitation.

6.3.3 Application of the new system

Application of the proposed system requires a sufficient amount of training material. Therefore, magnetic particles have been imaged using different pulse sequences with variations in echo type, echo time, voxel size and flip angle. Furthermore, particles of varying size and material have been studied. A General Electric Signa 3T and a SIEMENS Verio 3T scanner have been used to perform the experiments. Isolated metallic objects cannot be imaged using MRI, due to the absence of the MR effect. Therefore, an additional signal source is needed. Agarose gel has been chosen because it combines two beneficial properties. First, it is a good source of signal due to the high water content. Second, is shows sufficient stiffness for fixation of the samples when they are introduced into the B_0 field. Strong field gradients ($> 1T/m$) and hence force effect is observed there.

A collection of small solid metallic objects and Ferrofluids were used as samples for the systematic experiments. Ferrofluids are superparamagnetic nanoparticles

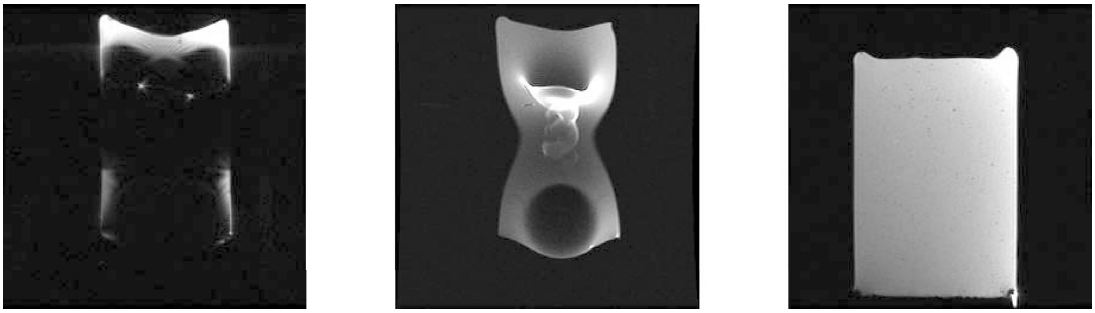


Figure 6.13: Heavy distortions caused by Ferrofluid, embedded into a 0.5l container of agar in gradient-echo (left) and spin-echo (middle). Right side: Spin-echo scan of agar phantom without Ferrofluid.

Object No.	Material	Shape	Dimensions	Total Vol.
1	Steel	Cube	1 mm × 1 mm × 0.5 mm	0.50 mm^3
2	Steel	Cube	1mm × 1mm × 1mm	1.00 mm^3
3	Steel	Cube	1mm × 1mm × 2mm	2.00 mm^3
4	NdFeB	Disc	Diam.: 2 mm, Ht.: 1 mm	3.14 mm^3
5	Steel	Sphere	Diameter: 1.0 mm	0.52 mm^3
6	Steel	Sphere	Diameter: 1.2 mm	0.91 mm^3
7	Steel	Sphere	Diameter: 1.5 mm	1.77 mm^3
8	Steel	Sphere	Diameter: 2.0 mm	4.19 mm^3
9	Steel	Sphere	Diameter: 2.5 mm	8.18 mm^3
10	Steel	Sphere	Diameter: 3.0 mm	14.10 mm^3

Table 6.1: Summary of solid objects used during experiments

of approximately 10nm in a suspension with a synthetic carrier oil. The solid samples are of steel and NdFeB, one of the strongest permanent magnets. The aim of this investigation was to study isolated susceptibility artifacts without superposition of tissue structure or additional objects. Therefore, each sample has been embedded into an individual agar container. Depending on the sample mass and material, 1l, 2l or 4l containers have been used. Figure 6.13 shows two sagittal scans of a 0.5l container of agar with approximately 0.5ml of Ferrofluid embedded and a container with pure agarose gel for comparison. The disproportion leads to a heavy distortion of the image in both cases. For a systematical study, the artifacts are supposed not to extend past the walls of the container. The solid objects are summarized in Table 6.1. A selection of typical artifacts for the three main sequence classes used during the experiments, can be seen in Figure 6.14. Three sequence types have been used: native 3D gradient-echo (GRE) and 2D spin-echo (SE) as the two sequence base-types, 2D single-shot fast spin-echo (SSFSE) as a representative real-time sequence. For the experiments, the real-time sequence was operated at 1 frame per second. Depending on the imaging parameters, 10 frames per second or more are possible.

For the detection of the susceptibility artifacts, a segmentation between artifact and image background is needed. In general, an artifact is defined to be a deviation between image and real structure that results from the imaging principle. Therefore, subtraction of a background image from the actual scan is an option to be considered. For the agarose phantoms, a homogeneous background signal may be assumed. On the other hand, usually no background information will be available in real-tissue scans. Instead of the background subtraction approach, the EM algorithm with a Gaussian mixture model (see Section 4.2.2)

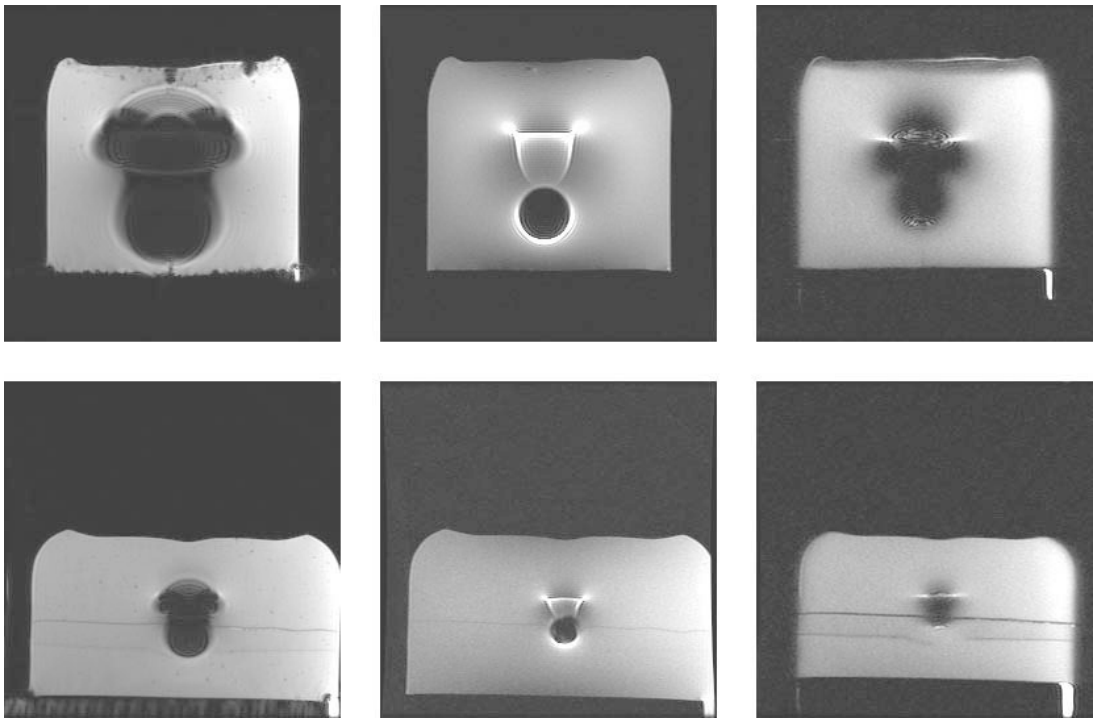


Figure 6.14: Typical artifacts in sagittal plane for gradient-echo (left), spin-echo (middle) and single-shot spin-echo (right) resulting from a 3mm steel sphere (upper row) and Ferrofluid (lower row), embedded into containers filled with agarose gel. The vertical lines result from the Ferrofluid sample preparation and are not an imaging artifact. The field of view width is 20cm for all scans.

is applied. A fixed number of segmentation zones is assumed to be present in the scan. The EM algorithm iteratively modifies the class centers, variances and proportions. In contrast to binary segmentation schemes, this procedure allows including the volume parts of signal loss into the artifact quantification as well as the high signal peaks caused by spatial misregistration. 3D objects are retrieved by 26-connected component search. Segmentation results for steel sphere samples can be seen in Figure 6.15. All sequence types show a characteristic shape of the artifact. Artifact size scales down with object size.

The artifact volume can be computed from the 3D object pixel mass and the voxel volume. Because different sequences have been used, the voxel volume needs to be derived for each examination using the following entries from the

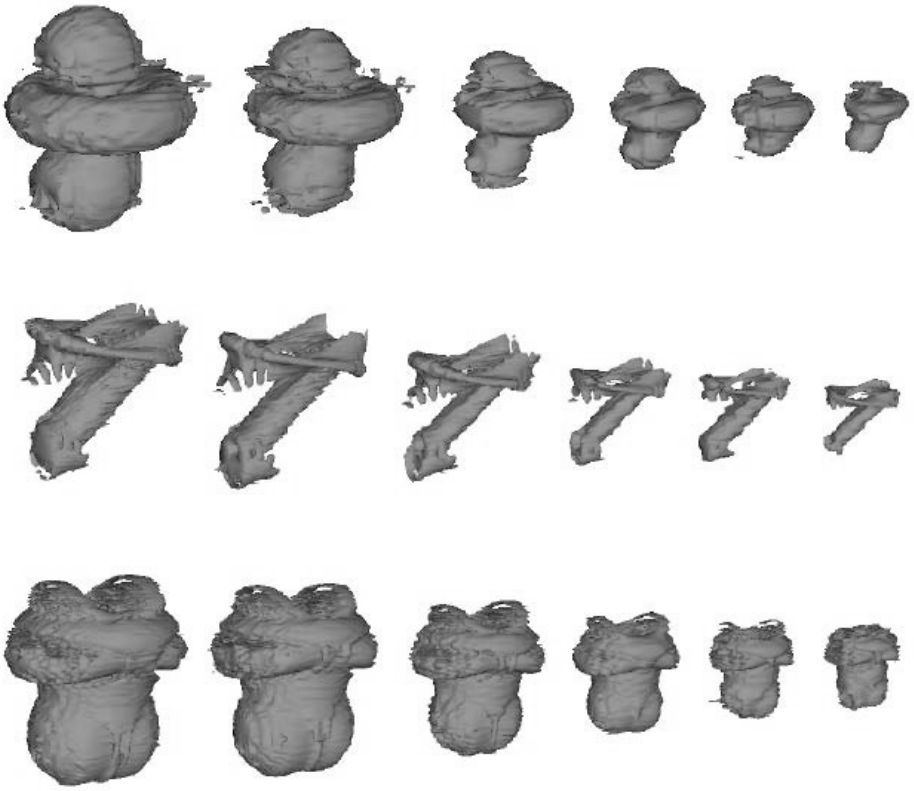


Figure 6.15: EM segmentation results for single-shot fast spin-echo (top) spin-echo (middle) and gradient-echo (bottom) for steel spheres. Diameter: 3mm, 2.5mm, 2.0mm, 1.5mm, 1.2mm, 1mm (left to right)

scan meta data:

$$VoxelVolume\,[mm^3] = PixelSpacing_1 \cdot PixelSpacing_2 \cdot SpacingBetweenSlices \tag{6.6}$$

SpacingBetweenSlides is supposed to be the sum of slice sickness and slice spacing. Measurement results for the 2mm steel sphere and all sequences can be seen in Figure 6.16. Gradient-echo artifacts are generally much larger than spin-echo artifacts. This is due to the general principle of echo formation of both techniques. As the refocusing pulse that is characteristic to all spin-echo sequences corrects for constant field inhomogeneities, generally smaller distortions can be expected. A long echo time increases artifact size for all spin-echo sequences. It has to be noted that Figure 6.16 gives a comparison in volume but the perceived difference in artifact size in a 2D slice is less severe.

Figure 6.17 provides further comparisons of artifact volume for selected se-

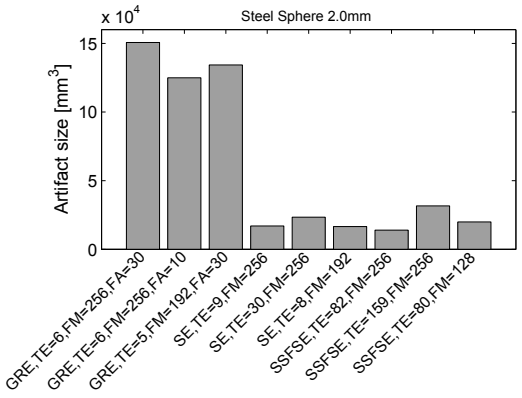


Figure 6.16: Artifact volume comparison of all sequences used for 2mm steel sphere with varying echo time (TE), flip angle (FA) and voxel size in frequency encoding direction (FM)

quences. As expected, the artifact volume is closely related to the magnetic object size and several orders of magnitude above the object volume. The artifacts observed for the Ferrofluid samples are comparable in shape and size to the solid samples. In contrast to medical contrast agents, they do not only lead to a local loss of signal but create heavy distortions. This is probably because of the high concentration of superparamagnetic particles and the high saturation magnetization (30 mT) of pure Ferrofluid.

For demonstrating the reproducibility of the results in a real-tissue environment, the steel sphere with a diameter of 1mm has been embedded into an animal tissue sample. Figure 6.18 shows corresponding MRI scans using different imaging sequences: fast low angle shot (FLASH), turbo spin-echo (TSE) and true fast imaging with steady state precession (TRUFI). The EM segmentation procedure can be applied in a similar way as for segmenting the agarose gel phantom scans. Intensity levels in MRI scans typically cover a larger range as compared to the usual 8-bit image data. Prior to display, a contrast stretch is normally performed. Instead, the segmentation procedure is working on the raw image data and can benefit from the high dynamic range. The areas of signal loss caused by a susceptibility artifact reliably produce low signal intensity values.

The proposed system for object classification has been tested with the help of a selection of scans from phantom experiments and full 3D head imaging. All scans have been acquired using gradient-echo sequences. After applying the segmentation procedure, three types of objects have been identified:

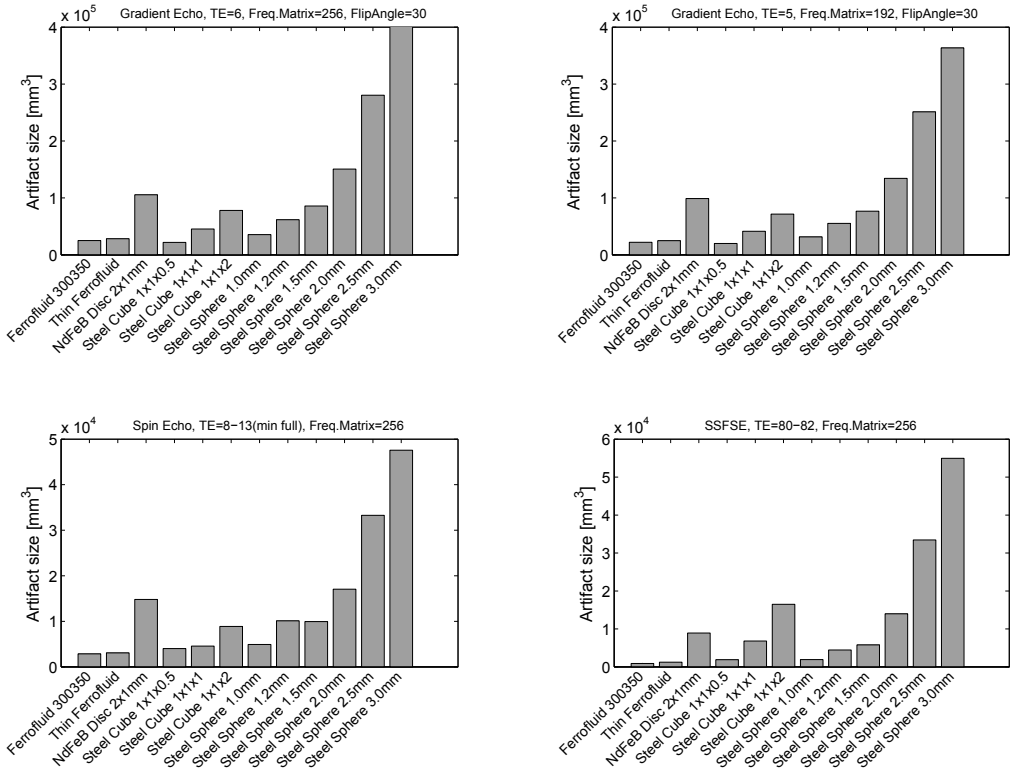


Figure 6.17: Artifact volume comparison for different sequences. The volume is derived by a fixed threshold 3D segmentation.

- The surrounding air produces low signal intensities and can be confused with the areas of signal loss caused by magnetic objects. Typically, the segmented background is the largest object in the volume of interest.
- Susceptibility artifacts caused by magnetic objects. The characteristics have been studied above.
- Cavities, bones and other anatomical structure can show object sizes comparable to those of susceptibility artifacts.

A scatter plot of all indentified objects can be seen in Figure 6.19, where positive samples (susceptibility artifacts) and negative samples (background and anatomical structure) have been marked. The feature vector for each sample is composed of the segmented object volume and the mean signal intensity inside the object. The plot shows that the EM segmentation can indeed converge

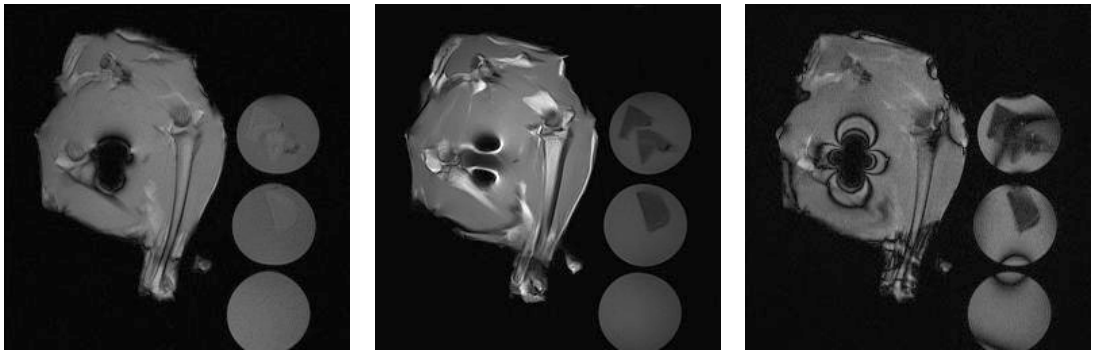


Figure 6.18: Steel sphere embedded into real-tissue environment. The image scenes have been captured using a FLASH (left), Turbo spin-echo (middle) and TRUFI (right). On the left side of each scan, a steel sphere with a diameter of 1mm has been embedded into two duck legs. The right side of each scan shows gel phantoms for comparison.

towards intensity levels above the mean level found in susceptibility artifacts. Nevertheless, the object classes can be distinguished based on the proposed feature vector and are in fact linearly separable. The decision boundaries of three classifiers have been indicated: a quadratic Bayes classifier, a back-propagation trained feed-forward neural network and a linear SVM. For the given classification task, the quadratic Bayes classifier is underfitted but ANN and SVM perform similarly well.

6.3.4 Closed-loop position control

A simplified setup has been used in order to demonstrate closed-loop position control. For this purpose, a swimming capsule has been filled with Ferrofluid and navigated along a parcours. The setup can be seen in Figure 6.20. The side length of the acrylic box is 30cm. In order to enhance optical visibility, a dark marker has been attached to the capsule. Propulsion and imaging has been carried out using a modified FLASH sequence. Position control has been performed by a two-dimensional PI controller. A list of waypoints has been set manually in order to define the travelling path of the capsule. The travelling speed of the capsule is mainly influenced by the gradient strength and the duty cycle. A maximal gradient strength of 20mT/m has been used. By optimizing the FLASH sequence parameters with respect to acquisition time, a complete cycle time around the parcours of 37s has been achieved.

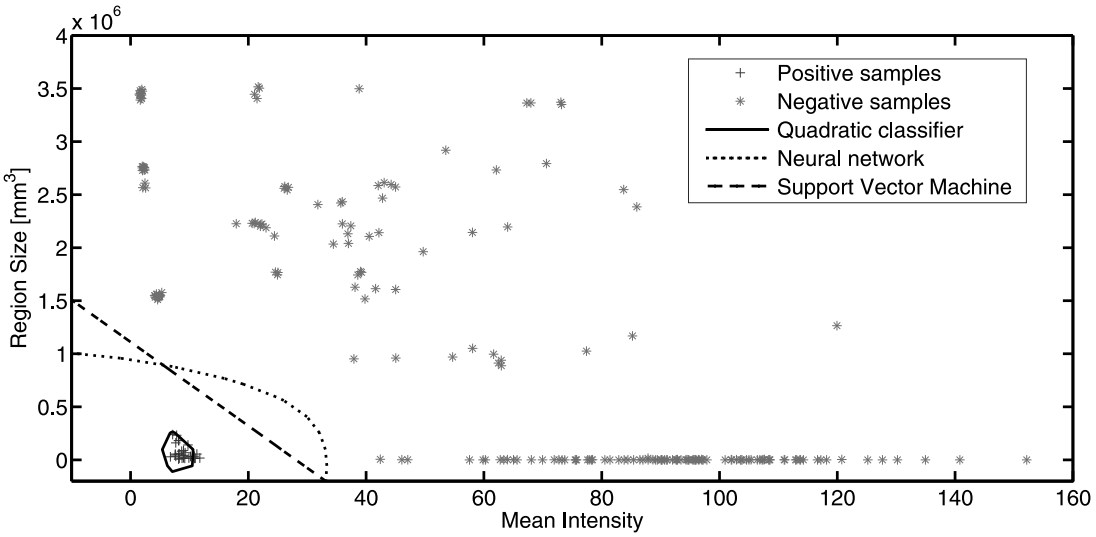


Figure 6.19: Scatter plot of segmented volumes with respect to size and mean intensity. The decision boundaries of three different classifiers are indicated.

6.4 Conclusion

This chapter presented three very dissimilar applications of the proposed system for micro- and nanoscale object classification. The applications cover three imaging modalities and also strongly different types of objects. Also, the classification tasks to be carried out are unequal. Workpiece detection has been demonstrated by performing the SEM-based CNT search. Defect detection of biological cells using the optical microscope implements a quality check. Localization of the magnetic particles in the MRI is a form of actuator detection. Another dissimilarity of the three setups are the real-time constraints. While the oocyte quality check is performed in motion and requires a system response within $<<0.5$s, the CNT search and magnetic capsule detection are performed in static image scenes, which do not impose strict timing requirements.

It can be concluded that the proposed system presented in Chapter 4 solves the three tasks successfully, whereas the state-of-the-art methods described in Chapter 2 are hardly applicable. The main reason for the success of the proposed system is the idea of not focusing on direct similarity between a new image and some training images. Instead, the problem characteristics are learned and also the decision boundaries are chosen automatically. The proposed system does not replace the state-of-the-art methods described in Chapter 2. It rather extends the capabilities of automatic image analysis in micro- and nanorobotic applications

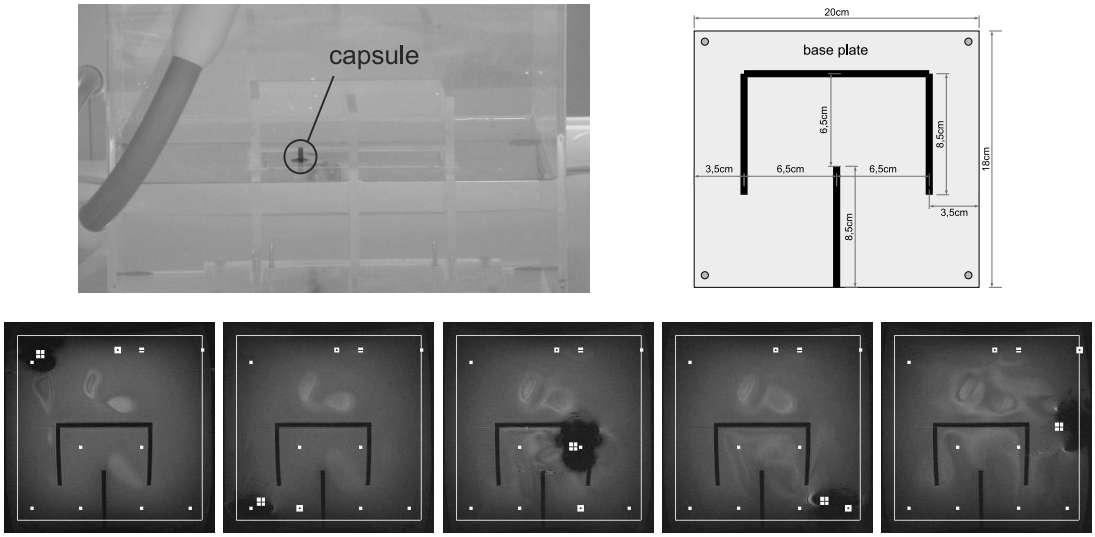


Figure 6.20: Experimental setup for closed-loop position control. Upright walls are mounted on the base plate (upper right), which has been placed inside a water-filled acrylic box (upper left, camera view). A capsule filled with Ferrofluids is travelling along a pre-planned path around the obstacles (lower row, MRI view).

and thereby increases the level of automation.

A clear benefit of the proposed system is the high level of integration into a control- and automation environment. By using a GUI, the proposed system can be configured flexibly for any new classification task. During automated micro- and nanohandling procedures, the control system then restores the configurations. The side-by-side use of the proposed system with other image processing routines and also other forms of sensory feedback enables fully automated handling or assembly procedure at the micro- and nanoscale.

7 Experimental validation of the image registration strategy

A new scheme for multimodal image registration has been presented in Chapter 5. The proposed registration scheme has been implemented and applied to the problem of registering images obtained from AFM and SEM. This chapter presents the implementation details and performance results obtained during the experimental validation of the proposed registration scheme. The results partly originate from the EU-funded research project *Building an Analyzing Focused Ion Beam for Nanotechnology* (FIBLYS) and have been partly published before [111].

7.1 Registration of AFM and SEM images

7.1.1 Motivation for combining AFM and SEM

AFM and SEM have been introduced already in Chapter 4 in the context of micro- and nanoscale object detection. Both modalities bring the necessary capability of imaging objects and structures at the micro- and nanoscale. On the other hand, working principle and imaging characteristics of AFM and SEM are strongly different. Many arguments motivate using AFM or SEM not alternatively but to combine the benefits of both imaging modalities. Therefore, AFM and SEM will be revisited with a focus on complementary properties.

The AFM probes the sample surface with the tip of a cantilever. From the deflection of a laser beam pointing at the cantilever, the force between tip and surface is derived. An alternative method of measuring the cantilever bending uses piezoresistive elements integrated into the cantilever. Generally, three modes of AFM operation can be distinguished based on the tip-sample interaction: contact mode, intermittent contact mode and non-contact mode. Besides the ability to reconstruct the specimen topography, the AFM can also be used to measure physical properties of a sample surface. These include magnetic and Coulomb forces, friction and chemical interaction. The working principle of the conventional AFM is depicted in Figure 7.1.

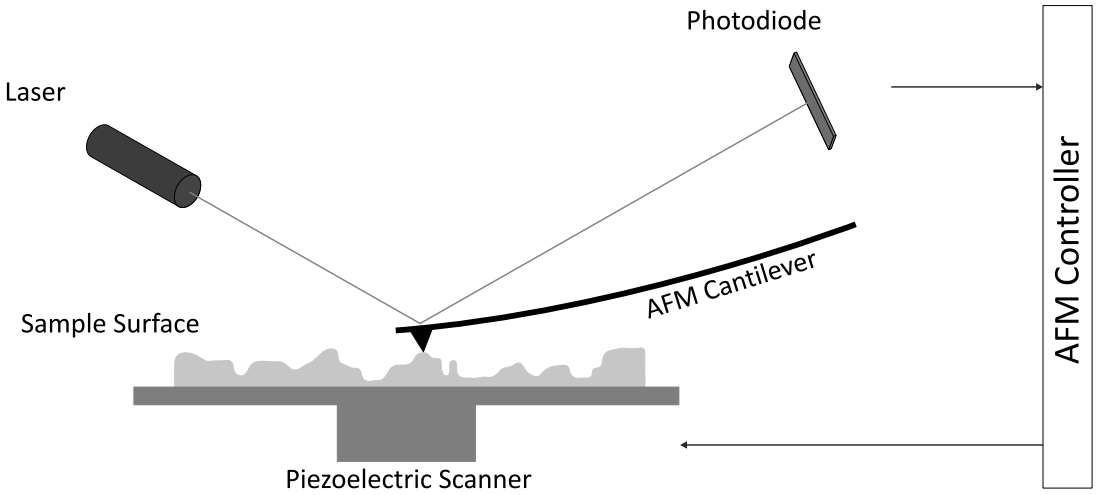


Figure 7.1: Working principle of the AFM. The sample is mounted onto a piezoelectrically driven stage and scanned below the tip of the cantilever. By measuring the laser beam deflection, the controller reconstructs the sample surface.

The SEM generates an electron beam which is used to scan the sample surface [84]. An electron gun equipped with a tungsten filament or a field emission gun is used as electron source. From the electrons emitted by the sample, a signal can be measured that is used to form an image. The resulting SEM image displays a mixture of different types of image contrast. Contrary to AFM, a pure representation of the specimen topography by image intensity is difficult in the SEM. Because the electron beam does not only interact with the specimen surface but also with subjacent material, the measured signal is influenced by the three-dimensional structure of the specimen. The imaging modes of the SEM can be distinguished depending on the type of the detected signal. The secondary electrons emitted by the sample are the most frequently used source of signal. Secondary electron (SE) scans are strongly influenced by the specimen topography. From the primary electrons that are backscattered from the sample, the backscattered electron (BSE) signal can be detected. These images are mostly influenced by the material composition of the sample. The working principle of the SEM is depicted in Figure 7.2.

Combined AFM and SEM studies can provide a thorough view of a specimen surface and material properties. Dual studies have been reported in several applications, indicating a clear benefit of the side-by-side use of AFM and SEM. Some examples are human hair analysis [81], imaging of Bacillus spores [102], nanofibres [107] and studies on porous anodic alumina [119]. An example of an

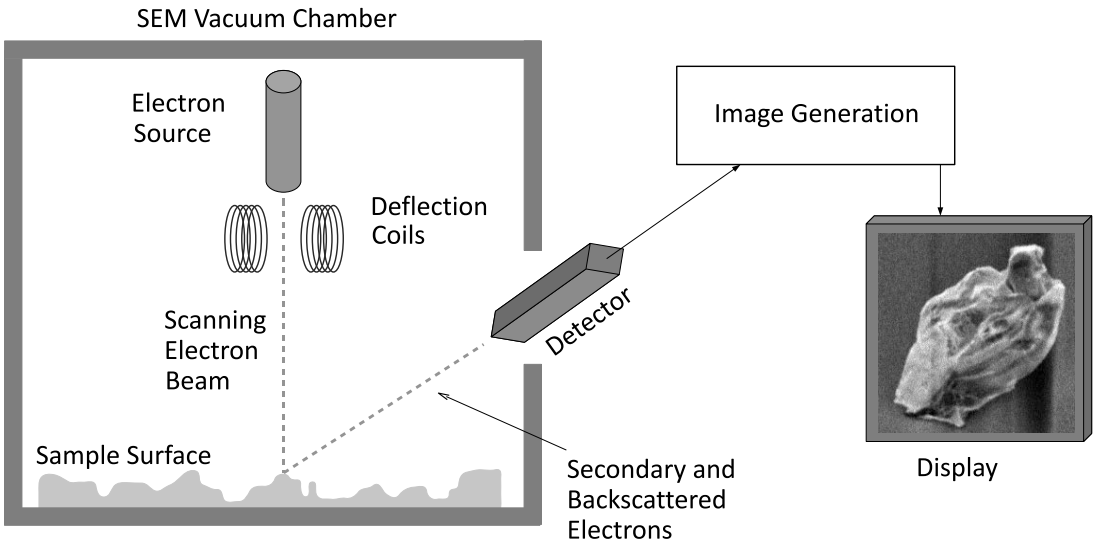


Figure 7.2: Working principle of the SEM. The focused electron beam scans the sample surface. In standard imaging mode, backscattered or secondary electrons are detected and the measured signal is used to form a two-dimensional image. More specialized analysis techniques detect X-rays or cathodoluminescence.

AFM and SEM image pair can be seen in Figure 7.3. The AFM can provide a higher vertical resolution than the SEM, ranging down to < 0.5 Å. With a number of precautions, true atomic resolution is possible with the AFM. The maximal lateral resolution of AFM and SEM is approximately equal and falls in the range of a few nanometers. On the other hand, SEMs can be operated at low magnifications with a field of view of several millimeters. The largest AFM scans typically cover an area of 100 µm $\times$ 100 µm. Due to the high depth of field of the SEM, which can be in the range of millimeters, it can image rough surfaces. The imaging height of the AFM is limited by the vertical range of the scanner, which is typically < 20 µm.

Due to the complementary nature of AFM and SEM in regard to maximal resolution and field of view, combined studies do not only benefit from successive but also from simultaneous AFM and SEM imaging. In this case, the SEM also helps to guide the AFM cantilever to the designated location. Some effort has been made towards the integration of an AFM inside the vacuum chamber of the SEM [28, 57], where the focus has been on the mechanical setup. The proposed registration scheme provides a method for automatically generating a more meaningful view of the obtained image data. This can be employed

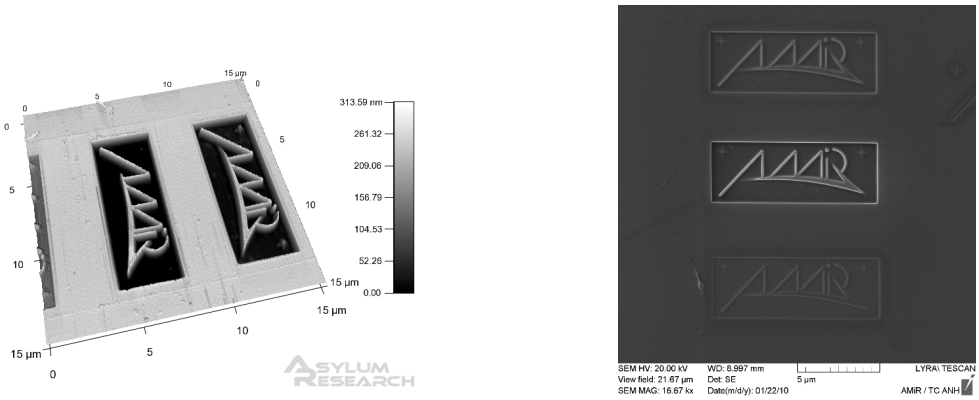


Figure 7.3: Example of AFM and SEM image pair, showing FIB-milled structures on a silicon wafer. The left scan shows the AFM topography view. The right scan shows a secondary electron SEM image, covering a larger scanning area than the AFM scan.

for either retroperspectively analyzing the outcome of a combined AFM and SEM study or for guiding successive AFM scans during the examination. The proposed registration scheme is applicable to hybrid AFM and SEM setups as well as successively acquired scans from separate devices.

7.1.2 Transformation model

The description of the spatial correspondence between AFM and SEM scan requires a transformation model $T(x, y)$. For any image coordinates (x, y) in the target image, $T(x, y)$ are the image coordinates in the base image. In many applications, the SEM scan covers a larger area than the AFM scan. This is a reason why the SEM scan has been chosen as the base image. However, this choice is arbitrary and all methods presented will also function with the AFM scan as the base image. A general consideration when chosing the transformation model is which aspects to include into the model and which to exclude from the optimization and tread as a preprocessing step. The only preprocessing step applied here is fitting of a polynomial curve, in order to assure a uniform ground level in the AFM scans. The curve

$$P(x, y) = \sum_{i=0}^{2} \sum_{j=0}^{2} \alpha_{ij}\, x^i\, y^j \tag{7.1}$$

has been fitted into the ground regions and subtracted from the original scan data. Additionally, the method proposed in [40] could be used to compensate for

local distortions in the SEM scan. Similar to [30], where a linear model is used to register contact- and intermittent contact mode AFM scans, the transform is chosen to be a combination of scaling $\mathbf{S}$, rotation $\mathbf{R}$ and translation $\mathbf{t}$:

$$T(x, y) = \mathbf{S} \cdot \mathbf{R} \cdot \begin{pmatrix} x \\ y \end{pmatrix} + \mathbf{t}, \tag{7.2}$$

where

$$\mathbf{S} = \begin{pmatrix} s_x & 0 \\ 0 & s_y \end{pmatrix}, \quad \mathbf{R} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \quad \textit{and} \quad \mathbf{t} = \begin{pmatrix} t_x \\ t_y \end{pmatrix}. \tag{7.3}$$

The use of this simple transformation model can be justified by two arguments. First, the rigid nature of specimens such as a silicon wafer makes deformations during or between the acquisition procedures unlikely. Second, global distortions originating from the acquisition procedure can be assumed to be mostly constant over time. Those can be corrected for prior to the registration procedure.

7.2 Selection of samples for performance evaluation

7.2.1 Requirements to the image material

Testing the proposed registration scheme requires a selection of samples which provide a large spectrum of image contents. This includes all basic geometric shapes such as corners, edges and blob-like structures. Some samples should contain deep structures and also rough surfaces. Ambiguous structures are simple to detect but hard to distinguish. Those structures are good for testing the descriptor performance. Another type of structures reproduces over changes in scale. This includes especially isolated corners and junctions. While these are usually not a problem to the area-based registration step, the scale detection in feature-based registration is likely to become unstable at these points.

Besides these requirements to the structures to be imaged, the scans should also include some sensor-specific imaging artifacts. For the SEM these can be the occurrence of shadows in the direction facing away from the detector. Also depending on the viewing angle, thin structures and edges tend to appear bright in the SEM scan. Frequent artifacts in AFM scans include line artifacts originating from a malfunction of the height control. Also, inaccuracies of fitting a ground level into the scan area can lead to an intensity ramp superimposed to the ground level.

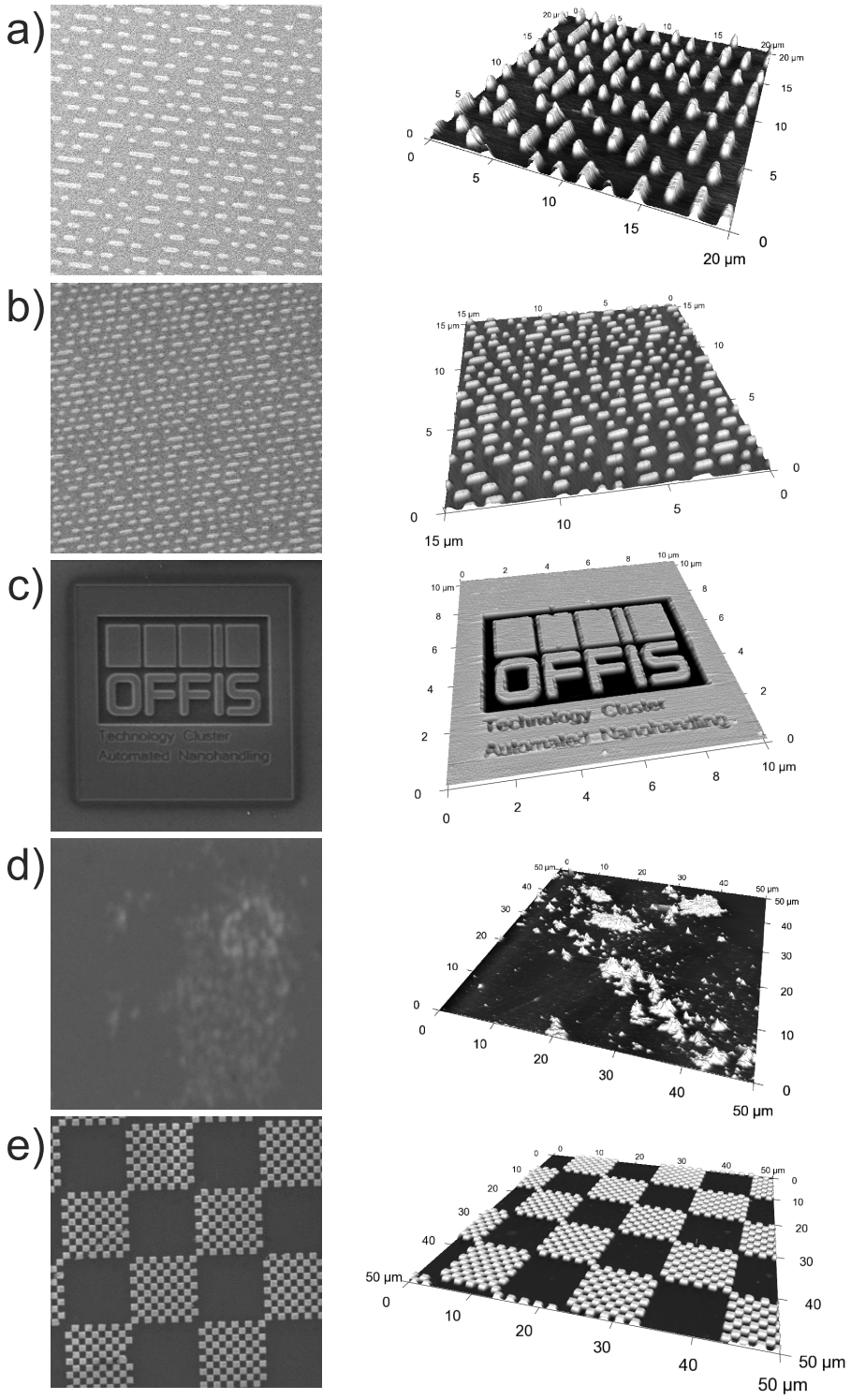


Figure 7.4: Samples used during systematical performance evaluation, scanned by SEM (left column) and AFM (right column): CD surface (a), DVD surface (b), FIB-milled pattern (c), gold nanoclusters (d), gold on silicon test pattern (e).

7.2.2 Sample preparation

The image material has been acquired using a Tescan LYRA 3 FEG/XMH (SEM), a Carl Zeiss LEO 1450 (SEM) and a custom-built AFM setup, which can be integrated into the vacuum chamber of the SEM [28]. If not stated otherwise, SEM scans are based on the SE detector signal. The AFM scans measure surface topography in intermittent contact mode. All types of samples used during systematical performance evaluation can be seen in Figure 7.4. The goal has been to cover a range of different image contents for obtaining a meaningful performance estimation of the proposed registration scheme. Compact Disc (CD) and Digital Versatile Disc (DVD) samples have been prepared for inspection by separating pieces of the data layer. The data layer surface shows structures in blob and dash shape. For obtaining edge-like structures, letters have been milled into a silicon substrate using the focused ion beam (FIB) column of the Tescan LYRA. Gold nanoclusters of varying size have been imaged on a silicon substrate. A test pattern made of gold on a silicon substrate exhibits a lot of square-shaped structure, which is highly ambiguous.

In addition to the samples used during systematical performance evaluation, the FIB-milling has been applied to produce a number of especially challenging samples. The procedure is the same applied in the production of the pattern seen in Figure 7.4 c), but the structures show a lack of distinctive image detail.

7.2.3 Artificial image material

Multimodal image registration can be regarded as a registration task with two different sensor characteristics. In unimodal image registration, the main difficulty is to handle geometric differences in the arrangement of sensor and image scene. The assumption is that without any change in scene arrangement or sensor settings, images acquired multiple times are either identical or eventually degraded by additive noise. However, this assumption does not hold in multimodal image registration and especially not in AFM and SEM registration. The registration problem can be understood more clearly by considering that there is only one physical sample surface both imaging modalities are working on. The AFM and SEM scan are influenced by the sample (which is identical), the sensor arrangement and imaging characteristics (which are both different). Imaging characteristics can be modeled by a number of geometric and photometric transformations. Therefore, not only the process of imaging the sample but also the transition from AFM to SEM scan can be modeled by a series of transformations.

The benefit of using artificially generated images for testing the registration scheme is that the effects of different imaging characteristics can be studied in

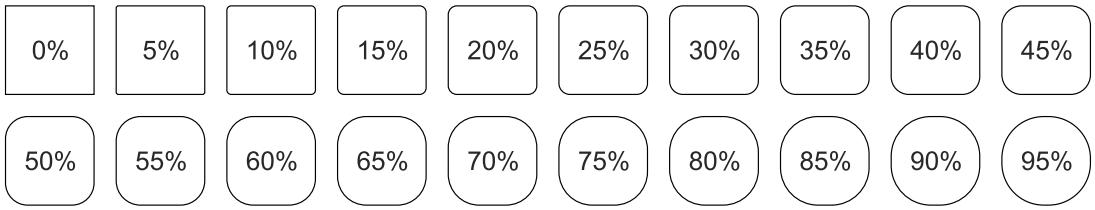


Figure 7.5: Construction of objects for generating artificial image material. The targets are parameterized by width, height and curvature. This series shows objects with a fixed and identical width and height and varying curvature ranging from 0% to 95%.

an isolated way. When using real AFM and SEM scans, the registration scheme needs to manage a number of differences in imaging characteristics simultaneously. Isolating these effects in synthetic images allows to study effects such as differences in contrast level separately. For this reason, a simple method of generating synthetic image pairs with well-defined properties has been established. A requirement to the image material is the diversity of image structures presented. Synthetic image material can be generated with the help of any pseudo-random number generator and the help of parameterized geometrical objects. The objects incorporated here are rectangles and ellipses. Object parameters are object height, width, location, intensity and curvature. The curvature parameter states the percentage of the object side formed by the quarter-pieces of ellipses. The remaining side length is a straight line. This means that 0% of curvature produces a rectangle and 100% of curvature produces an ellipse. The principle is depicted in Figure 7.5.

Figure 7.6 shows an example of an artificially generated image. Object positions have been specified using an equidistant grid, all other object parameters are obtained by the random process. This type of image material shows structures ranging from spherical to string-like and also different strength of background contrast. It should cause all feature detectors to respond and also provides enough structure for area-based registration. Image pairs for testing the registration scheme can be constructed by applying different transforms to an artificially generated image. Those can include morphological operations, variations in contrast or intensity level or also adding noise.

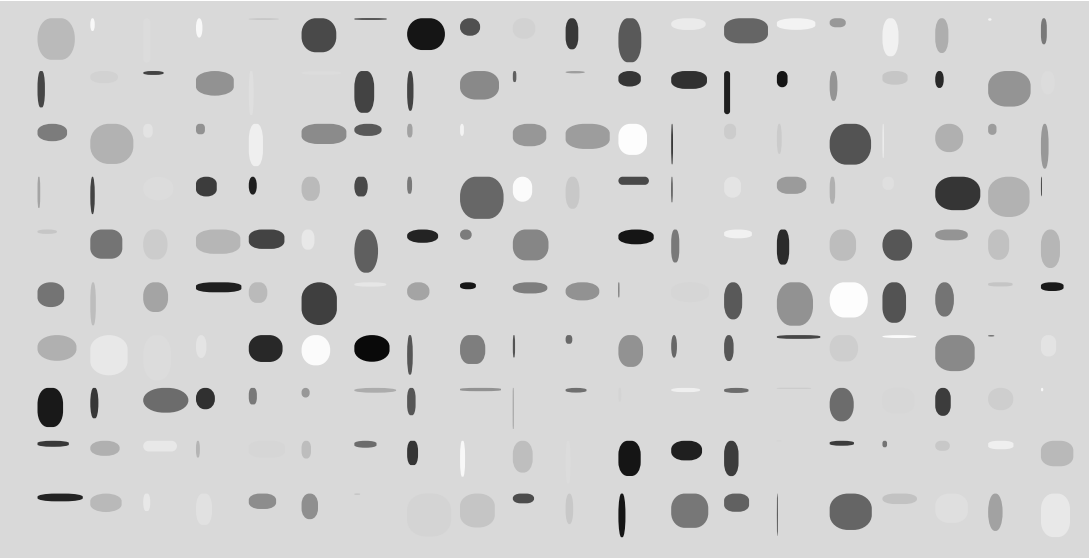


Figure 7.6: Artificial image showing 200 basis objects. Object parameters including width, height, curvature and intensity have been generated using a pseudo-random number generator. These artificially generated images are used in order to study the effect of isolated differences in imaging conditions on the registration performance.

7.3 Performance analysis of the new registration strategy

The proposed registration strategy has been validated component-wise and also as a complete system. Initially, the performance of the feature-based registration step has been studied systematically by registering scans obtained from AFM and SEM. Next, the gain in performance caused by combining detectors and applying the new feature matching strategy is discussed. It is shown, how area-based refinement of the result leads to an improvement of registration accuracy. Finally, the registration strategy is discussed in its entirety.

7.3.1 Feature-based registration

Performance criteria

In Section 5.2.2, the repeatability criterion and matching score have been introduced in the context of combining detectors. Repeatability is used here as a measure of how well a detector reproduces results in AFM and SEM scans. The

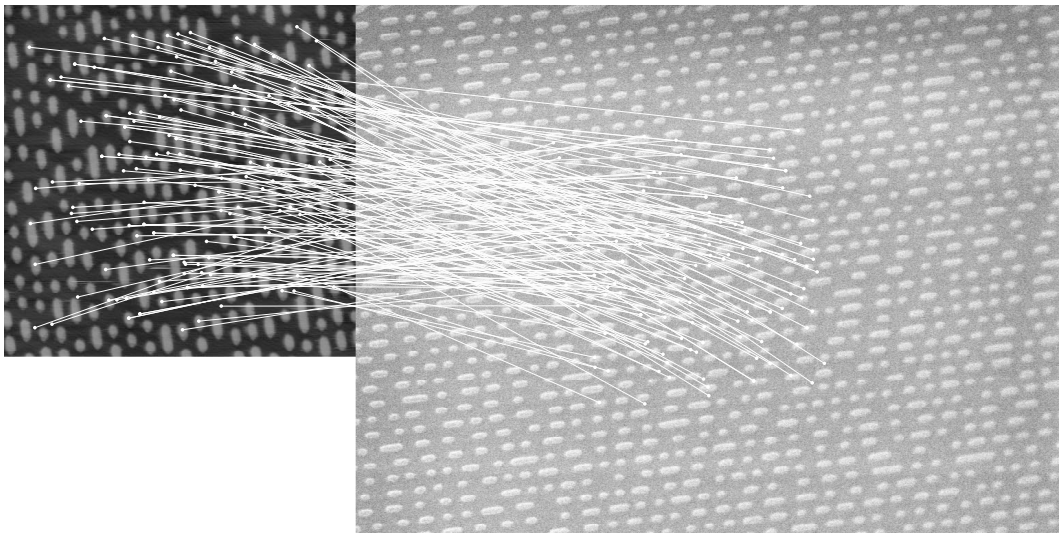


Figure 7.7: Feature correspondences between an image pair showing the surface of a DVD. Only correct matches are displayed. The left image shows the AFM scan; the right image shows the SEM scan.

matching score is a simple measure of how well a description scheme transfers initially detected regions into correct region correspondences. It is a good indicator for directly comparing multiple descriptors. A deeper insight into descriptor and matching performance can be obtained with the help of the recall versus 1-precision plot [75]. It measures the ability to detect correct feature correspondences and to reject incorrect matches. Recall and 1-precision are defined as follows:

$$recall = \frac{\#correct\ matches}{\#correspondences} \quad , \tag{7.4}$$

$$1 - precision = \frac{\#false\ matches}{\#correct\ matches + \#false\ matches}. \tag{7.5}$$

For a series of varying matching thresholds the values are plotted on a curve. A good descriptor and matching strategy obtains a high recall rate (detects many of the existing feature correspondences) and a low 1-precision value (returns few incorrect correspondences). Such a results allows model fitting tools such as RANSAC to identify the correct subset of matches and to accurately determine the transformation model parameters. Correct feature correspondences between AFM and SEM scan of a DVD surface can be seen in Figure 7.7.

Feature detectors

The result of the detector performance analysis is shown in Figure 7.8 for all detectors. The repeatability values are comparable to those obtained in natural photography under transformations such as a change in viewpoint. Random sampling is added for comparison only and is not considered as a detector actually used in the proposed registration scheme. It has to be noticed, that region sizes have been normalized prior to computing the repeatability score. Therefore, the region sizes produced by the detectors do not have an influence on the repeatability score.

For the DVD and FIB-pattern sample, the detector performance is comparably uniform. For the other samples, there are high variations in repeatability. The best detector in average repeatability is the Hessian-Laplace detector, followed by SIFT and Harris-Laplace detector. MSER and SURF detector reside in the medium range. The average results for the salient region detector and the IBR are noticeably low. It can be expected that this is due to violation of the basic assumptions these detectors are built on. For the IBR detector, this is the repeatability of peaks in image intensity. The salient regions detector is based on the idea that peaks in an entropy measure reproduce. It seems that the assumptions, the other detectors are based on better meet the actual conditions found in AFM and SEM image registration. Those are the repeatability of segmented regions (MSER) and basic geometric shapes (Hessian-Laplace, Harris-Laplace, SIFT, SURF).

Principally, feature-based registration can be performed with all detectors and no repeatability score is prohibitively low. However, applying the proposed registration scheme under exclusive usage of the salient region detector or IBR should be avoided.

Feature descriptors

The performance of the feature descriptors is analyzed in terms of the matching score. Because this measure is obtained by building a correct ratio, also the total number of correct matches is stated. A descriptor can only be tested if regions have been detected before. The question arises, which region detector to choose for descriptor evaluation. Feature detectors and descriptors can be tested independently in wide parts [75]. Nevertheless, the actual results obtained in descriptor evaluation will vary with the choice of a detector. What remains mostly untouched is the ranking of the descriptors. Here, the best (Hessian-Laplace), the second best (SIFT) and the fastest detector (SURF) have been

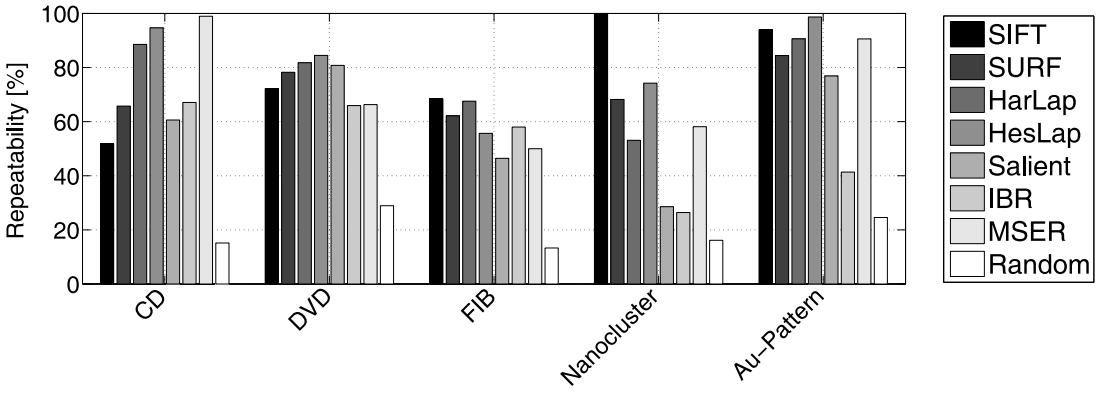


Figure 7.8: Detector performance in terms of repeatability for all samples. A high repeatability means that if a feature is detected in the AFM scan, it is likely to find a corresponding feature in the SEM scan. Feature correspondence is defined by an overlap error $< 50\%$. Due to the aspect of multimodality, a lower repeatability is obtained here as compared to unimodal registration (e.g. photograph stitching).

selected for comparing descriptor performance. The results can be seen in Figures 7.9 and 7.10.

It can be seen that the performance varies significantly between the different samples. This behavior is due to the different characteristics of the samples in terms of richness of image structure and also distinctiveness. In other words, the registration task is of different difficulty. It can also be seen, that the number of correct matches varies between the different detectors. This is caused by the different detection behavior and the different set of regions the descriptors are extracted from.

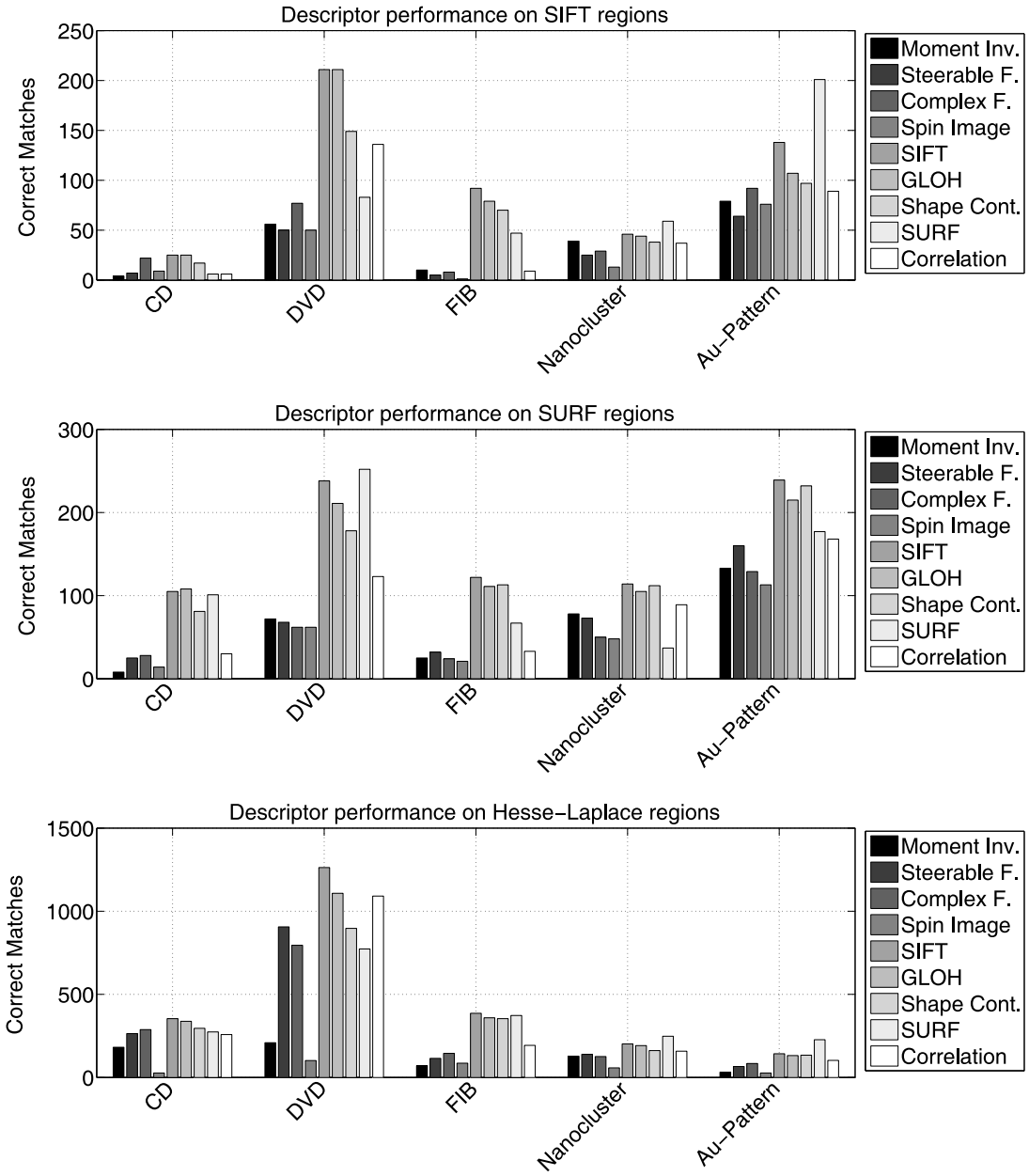


Figure 7.9: Performance of the feature descriptors in terms of total correct matches. Results for regions obtained by the SIFT (top), SURF (middle) and Hessian-Laplace (lower row) detector are shown.

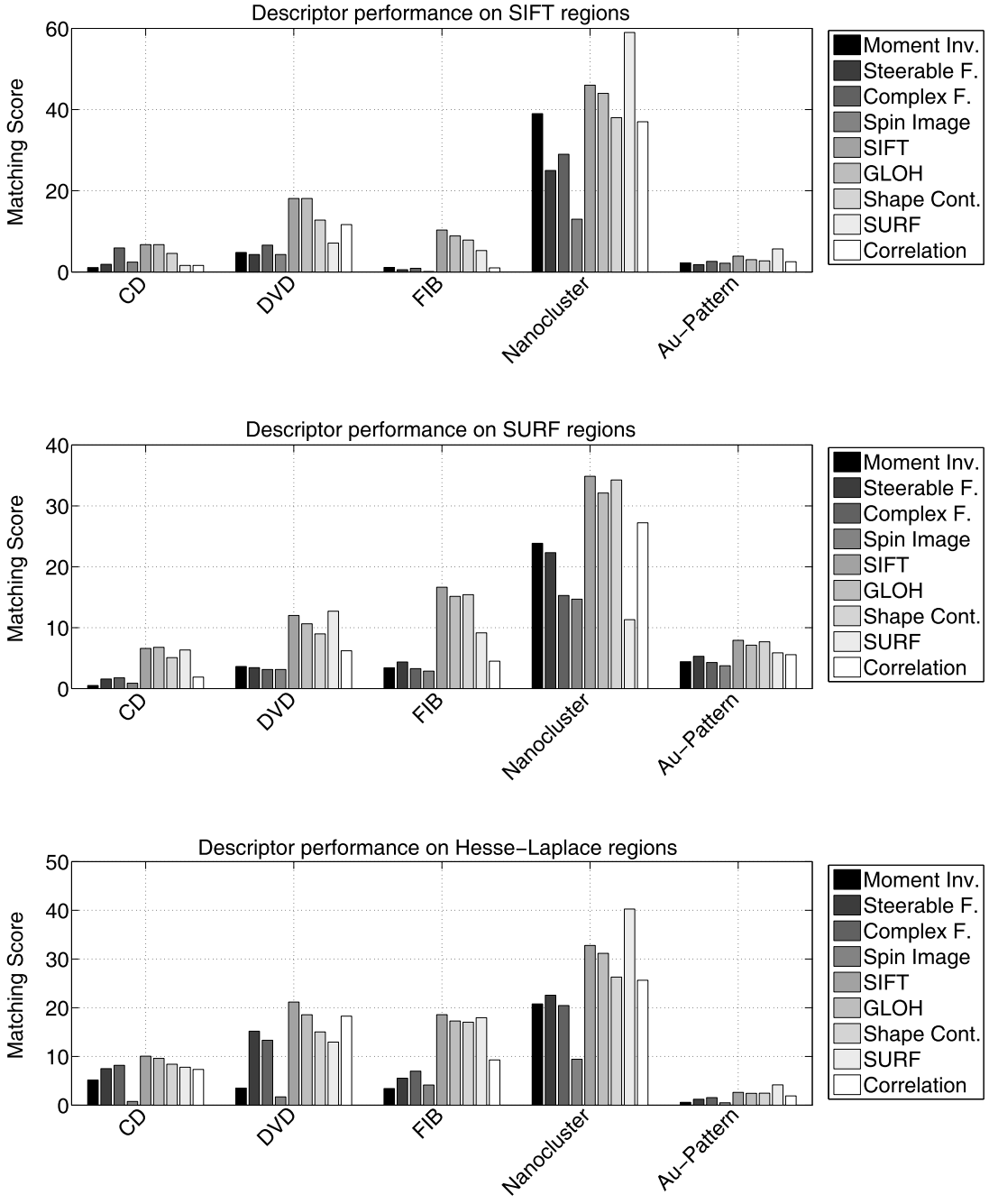


Figure 7.10: Performance of the feature descriptors in terms of matching score. Results for regions obtained by the SIFT (top), SURF (middle) and Hessian-Laplace (lower row) detector are shown.

However, when fixing the setup and detector, the ranking of the descriptors can be seen clearly. The performance of moment invariants, steerable filters, complex filters and spin images is far below the performance of the other descriptors. The information extracted by these descriptors does not reproduce sufficiently in corresponding regions scanned by AFM and SEM. In average, the SIFT descriptor performs best, followed by the GLOH descriptor. Shape context and the SURF descriptor show a slightly lower performance. Cross-correlation performs on a medium level, although it is the most simple descriptor.

In contrast to feature detectors, combination of multiple description schemes implies much computational effort and also the fusion of the results is unclear. If one exclusive descriptor is selected, the SIFT detector proved best performance in the test setup. However, the SURF algorithm execution speed benefits from reusing integral images. If those are available from the detection phase, the SURF descriptor is still a choice worth considering. The GLOH descriptor has been introduced as an improved SIFT descriptor mainly for photography. However, for the registration of AFM and SEM scans, only one example could be found where a marginal improvement has been achieved (CD sample, SURF regions). For this reason, usage of the GLOH descriptor is not considered for AFM and SEM registration.

Synthetic image material

Synthetic image material has been generated in order to further study the behavior of multimodal registration based on local features. Figure 7.11 shows examples of the image material used. The base image has been superimposed with an intensity ramp $\mathbf{I_R}$ with ramp intensity i_R. It is constructed with the help of the dyadic product of vector $\mathbf{v_R}$ which is equal in length to the side length of the base image.

$$\mathbf{v_R} = (0 \ldots 1)^T \qquad \mathbf{I_R} = i_R \mathbf{v_R} \mathbf{v_R}^T \tag{7.6}$$

Such a behavior can be found around exposed structures in the SEM or also in AFM scans where the surface fit has been carried out imperfectly. Another operation used is grey-level dilation with a rectangular structuring element. This mostly reflects the imaging capabilities of the AFM, which tends to expand deep edges. The third operation displayed in Figure 7.11 is a flip in image contrast. Initially, a random value between -1 and 1 is assigned to each object intensity. The background level is zero. This produces objects which are of higher or lower intensity than the background level. For generating a series of images, object intensities are multiplied with a varying contrast factor, which is also of range -1 to 1. Before analysis, each image is mapped to the positive range of image intensities. Images with contrast factor < 0 will show inverse contrast. This

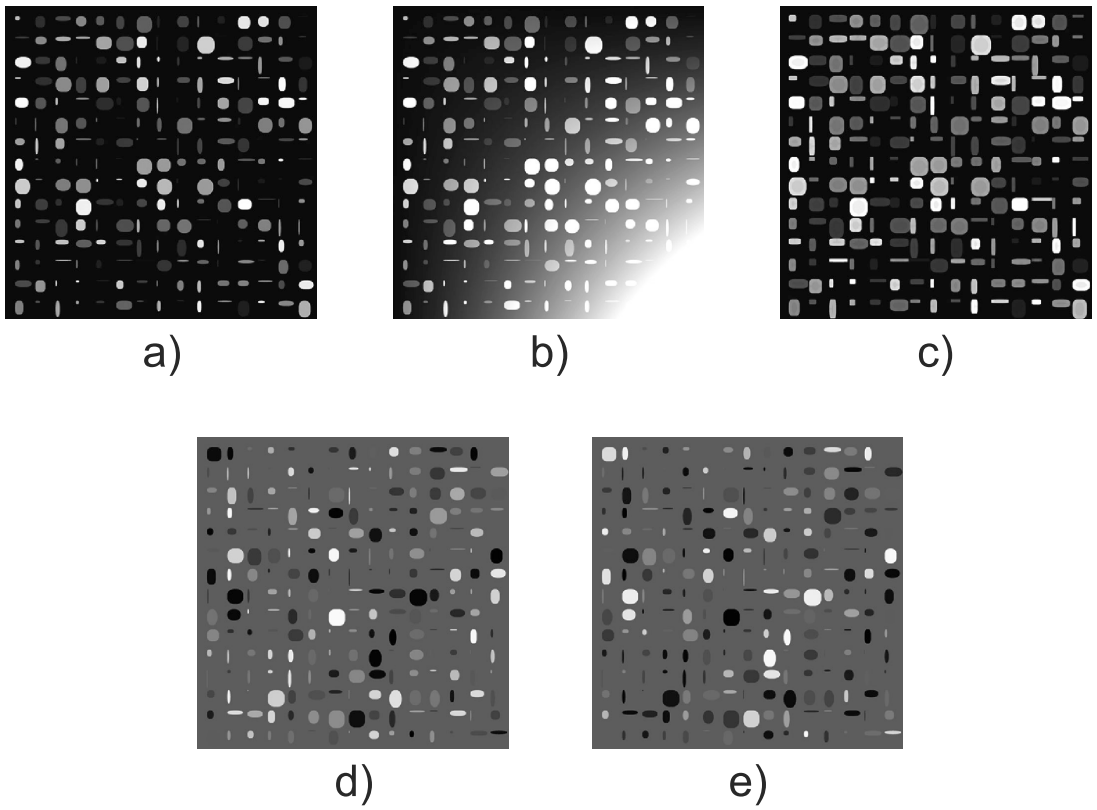


Figure 7.11: Synthetic image material for studying multiple effects. In (b), an intensity ramp has been added to image (a), simulating nonflat background regions. Image (c) shows a dilated version of (a), simulating the imaging characteristics of the AFM. Between (d) and (e), the contrast has been inverted.

means that gradient directions from object to the background are flipped by 180°. Such a behavior can be expected if special imaging modes of AFM or SEM are used and no monotonic mapping between AFM and SEM scan intensities can be assumed.

The performance results for the intensity ramp and the dilation experiment can be seen in Figure 7.12. Starting with a ramp intensity of 0, the matching score is 100 %. In this case, identical images are registered. As the ramp intensity increases, the matching score decreases moderately, almost monotonically. The dilation experiment shows a similar behavior, as dilation in x and y direction are increased. This behavior can be classified as beneficial, because the registration procedure is not over-sensitive to variations in background intensity or shapes of

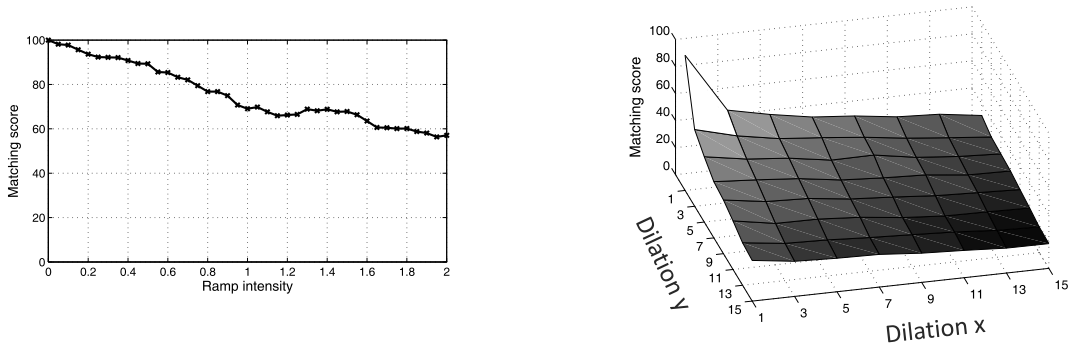


Figure 7.12: Synthetic image matching performance in terms of matching score. The left plot shows falling of the matching score as the gradient of the intensity ramp is increased. The right plot shows falling of the matching score as dilation in both directions is increased. SIFT detector and descriptor have been used.

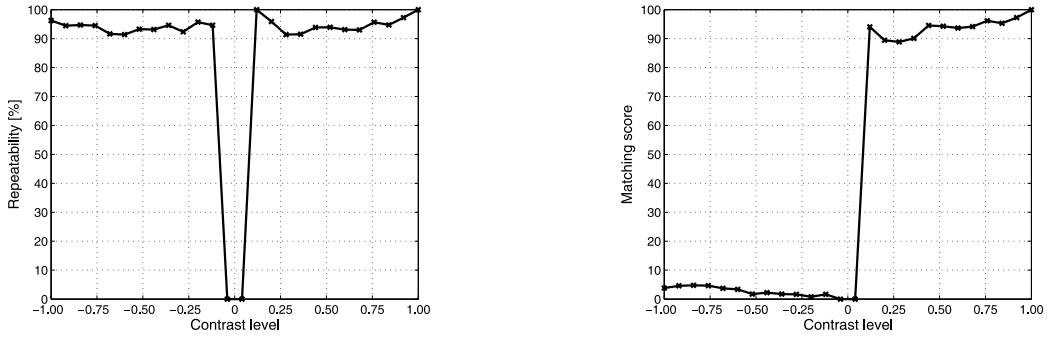


Figure 7.13: Results for the registration experiment with inverse contrast. A contrast level of 1 corresponds to identical images. -1 corresponds to inverse contrast. 0 corresponds to no contrast. The detector repeatability falls to zero where the image contrast vanishes. It is high for all other contrast strength. The matching score is low for all negative contrast values. SIFT detector and descriptor have been used.

corresponding objects.

On the other hand, Figure 7.13 depicts a general limitation of feature descriptors which make use of gradient directions. For a contrast level of 1, the images are identical and best repeatability and matching scores are obtained. As the level of contrast is decreased, both values are degraded only slightly. This shows

the tolerance of both, detector and descriptor to changes in contrast level. For a contrast level of 0, registration fails because no image information is available. For negative contrast values, two effects can be observed. High repeatability values are obtained. The detector is tolerant to inverse contrast. On the other hand, matching mostly fails. This is due to failure of the orientation estimation step of the description scheme. Descriptors are now computed in a coordinate system, which is flipped by 180° with respect to the base image features.

Additional application cases

The samples used in the systematical performance analysis (Figure 7.4) represent multiple realistic application scenarios for AFM and SEM image registration. Additionally, two samples have been prepared in order to further demonstrate capabilities and limitations of multimodal registration based on local features. These samples can be seen in Figure 7.14. They have been created as especially difficult cases, by FIB-milling patterns into the surface of a silicon substrate.

The upper row shows a sample which is composed of rectangles of varying side lengths. Starting from the upper left side, rectangle sidelengths are increased column and row-wise until the maximum size is reached in the lower right corner. This setup is especially difficult to register, because the only distinct local information about a rectangle is the size. Due to the different imaging characteristics of AFM and SEM, the rectangles can appear slightly resized. On the other hand, the exact size is the only way of distinguishing between the rectangles. A second difficulty is the repetition of patterns. If looking at any rectangular subpattern of the sample, an identical pattern with slightly smaller rectangles can be observed by moving the subpatch diagonally to the top left direction. This way, the sample produces not only a high number of incorrect matches but also incorrect geometrically consistent subsets of matches. Nevertheless, using the SIFT feature detector and descriptor, RANSAC manages to identify the correct subset of matches.

In the lower row of Figure 7.14, the second challenging sample can be seen. It is composed out of structures, which reproduce under changes of scale. Those are corners and junctions. A problem with these structures occurring in an isolated location is the functionality of the automatic scale selection. No stable peaks in the detector response can be expected around these structures. As a result, the information extracted by the descriptor is mostly modality-specific imaging artifact at an arbitrary scale. The descriptor hardly reproduces between AFM and SEM. SIFT-based registration of these scans leads to a high number of incorrect matches and failure of the RANSAC procedure.

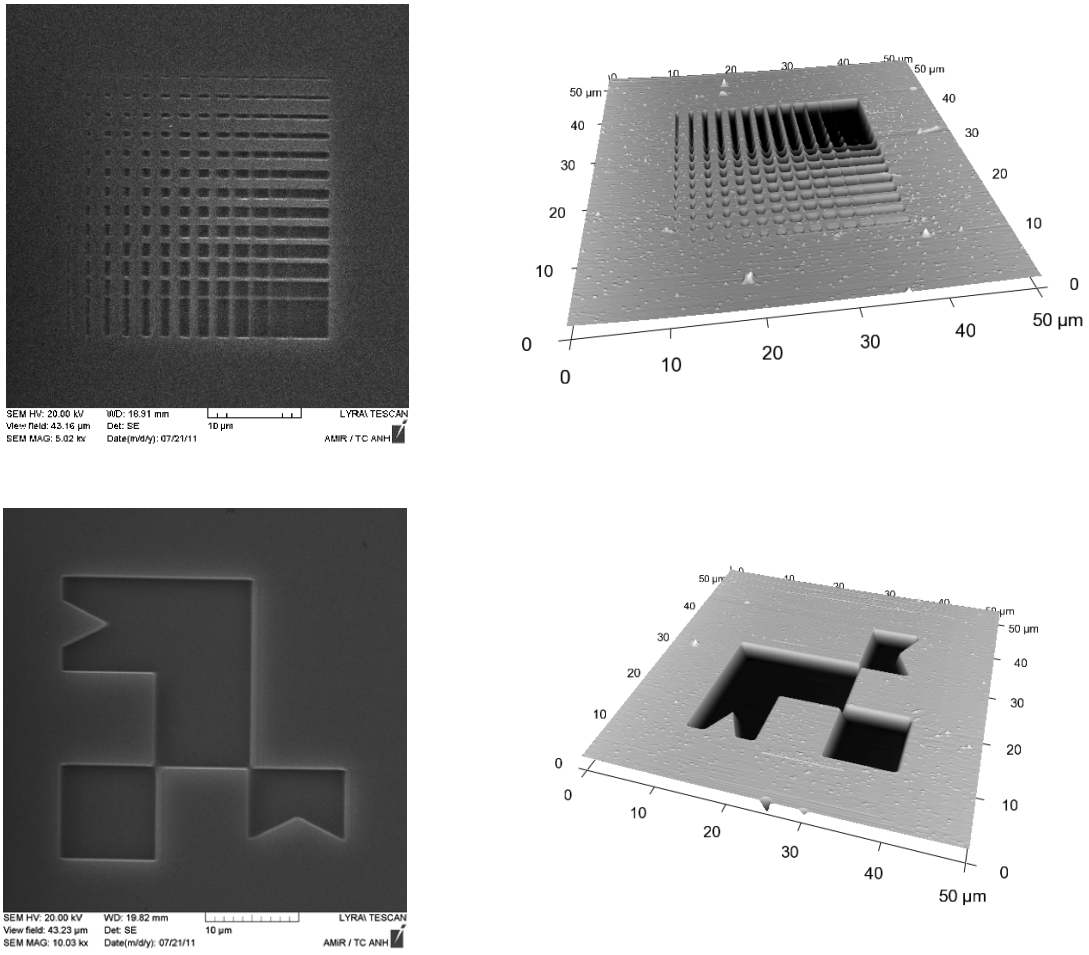


Figure 7.14: Samples imposing special difficulty on the feature-based registration step. The SEM scans (left side) and AFM scans (right side) show a lack of distinct image detail. In the upper sample, the rectangular structures vary only slightly in size. In the lower sample, most features cannot be located well over scale.

7.3.2 Combined detectors

In Section 5.2, a strategy for combining multiple region detectors has been presented. It is based on maximizing the repeatability of the detector responses in the different imaging modalities, while at the same time minimizing the similarity of the detectors. This section shows, how the proposed selection strategy can be applied in order to compose more robust region detectors.

The functionality of the proposed scheme can be demonstrated by applying it

to the regions detected on the DVD sample (see Figure 7.7). In the beginning, a list of candidate detectors is selected, which are the highest-ranked detectors of the previous section:

- Hessian-Laplace detector (Hes)
- Harris-Laplace detector (Har)
- SIFT feature detector (DoG)
- SURF feature detector (Surf)

In the next step, the inter-modality repeatabilities are determined:

$$Rep(AFM\|Hes, SEM\|Hes) = 84.48\%$$
$$Rep(AFM\|Har, SEM\|Har) = 81.77\%$$
$$Rep(AFM\|DoG, SEM\|DoG) = 72.20\%$$
$$Rep(AFM\|Surf, SEM\|Surf) = 78.26\%$$

Due to the lowest rank of the DoG detector, it is removed from the procedure. Next, all 2-combinations of the remaining detectors are identified. Those are Hes-Har, Hes-Surf and Har-Surf. The selection criterion is the detector dissimilarity. It is checked by computing the inter-detector repeatabilities for all detector combinations:

$$Rep(AFM\|Hes, AFM\|Har) = 83.43\%$$
$$Rep(SEM\|Hes, SEM\|Har) = 93.08\%$$

$$Rep(AFM\|Hes, AFM\|Surf) = 44.32\%$$
$$Rep(SEM\|Hes, SEM\|Surf) = 20.95\%$$

$$Rep(AFM\|Har, AFM\|Surf) = 43.18\%$$
$$Rep(SEM\|Har, SEM\|Surf) = 41.77\%$$

This means that there is a high degree of similarity between the Hessian-Laplace and Harris-Laplace detector. Using the combined detector Hes-Har will bring only a limited benefit. The remaining combinations are Hes-Surf and Har-Surf. It has been pointed out in Section 5.2 that the repeatability criterion is not sufficient in order to assure region informativeness. This criterion is tested

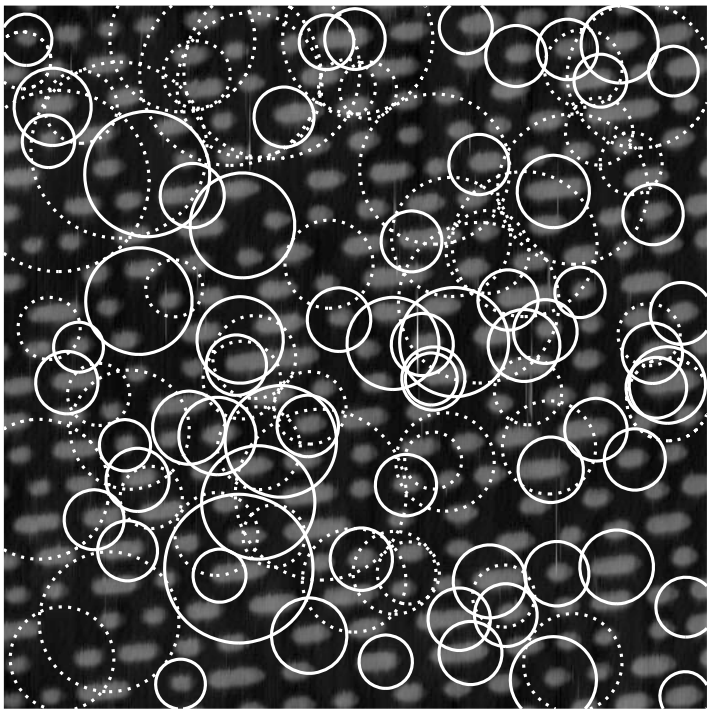


Figure 7.15: Selection of fused detector responses of Hessian-Laplace and SURF region detector, applied on the AFM scan of the DVD sample surface. The regions marked by solid lines originate from the Hessian-Laplace detector, while the dashed lines mark regions obtained using the SURF detector. The field of view width is 15µm.

by performing a matching experiment. Because matching can only be carried out between similar feature descriptors, the SIFT descriptor is chosen for the matching experiment. Following Equation 5.15, the fused detector responses are obtained. The Hes-Surf detector leads to a matching score of 20.6 and a total number of correct correspondences of 1635. The Har-Surf detector obtains a matching score of 10.8 and a total number of 1078 correct correspondences. Hence, Hes-Surf is the preferred detector combination. Hes-Surf regions are displayed in Figure 7.15.

For the task of AFM and SEM image registration, the fused Hes-Surf detector is not as such superior to the single detectors. The benefit is in the capability of better handling variations in image contents. Figure 7.16 shows corresponding AFM and SEM scans of a FIB-milled structure with Hessian-Laplace and SURF regions indicated. The Hessian-Laplace detector faces problems with reproducing responses in scale-space. Inter-modality repeatabilities are 10.3% (Hessian-

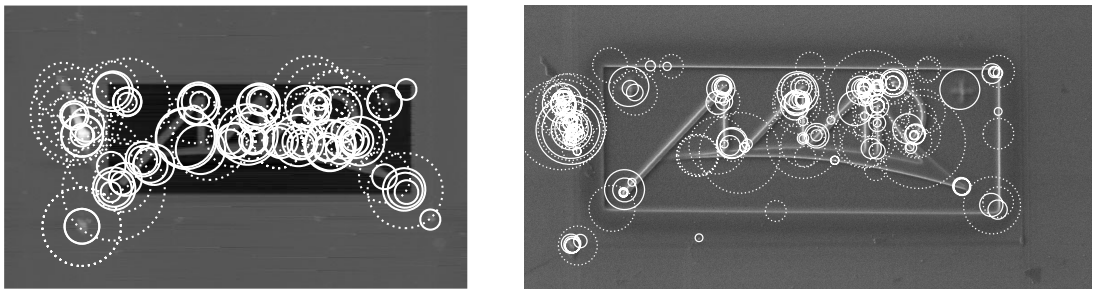


Figure 7.16: Selection of fused detector responses of Hessian-Laplace (solid lines) and SURF region detector (dashed lines), applied on the AFM scan (left) and SEM scan (right) of a FIB-milled structure. The FIB-milled region has a size of 10µm×3.5µm.

Laplace) and 38.4% (SURF). The fused Hes-Surf detector obtains 34.6% and thus falls between the single detector values. When repeating the matching experiment, a total number of correct matches of 8 (Hessian-Laplace), 20 (SURF) and 38 (Hes-Surf) are obtained. This means that correct correspondences have been established between Hessian-Laplace and SURF regions or in other words, the detectors complement each other.

Fusion of region detectors has some similarities with the selection of an appropriate classifier in SPR. For a set of input data, an optimal region detector can clearly be identified. The question is how it will perform under variations of the input data. It has been shown that the highest-ranked detector (Hessian-Laplace) on the DVD sample faces strong difficulties in reproducing responses on the FIB-milled structure. In this sense, the fused Hes-Surf detector is more general as it enables successful registration in a broader range of target samples. A dissimilarity between the selection of a region detector and the selection of an appropriate classifier in SPR is the possibility of testing the result during application time. While the classifier output in a classification task must be accepted, the geometric consistency of a registration result can be tested. This enables the automatic profile selection introduced in Section 5.2.2, which is a type of trial and error method.

7.3.3 The new feature matching strategies

The performance of the new feature matching strategy has been evaluated using the SIFT and SURF feature detector and descriptor packages. This selection is representative because most other descriptors use the SIFT procedure for computing the dominant feature orientation. Therefore, all results concerning

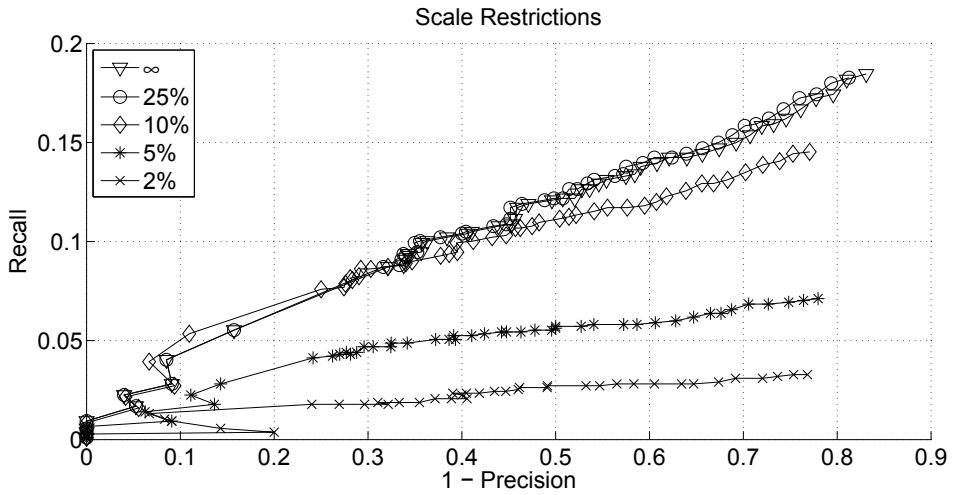


Figure 7.17: Performance of the scale ratio restriction approach, depicted by the recall 1-precision plot for SIFT features computed on the DVD sample. A moderate restriction allowing 25% of deviation from the ground truth reproduction scale slightly improves the performance in comparison to the unrestricted case (allowing ∞ deviation). By further increasing the restrictions, the performance decreases, because correct matches with imprecisely detected scale are incorrectly rejected then.

matching restrictions on feature orientation are directly transferable. In addition to the regular SURF descriptor (SURF64), also the extended SURF descriptor (SURF128) has been used for comparison. For the experiments it has been assumed that the registration task has to be carried out with a varying level of prior knowledge. Without applying the postmatching step by imposing matching restrictions, the application case is similar to the experiments presented in Section 7.3.1. Additionally, the transformation model components scale **S**, rotation **R** or both can be given prior to the registration step. Initially, both types of matching restrictions have been studied separately. The effect of the window size for matching restrictions is critical for the success of the postmatching step.

Figure 7.17 studies the effect of restricting differences in the scale ratio during the matching procedure. SIFT-features computed on the DVD sample have been used for this experiment. Although the 1-precision values can be improved by this postmatching step, the recall rate decreases noticably. An improvement of the overall performance is only observed for a moderate restriction of scale ratios by allowing 25% of deviation from the ground truth scale ratio. On the other

Figure 7.18: Performance of orientation difference restriction approach, depicted by the recall 1-precision plot for SIFT features computed on the DVD sample. Even low restrictions such as a deviation of $\pi/2$ from the ground truth rotation result in a significant gain in performance. For very high restrictions such as a deviation of $\pi/32$ the performance falls below the restriction-free case.

hand, introducing restrictions on the allowed difference in orientation leads to a strong gain of performance. Figure 7.18 shows the recall vs. 1-precision plot for a different amount of deviation from the ground truth difference in angle. The level of π corresponds to the absence of any restrictions. At the very restrict level of $\pi/32$ the performance falls below the unrestricted case. It has to be noted that the ground truth rotation $\mathbf{R}$ is accurate in this case. Using a biased estimate of $\mathbf{R}$ the performance gain caused by orientation restrictions will be smaller.

In the following, a moderate restriction level has been chosen to study the effect of the new feature matching strategy in more detail for each sample. The overall matching performance is compared in Figures 7.19 - 7.23. Experiments have been carried out using regions detected by the SIFT and SURF detector. Therefore, the number of ground truth correspondences used to compute the recall rate is different for SIFT and SURF matching. Generally, the SIFT algorithm outperforms the SURF algorithm. The performance is best for the CD, DVD and nanocluster matching. These are the samples with blob- and dash-shaped structures. The FIB-milled pattern and the gold on silicon test pattern exhibit mostly edge- or corner-like structure and the matching performance is comparably low in both cases. For the FIB-milled pattern, the 1-precision can be

improved significantly by using matching restrictions. In contrast, the improvement is moderate for the gold on silicon test pattern sample. In this case, the algorithm performance suffers from the highly ambiguous image structure. No application could be identified where the extended SURF descriptor is superior to the regular descriptor.

Both algorithms benefit from the reuse of computation results between descriptor and detector and are among the fastest and at the same time most accurate algorithms. Although the SIFT algorithm is superior to SURF under many aspects, SURF clearly shows shorter execution times. Table 7.1 shows the execution times measured on an Intel Core i5-750 with 4GB RAM. Only the regular SIFT and SURF algorithm have been evaluated under the aspect of execution time. The values strongly depend on the scene contents and also the scan size. In average, the SIFT-based registration takes 3.21 times more computation time. The values have been split up into feature extraction (detection and description), initial matching and refinement (RANSAC). The computational burden of matching restrictions is negligible and has not been taken into account.

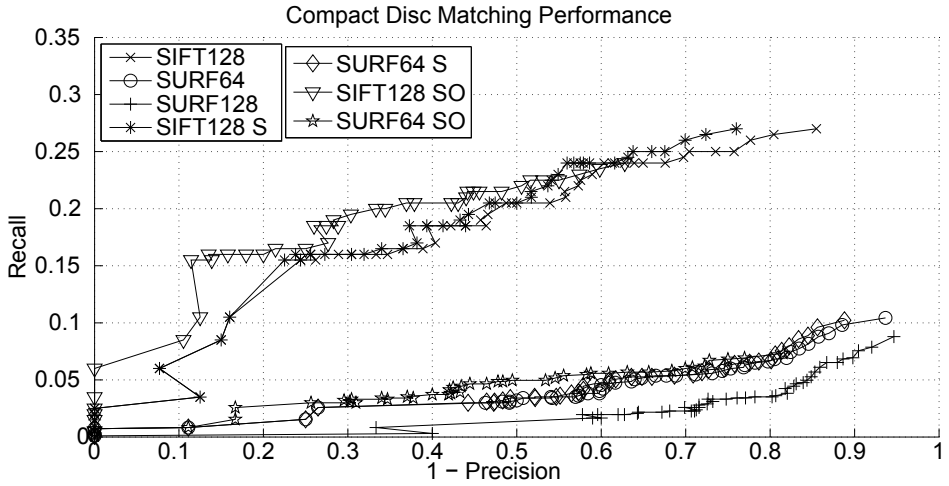


Figure 7.19: Matching performance for the CD sample and multiple detector/descriptor combinations and matching restrictions on scale ratio (S, 25%) and orientation (O, $\pi/16$). The SIFT algorithm clearly outperforms the SURF algorithm. The regular-sized SURF descriptor (64) outperforms the extended (128) SURF descriptor.

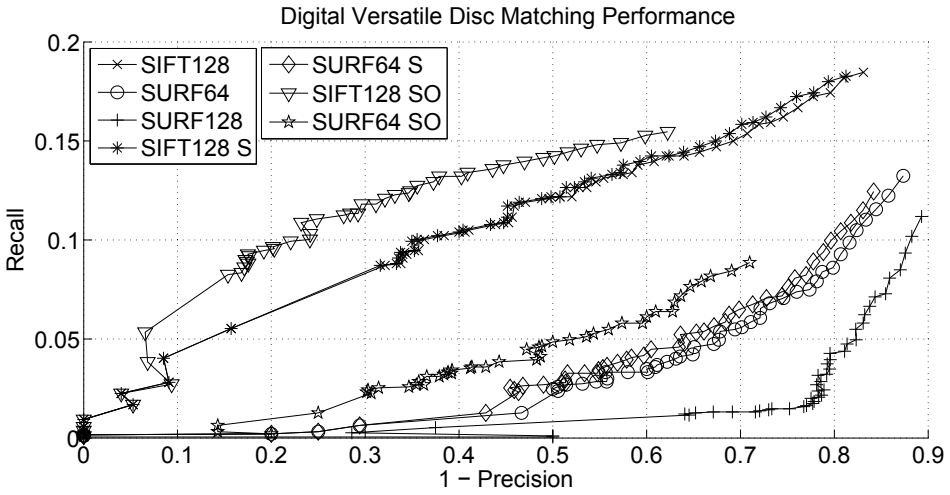


Figure 7.20: Matching performance for the DVD sample and multiple detector/descriptor combinations and matching restrictions on scale ratio (S, 25%) and orientation (O, $\pi/16$).

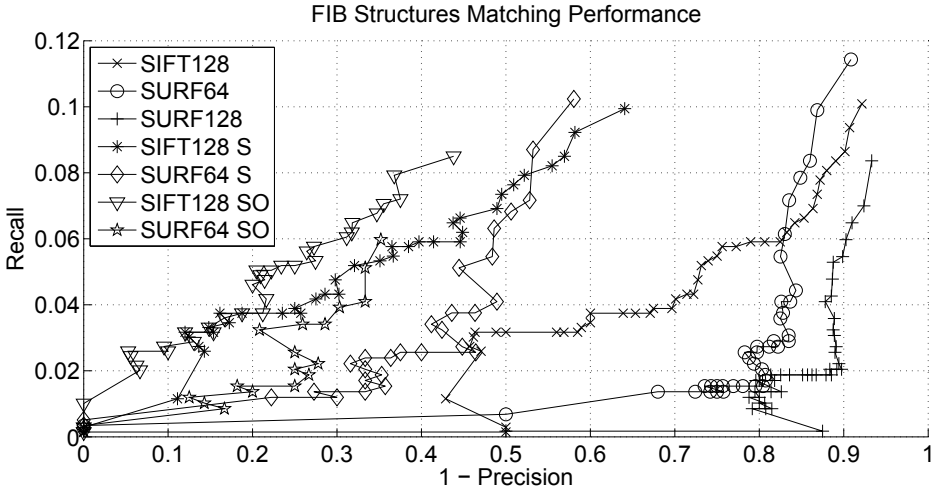


Figure 7.21: Matching performance for the FIB-milled pattern and multiple detector/descriptor combinations and matching restrictions on scale ratio (S, 25%) and orientation (O, $\pi/16$).

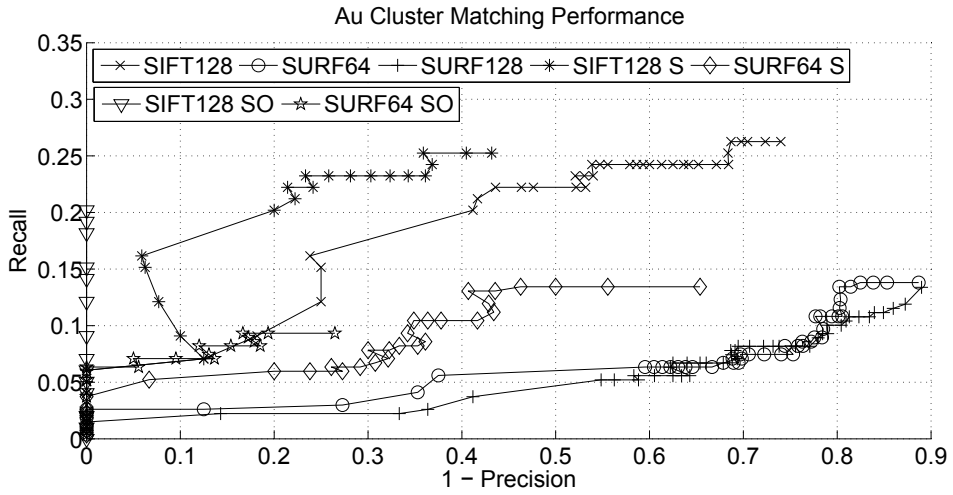


Figure 7.22: Matching performance for the nanocluster sample and multiple detector/descriptor combinations and matching restrictions on scale ratio (S, 25%) and orientation (O, $\pi/16$).

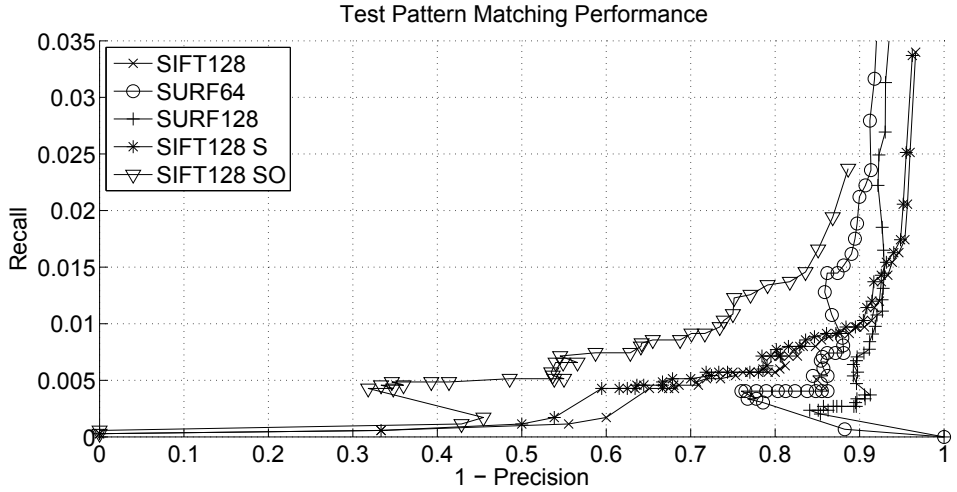


Figure 7.23: Matching performance for the gold on silicon test pattern and multiple detector/descriptor combinations and matching restrictions on scale ratio (S, 25%) and orientation (O, $\pi/16$). Matching restrictions on SURF features bring moderate gains in performance but have been left out for clarity reasons.

Table 7.1: Average computation times of the different setups for SIFT and SURF feature extraction, matching and refinement step.

Setup	SIFT extr.	SIFT match	SIFT refine	SURF extr.	SURF match	SURF refine
CD	5.33s	0.11s	0.14s	1.81s	0.60s	0.14s
DVD	7.55s	1.07s	0.11s	3.53s	1.05s	0.17s
FIB	22.3s	1.21s	0.44s	2.66s	0.08s	0.49s
Nanocluster	4.84s	0.02s	1.03s	0.69s	0.01s	0.38s
Au-Pattern	8.71s	5.61s	1.08s	1.19s	0.44s	5.30s

7.3.4 Area-based refinement step

Once, feature-based registration including optional improvement steps is completed, an initial set of transformation model parameters is available. The aim of the area-based registration step is to further improve the accuracy of the registration. Area-based registration optimizes a similarity measure between AFM and SEM scan, which is a type of quality criterion. However, since this quality criterion is objective of the optimization procedure it cannot be used for judging the registration performance at the same time. Also the question would arise, which similarity measure to select as the quality criterion. Instead, an external ground truth transformation $\underline{T}(x, y)$ is used for determining the accuracy of a set of transformation model parameters. The ground truth transformation is obtained by manually labeling landmark points in corresponding AFM and SEM scans.

The accuracy of the area-based registration is examined by comparing the obtained transformation with the ground truth transformation. A direct comparison of the transformation parameters in terms of absolute differences does not provide a good measure of the expected error, due to the different nature of the single model parameters. Another problem is the dependency between the alignment error and the scan area: In the center of rotation and scale, errors in $\mathbf{R}$ and $\mathbf{S}$ have no impact. A better performance criterion is described in [14]. The displacement errors between an estimated transformation $\hat{T}(x, y)$ and the ground truth transformation $\underline{T}(x, y)$ can be computed from:

$$\delta_{\mathbf{1}} = \underline{T}(x_1, y_1) - \hat{T}(x_1, y_1), \tag{7.7}$$

$$\delta_{\mathbf{2}} = \underline{T}^{-1}(x_2, y_2) - \hat{T}^{-1}(x_2, y_2). \tag{7.8}$$

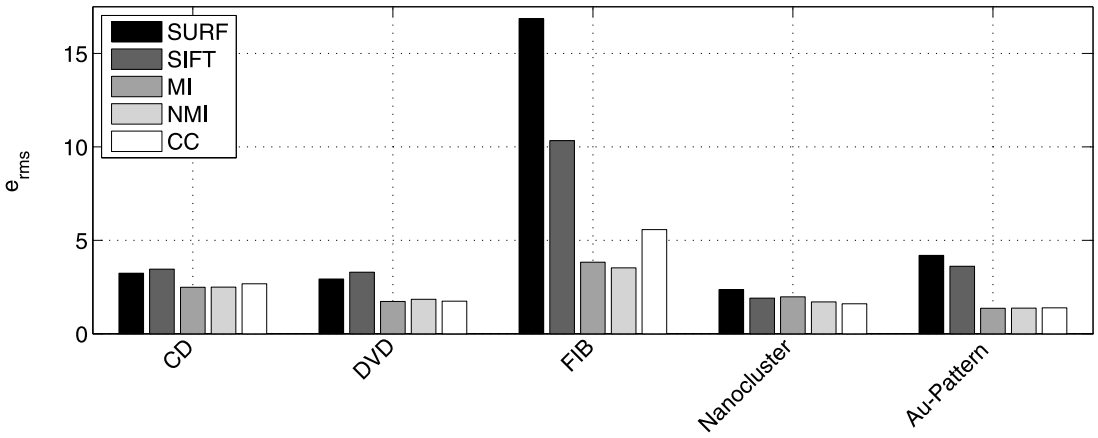


Figure 7.24: Performance results of the area-based registration in terms of e_{rms}. Values for SIFT and SURF show the remaining error after feature-based registration. MI, NMI and CC are the errors after area-based refinement of the registration.

For symmetry reasons, the displacement error is computed in forward ($\delta_\mathbf{1}$) and inverse ($\delta_\mathbf{2}$) direction. Over a region of interest Ω, the vector field of displacement errors is averaged, which results in the root mean square transfer error e_{rms} of $\hat{T}(x, y)$:

$$e_{rms} = \sqrt{\frac{1}{2\Omega} \int_\Omega \|\delta_\mathbf{1}\|^2 + \|\delta_\mathbf{2}\|^2 d\Omega} \quad . \tag{7.9}$$

Ω is selected to be the entire overlap area between AFM and SEM scan. For the experiments, the transformation parameters obtained after feature-based registration using SIFT detector and descriptor with additional RANSAC refinement have been used as initialization. However, convergence towards identical optima has been observed using the output of different detector/descriptor combinations as initialization. The results of the area-based registration refinement can be seen in Figure 7.24. For comparison, the normalized cross-correlation (see Equation 2.2) has also been used as optimization criterion.

The results show that the area-based registration step helps to increase the registration accuracy. This gain in accuracy ranges from a moderate improvement to a substantial improvement. Nevertheless, all procedures leave a residual registration error. The difference between MI and NMI is minimal for all scenarios tested. However, for the nanocluster sample, MI brings a marginal degradation while NMI leads to a moderate improvement of the SIFT registration result. The performance of the normalized cross-correlation measure is surprisingly high in

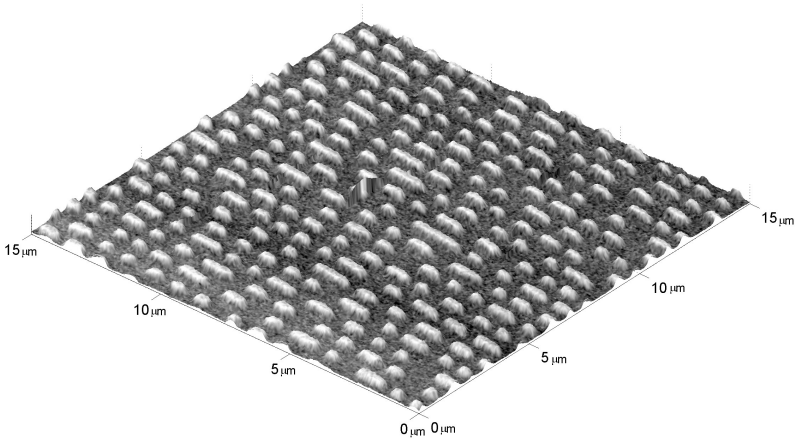


Figure 7.25: Fusion result for the DVD sample. The AFM scan has been rendered as a three-dimensional surface and is textured by the correctly registered SEM scan.

most cases. For the FIB-milled pattern on the other hand, the MI and NMI performance is clearly better.

7.3.5 Output of the proposed registration scheme

The final output of the system is the fused representation of AFM and SEM scan. It enables the inspection of corresponding surface scans in a single view. Figure 7.25 shows a three-dimensional view of the DVD surface, textured with the correctly registered SEM scan of the corresponding area. In Figure 7.26, a similar results is shown for the FIB-milled pattern. A high-intensity SE signal can be seen at the edges of the FIB-milled structure, which are facing the direction of the SEM electron detector. Figures 7.25 and 7.26 show the fusion results of standard SE SEM scan and intermittent contact mode AFM scan. Nevertheless, the same methods can also be used in order to create fused views of special imaging modes for material inspection.

7.4 Conclusion

The proposed procedure has been implemented and tested on a variety of different sample and equipment combinations. In summary, the registration succeeded in all application scenarios. A major benefit of the proposed method over the

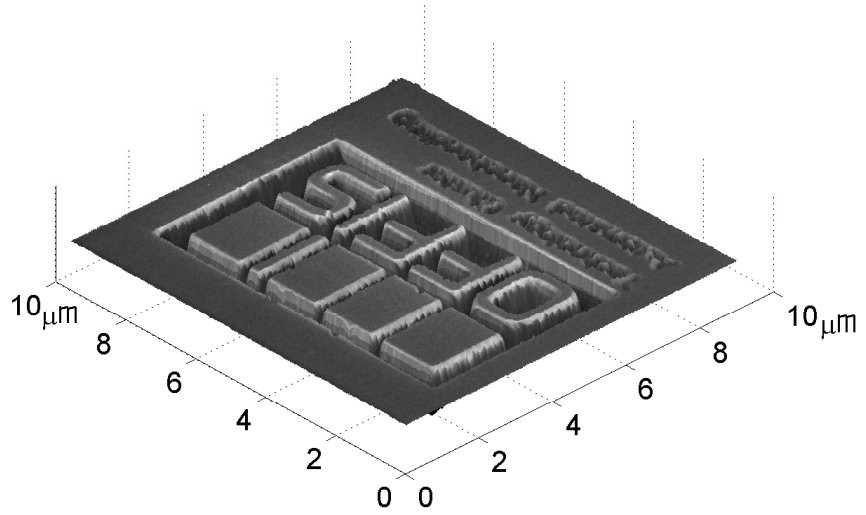


Figure 7.26: Fusion result for the FIB-milled structure. At the edges of the structure which are facing the SEM's electron detector, a high-intensity SE signal is visible.

registration scheme described in [94] is the total absence of manual working steps and a minimum number of parameters. In the application scenarios presented here, the feature-based registration faces two imaging modalities with strongly different imaging characteristics and artifacts. These include different intensity levels, morphological changes, shadows, AFM cantilever control-related artifacts and scanner noise. Under these conditions, generally weaker matching results must be expected as compared to those reported by [76] or [68] for natural photography. Nevertheless, the correspondence analysis of pairs of scans is still successful, despite the lower absolute number of correct feature matches.

All feature detectors and descriptors included in the investigation have been designed mainly for the registration of natural photographs or video, where the main challenge is to handle changes in illumination or viewpoint and occlusions. It seems that for the higher-ranked detectors and descriptors, these requirements correspond well with the requirements of AFM and SEM image registration. The aspect of different intensity levels between AFM and SEM is compensated by the normalization of feature vectors. A reduction of the level of sensor noise is integrated in the scale-space approach. Sensor-specific artifacts such as local charging in the SEM scan show the same effect as occlusions in natural photography: Features cannot be matches in this local neighborhood, but the performance in

the remaining image area is stable.

It has been shown, that incorporating prior knowledge on the difference in scan orientation or magnification ratio helps to improve the matching performance significantly. Also, the performance of the detectors and descriptors for different target structures has been analyzed extensively and the possibility of combining detectors has been pointed out. It must be decided in the actual application case, how many of these additions are implemented and how much effort is spent on identifying optimal detectors and descriptors. Although not superior under all aspects, the SIFT and SURF detector and descriptor bundles perform well in many applications cases. Additionally, highly optimized implementations are available, making SIFT and SURF superior to most other algorithms under the aspect of execution time.

A limitation of most description schemes in the presence of multiple forms of contrast is the computation of the local gradients or wavelet responses respectively (cf. Figures 5.9 and 5.10). The determination of the dominant feature orientation and local gradient orientations assume a nearly monotonically increasing mapping between AFM and SEM intensity values and therefore identical gradient orientations in both modalities. Violations of this assumption lead to a total failure of the matching procedure, as the feature orientation and therefore the assignment of local gradients or wavelet responses are incorrect. This limitation can be overcome by adding descriptor copies with inverse orientation to the feature set. However, the negative effect in terms of 1-precision has not been studied yet.

Due to shadowing and morphological differences between the modalities, the feature detectors do not reproduce exact feature locations for an image pair. The area-based refinement of the registration result brings an improvement but still leaves a significant transfer error. This can potentially be compensated for by implementing alternative similarity measures or means of regularization. However, all methods considered here work directly on the scan data and it is possible that the residual error cannot be removed by means of a direct method. An alternative approach is to model the process of AFM and SEM image formation and to compensate for sensor specific artifacts in order to create an alike image pair for registration. In trade for the potential gain in registration accuracy, a multitude of assumptions have to be made about the imaging process and equipment parameters.

8 Summary and outlook

8.1 Summary

This thesis presented two new procedures which help to increase the level of automation in robotic tasks on the micro- and nanoscale. Image analysis can be regarded as one of the most important forms of sensory feedback in micro- and nanorobotics. Due to the small dimensions of the target objects and structures, the image material is acquired using microscopes and miniature cameras. In the past, visual feedback in micro- and nanorobotics has been used mainly for continuously following the movement of individual objects and also for depth estimation. This helped to increase the level of automation but nevertheless left initialization and labeling steps for manual user interaction.

Two tasks have been identified which can help to further increase the level of automation. The first task is the localization and classification of micro- and nanoscale objects from a new image scene. Examples of such objects are tools or workpieces for characterization or assembly operations. The second task is the determination of the spatial relationship between multiple images, originating from heterogeneous image sensors. Such a procedure is also referred to as image registration and allows to automatically acquire and fuse micrographs of different imaging modalities. The fused representation benefits from complementary imaging characteristics, providing a more thorough view on specimen geometry and material properties. Both tasks are different by nature and therefore demand for distinct solutions.

The first task has been faced by developing a new system for micro- and nanoscale object classification. Instead of extending earlier object tracking procedures to the task of object classification, the new system is based on statistical pattern recognition. This brings the advantage that decision rules are learned automatically. Also, the classification rules are not limited to geometric object properties. The system is composed out of four processing steps. It starts with the image acquisition and preprocessing step which provides a unique interface to the different type of microscopes used. In the next step, objects are segmented from the image background and a list of connected objects is established. A meaningful descriptor is extracted from each object in the next step, taking into

account the characteristic object properties. In the last step, the class membership of each object is determined with the help of a SVM classifier. The system is designed for being highly integrated into setups for automated micro- and nanorobotic operations.

Validation of the proposed system for micro- and nanoscale object classification has been carried out in three different application scenarios and imaging modalities. The first application is the localization of carbon nanotubes on the surface of a silicon wafer. Due to the dimensions of the nanotubes, the SEM is used as image sensor. Other objects introduced by contamination of the vacuum must be rejected. The second application is a quality check for biological cells, carried out using an optical microscope. Defect cells are recognized automatically and excluded from further processing steps. The last application is the detection of magnetic particles using MRI. Magnetic objects can potentially be used in medical interventions by carrying drugs or guiding catheter operations. The proposed system succeeds in all three applications for which the state-of-the-art methods can hardly be adapted for.

The second task targeted in this thesis has been faced with the help of a newly developed scheme for multimodal image registration. Generally, two types of registration schemes are distinguished. Feature-based registration establishes correspondences between image features and derives parameters for a coordinate transformation model. Area-based registration varies the model parameters and optimizes a similarity measure between the images. Both strategies typically have differences in accuracy, execution time and properties of convergence. The proposed scheme initially performs feature-based registration and refines the resulting model parameters in an area-based registration step. Thereby, the benefits of both methods are combined. Two optional strategies for improving the feature-based registration step are introduced. It is shown, how multiple feature detectors can be combined in order to avoid too much specialization on a particular kind of image contents. The second improvement integrates prior knowledge about the scene alignment into the feature matching procedure.

For validation of the proposed registration scheme, surface scans obtained from the SEM and AFM have been registered. The experiments included multiple specimens with strongly different surface structures. Additionally, artificial image material has been generated. A total number of eight feature detectors and nine feature descriptors have been tested. Area-based registration has been carried out using three different similarity measures. The performance data clearly show a preferred subset of feature detection and description methods, although none is superior for all applications. Area-based registration further improves the registration accuracy in all applications, thereby justifies the multi-stage registration scheme. Combining multiple feature-detectors can lead to increased

performance under varying image contents. Also prior knowledge about the transformation model parameters should be incorporated into the feature-based registration step, leading to a considerable gain in performance.

The goals stated in Section 1.1 have been attained entirely. The new methods presented in this thesis have been applied successfully in the context of multiple research projects and helped to increase the level of automation significantly.

8.2 Outlook

Both approaches presented in this thesis have been tested in multiple scenarios and will support automation in future applications of micro- and nanorobotics as well. Nevertheless, some limitations have been indicated already and both approaches leave opportunities for future research.

The system for micro- and nanoscale object classification provides sensor data to the control system and works side-by-side with other sensors. One possibility of improving the classification capabilities of the system is to include additional sensor data as objects features. For instance, such object features could be provided by electrical or mechanical measurements and not be extracted from the image contents. From the perspective of the SVM, the origin of the features is transparent. Therefore, features extracted from images and further features can be used simultaneously. Another promising approach is to use object properties in order to assign unequal weights to each training sample [10]. Following this idea the training procedure is told, which samples are more or less important for determining the decision boundary.

Until now, the new strategy for multimodal image registration has been applied to scans obtained from the AFM and SEM directly. It has been pointed out by experiments with synthetic image data that the registration compensates for differences in the imaging characteristics of both modalities to a certain degree. However, knowledge about the imaging characteristics can also be used in order to create a more similar pair of scans in a preprocessing step. This approach could help to further decrease the residual error observed after registration. On the other hand, additional assumptions and knowledge about the process of image acquisition must be incorporated, changing the registration scheme from a general solution to a more application-specific procedure.

In the future, the proposed registration scheme should be applied in other areas of microscopy. For applications with limited requirements to optical resolution, optical microscopy can be included. Also, alternative imaging modes of AFM and SEM can be incorporated in order to obtain a more thorough view on material properties such as magnetic or electrical properties. As an example, Figure 8.1

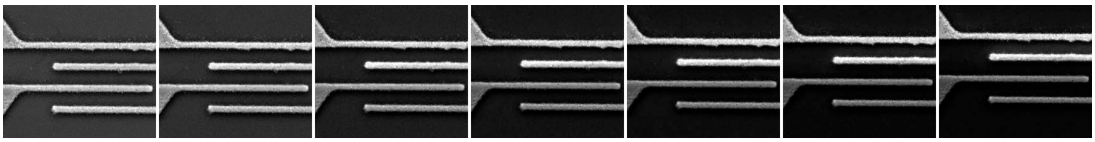


Figure 8.1: Series of SEM scans of electrodes (field of view width = 28µm). The voltage between upper and lower two electrodes is rising from $0V$ (left) to $30V$ (right).

shows an SEM scan of four electrodes with different voltages applied. However, functionality of the proposed registration scheme under special imaging modes has not been studied extensively. It has been shown, that the feature-based registration step is robust to changes in contrast level but sensitive to flips in gradient direction. The occurrence of such effects and the construction of robust features is left as a topic for future investigations.

Bibliography

[1] J.J. Abbott, Z. Nagy, F. Beyeler, and B.J. Nelson. Robotics in the small, part I: Microrobotics. *IEEE Robotics and Automation Magazine*, 14(3):92–103, 2007.

[2] P. Avouris, J. Appenzeller, R. Martel, and S.J. Wind. Carbon nanotube electronics. *Proc. of the IEEE*, 91 (11):1772 – 1784, 2003.

[3] Y. Bastanlar, A. Temizel, and Y. Yardimci. Improved sift matching for image pairs with a scale difference. *IET Electronics Letters*, 46(5):346–348, 2010.

[4] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool. Speeded-up robust features (SURF). *Comput. Vis. Image Underst.*, 110(3):346–359, 2008.

[5] J.S. Beis and D.G. Lowe. Shape indexing using approximate nearest-neighbour search in high-dimensional spaces. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pages 1000–1006, 1997.

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

[7] Y. Bentoutou, N. Taleb, K. Kpalma, and J. Ronsin. An automatic image registration for applications in remote sensing. *IEEE Transactions on In Geoscience and Remote Sensing*, 43(9):2127–2137, 2005.

[8] G. Binnig, C.F. Quate, and C. Gerber. Atomic force microscope. *Phys. Rev. Letters*, 56 (9):930–933, 1986.

[9] L. Bonetta. Flow cytometry smaller and better. *Nature methods*, 2(10):785–795, 2005.

[10] U. Brefeld, P. Geibel, and F. Wysotzki. Support vector machines with example dependent costs. In *Proceedings of the European Conference on Machine Learning*, 2003.

[11] M. Brown and D.G. Lowe. Automatic panoramic image stitching using invariant features. *Int. J. Comput. Vision*, 74(1):59–73, 2007.

[12] C.J.C. Burges. A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, 2:121–167, 1998.

[13] J. Canny. A computational approach to edge detection. *IEEE Trans. Pattern Anal. Mach. Intell.*, 8:679–698, November 1986.

[14] D. Capel. *Image Mosaicing and Super-Resolution*. Springer, London, UK, 2004.

[15] H.S. Cho. *Opto-mechatronic systems handbook: techniques and applications*. The mechanical engineering handbook series. CRC Press, 2002.

[16] T.F. Cootes and C.J.Taylor. Active shape models - smart snakes. In *In British Machine Vision Conference*, pages 266–275. Springer-Verlag, 1992.

[17] C. Dahmen and T. Wortmann. Antrieb und Verfolgung von magnetischen Partikeln im MRT. In *Bildverarbeitung für die Medizin 2011, Algorithmen - Systeme - Anwendungen*, 2011.

[18] C. Dahmen, T. Wortmann, and S. Fatikow. Olvis: A modular image processing software architecture and applications for micro- and nanohandling. In *Proc. of the Eighth IASTED Int. Conference on Visualization, Imaging and Image Processing (VIIP)*, pages 977–1000, 2008.

[19] C. Dahmen, T. Wortmann, and S. Fatikow. Magnetic resonance imaging of magnetic particles for targeted drug delivery. In *ASME 2010 First Global Congress on NanoEngineering for Medicine and Biology (NEMB2010)*, 2010.

[20] C. Dahmen, T. Wortmann, and S. Fatikow. Actuation and tracking of ferromagnetic objects using MRI. In *Proc. of Int. Symposium on Optomechatronic Technologies (ISOT)*, 2011.

[21] C. Dahmen, T. Wortmann, and S. Fatikow. MRI-based nanorobotics. In C. Mavroidis and A. Ferreira, editors, *NanoRobotics: Current Approaches and Techniques*. Springer, 2012.

[22] F. De Guio, H. Benoit-Cattin, and A. Davenel. Quantitative study of signal decay due to magnetic susceptibility interfaces: MRI simulations and experiments. In *29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS)*, 2007.

[23] S. DiMaio, D. Kacher, R. Ellis, G. Fichtinger, N. Hata, G. Zientara, L. Panych, R. Kikinis, and F. Jolesz. Needle artifact localization in 3T MR images. Technical report, Brigham and Women's Hospital, Harvard Medical School, Boston, USA., January 2006.

[24] L.X. Dong and B.J. Nelson. Robotics in the small, part II: Nanorobotics. *IEEE Robotics and Automation Magazine*, 14(3):111–121, 2007.

[25] R.O. Duda, P.E. Hart, and D.G. Stork. *Pattern Classification*. Wiley-Interscience Publication, 2000.

[26] R.R. Edelman, J.R. Hesselink, M.B. Zlatkin, and J.V. Crues, editors. *Clinical Magnetic Resonance Imaging, 3rd edition*. Saunders Elsevier, 2006.

[27] V. Eichhorn, S. Fatikow, T. Wortmann, C. Stolle, C. Edeler, D. Jasper, O. Sardan, P. Bøggild, G. Boetsch, C. Canales, and R. Clavel. NanoLab: A Nanorobotic System for Automated Pick-and-Place Handling and Characterization of CNTs. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA 2010)*, pages 1826–1831, 2009.

[28] V. Eichhorn, D. Jasper, C. Dahmen, and S. Fatikow. Automatisierte nanorobotische Mikro-Nano-Integration zur Herstellung prototypischer Mikrosysteme mit Nanostrukturen. In *Proceedings of the 2. GMM Workshop Mikro-Nano-Integration, Erfurt, Germany*, 2010.

[29] T. Ernstberger, G. Buchhorn, and G. Heidrich. Artifacts in spine magnetic resonance imaging due to different intervertebral test spacers: an in vitro evaluation of magnesium versus titanium and carbon-fiber-reinforced polymers as biomaterials. *Neuroradiology*, 2009.

[30] Y. Fan, Q. Chen, S. Arun-Kumar, A.D. Baczewski, N.V. Tram, V.M. Ayres, L. Udpa, and A.F. Rice. Registration of tapping and contact mode atomic force microscopy images. In *Sixth IEEE Conference on Nanotechnology (IEEE-NANO)*, volume 1, pages 193–196, 2006.

[31] S. Fatikow, T. Wortmann, M. Mikczinski, C. Dahmen, and C. Stolle. Towards automated robot based nanohandling. In *Proc. of IEEE Chinese Control and Decision Conference (CCDC)*, 2009.

[32] O. Felfoul, J.-B. Mathieu, G. Beaudoin, and S. Martel. Mr-tracking based on magnetic signature selective excitation. *IEEE Transactions on Medical Imaging*, 27 (1):28–35, 2008.

[33] B. Fischer and J. Modersitzki. Ill-posed medicine: an introduction to image registration. *Inverse Problems*, 24(3):034008, 2008.

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

[35] D. Folio, C. Dahmen, T. Wortmann, A.M. Zeeshan, K. Shou, S. Pane, B.J. Nelson, A. Ferreira, and S. Fatikow. MRI magnetic signature imaging, tracking and navigation for targeted micro/nano capsule therapeutics. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2011.

[36] D.A. Forsyth and J. Ponce. *Computer Vision: A Modern Approach*. Prentice Hall, 2002.

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

[38] T. Fuchs, U. Hassler, B. von Stackelberg, S. Hübner, J. Mühlbauer, and L. von Bernus. Multi-modality approaches for complex test requirements. In *International Symposium on NDT in Aerospace*, 2008.

[39] Carl Zeiss NTS GmbH. Correlative microscopy. Online source, accessed on October 24, 2011: *http://www.zeiss.de/corrmicforma*.

[40] D. Gnieser, C.G. Frase, H. Bosse, and R. Tutsch. Model-based correction of image distortion in scanning electron microscopy. In *Proc. of 9th Int. Symp. on Measurement Technology and Intelligent Instruments (ISMTII)*, volume 1, pages 1–147 – 1–151, 2009.

[41] R.C. Gonzalez, R.E. Woods, and S.L. Eddins. *Digital Image Processing Using MATLAB*. Prentice Hall, Upper Saddle River, NJ, USA, 2004.

[42] A. Goshtasby. *2-D and 3-D Image Registration*. Wiley-Interscience, Hoboken, NJ, USA, 2005.

[43] A. Goshtasby, G. Stockman, and C. Page. A region-based approach to digital image registration with subpixel accuracy. *IEEE Transactions on Geoscience and Remote Sensing*, 24(3):390–399, 1986.

[44] I. Guyon, S. Gunn, M. Nikravesh, and L.A. Zadeh, editors. *Feature extraction: foundations and applications*. Studies in fuzziness and soft computing. Springer-Verlag, 2006.

[45] J.V. Hajnal and D.L.G. Hill, editors. *Medical Image Registration*. CRC Press, 2001.

[46] H. Handels. *Medizinische Bildverarbeitung*. Vieweg + Teubner, 2009.

[47] R. Haralick. Statistical and structural approaches to texture. *Proceedings of the IEEE*, 67(5):786–804, 1979.

[48] C. Harris and M. Stephens. A combined corner and edge detector. In *Proceedings of the 4th Alvey Vision Conference*, pages 147–151, 1988.

[49] M. L. Hentschel and N. W. Page. Selection of descriptors for particle shape characterization. *Particle & Particle Systems Characterization*, 2002.

[50] W. Hoenlein, F. Kreupl, G.S. Duesberg, A.P. Graham, M. Liebau, R.V. Seidel, and E. Unger. Carbon nanotube applications in microelectronics. *IEEE Transactions on Components and Packaging Technologies*, 27 (4):629 – 634, 2004.

[51] J.-W. Hsieh. Fast stitching algorithm for moving object detection and mosaic construction. *IEEE International Conference on Multimedia and Expo*, 1:85–88, 2003.

[52] J. Inglada and A. Giros. On the possibility of automatic multisensor image registration. *IEEE Transactions on Geoscience and Remote Sensing*, 42(10):2104–2120, 2004.

[53] D. Jasper and S. Fatikow. Line scan-based high-speed position tracking inside the SEM. *International Journal of Optomechatronics*, 4(2):115–135, 2010.

[54] H. Jegou, M. Douze, and C. Schmid. Hamming embedding and weak geometric consistency for large scale image search. In D. Forsyth, P. Torr, and A. Zisserman, editors, *Computer Vision - ECCV 2008*, volume 5302 of *Lecture Notes in Computer Science*, pages 304–317. Springer Berlin / Heidelberg, 2008.

[55] B. Jähne. *Digital Image Processing*. Springer, 1991.

[56] T. Joachims. Making large-scale svm learning practical. *Advances in Kernel Methods - Support Vector Learning*, 1999.

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

[58] T. Kadir, A. Zisserman, and J. M. Brady. An affine invariant salient region detector. In *European Conference on Computer Vision*. Springer-Verlag, 2004.

[59] J.N. Kapur, P.K. Sahoo, and A.K.C. Wong. A new method for gray-level picture thresholding using the entropy of the histogram. *Computer Vision, Graphics, and Image Processing*, 29(3):273–285, 1985.

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

[61] C. Keck, M. Berndt, and R. Tutsch. Stereophotogrammetry in microassembly. In S. Büttgenbach, A. Burisch, and J. Hesselbach, editors, *Design and Manufacturing of Active Microsystems*, pages 309–326. Springer, 2011.

[62] B.E. Kratochvil, L.X. Dong, and B.J. Nelson. Real-time rigid-body visual tracking in a scanning electron microscope. *International Journal of Robotics Research*, 28(4):498–511, March 2009.

[63] B.V.K.V. Kumar, A. Mahalanobis, and R.D. Juday. *Correlation Pattern Recognition*. Cambridge University Press, New York, NY, USA, 2005.

[64] J.C. Lagarias, J.A. Reeds, M.H. Wright, and P.E. Wright. Convergence properties of the Nelder–Mead simplex method in low dimensions. *SIAM J. on Optimization*, 9(1):112–147, 1998.

[65] S. Lazebnik, C. Schmid, and J. Ponce. Sparse texture representation using affine-invariant neighborhoods. In *International Conference on Computer Vision & Pattern Recognition*, volume 2, pages 319–324, 2003.

[66] M.J. Lee, S. Kim, S.A. Lee, H.T. Song, Y.M. Huh, D.H. Kim, S.H. Han, and J.S. Suh. Overcoming artifacts from metallic orthopedic implants at high-field-strength MR imaging and multi-detector CT. *RadioGraphics*, 27:791–803, 2007.

[67] Y.-S. Lee, H.-S. Koo, and C.-S. Jeong. A straight line detection using principal component analysis. *Pattern Recogn. Lett.*, 27(14):1744–1754, 2006.

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

[69] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens. Multimodality image registration by maximization of mutual information. *IEEE transactions on Medical Imaging*, 16(2):187–198, 1997.

[70] J. Matas, O. Chum, U. Martin, and T. Pajdla. Robust wide baseline stereo from maximally stable extremal regions. In *Proc. of the British Machine Vision Conference*, pages 384–393, 2002.

[71] H. Matsuura, T. Inoue, H. Konno, M. Sasaki, K. Ogasawara, and A. Ogawa. Quantification of susceptibility artifacts produced on high-field magnetic resonance images by various biomaterials used for neurosurgical implants: Technical note. *Journal of neurosurgery*, 97(6):1472–1475, 2002.

[72] E. De Castro and C. Morandi. Registration of translated and rotated images using finite fourier transforms. *IEEE Trans. Pattern Anal. Mach. Intell.*, 9:700–703, September 1987.

[73] K. Mikolajczyk. *Detection of Local Features Invariant to Affine Transformations, Application to Matching and Recognition.* PhD thesis, Institut National de Polytechniques de Grenoble, 2002.

[74] K. Mikolajczyk and C. Schmid. Scale & affine invariant interest point detectors. *International Journal of Computer Vision*, 60 (1):63 – 86, 2004.

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

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

[77] T.K. Moon. The expectation-maximization algorithm. *IEEE Signal Process. Mag.*, 13(11):47–60, 1996.

[78] M.S. Nixon and A.S. Aguado. *Feature extraction and image processing.* Academic Press. Academic, 2008.

[79] N. Otsu. A threshold selection method from gray level histograms. *IEEE Trans. Syst. Man Cybern.*, 9(1):62–66, 1979.

[80] P. Patel-Predd. Carbon-nanotube wiring gets real. *IEEE Spectrum*, 45(4):14, 2008.

[81] G. Poletti, G. Orsini, C. Lenardi, and E. Barborini. A comparative study between AFM and SEM imaging on human scalp hair. *Journal of Microscopy*, 211:249–255, 2003.

[82] J.D. Port and M.G. Pomper. Quantification and minimization of magnetic susceptibility artifacts on GRE images. *Neuroradiology*, 24(6):958–964, 2000.

[83] S. Rathmann, K. Schöttler, M. Berndt, G. Hemken, A. Raatz, R. Tutsch, and S. Böhm. Sensor-guided micro assembly of active micro systems by using a hot melt based joining technology. *Microsystem Technologies*, 14(12):1975–1981, 2008.

[84] L. Reimer. *Scanning Electron Microscopy: Physics of Image Formation and Microanalysis*. Springer, 1998.

[85] R. Rifkin and A. Klautau. In defense of one-vs-all classification. *Journal of Machine Learning Research*, 5:101–141, 2004.

[86] A. Roche, G. Malandain, X. Pennec, and N. Ayache. The correlation ratio as a new similarity measure for multimodal image registration. In W. Wells, A. Colchester, and S. Delp, editors, *Medical Image Computing and Computer-Assisted Intervention*, volume 1496 of *Lecture Notes in Computer Science*, pages 1115–1124. Springer Berlin / Heidelberg, 1998.

[87] R. Rodrigo, W. Shi, and J. Samarabandu. Energy based line detection. In *Canadian Conference on Electrical and Computer Engineering*, 2006.

[88] A. Rosenfeld and G.J. Vanderbrug. Coarse-fine template matching. *IEEE Trans. Systems, Man, and Cybernetics*, 7(2):104–107, 1977.

[89] O. Sardan, D. H. Petersen, K. Mølhave, O. Sigmund, and P. Bøggild. Topology optimized electrothermal polysilicon microgrippers. *Microelectronic Engineering*, 85(5-6):1096–1099, 2008.

[90] A. Sartori, R. Gatz, F. Beck, A. Rigort, W. Baumeister, and J.M. Plitzko. Correlative microscopy: Bridging the gap between fluorescence light microscopy and cryo-electron tomography. *Journal of Structural Biology*, 160(2):135–145, 2007.

[91] F. Schaffalitzky and A. Zisserman. Multi-view matching for unordered image sets. In *Proc. Seventh European Conf. Computer Vision*, pages 414–431, 2002.

[92] O. Scherzer, editor. *Mathematical Models for Registration and Applications to Medical Imaging*. Springer Verlag, 2006.

[93] J. Schürmann. *Pattern Classification: A Unified View of Statistical and Neural Approaches*. Wiley-Interscience, 1996.

[94] A. Seeger. *Surface Reconstruction From AFM and SEM Images*. PhD thesis, University of North Carolina, 2004.

[95] M. Sezgin and B. Sankur. Survey over image thresholding techniques and quantitative performance evaluation. *Journal of Electronic Imaging*, 13 (1):146–165, 2004.

[96] F. Shafiei, E. Honda, H. Takahashi, and T. Sasaki. Artifacts from dental casting alloys in magnetic resonance imaging. *Journal of Dental Research*, 82(8):602–606, 2003.

[97] T. Sievers. *Echtzeit Objektverfolgung im Rasterelektronenmikroskop*. PhD thesis, Universität Oldenburg, 2007.

[98] T. Sievers and S. Fatikow. Visual servoing of a mobile microrobot inside a scanning electron microscope. In *Proc. of IEEE Int. Conference of Intelligent Robots and Systems*, 2005.

[99] T. Sievers and S. Fatikow. Real-time object tracking for the robot-based nanohandling in a scanning electron microscope. *Journal of Micromechatronics - Special Issue on Micro/Nanohandling*, 3(3-4):267–284(18), 2006.

[100] I. Steinwart and A. Christmann. *Support Vector Machines*. Information Science and Statistics. Springer, 2008.

[101] C. Studholme, D.L.G. Hill, and D.J. Hawkes. An overlap invariant entropy measure of 3D medical image alignment. *Pattern Recognition*, 32(1):71–86, 1999.

[102] O. Tarasenko, S. Nourbakhsh, S.P. Kuo, A. Bakhtina, P. Alusta, D. Kudasheva, M. Cowman, and K. Levon. Scanning electron and atomic force microscopy to study plasma torch effects on b. cereus spores. *IEEE Transactions on Plasma Science*, 34(4):1281–1289, 2006.

[103] T. Tuytelaars and K. Mikolajczyk. Local invariant feature detectors: a survey. *Computer Graphics and Vision: Foundations and trands*, 3(3):177–280, 2008.

[104] T. Tuytelaars and L. Van Gool. Matching widely separated views based on affine invariant regions. *International Journal of Computer Vision*, 59(1):61–85, 2004.

[105] L.J. Van Gool, T. Moons, and D. Ungureanu. Affine / photometric invariants for planar intensity patterns. In *Proceedings of the 4th European Conference on Computer Vision*, pages 642–651. Springer-Verlag, 1996.

[106] V. N. Vapnik. An overview of statistical learning theory. *Neural Networks, IEEE Transactions on*, 10(5):988–999, 1999.

[107] Q. Wei, D. Tao, W. Gao, and Y. Huang. Scanning electron microscopy and atomic force microscopy of composite nanofibres. *Microscopy and Analysis*, 22(2):11–12, 2008.

[108] C. Wählby. *Algorithms for Applied Digital Image Cytometry*. PhD thesis, Uppsala University, Centre for Image Analysis, 2003.

[109] T. Wortmann. Fusion of AFM and SEM scans. In *Proc. of Int. Symposium on Optomechatronic Technologies (ISOT)*, pages 40–45, 2009.

[110] T. Wortmann. Automatic stitching of micrographs using local features. In *Proc. of Int. Symposium on Optomechatronic Technologies (ISOT)*, 2010.

[111] T. Wortmann. Registration of AFM and SEM scans using local features. *International Journal of Optomechatronics*, 5(3):249–270, 2011.

[112] T. Wortmann, C. Dahmen, and S. Fatikow. Study of MRI susceptibility artifacts for nanomedical applications. *Journal of Nanotechnology in Engineering & Medicine*, 1(4):041002, 2010.

[113] T. Wortmann, C. Dahmen, C. Geldmann, and S. Fatikow. Recognition and tracking of magnetic nanobots using MRI. In *Proc. of Int. Symposium on Optomechatronic Technologies (ISOT)*, 2010.

[114] T. Wortmann, C. Dahmen, R. Tunnell, and S. Fatikow. Image processing architecture for real-time micro- and nanohandling applications. In *Proc. of the Eleventh IAPR Conference on Machine Vision Applications (MVA)*, pages 418–421, 2009.

[115] T. Wortmann and S. Fatikow. Carbon nanotube detection by scanning electron microscopy. In *Proc. of the Eleventh IAPR Conference on Machine Vision Applications (MVA)*, pages 370–373, 2009.

[116] J. Zhang, M. Marszałek, S. Lazebnik, and C. Schmid. Local features and kernels for classification of texture and object categories: A comprehensive study. *International Journal of Computer Vision*, 73:213–238, 2007.

[117] Y. Zhang, B.K. Chen, X. Liu, and Y. Sun. Autonomous robotic pick-and-place of microobjects. *IEEE Transactions on Robotics*, 26(1):200 – 7, 2010.

[118] Y. Zheng, editor. *Image Fusion and Its Applications*. InTech, 2011.

[119] Y.Y. Zhu, G.Q. Ding, J.N. Ding, and N.Y. Yuan. AFM, SEM and TEM studies on porous anodic alumina. *Nanoscale Research Letters*, 5(4):725–734, 2010.

[120] B. Zitova and J. Flusser. Image registration methods: a survey. *Image and Vision Computing*, 21(11):977–1000, 2003.