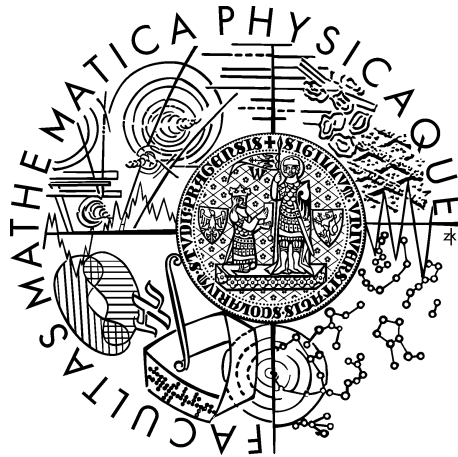


Charles University in Prague
Faculty of Mathematics and Physics

DOCTORAL THESIS



Jiří Sedlář

Image Analysis in Microscopy and Videokymography

Institute of Information Theory and Automation
Academy of Sciences of the Czech Republic

Supervisor of the doctoral thesis: Prof. Jan Flusser

Study programme: Computer Science

Specialization: Software Systems

Prague 2012

I would like to express my gratitude to Prof. Jan Flusser for supervision of my doctoral studies; Assoc. Prof. Michaela Sedlářová, Dr. Jaromír Kopeček, and Dr. Jan G. Švec for consultations on phytopathology, atomic force microscopy, and videokymography, respectively; members of the Department of Image Processing, namely Dr. Barbara Zitová, M.Sc. Jiří Dvořák, M.Sc. Michal Bartoš, and M.Sc. Miroslav Beneš, for their advice; M.Sc. David Hauzar and M.Sc. Adam Novozámský for permission to use their implementation of algorithms for detection of glottal features and thresholding by graph cuts, respectively; teachers at the Faculty of Mathematics and Physics for consultations; chairman and members of my Thesis defence committee and opponents of my Thesis for their time; my family for their support; and my friends for their encouragement.

Thank you.

I declare that I carried out this doctoral thesis independently, and only with the cited sources, literature and other professional sources.

I understand that my work relates to the rights and obligations under the Act No. 121/2000 Coll., the Copyright Act, as amended, in particular the fact that the Charles University in Prague has the right to conclude a license agreement on the use of this work as a school work pursuant to Section 60 paragraph 1 of the Copyright Act.

In Prague date

signature of the author

Název práce: Analýza obrazu v mikroskopii a videokymografii

Autor: Jiří Sedlář

Katedra: Ústav teorie informace a automatizace, Akademie věd České republiky, v.v.i.

Vedoucí disertační práce: Prof. Ing. Jan Flusser, DrSc., Ústav teorie informace a automatizace, Akademie věd České republiky, v.v.i.

Abstrakt: Tato disertační práce popisuje nové metody pro automatické zpracování obrazových dat v biologii, fyzice a medicíně. Navržené metody slouží k rekonstrukci mikroskopických snímků rostoucích mikroorganismů v intervalech mezi pozorováními, k měření částic na snímcích mikroskopu atomárních sil a k vyhodnocení parametrů kmitání hlasivek na videokymografických snímcích. Příslušná obrazová data byla zatím zpracovávána převážně vizuálně. Navržené metody umožňují automatické či počítačem podporované zpracování těchto dat, čímž usnadňují jejich vyhodnocování. Vyvinuté metody byly testovány na reálných snímcích. Výsledky automatického zpracování byly srovnatelné s reálnými hodnotami nebo s výsledky vizuálního vyhodnocení. Aplikace navržených metod není omezena na konkrétní typ obrazových dat. Metody mohou být použity obecně ke zpracování snímků s podobnými vlastnostmi.

Klíčová slova: analýza obrazu, počítačem podporované vyhodnocování, světelná mikroskopie, mikroskopie atomárních sil, videokymografie

Title: Image Analysis in Microscopy and Videokymography

Author: Jiří Sedlář

Department: Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic

Supervisor: Prof. Ing. Jan Flusser, DrSc., Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic

Abstract: The Thesis describes new methods for automatic processing of image data in biology, physics and medicine. The developed methods reconstruct light microscopy images of growing microorganisms in intervals between observations, measure particles in atomic force microscopy images, and evaluate parameters of vocal fold vibrations in videokymographic images. All three problems have been hitherto solved primarily visually. The proposed methods allow automatic or computer-aided processing of the image data, and thus facilitate the evaluation process. Performance of the developed methods was tested on real images; the results were comparable with ground truth or results of visual evaluation. Application of the developed methods is not limited to the specific type of image data; the methods can be used in general for processing of images with similar characteristics.

Keywords: image analysis, computer-aided evaluation, light microscopy, atomic force microscopy, videokymography

Contents

1	Introduction	3
1.1	Objectives	4
1.1.1	Reconstruction of the growth of filamentous specimen over time	4
1.1.2	Measurement of elliptical particles in atomic force microscopy images	6
1.1.3	Measurement of vocal fold vibration parameters in videokymographic images	8
1.2	Contribution	11
2	Reconstruction of the growth of settled filamentous specimens in light microscopy images	12
2.1	Introduction	12
2.2	Applied image processing tools	14
2.2.1	Morphological skeleton	15
2.2.2	Thin-plate splines	16
2.3	Method	17
2.4	Results	23
2.5	Discussion	30
2.6	Conclusion	31
3	Measurement of particles in atomic force microscopy images	32
3.1	Introduction	32
3.2	Applied image processing tools	34
3.2.1	Watershed transform	35
3.2.2	Image moments	36
3.3	Method	38
3.3.1	Denoising	38
3.3.2	Segmentation	38
3.3.3	Approximation by ellipses	40
3.3.4	Approximation by ellipsoids	43
3.3.5	Measurement	45
3.4	Results	46
3.5	Discussion	55
3.6	Conclusion	57
4	Evaluation of vocal fold vibration parameters in videokymographic images	59
4.1	Introduction	59
4.2	Applied image processing tools	64
4.2.1	Thresholding by normalized graph cuts	64
4.3	Methods	66
4.3.1	Reflections	66
4.3.2	Rima glottidis	68
4.3.3	Mucosal waves	75

4.3.4	Parameters of vocal fold vibrations	80
4.3.5	Comparison of automatic and visual evaluations	87
4.4	Results	89
4.5	Discussion	91
4.6	Conclusion	93
5	Conclusion	94
	Bibliography	96
	List of Abbreviations	101

1. Introduction

Imaging technologies are an indispensable tool for examination of samples in biology, physics and medicine. Their rapid development has allowed observation of previously unobservable image data. In current practice, however, most of the acquired image data are evaluated visually, with use of image methods only for preprocessing and visualization. Development of automatic methods for evaluation of such image data would not only decrease demands on evaluators' time but also increase precision and consistency of results, and thus increase effectiveness of the evaluation process.

New imaging technologies have allowed acquisition of previously unobservable details and processes. Progress of research in physics, optics and computer science has led to significant increase in feasible spatial and temporal resolution as well as to development of a number of new imaging technologies. Microscopy, for example, has allowed observation of samples in very high resolution. High-speed imaging, in another example, has allowed observation of very fast processes. The fact that such image data cannot be observed directly makes their interpretation more difficult; the evaluation process usually requires prior information about known properties of the observed sample or process.

Evaluation of biomedical image data is usually quite complex. The visual evaluation process is often time-consuming and tedious, and thus prone to error. Moreover, it is subjective and inconsistent, since the results may differ for different evaluators and even for different evaluations by the same evaluator. Visual evaluation also limits the precision of results. Automation of the evaluation process could increase its efficiency, accuracy and consistency.

Demand for automatic image processing methods has been increasing with development of new imaging technologies. Imaging technologies have rapidly developed recently, while their decreasing costs have increased their availability. With progressively improving parameters of imaging technologies, namely higher resolution and faster acquisition time, the production of image data has been growing exponentially; for example, the amount of medical image data alone is estimated to constitute one third of the overall amount of all data [49]. Visual evaluation is not sufficient for such amounts of image data, so development of new automatic evaluation methods is of great interest.

Accuracy of automatic evaluation methods is limited by variability of observed data and precision of acquisition techniques. Real data, particularly in biology and medicine, are quite variable. Evaluation of such data should thus use prior knowledge about their parameters to a limited extent; the evaluation methods should be able to deal with images of samples with unusual parameters. Correspondence between real and acquired image data is limited by properties of the imaging device. Precision of digital image data is limited by diffraction, which restricts acquisition of high frequencies; this complicates mainly observation of very small objects and very fast processes. Moreover, accuracy of acquired data is distorted by noise. Type and level of noise in image data depends mostly on parameters of the imaging device and on acquisition conditions. High levels of noise can lead to imprecisions or errors in evaluation. Due to these limitations, methods for evaluation of real image data cannot be errorless.

In biomedical practice, image processing methods [5, 22] are often used as a preprocessing step for visual evaluation or in combination with subsequent visual verification. High complexity and variability of image data, particularly in medicine and biology, makes full automation of the evaluation process difficult. Accuracy of evaluation results is often critical, especially if errors in evaluation would lead to serious consequences, e.g. for the health of the examined patient. Results of automatic evaluation are thus usually visually verified by experts. Experts can also adaptively modify parameters of the automatic method and correct the results if necessary. This approach combines efficiency of the automatic method with knowledge and experience of experts. It can also significantly reduce demand on experts' time and effort and possibly increase accuracy of evaluation results.

1.1 Objectives

The objective of this research was to develop methods for automatic processing of specific biological, medical and physical image data that have been thus far evaluated mostly visually.

The Thesis focuses on three image processing projects recently researched at the Department of Image Processing of the Institute of Information Theory and Automation of the Academy of Sciences of the Czech Republic, and describes the developed methods. The first project aims to reconstruct light microscopy images of growing settled filamentous specimens in intervals between consecutive observations, the second project measures salient elliptical particles in atomic force microscopy images, and the third project computes parameters of vocal fold vibrations in videokymographic images. In all three cases, the data—due to their small size or high temporal frequency—cannot be observed without special imaging technologies. Motivation and objectives of the respective projects are introduced in Subsections 1.1.1, 1.1.2, and 1.1.3.

1.1.1 Reconstruction of the growth of filamentous specimens over time

The first project addresses the problem of missing information about the growth of microorganisms in intervals between consecutive observation sessions. This is a common issue in phytopathology, where examined microorganism have to be cultivated separately in optimal conditions and their growth is thus documented usually in sparsely repeated observation sessions. However, the documented growth pattern is often incomplete because the intervals between consecutive sessions are relatively long. The objective of this project was to develop a method that would reconstruct light microscopy images of settled filamentous specimens, namely *Alternaria* sp. and *Fusarium oxysporum* (see Section 2.4), corresponding to intervals between consecutive observation sessions.

Sparsity in documentation of growing microorganisms in phytopathology is caused by special requirements on cultivation conditions and by complexity of the documentation process. Most phytopathogenic microorganisms need special environmental conditions, particularly temperature and humidity, for cultivation.

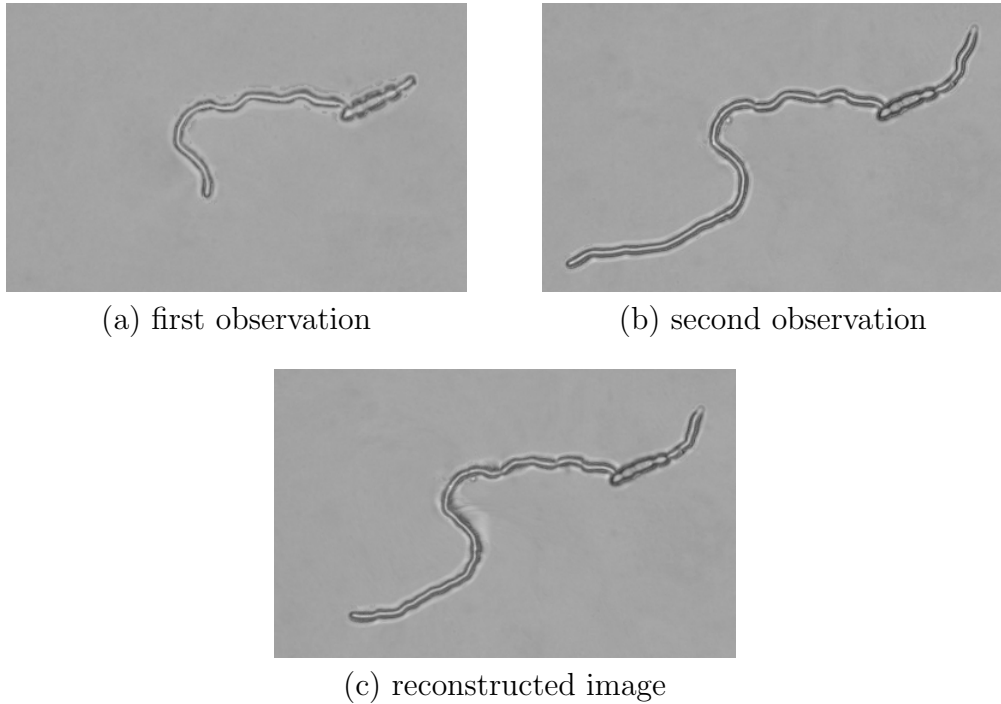


Figure 1.1: (a, b) Light microscopy images of a *Fusarium oxysporum* f.sp. *pisi* specimen acquired at two consecutive observation sessions. (c) A reconstructed image corresponding to 3/5 of the interval between the two sessions.

However, the temperature range of current life-imaging systems with controlled environment parameters is mostly inconvenient for phytopathogenic fungi. The microorganisms are thus cultivated separately, in special conditions, and repeatedly fetched for microscopical documentation. The documentation process is complex and quite demanding. In each observation session, the examiner has to fetch the analyzed specimens from the cultivation environment, place them sequentially under the microscope, position them appropriately, focus the objective, photograph them¹, return them to their cultivation environment, save the acquired images and describe them properly, including the time and parameters of acquisition and the label of each specimen. The process is time-consuming and laborious, so the intervals between consecutive sessions are often relatively long. As a result, the documented growth pattern may be incomplete.

Reconstructed images showing the gradual growth of phytopathogenic microorganisms between consecutive sessions could be used for computation of their growth parameters as well as for smooth visualization of their development. Phytopathogenic fungi cause severe plant diseases, which can significantly decrease production of many economically important crops, so detailed understanding of the growth pattern could increase effectiveness of plant disease control.

The reconstruction problem has not been satisfactorily addressed. The current approach in biology is development of life-imaging systems that allow continuous observation of phytopathogenic fungi in convenient environment conditions. The systems are, however, quite expensive. Moreover, this approach requires cultiva-

¹ If the specimen is thicker than the depth of field, it is photographed in several focal lengths, which can be later digitally fused into one all-focus image.

tion, observation and documentation of new specimens instead of using available sparsely documented data. The current software approach is based on description of the specimen by a geometrical model and visualization of its simulated growth by rendering techniques. It constructs the model according to prior information about morphology and growth pattern of the specimen and adapts its parameters according to specific data. The growth of settled biological specimens such as fungal pathogens [40] has been successfully modeled by L-system grammars [35]. The synthesized images, however, appear artificial.

The objective of this research project was to reconstruct the missing images without significant artifacts from microscopy images acquired at observation sessions, i.e. without additional observations. The method should use information about known parameters of observed specimens. The project focused on light microscopy images of settled filamentous specimens, e.g. filamentous fungi and oomycetes; they grow by elongation of their filaments while the shape of already developed parts does not change. The method should be able to reconstruct images corresponding to arbitrary times within intervals between observation sessions and thus gradually visualize the growth. The reconstructed images should not contain artificial deformations that would lead to false conclusions about properties of the biological specimen.

The developed method geometrically transforms images acquired at two consecutive observation sessions. It computes a growth model that estimates the trajectory of salient control points during the interval. The method, however, reconstructs the missing images directly from the acquired images, and thus outperforms the rendering-based methods. The method computes the morphological skeleton of the shape of the segmented specimen in the images from consecutive observation sessions. Then it selects a number of corresponding control points on the skeletons. The method then tracks the movement of the control points along the morphological skeleton during the interval between the sessions; it utilizes the prior knowledge that the specimen grows by elongation of its filaments. For an arbitrary time within the interval, the method geometrically transforms the temporally closer image by mapping the positions of its control points to positions corresponding to the time. The result is a sequence of reconstructed light microscopy images showing the gradual growth of the specimen during the interval between the two sessions. Chapter 2 describes the method in detail.

1.1.2 Measurement of elliptical particles in atomic force microscopy images

The second project addresses the problem of detection and measurement of nanoparticles in atomic force microscopy (AFM) images. This is a common task in physics since the distribution of particle sizes indicates various properties of the analyzed material. Manual measurement of particles in AFM images is, however, time-consuming and inconsistent. The objective of this project was to develop an automatic method that would detect salient ellipsoidal particles in atomic force microscopy images, namely of phenylpyridyldiketopyrrolopyrrole (PPDP) samples, and measure their average length and width.

Surface topography in AFM images is distorted by several factors. Atomic force microscopy estimates topography of solid-body surfaces by measuring in-

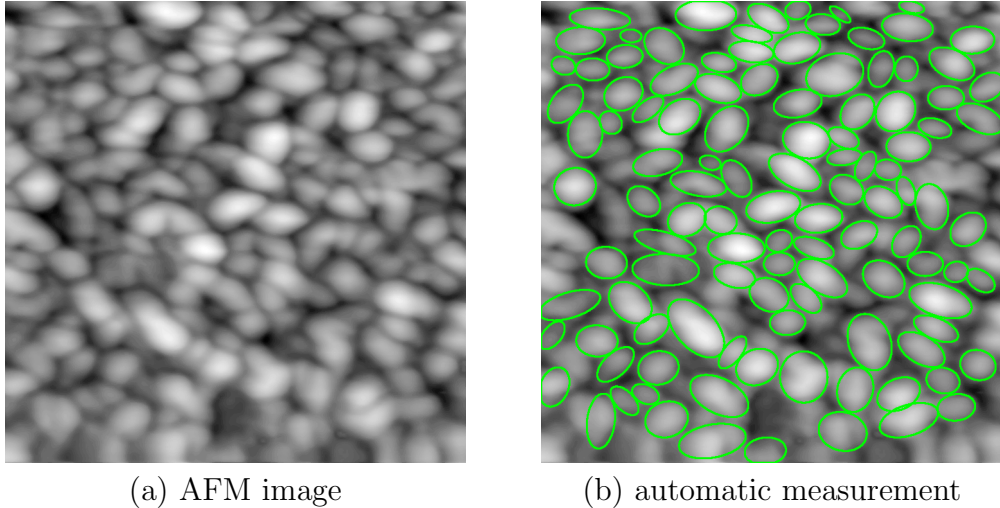


Figure 1.2: (a) Atomic force microscopy (AFM) image of a phenylpyridyldiketopyrrolopyrrole (PPDP) sample. (b) Automatically detected shapes of particles.

teractive forces between the surface and a scanning tip. The observed image is a convolution between the surface and the tip; this means that objects smaller than the tip appear significantly distorted. The shape of the tip is usually unknown and may even change during measurements. Moreover, atomic force microscopy is highly sensitive to noise, so the acquired images contain noise artifacts. These distortions complicate analysis of AFM images.

The average length and width of PPDP particles in AFM images characterize physical properties of the sample. PPDP is a perspective organic semiconducting material used for hydrogen sensors. Deposited layers of PPDP contain similarly sized, mostly spatially separated ellipsoidal particles. PPDP particles are relatively large in comparison to the scanning tip, so their surface can be observed by AFM with relatively little deformation. Lateral size distribution of particles in a sample determines its properties. Using the prior knowledge about ellipticity and little size variation of PPDP particles, the properties of PPDP samples can be effectively estimated just from the average length and width of particles in AFM images.

Manual measurement of PPDP particles in AFM images is inconvenient for practical use. AFM images of PPDP samples typically contain hundreds to thousands of particles. The physicist thus usually measures just a small subset of particles and computes the average lateral length and width from them. This, however, decreases accuracy of measured results; moreover, the manual measurement process remains time-consuming and tedious. Manual measurements are also imprecise and inconsistent: the measured average length and width depend on which particles and how accurately were measured. The results of manual measurement in an AFM image may thus differ not only for different evaluators but also for different measurements by the same physicist. Automation of the measurement process would increase its speed, accuracy and consistency.

Existing methods for automatic processing of AFM images are inconvenient for measurement of PPDP particles. Most methods for segmentation of particles in AFM images are quite sensitive to noise. Although the noise artifacts can be

suppressed [60], the segmented particles are often distorted [51] as the methods do not use prior information about their real shape. Methods using prior information about topography of scanned samples are usually designed for samples with densely packed particles [11] and thus cannot be used for segmentation of PPDP particles.

The objective of this project was to automatically detect salient ellipsoidal particles in AFM images and measure their lateral length and width. The method should use prior information about known properties of the scanned material to compensate for limitations of the imaging technique; the project focused specifically on AFM images of PPDP samples. The method should be robust to distortions typical of AFM images. It should measure only salient particles because neither partially occluded particles nor particles significantly distorted by noise artifacts can be measured with high accuracy. Results of the method, namely the average measured lateral length and width, should be statistically comparable with results of manual measurements.

The developed method detects salient ellipsoidal particles by watershed segmentation and approximates their shapes by ellipses. To increase accuracy of results, it does not measure partially occluded or otherwise distorted particles. The method first both smooths topography and suppresses high-frequency noise in the AFM image by convolution with a Gaussian mask. Then it roughly segments particles by watershed transform and removes particles intersecting watershed lines because such particles are probably partially occluded. The method approximates the shapes of particles in watershed regions by ellipses; it computes parameters of the approximating ellipses by image moments. Then it approximates topography of each separated particle by the upper half of an ellipsoid above the approximating ellipse, and removes non-ellipsoidal particles, which usually correspond to particles distorted by noise artifacts, vertically tilted particles and partially occluded particles not separated by the watershed transform. The method estimates the lengths and widths of ellipsoidal particles by the major and minor axes, respectively, of corresponding approximating ellipses. Finally, it respectively averages the estimated lateral lengths and widths of particles in the analyzed AFM image. Chapter 3 describes the method in detail. The automatic method was tested on AFM images of PPDP samples (see Figure 1.2); its results were comparable with results of manual measurements.

1.1.3 Measurement of vocal fold vibration parameters in videokymographic images

The third project addresses the problem of evaluation of vocal fold vibration parameters in videokymographic images (videokymograms). Videokymography (VKG) is a high-speed imaging technique for observation of fast movements. Accurately evaluated parameters of vibratory patterns in videokymograms can be used in laryngology and phoniatrics for diagnosis of voice disorders. In current practice, clinicians evaluate videokymograms of vocal fold vibrations visually; the visual evaluation process is, however, time-consuming, and its results are often inconsistent. The objective of this project was to develop automatic methods that would detect important vocal fold vibration features in videokymograms and compute corresponding vibration parameters.

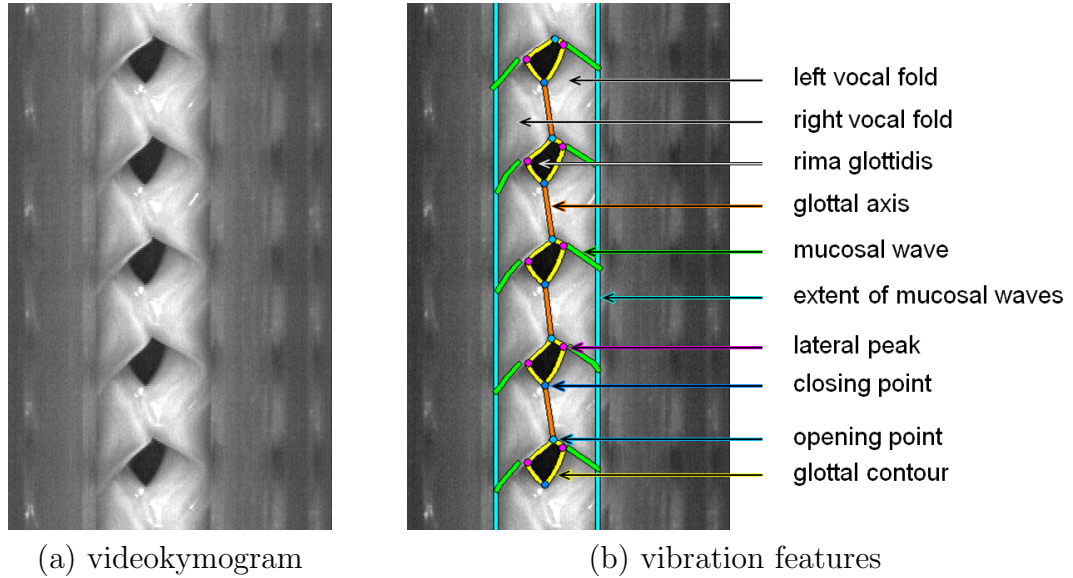


Figure 1.3: (a) Videokymogram of vocal fold vibrations. (b) Automatically detected vibration features.

Videokymography is a novel imaging technique used for examination of vocal fold vibrations [58]. Videokymographic camera is a specially adapted video camera that scans frames in only one row but with frequency c. 8000 Hz. This allows observation of very fast processes, including vibrations of vocal folds. The acquired videokymogram is a spatiotemporal image composed of consecutive scans of the row. A videokymogram of vibrating vocal folds shows gradual movements of vocal folds along the scanned row. It contains a number of well-defined features that determine various parameters of vocal fold vibrations. The parameters are examined in laryngology and phoniatrics for purposes of voice disorder diagnostics.

Evaluation of videokymograms is quite complex. The VKG evaluation sheet [59] defines parameters describing various properties of the vibratory pattern. In current practice, clinicians estimate the value or category of each parameter visually from the videokymogram. The evaluation process is, however, time-consuming and tedious, since the sheet contains a number of parameters. The fact that videokymograms are artificially composed, spatiotemporal images further complicates their evaluation. For most parameters, the evaluator just selects one of predefined categories instead of computing the exact value, which limits the accuracy of results. Results of visual evaluations are often inconsistent since they may differ even for the same evaluator over time [15]. Development of automatic evaluation methods would increase precision and consistency of results and decrease demand on evaluators' time. Results of automatic evaluation could be straightforwardly verified by clinicians and compared with results of visual evaluations.

Automation of the evaluation process has not been satisfactorily solved. Low contrast and high level of noise, as well as irregularities in periodicity of vocal fold vibrations complicate automatic processing of videokymograms. In 2003 Qiu *et al.* [36] introduced an algorithm for detection of glottal contour, i.e. the boundary

of rima glottidis, by Otsu's bi-thresholding [34] and active contour fitting, and an algorithm for quantification of corresponding vibration parameters by measurement of periodical structures in the signal. In 2006 Manfredi *et al.* [29] developed an algorithm for approximation of glottal contour by active contours. In 2008 Jiang *et al.* [23] proposed a curve-fitting algorithm for detection of the trajectories of upper and lower vocal fold lips, which determine the glottal contour. In 2010 Zhang *et al.* [67] introduced a method for segmentation of rima glottidis by Lagrange interpolation, differentiation, and Canny's edge detection. In the same year Hauzar [17] developed methods for hierarchical extraction of rima glottidis features from videokymograms. In 2011 Zita [68] proposed a method for detection of mucosal waves in videokymograms of healthy vocal folds, which uses the extracted rima glottidis features. The algorithms are, however, sensitive to irregularities and other pathologies of the vibratory pattern, as well as to noise and low contrast in videokymographic images. Moreover, the problem of automatic computation of vibration parameters in the VKG evaluation sheet from features in videokymograms has been addressed only in 2012 [44].

The objective of this project was to automatically detect important vocal fold vibration features in videokymograms and evaluate corresponding parameters in the VKG evaluation sheet. The methods should focus on vibration parameters defined by the shape of rima glottidis and also on trajectories of mucosal waves. The methods should be applicable to videokymograms of not only physiologic vibratory patterns but also patterns corresponding to various types and degrees of voice disorders. Results of the automatic evaluation should be comparable with results of visual evaluations.

The developed methods detect specular reflections, the shape of rima glottidis and directions of mucosal waves, as well as vibration features (see Figure 1.3), and evaluate corresponding vibration parameters in the VKG evaluation sheet. One method detects specular reflections on vocal folds by thresholding and region growing, and removes them by diffusion from outer boundaries. Another method segments the shape of rima glottidis by thresholding based on graph cuts in combination with binary morphology and column operations. Another method detects the shape of mucosal waves by Fourier transform. Combination of these methods with existing detection methods can increase accuracy of results, especially for vibratory patterns corresponding to various voice disorders. The last method computes glottal vibration parameters in the evaluation sheet from glottal features extracted from the computed rima glottidis shape using the algorithm developed by Hauzar [17]. Chapter 4 describes the developed methods in detail. Performance of the automatic methods was tested on a representative set of videokymograms covering a wide range of vibratory patterns. The results were comparable with results of visual evaluations.

1.2 Contribution

This section summarizes main contributions of the Thesis. The developed methods successfully solved objectives of the researched projects. Their performance was tested on real data; the results were comparable with ground truth or results of visual evaluations. Moreover, the results can be straightforwardly verified by experts or combined with results of visual evaluation. Although the methods were designed for specific type of data, they can be applied generally to any data with similar properties.

The method developed in the first project (see Subsection 1.1.1) enables examination of gradual growth of settled filamentous specimens from light microscopy images acquired at sparse observation sessions. It thus avoids the need for expensive continuously monitoring systems convenient for cultivation of phytopathogenic fungi. The main contribution was development of a new growth-tracking method that simulates elongation of filaments by elongation of the morphological skeleton. Straightforward extension of points on the skeleton to object boundary defines control points for geometrical transformation of the temporally closer image to an image representing an arbitrary time within the interval between observations.

The method created in the second project (see Subsection 1.1.2) enables automatic detection and measurement of ellipsoidal particles in atomic force microscopy (AFM) images. The method can be used instead of—or in combination with—manual measurements: it increases not only efficiency of the measurement process but also consistency and precision of results. The main contribution was development of an algorithm that approximates the shape of particles by ellipses and thus makes the segmentation robust to high-frequency noise. The method also allows reconstruction of the topography of ellipsoidal particles in AFM images without the scanning tip convolution.

The methods developed in the third project (see Subsection 1.1.3) enable automatic detection of important vocal fold vibration features in videokymographic (VKG) images and measurement of corresponding vibration parameters. The methods can be used—in combination with verification and visual evaluation of other parameters—for medical examination of vocal folds and diagnosis of voice disorders. The automatic methods increase efficiency, precision and consistency of the evaluation process. The main contributions include development of a new methods for detection and removal of specular reflections, segmentation of the shape of rima glottidis, detection of mucosal waves by Fourier transform, and automatic evaluation of both numerical values and evaluation sheet categories of vocal fold vibration parameters. The evaluation method enables comparison of automatic and visual evaluations by categories in the VKG evaluation sheet.

The rest of the Thesis is organized as follows: Chapters 2, 3, and 4 describe the respective projects and detail the developed methods; Chapter 5 summarizes results and contributions of the Thesis, and outlines possible future perspectives.

2. Reconstruction of the growth of settled filamentous specimens in light microscopy images

Abstract:

This chapter presents a new method for reconstruction of light microscopy images of settled filamentous specimens, such as fungi and oomycetes, within intervals between observations. In phytopathology, growing microorganisms are usually observed repeatedly in a defined time sequence, leaving intervals between consecutive observation sessions undocumented. The objective of this project was to develop a method that would reconstruct images of the microorganisms during the intervals. The proposed method is based on warping of the available images by thin-plate splines; it estimates parameters of the geometric transformation by tracking of the growth of filaments in length along their morphological skeletons. The developed method was tested on real data; the comparison showed little difference between the reconstructed images and ground truth.

Keywords: filamentous fungi, growth tracking, image warping, morphological skeleton, thin-plate splines

2.1 Introduction

Phytopathogenic fungi cause severe diseases of economically important crop plants. Detailed understanding of their growth pattern could lead to increased effectiveness of the disease control. Their development is usually documented in temporally sparse observation sessions. The objective of this project was to develop a method for reconstruction of light microscopy images of settled filamentous specimens corresponding to the undocumented intervals.

In phytopathology, the growth of microorganisms is usually examined in observation sessions (see Figure 2.1), leaving the intervals between consecutive sessions undocumented (see Figure 2.2). Equipment for continuous monitoring of specimens by light microscopy over long periods of time is not only expensive, but also inconvenient for use in phytopathology: although life-imaging microscopy systems with controlled environment parameters have been introduced, they are adapted mostly for human and animal cells research, with temperature range¹ RT-55 °C inconvenient for phytopathogenic fungi. Such specimens are thus cultivated separately in optimal conditions and observed repeatedly in a defined time sequence.

The time sequence of acquired images lacks information about gradual changes in the shape of the examined specimen. In practice, the intervals between consecutive sessions are usually relatively long with respect to the speed of growth of the specimens; this is often deliberate because the manual documentation process is laborious and time-consuming. The missing parts of the growth pattern may be important for the purpose of the study, however.

¹ RT denotes room temperature.

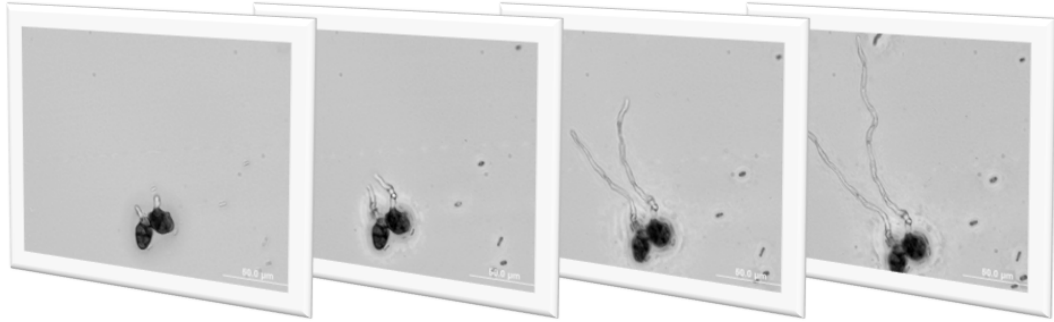


Figure 2.1: Fused light microscopy images of an *Alternaria* sp. specimen acquired at consecutive observation sessions.



Figure 2.2: Light microscopy images of an *Alternaria* sp. specimen acquired at two consecutive observation sessions showing the growth of a hypha from a conidium. Gradual elongation of the hypha during the interval between sessions was not documented.

The problem of incomplete growth pattern can be solved by two approaches. The pattern could be completed by adding new observations so as to shorten the length of intervals. However, as the growth of each microorganism is unique, this would require new experiments with appropriately short intervals. The problem can be solved by reconstructing images corresponding to the intervals from the available images. The reconstruction process consists of two steps: growth modeling and visualization.

The first reconstruction step constructs a geometric model of the growth of a specimen. The growth of settled biological specimens such as plants [48] and fungal pathogens [40] has been successfully modeled by L-system grammars [35]; L-systems are, however, convenient for synthesis of artificial specimens rather than for reconstruction of real images. The method proposed in this chapter computes a growth model that estimates the trajectory of salient control points over the interval directly from the acquired images.

Tracking of control points has to preserve the growth pattern of the specimen over the whole interval. Most methods that track control points directly on object boundaries, such as active contours [65], do not preserve the anisotropic growth pattern of microorganisms. They significantly distort positions of control points on curved boundaries, so the estimated trajectories are inaccurate. The resulting warped images thus contain unnatural deformations of curved filaments; these

errors can lead to false estimates of the growth pattern. The proposed method tracks control points along the morphological skeleton (see Subsection 2.2.1) of the specimen, instead.

The second reconstruction step visualizes the computed growth model for arbitrary times within the interval. There are principally two visualization techniques: image rendering and image warping. Whereas images synthesized by standard rendering methods appear artificial, warped images preserve the natural appearance of real images. The visualization step in the presented method is based on image warping.

Image warping can geometrically transform an image from either of the two consecutive sessions to correspond to an arbitrary time within the interval between them. The gradual growth of the specimen over the whole interval can be visualized by a sequence of warped images representing successive time frames within the interval. Image warping is used in image processing for rectification of geometric distortions [69].

The gradual growth could be smoothly visualized by image morphing [64]. Morphing first warps images from two consecutive observation sessions so that they correspond to a particular time within the interval between them. Then it composes the warped images into a single blended image, in ratio corresponding to the relative time within the interval. Image morphing is used in computer graphics for generating artificial motion sequences [13] or smooth transitions between objects [1], as well as for mapping image textures onto 3D objects [18].

This project focuses on the growth of settled specimens with filamentous growth patterns, such as fungi and oomycetes (see Figures 2.4 and 2.11). Whereas their filaments elongate over time, their growth in width is negligible, and the shape of the already developed parts remains mostly unchanged. Moreover, the filaments develop practically independently of each other. The proposed method is based on these assumptions.

The rest of this chapter is organized as follows: Section 2.2 contains definitions of the morphological skeleton and thin-plate splines; Section 2.3 describes steps of the proposed method; Section 2.4 presents results of the proposed method; Section 2.5 compares the results with ground truth, discusses properties and limitations of the method and outlines its possible modifications; and Section 2.6 summarizes this chapter.

2.2 Applied image processing tools

This section describes details of two standard image processing tools employed by the proposed method. Subsection 2.2.1 describes computation of the morphological skeleton by parallel thinning; Guo and Hall [14, Section 3] describe the skeletization algorithm in further detail. The method uses morphological skeleton for tracking of the growth of filaments in length (see Section 2.3). Subsection 2.2.2 describes image warping by thin-plate splines; Rohr [38] details theory and application of thin-plate splines in image processing. The method uses thin-plate splines for geometrical transformation of images from consecutive observation sessions to reconstruct images corresponding to the interval between them (see Section 2.3).

2.2.1 Morphological skeleton

Morphological skeleton (see Figure 2.6), or centerline, is a line representation of the shape of a segmented object (see Figure 2.5). The exact shape of the morphological skeleton depends on its definition². A common intuitive definition is based on the principle of prairie fire: if the object starts burning at boundaries and the fire spreads with uniform speed within the object, the skeleton is defined by points where fires from different directions meet. The morphological skeleton thus corresponds to the medial axis of the object.

The morphological skeleton efficiently represents the shape of the object in three aspects. Firstly, it reduces dimension of the object. Secondly, it is topologically equivalent to the object. Thirdly, it is invariant to similarity geometric transformation, i.e. a geometrically transformed skeleton of an object equals the skeleton of the geometrically transformed object

$$T(S(f)) = S(T(f)),$$

where $f : \mathbb{R}^2 \rightarrow \{0, 1\}$ is a binary image, $T : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ is a similarity transformation, and S is a skeletization transformation.

Points of an object and its morphological skeleton are categorized according to their properties. An object consists of simple and non-simple points; removal of a non-simple point changes the topology of the object, whereas removal of a simple point preserves the topology. A morphological skeleton consists of line points and end points; line points and end points are skeleton points with at least two neighbors and just one neighbor, respectively.

The morphological skeleton can be computed by several skeletization methods³. Skeletization by morphological thinning [26] is currently considered the most effective method. Thinning can be defined as sequential or parallel; parallel thinning algorithms are preferred because they allow fast, parallel implementation. Parallel thinning iteratively erodes object boundaries while preserving the object topology. In each direction it deletes a point on the current boundary if and only if the point is a simple point and not an end point.

Most skeletization algorithms are sensitive to irregularities on the boundary of the segmented object. The resulting skeleton may thus contain short spurs corresponding to coarseness on boundary. Such spurs can be suppressed by three main approaches. The first approach smooths object boundaries before skeletization. The second approach prunes the spurs using knowledge about their length and number of iterations. The third approach suppresses such spurs already during iterations of the skeletization algorithm; this is considered the best approach.

Parallel thinning algorithm developed by Guo and Hall [14, Section 3] suppresses possible spurs already during the iteration process. The algorithm divides image pixels into two disjoint subsets in a checkerboard pattern. In each iteration it erodes the current boundary in two subiterations. In the first subiteration it eliminates pixels from the first subset if and only if they satisfy conditions (2.1),

² The morphological skeleton is typically defined by prairie fire, inscribed circles or two closest points on boundary.

³ Most skeletization algorithms are based on thinning, distance map or Voronoi diagram.

(2.2) and (2.3)

$$\sum_{i=1}^4 b_i = 1, \quad (2.1)$$

where

$$b_i = \begin{cases} 1, & x_{2i-1} = 0 \ \& \ (x_{2i} = 1 \vee x_{2i+1} = 1), \\ 0, & \text{otherwise} \end{cases}$$

and x_i ($i = 1, \dots, 8$) denote values of the 8 neighbors of the pixel from the right neighbor in counter-clockwise direction,

$$2 \leq \min \left\{ \sum_{k=1}^4 x_{2k-1} \vee x_{2k}, \sum_{k=1}^4 x_{2k} \vee x_{2k+1} \right\} \leq 3, \quad (2.2)$$

and

$$(x_2 \vee x_3 \vee \overline{x_8}) \ \& \ x_1 = 0. \quad (2.3)$$

In the second subiteration it eliminates pixels from the second subset if and only if they satisfy conditions (2.1), (2.2) and (2.4)

$$(x_6 \vee x_7 \vee \overline{x_4}) \ \& \ x_5 = 0. \quad (2.4)$$

The skeletization algorithm terminates when an iteration does not eliminate any pixel. The algorithm is relatively robust to coarseness on object boundaries.

The morphological skeleton consists of branches. The branches are divided by line points called branch points. The position of branch points can be located by standard morphological operations [16, 25]. Removal of branch points divides the skeleton into separate branches; if the skeleton contained spurs caused by coarseness of object boundaries, they would be falsely identified as branches.

2.2.2 Thin-plate splines

Thin-plate splines [38] define a locally varying geometric transformation that smoothly warps an image by mapping control points from their original positions $\{\mathbf{p}_i\}_{i=1}^n$ to new positions $\{\mathbf{q}_i\}_{i=1}^n$. The transformations are restricted to the space of functions on \mathbb{R}^2 for which all second order partial derivatives are square integrable⁴. Interpolating thin-plate splines define a continuous geometric transformation

$$\mathbf{u} : \mathbb{R}^2 \rightarrow \mathbb{R}^2$$

that interpolates control points

$$\mathbf{q}_i = \mathbf{u}(\mathbf{p}_i), \quad i = 1, \dots, n,$$

so that their transformed original positions $\mathbf{u}(\mathbf{p}_i)$ exactly match the corresponding new positions \mathbf{q}_i , and that minimizes the quadratic variation “bending energy” functional

$$\min_{\mathbf{u} \in W^{2,2}(\mathbb{R}^2)} J(\mathbf{u}) = J(u_1) + J(u_2),$$

⁴ $\mathbf{u} \in W^{2,2}(\mathbb{R}^2)$, i.e. $\iint_{\mathbb{R}^2} \left(\frac{\partial^2 \mathbf{u}}{\partial x^2} \right)^2 + \left(\frac{\partial^2 \mathbf{u}}{\partial x \partial y} \right)^2 + \left(\frac{\partial^2 \mathbf{u}}{\partial y^2} \right)^2 dx dy < \infty$

where

$$J(u_j) = \iint_{\mathbb{R}^2} \left(\frac{\partial^2 u_j}{\partial x^2} \right)^2 + 2 \left(\frac{\partial^2 u_j}{\partial x \partial y} \right)^2 + \left(\frac{\partial^2 u_j}{\partial y^2} \right)^2 dx dy, \quad j = 1, 2.$$

Its solution consists of an affine and an elastic part and can be expressed as

$$u_j(\mathbf{x}) = \mathbf{A} \begin{pmatrix} 1 \\ \mathbf{x} \end{pmatrix} + \sum_{i=1}^n w_i \phi(\mathbf{x}, \mathbf{p}_i), \quad j = 1, 2, \quad (2.5)$$

where \mathbf{x} is a 1×2 vector of coordinates in the warped image, \mathbf{A} is a 2×3 matrix of affine transform coefficients, n is the number of control points, $\phi(\mathbf{x}, \mathbf{p}_i)$ are basis functions

$$\phi(\mathbf{x}, \mathbf{p}_i) = \phi(r_i) = \begin{cases} r_i^2 \ln r_i, & r_i \neq 0, \\ 0, & r_i = 0, \end{cases}$$

where $r_i = \|\mathbf{x} - \mathbf{p}_i\|$ is the Euclidean distance between points \mathbf{x} and \mathbf{p}_i , and w_i are weights of the basis functions. The value of a basis function $\phi(\mathbf{x}, \mathbf{p}_i)$ depends solely on the Euclidean distance r_i of its argument \mathbf{x} from its center \mathbf{p}_i . The thin-plate splines are thus radially symmetric; they belong to the more general class of radial basis functions. The minimum number of control points is three. In case of three control points, however, the transformation is defined only by the affine part; the elastic part is active only in case of at least four control points. The solution is unique under the condition that the points do not lie on a single line. Coefficients \mathbf{A} and w_i of the solution can be computed by placing control points \mathbf{q}_i into (2.5) and solving the resulting set of linear equations.

2.3 Method

The objective of the method described in this section is to reconstruct light microscopy images of a growing specimen within an interval between consecutive observations. The method (see Figure 2.3) reconstructs the missing images from the available ones by appropriate geometric transformations. In order to establish parameters of the transformations, the method selects a number of salient control points in the images acquired at the beginning and the end of the examined interval and estimates their movement during the interval. The images are then geometrically transformed so that the selected control points are mapped to their estimated positions in the interval. The result is a sequence of images showing the growth of the specimen during the undocumented interval.

The method works with light microscopy images from consecutive observation sessions (see Figures 2.4 and 2.11) and corresponding binary images of the shape of the specimen (see Figures 2.5 and 2.12). The shape can be usually computed from the microscopy images by standard segmentation algorithms; the segmentation is highly dependent on the processed data, however.

First, the method computes morphological skeletons (see Figures 2.6 and 2.13) of the segmented binary images. The skeletonizing algorithm should be robust to boundary coarseness so that the skeleton does not contain short spurs and

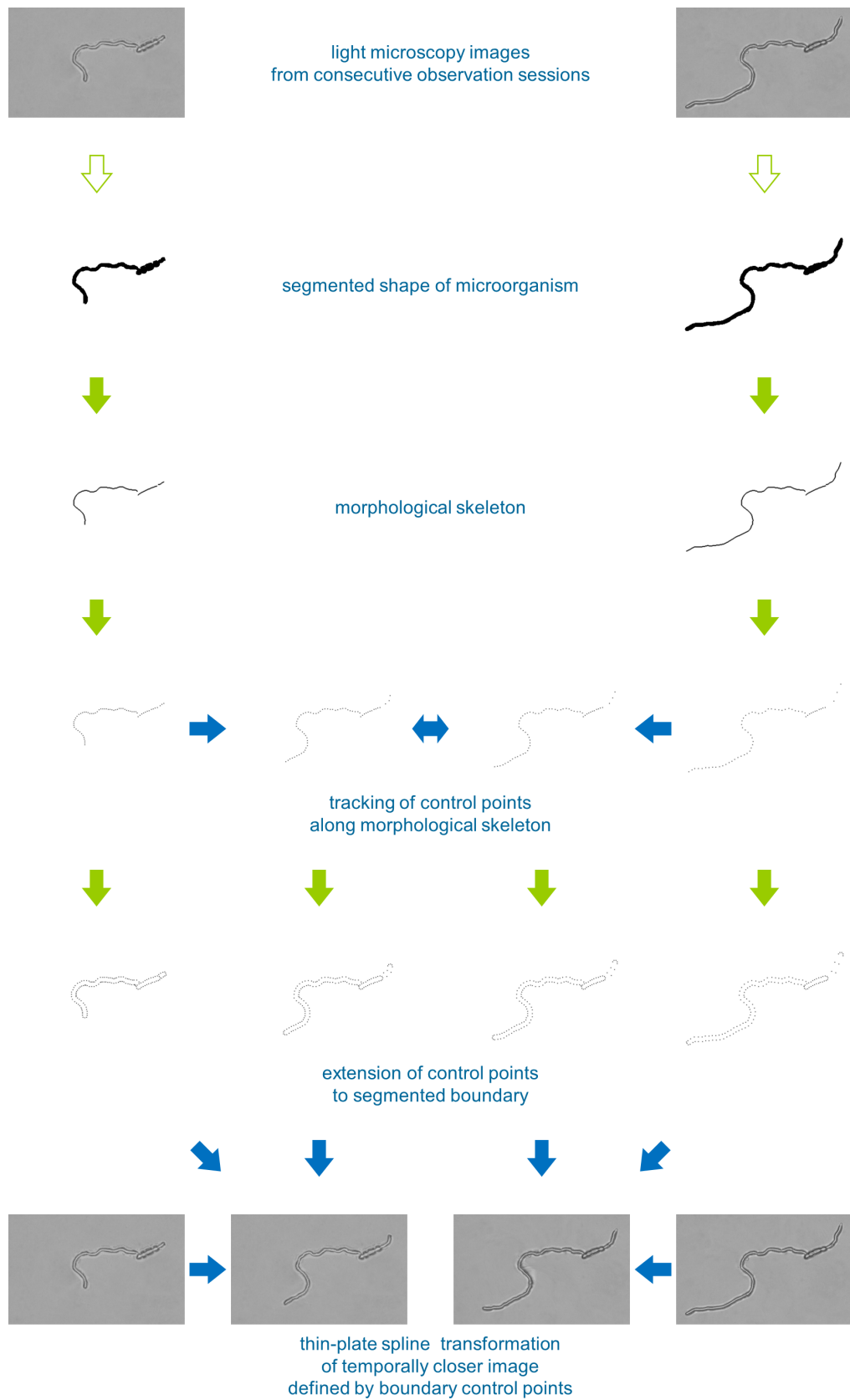


Figure 2.3: Steps of the developed method for reconstruction of light microscopy images of growing specimens corresponding to times within the interval between two consecutive observation sessions.

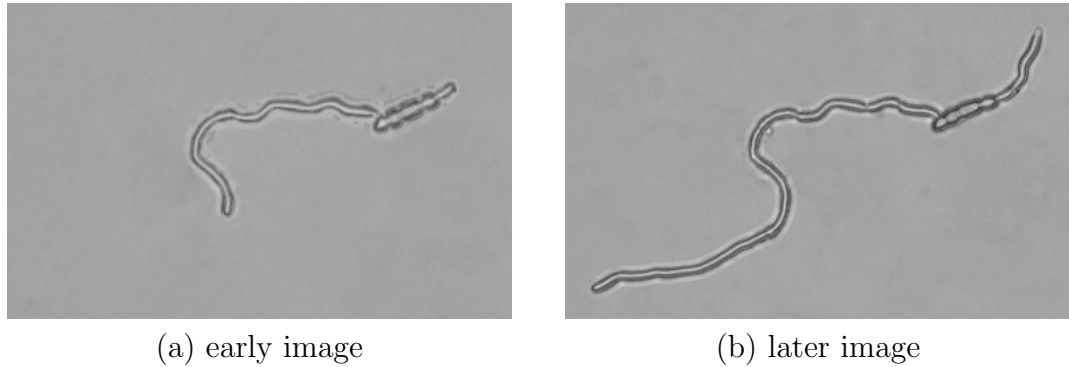


Figure 2.4: Preprocessed light microscopy images of a *Fusarium oxysporum* f.sp. *pisi* specimen acquired at two consecutive observation sessions, showing the development of two hyphae from a conidium and their growth in length; the preprocessing steps included flat-field correction, displacement rectification, multi-focal fusion, and debris suppression.

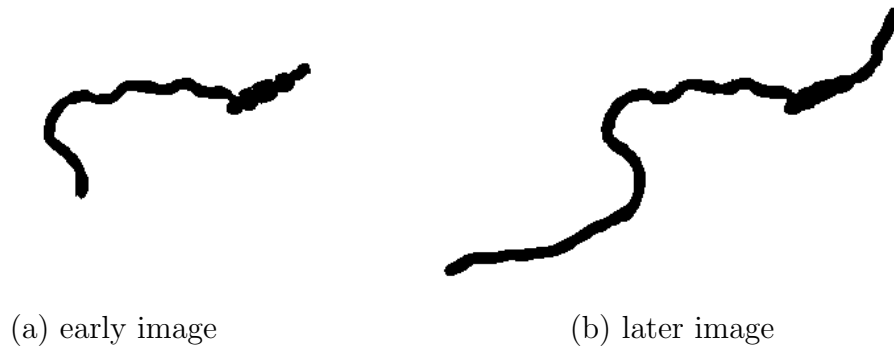


Figure 2.5: Binary images of the shape of the specimen from Figure 2.4 segmented by thresholding of local brightness variance.

distorted line endings. Branches of the skeleton correspond to filaments of the specimen.

Second, the method divides the skeletons into branches. Branches are parts of the skeleton between branch points that correspond to points of hypha branching or to points between a conidium and a hypha.

The method makes three assumptions on the skeletons from two consecutive observation sessions. Firstly, the number of branches in both skeletons should be the same. This means that no new filaments evolved during the interval between the observations; short branches in the second skeleton with no counterparts in the first skeleton can be neglected. Secondly, since the already grown parts do not move or bend, a part of the second skeleton should roughly overlap with the whole first skeleton. Thirdly, each branch elongates uniformly over time. For settled specimens with filamentous growth patterns, these assumptions are usually satisfied.

Third, the method selects control points on the morphological skeletons (skeleton CPs) and estimates their trajectories over the interval between observations (see Figures 2.7 and 2.14). The morphological skeleton is less sensitive to curving

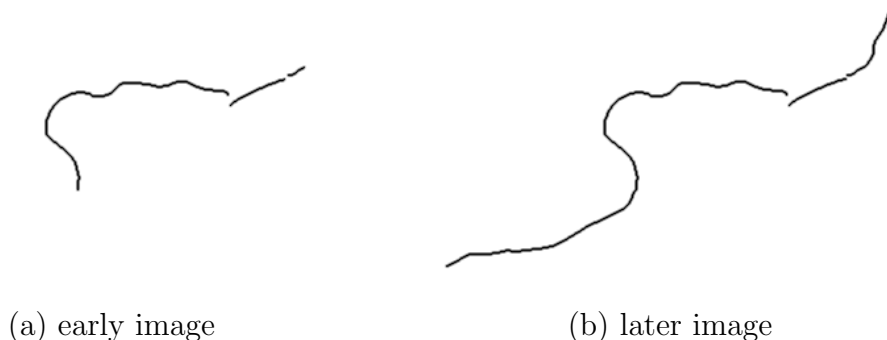


Figure 2.6: Morphological skeletons computed from the segmented images in Figure 2.5 by parallel thinning and separated into branches by morphological operations.

deformations than the boundary. The skeleton CPs are selected equidistantly along the branches of the second skeleton (see Figures 2.7f and 2.14f). Each branch is then gradually shrunk from its loose end along its shape to the length of the corresponding branch in the first skeleton (see Figures 2.7a and 2.14a); the skeleton CPs are moved correspondingly, keeping the Euclidean distance between them uniform. This process estimates trajectories of skeleton CPs from the end of the interval to its beginning. It simulates, in reverse, the growth of the specimen in length during the interval; the reverse direction of tracking solves the problem of inaccurate segmentation. In practice, the two skeletons usually do not overlap precisely; the proposed method, however, does not take the exact shape of the first skeleton into account.

Fourth, the method computes control points on the boundary of the specimen (see Figures 2.8 and 2.15). A geometric transformation defined just by skeleton CPs cannot preserve the shape of the specimen outside the skeleton. The method thus replaces each skeleton CP by two control points on the boundary of the specimen (boundary CPs), in the direction perpendicular to the local direction of the skeleton; each skeleton CP thus approximately bisects the line segment between two corresponding boundary CPs. The length of the line segment corresponds to the local thickness of the filament; it is interpolated over the interval as a weighted average of the lengths of the segment at the two consecutive sessions. A geometric transformation defined by boundary CPs can preserve the shape of the whole specimen.

Fifth, the method warps the temporally closer image to represent an arbitrary time within the interval (see Figures 2.9 and 2.16). The aim of the geometric transformation is to map the boundary CPs from the available image to their positions at the selected time. Due to the spatially local character of the growth process, the transformation should be sensitive to local changes. Elastic types of geometric transformations, such as thin-plate splines (see Subsection 2.2.2), can be used for this purpose. The transformation is defined by the positions of boundary CPs in the selected time within the interval and in the temporally closer image. The warped image is an estimate of the light microscopy image that would be acquired were the specimen observed at the selected time (see Figures 2.10 and 2.17).

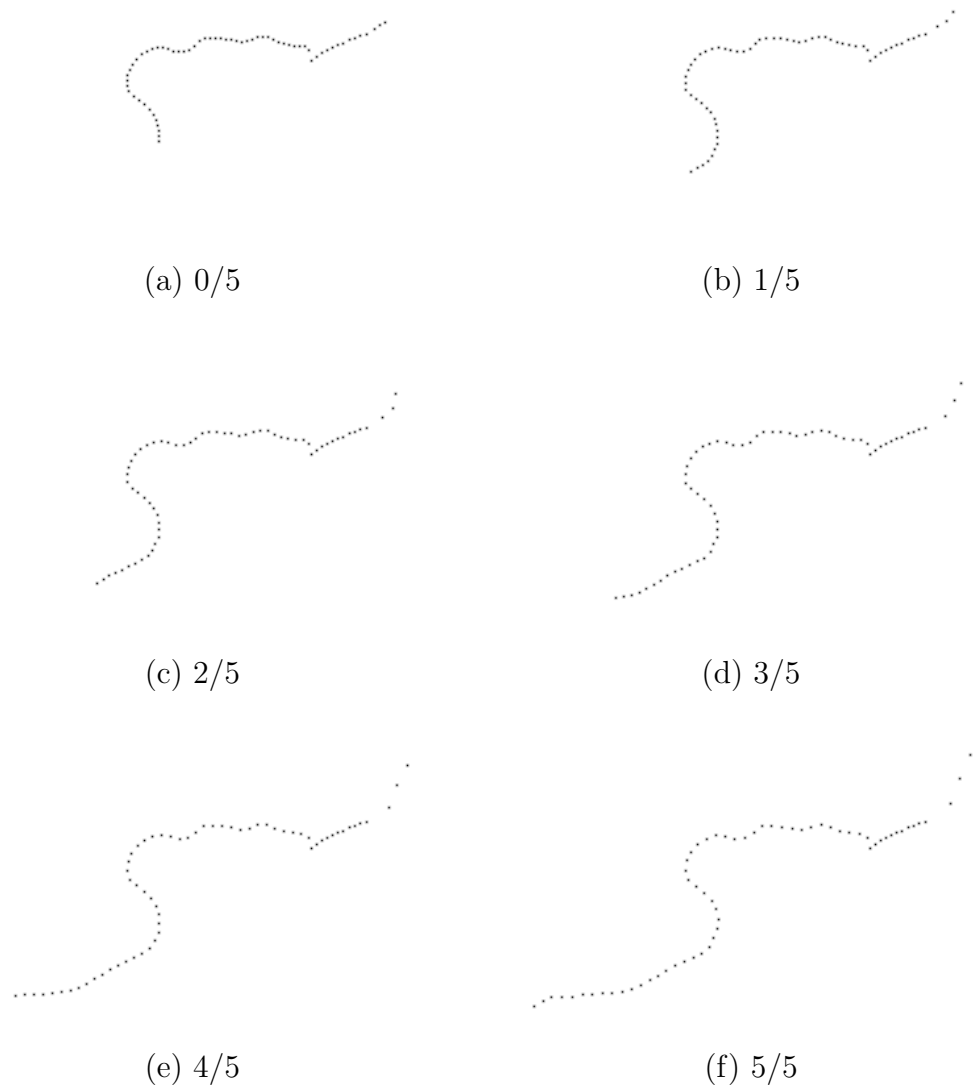
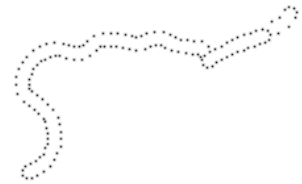


Figure 2.7: Tracking of skeleton control points within the interval between acquisitions of images in Figure 2.4; their positions were computed by shrinking of the filaments of the morphological skeleton from (f) the latter observation session, see Figure 2.6b, to the length of the corresponding filaments from (a) the early session, see Figure 2.6a.



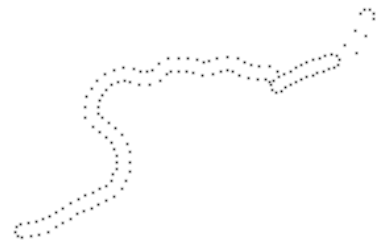
(a) 0/5



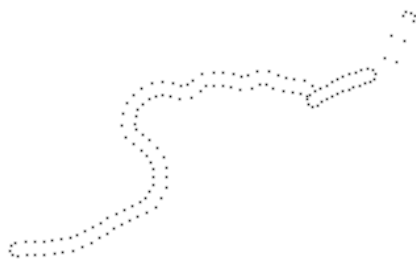
(b) 1/5



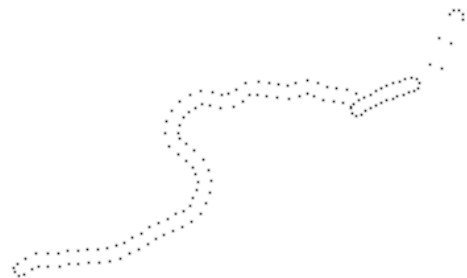
(c) 2/5



(d) 3/5



(e) 4/5



(f) 5/5

Figure 2.8: Boundary control points computed for the purpose of image warping from the skeleton control points in Figure 2.7.

2.4 Results

Performance of the proposed method was tested on a set of light microscopy images of the early development of *Fusarium oxysporum* f.sp. *pisi* and *Alternaria* sp.⁵ *Fusarium* [66] and *Alternaria* [6] spp. (Hyphomycetes, Deuteromycotina) are phytopathogenic fungi with a worldwide distribution. They cause severe diseases in a wide range of economically important crop plants. Both species spread by asexual spores—conidia. In proper environmental conditions, particularly temperature and humidity, a conidium germinates by hyphae to form a mycelium. The growth rate of a mycelial colony in optimal conditions is approximately 5–10 millimeters per day. Detailed understanding of the growth parameters could contribute to the increasing effectiveness of the disease control.

Image degradations introduced in the acquisition process were eliminated by several preprocessing steps; a detailed description of these steps is given in [41]. First, flat-field correction was applied to the images. Flat-field is a degradation caused by a non-uniformity in illumination of light microscopy samples; it appears as a gradual decrease in brightness from the center of an image to its borders. Since a microscopy image without a sample was not available, the shape of flat-field was roughly estimated from the background in available images, and the computed deficiency in brightness was added to the degraded images. Second, images from different focal planes were fused. As the observed specimens were thicker than the attainable depth of field—and thus always partly out of focus—each of them was photographed at several focal planes. These images were composed by means of a multi-focal fusion method [62] into one all-focused image. Third, a geometrical alignment was applied to displaced images from consecutive sessions. Since microscope slides with the specimens were positioned manually, a temporal image registration [69] was necessary to compensate for the resulting shift and rotation between images from different observation sessions. The displacements were rectified by a similarity transformation; its parameters were computed from a small set of manually placed control points. Fourth, debris was detected and removed from the images. Debris particles were segmented by thresholding of the local variance of brightness and their shapes adjusted by standard morphological operations [46]; detected particles were removed by diffusion from their boundary [52]. The result of the preprocessing process were mutually aligned images with uniform illumination and the whole specimen in focus (see Figures 2.4 and 2.11).

Shapes of the specimens were segmented from image background in the pre-processed images by thresholding of local variance in brightness. The observed specimens were semi-transparent and thus differed from image background in local changes of brightness rather than in brightness itself. Small irregularities in the shape were rectified by morphological closing [46]. The result were binary images of the segmented shapes of the specimens (see Figures 2.5 and 2.12).

The reconstruction of missing images was performed by the method described in Section 2.3 (see Figures 2.3, 2.6, 2.7, 2.8, 2.9a, 2.13, 2.14, 2.15, and 2.16a).

⁵ The specimens were incubated on Czapek-Dox agar at 4 and 20 °C and their growth documented in intervals of approximately 6 and 2 hours, respectively. The images were acquired by a CCD digital camera attached to a conventional light microscope with a 100× and 20× magnification, respectively.

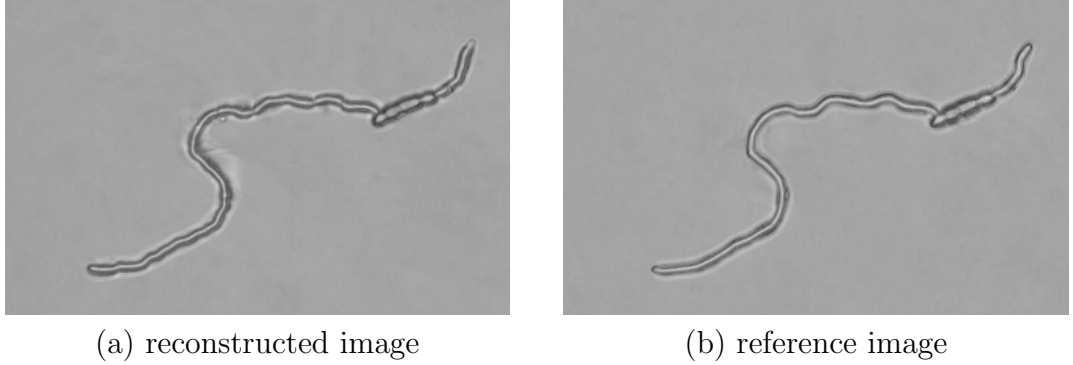


Figure 2.9: (a) Image from Figure 2.4b warped by thin-plate splines that mapped boundary control points from Figure 2.8f to the corresponding boundary control points from Figure 2.8d. (b) A preprocessed light microscopy image acquired for the purpose of comparison at the corresponding stage of growth.

It used preprocessed images from two consecutive observation sessions and the corresponding binary images to generate images representing the growth of the observed specimen during the interval. The morphological skeletons were acquired from the segmented images by means of the parallel thinning algorithm developed by Guo and Hall [14, Section 3] (see Subsection 2.2.1). The distance between the skeleton CPs was approximately a half of the average width of the filament. In order to preserve the shape of non-growing round objects, such as conidia, additional control points were automatically selected on their boundary and added to the set of boundary CPs (see Figure 2.15). The preprocessed images were geometrically transformed by means of thin-plate splines (see Subsection 2.2.2). Because of different acquisitions conditions, e.g. use of different focal planes in different observation sessions, and inaccuracies in registration and segmentation, a blended image would include double-exposure effects. Hence just the temporally closer image (see Figures 2.4b and 2.11b) was transformed (see Figures 2.9a and 2.16a).

The reconstructed images were compared with reference light microscopy images acquired for the purposes of comparison at corresponding times (see Figures 2.9b and 2.16b).

The reconstructed and reference images were visually compared in checkerboard (see Figures 2.17a and 2.10a) and difference images (see Figures 2.17b and 2.10b). Checkerboard image

$$f^{board}(x, y) = \begin{cases} f^{rec}(x, y), & \max(x \bmod 2s, y \bmod 2s) < s \\ \vee \min(x \bmod 2s, y \bmod 2s) \geq s, \\ f^{ref}(x, y), & \text{otherwise,} \end{cases}$$

is an image composed of tiles of $s \times s$ pixels alternately taken from reconstructed image f^{rec} and reference image f^{ref} . Difference image

$$f^{diff}(x, y) = |f^{rec}(x, y) - f^{ref}(x, y)|$$

is an image of the absolute pixel-wise difference of brightness values between reconstructed image f^{rec} and reference image f^{ref} .

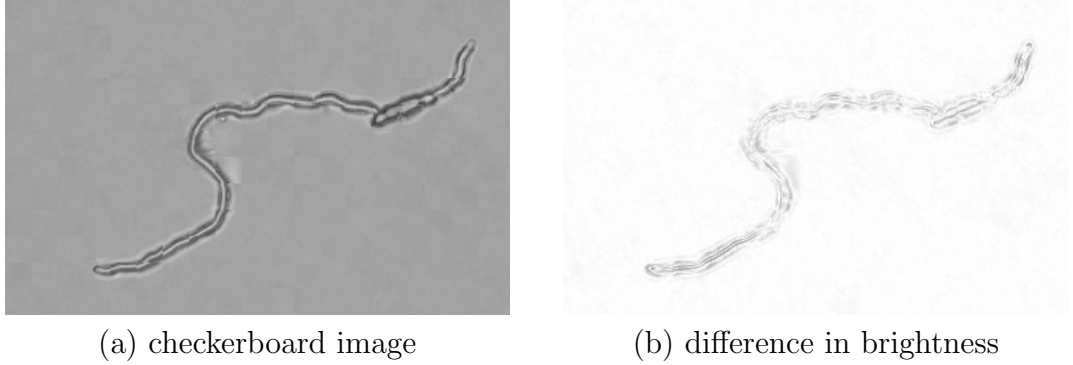


Figure 2.10: Comparisons of the warped image in Figure 2.9a with the reference image in Figure 2.9b: (a) a checkerboard image composed of square tiles alternately taken from the compared images, (b) an image of pixel-wise difference in brightness.

The reconstructed and reference images were numerically compared by differences between positions of corresponding boundary CPs. Boundary CPs in the reconstructed image were acquired by the proposed method from the light microscopy images acquired at the consecutive observation sessions. Boundary CPs in the reference image were acquired by the same method from the reference image itself, using the same number of boundary CPs as in the reconstructed image. The error of match between corresponding boundary CPs in the reconstructed and reference images was expressed by their Euclidean distance

$$\|\mathbf{x}_i^{rec}, \mathbf{x}_i^{ref}\| = \sqrt{(x_i^{rec} - x_i^{ref})^2 + (y_i^{rec} - y_i^{ref})^2}, \quad i = 1, \dots, n,$$

where $\mathbf{x}_i = (x_i, y_i)$ are coordinates and n the number of boundary CPs in the images, and the overall error of match was expressed by mean value

$$\frac{1}{n} \sum_{i=1}^n \|\mathbf{x}_i^{rec}, \mathbf{x}_i^{ref}\|.$$

(see Table 2.1). Note, however, that the error depends on the accuracy of segmentation and on computation of the morphological skeleton in discrete domain.

number of boundary CPs	average difference	median difference
63	0.9 ± 0.83	0.74

Table 2.1: Euclidean distance (in pixels) between corresponding boundary control points (boundary CPs) in the reconstructed image (see Figure 2.16a) and the corresponding reference image (see Figure 2.16b).

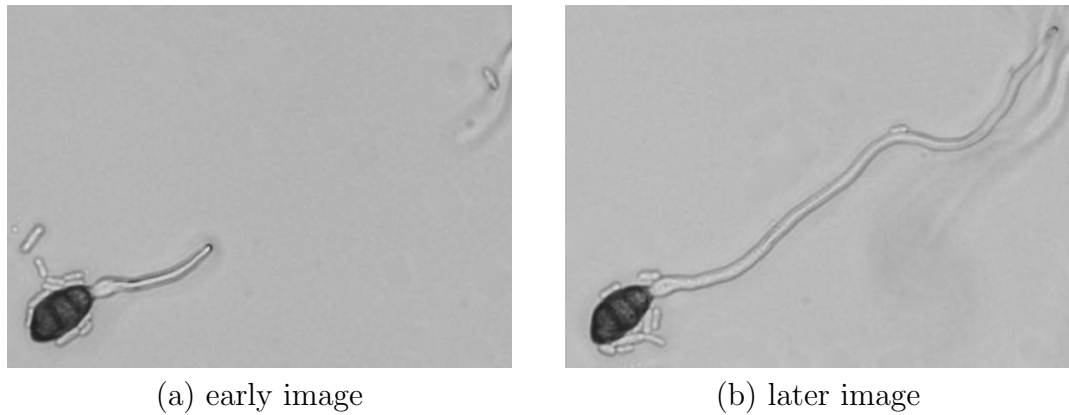


Figure 2.11: Light microscopy images from Figure 2.2 after preprocessing, namely displacement rectification and debris suppression.

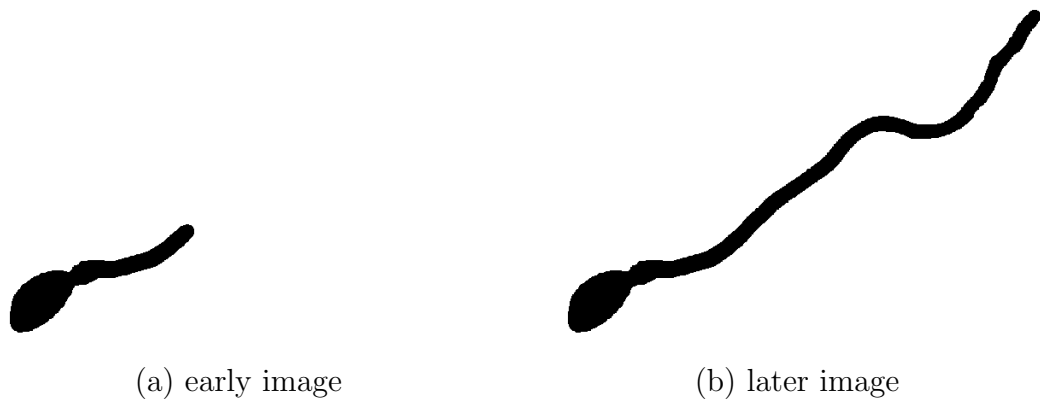


Figure 2.12: Binary images of the shape of the specimen from Figure 2.11 segmented by thresholding of local brightness variance.

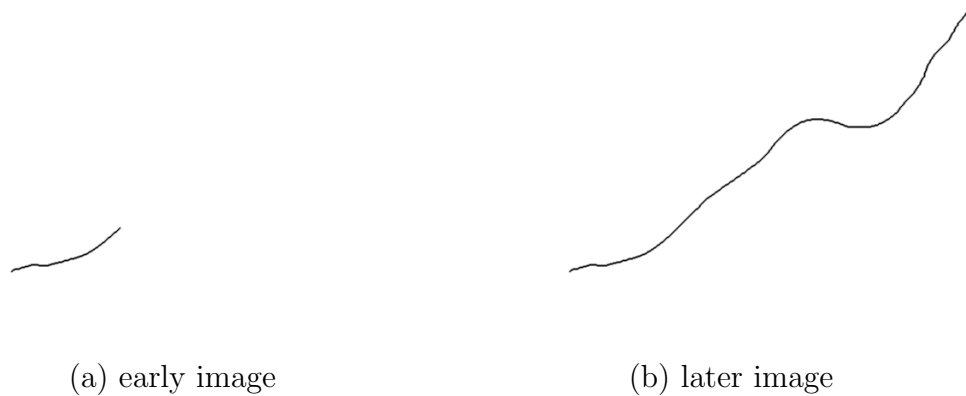


Figure 2.13: A branch corresponding to the elongating hypha in morphological skeletons computed from the segmented images in Figure 2.12.



Figure 2.14: Tracking of skeleton control points within the interval between acquisitions of images in Figure 2.11; their positions were computed by shrinking of the filament of the morphological skeleton from the latter observation session (f), see Figure 2.13b, to the length of the filament from the early session (a), see Figure 2.13a.

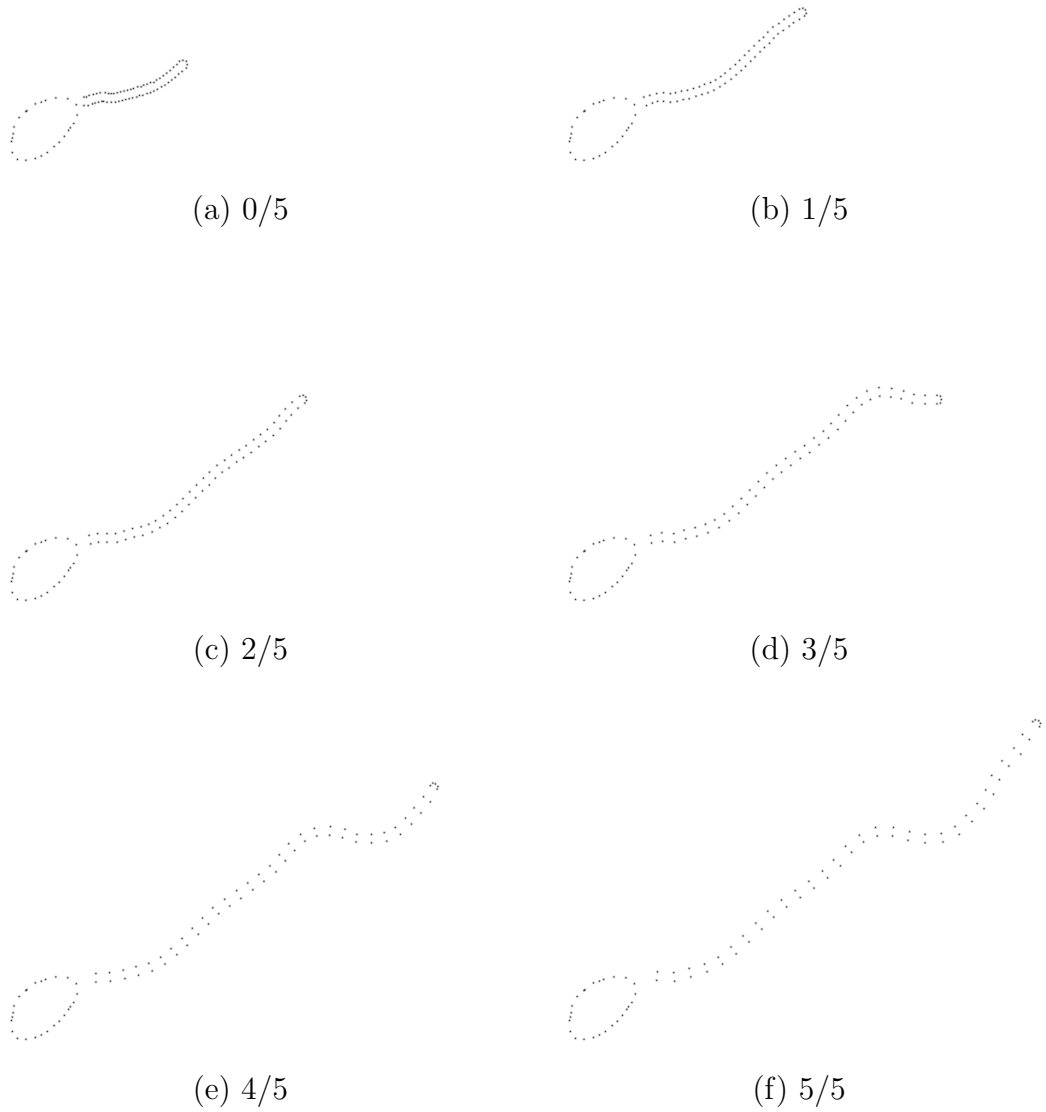


Figure 2.15: Boundary control points computed for the purpose of image warping from the skeleton control points in Figure 2.14 and control points automatically selected on the boundary of the conidium.

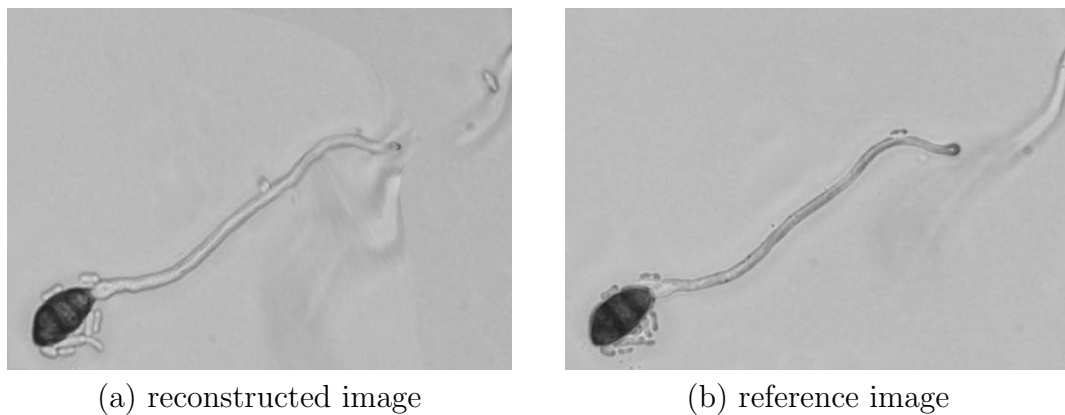


Figure 2.16: (a) Image from Figure 2.11b warped by thin-plate splines that mapped boundary control points from Figure 2.15f to the corresponding boundary control points from Figure 2.15d. (b) A preprocessed light microscopy image acquired for the purpose of comparison at the corresponding stage of growth.

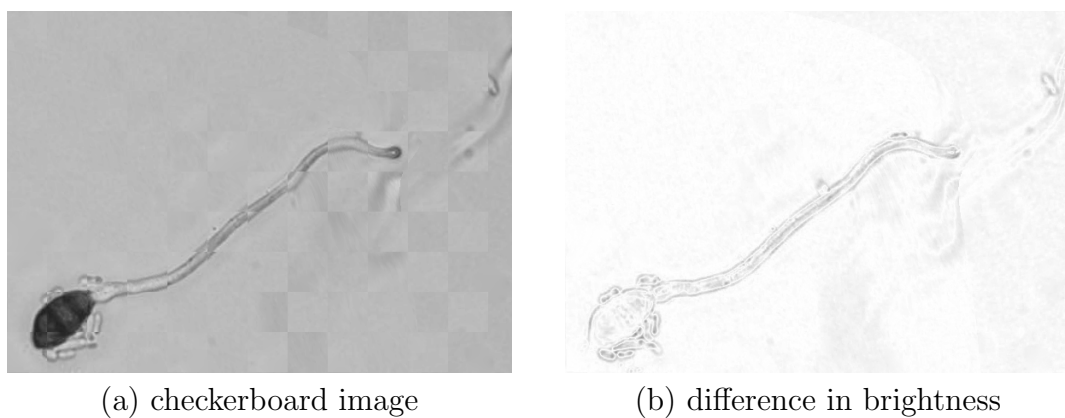


Figure 2.17: Comparisons of the warped image in Figure 2.16a with the reference image in Figure 2.16b: (a) a checkerboard image composed of square tiles alternately taken from the compared images, (b) an image of pixel-wise difference in brightness.

2.5 Discussion

Effectiveness of the developed method was illustrated by comparison of the reconstructed images with corresponding light microscopy images. The checkerboard and difference images showed that the synthetically computed images closely matched the corresponding references image without significant deformations (see Figures 2.10 and 2.17). The numerical comparison of errors in match of corresponding boundary CPs in the reconstructed and reference images also indicated good geometrical match.

The method allows to generate a sequence of images showing the gradual growth of the specimen over the whole interval. The warped images should be morphed only if the geometric transformation is estimated with high accuracy, however. If the features are not aligned precisely, the combination of two images might contain disturbing double-exposure effects; in such a case, just the temporally closer image is warped.

The accuracy of warping is affected by two main factors. Firstly, higher accuracy can be achieved by short distances between CPs. The more CPs are used, the more accurate the mapping function is; on the other hand, if the spaces between CPs are too short, the warped image may include local distortions. Secondly, the appearance of warped images is affected by the type of the geometric transformation. Although radial basis functions (see Subsection 2.2.2) are formally of a global nature, i.e. for every pixel in the warped image all basis functions ϕ_i ($i = 1, \dots, n$) are taken into account, they can successfully model also local deformations. The result depends on the type of the radial basis functions ϕ . Thin-plate splines proved effective for the warping.

Although the method was designed for images of settled filamentous specimens uniformly elongating over time, it could be modified to deal with more complicated growth patterns. Small growth in width can be simulated by interpolating the local width of filaments. The apparition of new branches could be partly modeled as well. If a new branch is very short in one of the acquired images, it is neglected for reconstruction of the interval before the acquisition but taken into account in the interval after the acquisition. If not, the time when the new branch started to develop must be somehow estimated; its growth in the interval before its apparition is reconstructed by warping of just the later image so that from the beginning of the interval to the estimated time, the length of the new branch equals half the width of the filament.

If the speed of growth is not uniform, additional information is necessary to estimate the changes in the speed during the examined interval for each filament separately. This can be estimated from images from more than two observation sessions. As only the length of branches is utilized from the first skeleton, the method could be used, to some extent, to generate images of the specimen in times before the acquisition of the first image; the resulting images could appear distorted, however.

In some cases, however, the results of the proposed method may include significant deformations. The method can be used for realistic modeling only if the growth consists of stretching and curving of filaments and the already developed parts do not change in shape. If not only elongation but also significant movement is a part of the growth process, a more complex method should be applied.

The problem of occlusion, such as overlapping of filaments, has not been solved either. The ability of the proposed method to preserve textures is also limited. If a significant texture is present, it may appear deformed in the warped image.

2.6 Conclusion

This chapter has introduced a new method for reconstruction of light microscopy images showing the growth of settled filamentous specimens in intervals between observations. It was developed for the purpose of completing time studies of settled and relatively slow-growing specimens with filamentous growth patterns, such as fungi and oomycetes. In principle, it can be used for any objects with such characteristics. The method is based on growth tracking along the morphological skeleton and image warping by thin-plate splines. It generates realistic images corresponding to arbitrary times within the undocumented intervals. The method is suitable for biomedical data as it does not introduce significant unnatural deformations. Its performance was successfully tested on light microscopy images of *Fusarium oxysporum* and *Alternaria* sp. germination and mycelium growth. The natural appearance of the artificially generated images and their high correlation with ground truth proved suitability of the method for the purposes of the study. The method was presented at the *Analysis of Biomedical Signals and Images BIOSIGNAL 2006* conference in Brno, Czech Republic [42], and published in 2008 in *EURASIP Journal on Advances in Signal Processing* [43].

Acknowledgment

The work was partly supported by grant projects GAUK 148207, MŠMT 1M0572, and MSM 6198959215. Fungal strain UPOC-FUN-161 was kindly provided by and its cultivation performed and documented in the Laboratory of Phytopathology, Department of Botany, Faculty of Science, Palacký University, Olomouc, Czech Republic, with special acknowledgement to Michaela Sedlářová and Drahomíra Vondráková for cultivation and documentation.

3. Measurement of particles in atomic force microscopy images

Abstract: The objective of this project was to measure the average size of particles in atomic force microscopy (AFM) images of phenylpyridyldiketopyrrolopyrrole (PPDP), an organic semiconducting material. Atomic force microscopy is an imaging technique convenient for measurement of solid body surfaces with very high resolution; pixel values in the acquired images represent the measured topography. PPDP samples consist of similarly sized elliptical particles; the size of particles in an image can be thus characterized by the average length and width computed from a relatively small subset of particles. This chapter describes a new method for automatic detection and measurement of PPDP particles in AFM images. The method is based on detection of salient particles and approximation of their shapes by ellipses; it estimates the length and width of each detected particle by the major and minor axes, respectively, of the corresponding approximating ellipse. The method is robust to distortions in AFM images. Results of the automatic method were compared with results of manual measurements. The comparisons indicated that the proposed automatic method could be used in place of time-consuming manual measurements.

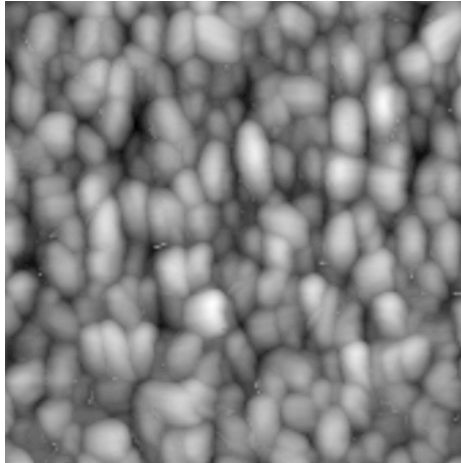
Keywords: atomic force microscopy, ellipsoidal approximation, image moments, particle measurement, watershed segmentation

3.1 Introduction

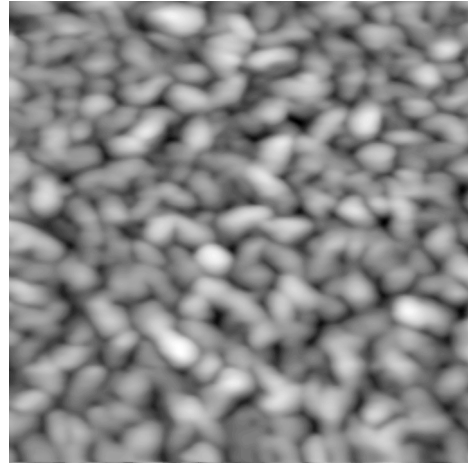
Phenylpyridyldiketopyrrolopyrrole (PPDP) is a perspective organic semiconducting material for hydrogen sensors [32, 39]. Deposited layers of PPDP consist of similarly sized, mostly spatially separated ellipsoidal particles. Topography of PPDP particles can be observed by atomic force microscopy (see Figure 3.1). Properties of PPDP samples are mainly characterized by the size of particles. Currently used manual measurement is, however, laborious and time-consuming. The objective of this project was to develop a method for automatic detection and measurement of salient particles in atomic force microscopy images of PPDP samples.

Atomic force microscope (AFM) [4] is an imaging device that allows measurement of topography of solid body surfaces with very high resolution. It gradually moves an elastic cantilever with a sharp tip close above the surface. The surface applies interactive force to the tip, which deflects the cantilever. The height of the surface is estimated from the measured deflection [31]. Atomic force microscopy allows measurement of topography without damage to the scanned surface.

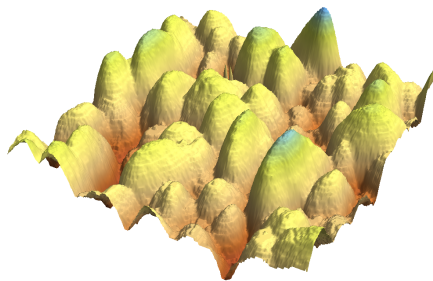
Atomic force microscopy images represent just an approximation of the surface topography. The observed image is, in fact, a convolution between the contacting tip and the surface of the sample [60]. As a result, only details larger than the diameter of the tip are observed. The exact shape of the tip is, however, unknown; it can even change during measurement, particularly due to breaking or adhesion of parts of the sample or dust particles. Furthermore, the acquired images may contain noise artifacts [31] (see Figure 3.2). Analysis of AFM images



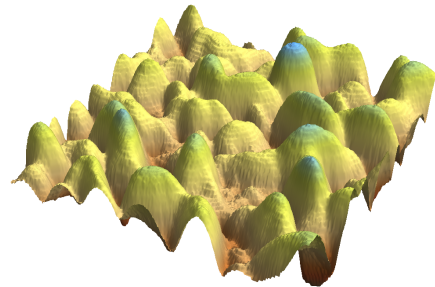
(a) AFM image 5



(b) AFM image 14



(c) AFM image 5 (detail)



(d) AFM image 14 (detail)

Figure 3.1: Atomic force microscopy (AFM) images of phenylpyridyldiketopyrrolopyrrole (PPDP) samples (a, b); pixel values linearly correspond to measured height (c, d).

should account for such distortions and use prior information about the scanned samples to compensate for them.

The objective of this project was to automatically measure the size of PPDP particles in each AFM image. PPDP samples contain similarly sized ellipsoidal particles, so the size of each particle can be characterized by its length and width, and the size of particles in an image can be characterized by their average size. Physicists thus measure just the average length and width of particles in a representative subset.

Currently used manual measurement of PPDP particles in AFM images is difficult. Because the images typically contain hundreds to thousands of particles, manual measurements are time-consuming and laborious. Physicist therefore manually measure only a relatively small subset of particles; this, however, increases inconsistency of results. Development of an automatic measurement method is thus of great interest.

Existing methods for automatic segmentation of particles in AFM images are not appropriate specifically for PPDP samples. In 1997 Villarrubia [60] developed methods for surface reconstruction in AFM images; the generally designed

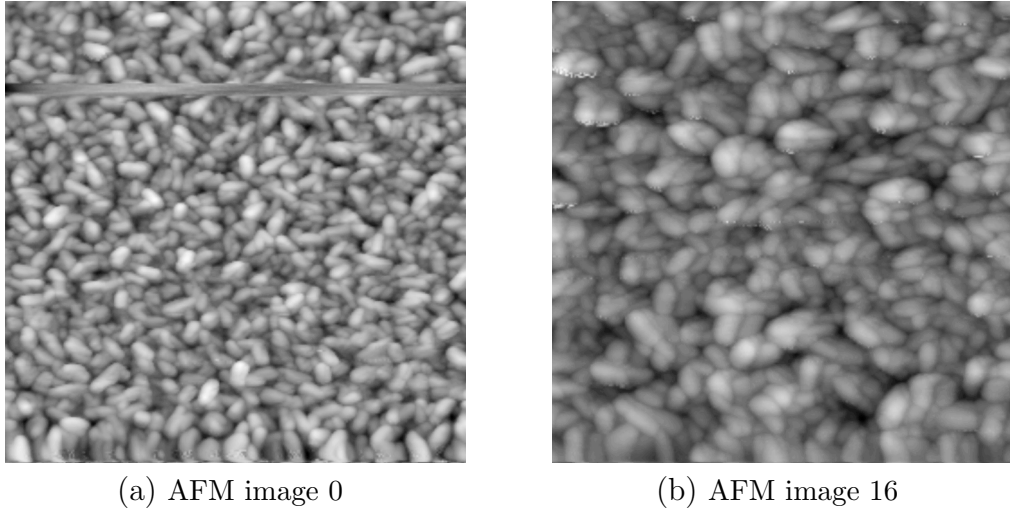


Figure 3.2: Atomic force microscopy images with typical noise artifacts.

methods, however, do not use prior information about known shape of scanned particles. Existing automatic methods for processing of AFM images [11, 51] are either sensitive to noise or designed for samples with densely packed particles; they are thus inconvenient for robust segmentation of spatially separated particles.

This chapter presents a new method for automatic measurement of ellipsoidal particles in AFM images. The developed method is based on rough segmentation of particles by watershed transform and approximation of their shapes by ellipses using image moments. It estimates the length and width of salient particles by the major and minor axes, respectively, of the approximating ellipses. The analyzed images contain hundreds of salient particles, so the method excludes partially occluded or otherwise distorted particles from measurements. The method utilizes prior information about elliptical shape of particles, increases its robustness to noise distortions.

The rest of this chapter is organized as follows: Section 3.2 describes watershed transform by immersion and image moments; Section 3.3 describes in detail the proposed method; Section 3.4 presents results of the developed method and compares them with results of manual measurements; Section 3.5 discusses properties and limitations of the developed method, and outlines its possible modifications and applications; and Section 3.6 summarizes this chapter.

3.2 Applied image processing tools

This section describes two image processing methods used by the proposed method. Subsection 3.2.1 describes watershed transform by immersion [61]; the developed method uses watershed transform for rough segmentation of particles in AFM images (see Subsection 3.3.2). Subsection 3.2.2 describes basics of image moments [12]; the developed method uses image moments for approximation of PPDP particles in watershed regions by ellipses (see Subsection 3.3.3).

3.2.1 Watershed transform

Image segmentation isolates objects in images from background. It partitions the image into disjoint regions so that they are homogeneous with respect to some property, e.g. brightness or texture. Watershed transform is a region-based segmentation method with several possible definitions. The developed method uses the watershed transform by immersion.

The principle of watershed transform by immersion can be explained on a geographic analogy. Let brightness function represent a topographic relief (see Figures 3.1c and 3.1d) with holes in local minima. The watershed transform gradually immerses the relief into a lake. As the water gradually fills up regions (basins) in the image from the local minima, the watershed transform constructs dams at points where water from different basins would meet. The result is a relief partitioned into basins separated by the dams. In the segmented image, the basins are called watershed regions and the dams are called watershed lines or watersheds.

Watershed transform by immersion can be formally defined recursively. Let $d_A(a, b)$ denote the geodesic distance between points a and b within set A , i.e. the infimum of path lengths between a and b within A . Geodesic distance of point a from subset $B \subseteq A$ within set A is defined as

$$d_A(a, B) = \min_{b \in B} d_A(a, b).$$

Let $f : D \rightarrow \{0, \dots, l\}$ denote a digital grayscale image with domain $D \subseteq \mathbb{R}^2$, and X_h denote union of basins at brightness level $h \in \{0, \dots, l\}$. The watershed lines are defined as

$$W(f) = D \setminus X_l.$$

X_h is defined recursively as

$$\begin{cases} X_0 = \{\mathbf{x} \in D \mid f(\mathbf{x}) = 0\}, \\ X_{h+1} = M_{h+1} \cup \bigcup_{i=1}^{n_h} \{\mathbf{x} \in T_{h+1} \mid d_{T_{h+1}}(\mathbf{x}, X_h^i) < d_{T_{h+1}}(\mathbf{x}, X_h^j) \forall j \in \{1, \dots, n_h\} \setminus \{i\}\} \end{cases}$$

for $h \in \{0, \dots, l-1\}$; M_h denotes the union of local minima at level h , T_h denotes threshold set

$$T_h = \{\mathbf{x} \in D \mid f(\mathbf{x}) \leq h\},$$

and X_h denotes the union of basins computed at level h

$$X_h = \bigcup_{i=1}^{n_h} X_h^i,$$

where X_h^i denotes the i^{th} watershed basin and n_h the number of basins at level h .

The algorithm is described in detail in [61]. Properties of the algorithm and its formal definitions are discussed in [37].

3.2.2 Image moments

Image moments [12] are projections of a grayscale image function $f : \mathbb{R}^2 \rightarrow \mathbb{R}$, i.e. a finite piece-wise continuous real function of two variables $\{x, y\} \in \mathbb{R}^2$ and with compact support, to polynomial basis $\{x^p, y^q\}_{p,q=0}^{\infty}$. They have many applications in image processing, particularly in object representation and recognition.

Geometric moment

$$m_{pq} = \iint_D f(x, y) x^p y^q dx dy, \quad p, q \in \mathbb{N}_0$$

is a projection of image function f with compact support $D \subseteq \mathbb{R}^2$ onto polynomial basis (x^p, y^q) ; the order of the moment is $(p + q)$. Zero-order moment m_{00} equals the volume of f and $x_t = m_{10}/m_{00}$ and $y_t = m_{01}/m_{00}$ are coordinates of its centroid; translation of the image by vector $(-x_t, -y_t)$, centers the image in zero.

Central geometric moment

$$\mu_{pq} = \iint_D f(x, y) (x - x_t)^p (y - y_t)^q dx dy$$

of order $(p + q)$ equals the geometric moment m_{pq} for a centered image; it is thus invariant to translation of image f . Central moments μ_{20} and μ_{02} represent the variance of f in x and y axes, respectively, and central moment μ_{11} represents the co-variance between them.

A normalized position of image f is acquired by rotation of the centered image around zero by angle

$$\alpha = \begin{cases} \frac{1}{2} \arctan \left(\frac{2\mu_{11}}{\mu_{20} - \mu_{02}} \right), & \mu_{11} \neq 0 \vee \mu_{20} \neq \mu_{02}, \\ 0, & \text{otherwise} \end{cases}$$

(if $\mu_{11} = 0$ or $\mu_{20} = \mu_{02}$, the image is already in a normalized position). The ambiguity of a normalized position can be removed by flipping the rotated image horizontally or vertically to satisfy conditions $\mu_{20} > \mu_{02}$ and $\mu_{30} \geq 0$, for example. Second order moments of an image in a normalized position satisfy conditions $\mu'_{11} = 0$ and $\mu'_{20} \geq \mu'_{02}$. Their values can be expressed as

$$\begin{aligned} \mu'_{20} &= \frac{1}{2} \left(\mu_{20} + \mu_{02} + \sqrt{(\mu_{20} - \mu_{02})^2 + 4\mu_{11}^2} \right), \\ \mu'_{02} &= \frac{1}{2} \left(\mu_{20} + \mu_{02} - \sqrt{(\mu_{20} - \mu_{02})^2 + 4\mu_{11}^2} \right). \end{aligned}$$

A reference ellipse of image f in a normalized position is an ellipse

$$\frac{x^2}{a_{ref}^2} + \frac{y^2}{b_{ref}^2} \leq 1, \quad a_{ref} \geq b_{ref} > 0,$$

with the same image moments μ_{pq} as image moments of f up to order two, i.e. $(p + q) \leq 2$; a_{ref} and b_{ref} denote the major and minor semiaxes, respectively.

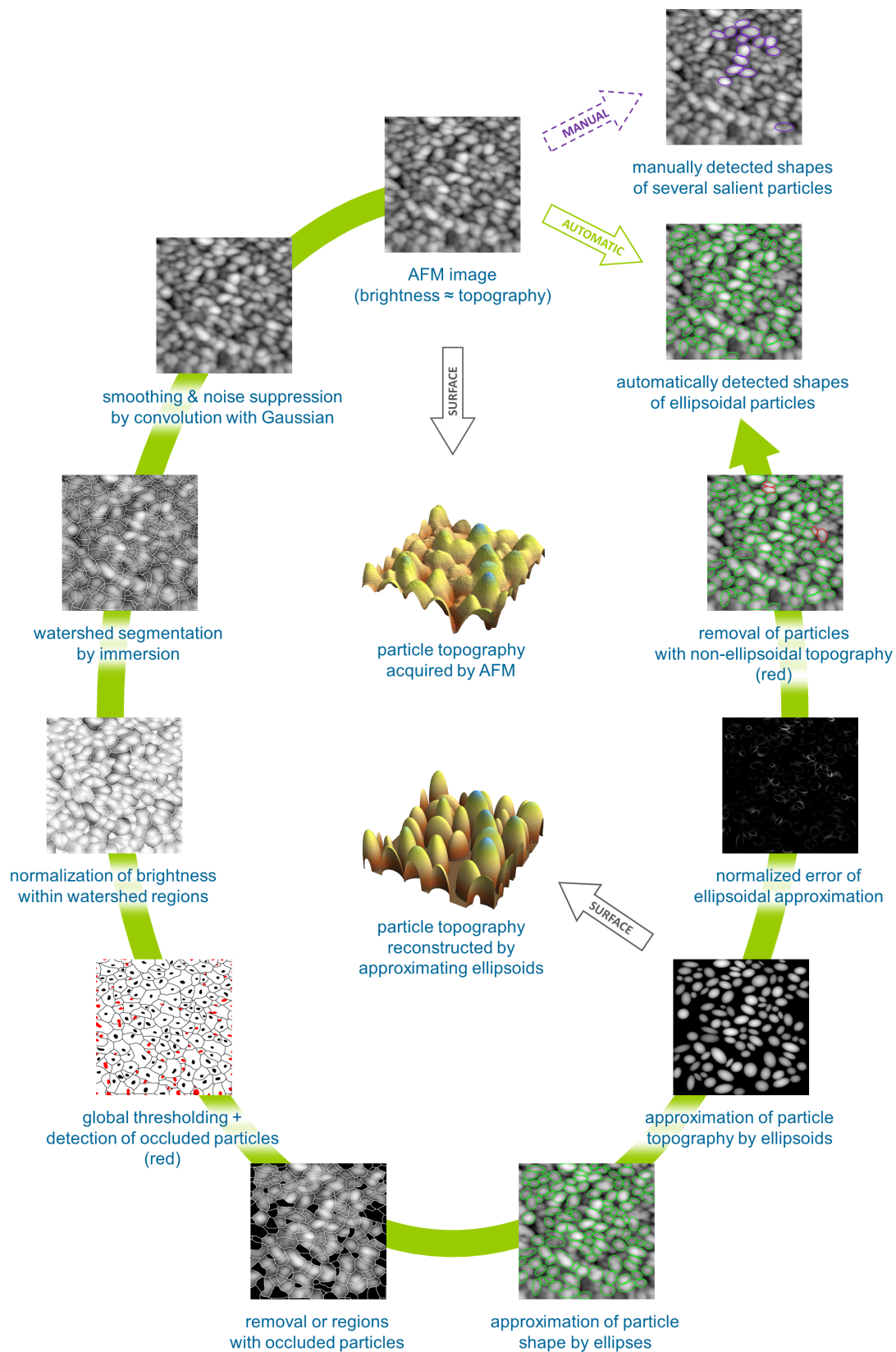


Figure 3.3: Steps of the developed method for automatic measurement of ellipsoidal particles in AFM images.

3.3 Method

The objective of the method described in this section is to compute the average length and width of salient particles in an AFM image of a PPDP sample. AFM image is a grayscale image with pixel values linearly corresponding to the measured topography of the scanned surface (see Figure 3.1); let f denote the image function and D its support.

The method is based on two assumptions. Firstly, the image contains particles of similar sizes. This means that their average length and width can be estimated from a subset of particles. Secondly, the particles are ellipsoidal. This means that their shapes can be approximated by ellipses and their lengths and widths estimated by the corresponding major and minor axes, respectively. The assumptions follow from known physical properties of PPDP samples.

The rest of this section details the proposed method. The method consists of the following steps (see Figure 3.3): first, it reduces high-frequency noise in the AFM image (see Subsection 3.3.1); then it segments the denoised image by watershed transform and removes regions containing noise artifacts or partly occluded particles (see Subsection 3.3.2); in each remaining region it approximates the shape of the corresponding particle by an ellipse and estimates its length and width by the major and minor axes, respectively (see Subsection 3.3.3); the method approximates topography within the approximating ellipse by the upper half of an ellipsoid and includes in measurement only particles which closely match the approximating ellipsoid (see Subsection 3.3.4); finally, it averages the lengths and widths, respectively, of the measured particles (see Subsection 3.3.5).

3.3.1 Denoising

The method first reduces high-frequency noise in the AFM image (see Figure 3.4). It convolves the grayscale image with a radially symmetric Gaussian kernel

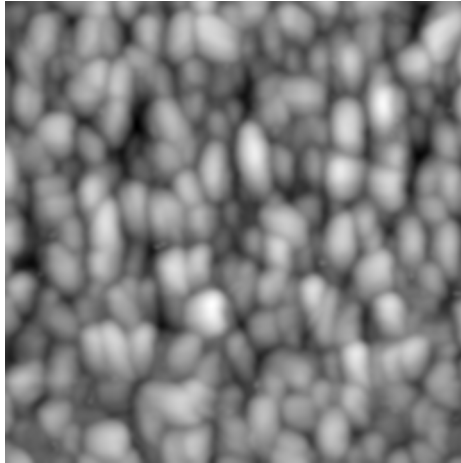
$$h(x, y) = h(r) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{r^2}{2\sigma^2}},$$

where σ is the standard deviation of the Gaussian and $r = (x^2 + y^2)^{\frac{1}{2}}$ the distance from center of the mask. The convolution not only suppresses high-frequency AFM noise (see Section 3.1) but also smooths the image function to avoid oversegmentation in the next step (see Subsection 3.3.2). The convolution reduces high-frequency information in the image in general but the proposed method is robust to this loss.

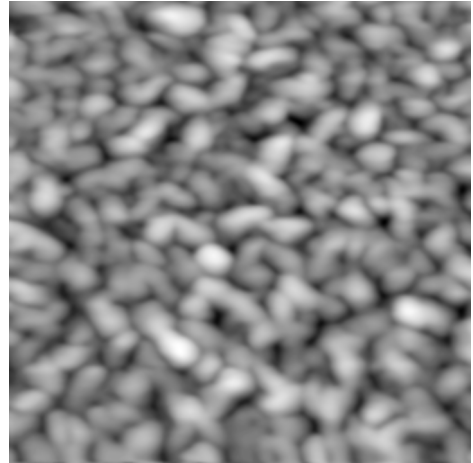
3.3.2 Segmentation

The method then segments the denoised image into regions corresponding to separate particles (see Figure 3.5) and selects regions containing salient particles appropriate for measurement.

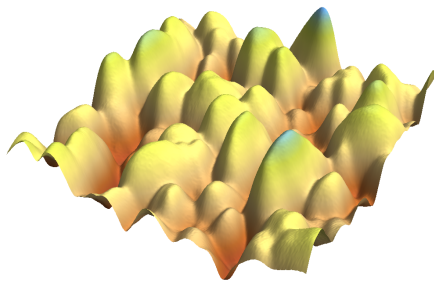
The method segments the denoised image by the watershed transform by immersion (see Subsection 3.2.1). Watershed transform performed on the original AFM image would lead to oversegmentation, with watershed lines dividing many particles into multiple regions. Although the segmentation may not separate all



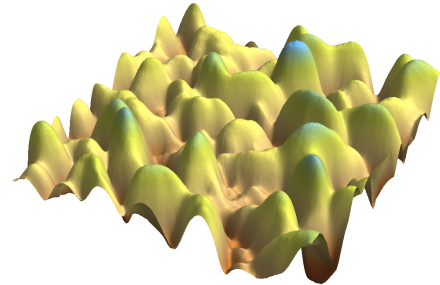
(a) AFM image 5



(b) AFM image 14



(c) AFM image 5 (detail)



(d) AFM image 14 (detail)

Figure 3.4: Convolution of AFM images with a Gaussian kernel; the convolution (a, b) both suppresses high-frequency noise and smooths the measured topography (c, d).

overlapping particles—the method solves this problem later (see Subsections 3.3.4 and 3.3.5)—a vast majority of watershed regions contain exactly one particle.

The method then detects watershed regions containing particles inappropriate for measurements and excludes them from further processing. This includes regions containing AFM noise artifacts or partly occluded particles. For this purpose, the method estimates central parts (cores) of particles. It first separately normalizes brightness in each region D_i ($i = 1, \dots, n$) of denoised image f to

$$f_{norm}(x, y) = f(x, y) + (\max_D f - \max_{D_i} f), \quad \{x, y\} \in D_i,$$

where $D = \bigcup_{i=1}^n D_i$ is the domain of f and n is the number of watershed regions (see Figure 3.6). Then it segments normalized image f_{norm} by thresholding (see Figure 3.7) with convenient global threshold t_0 ; this approach is equivalent to local thresholding by region-specific thresholds. Cores on watershed lines or image borders (see Figure 3.7) correspond to AFM noise artifacts or particles partly occluded by other particles or image borders, respectively. The method excludes regions corresponding to such cores from further processing (see Figure 3.8).

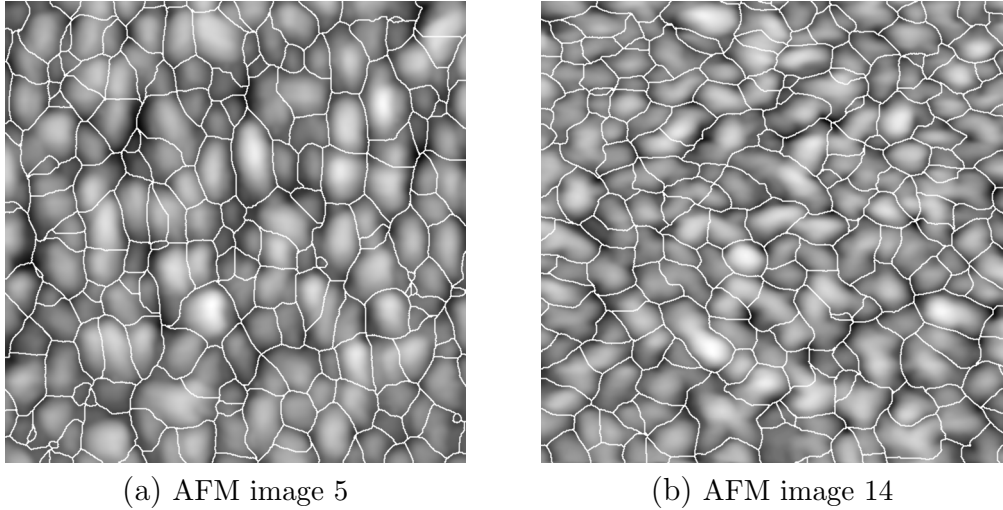


Figure 3.5: Segmentation of denoised AFM images by watershed transform by immersion. The watershed regions roughly separate individual particles.

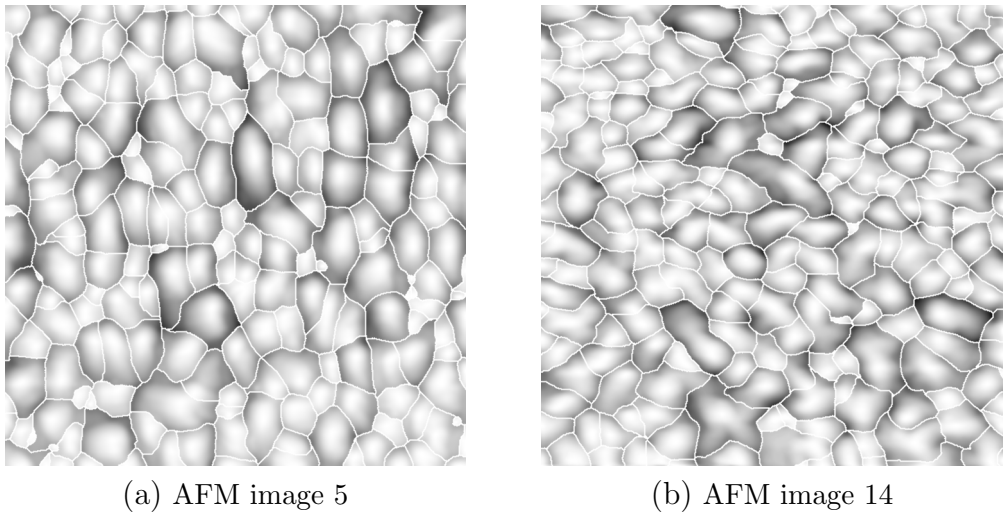


Figure 3.6: Normalization of brightness within watershed regions is a preprocessing step for global thresholding (see Figure 3.7).

3.3.3 Approximation by ellipses

The method approximates the shape of particles in remaining watershed regions by ellipses (see Figure 3.9). It approximates the shape of each particle by ellipse

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} \leq 1, \quad a \geq b > 0.$$

in coordinates of a normalized position of the corresponding region (see Subsection 3.2.2). The method estimates the major and minor semiaxes of the approximating ellipse by

$$a = 2\sqrt{\mu'_{20}}, \quad b = 2\sqrt{\mu'_{02}},$$

respectively; μ'_{20} and μ'_{02} are image moments of brightness function f of the denoised image within region D in a normalized position. The concept of ap-

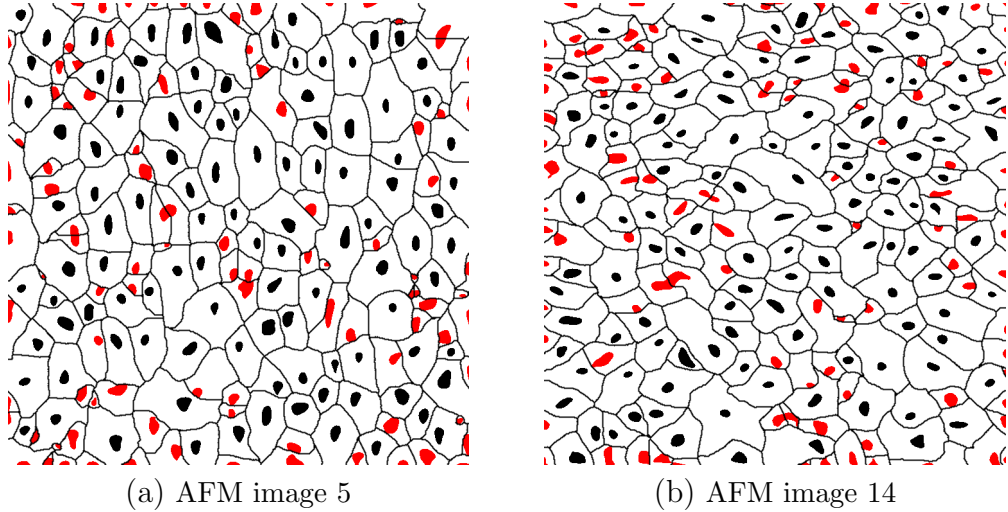


Figure 3.7: Cores of particles computed by thresholding of normalized watershed regions (see Figure 3.6) with global threshold $t_0 = 0.97$. Cores intersecting watershed lines or image borders (red) indicate partly occluded particles.

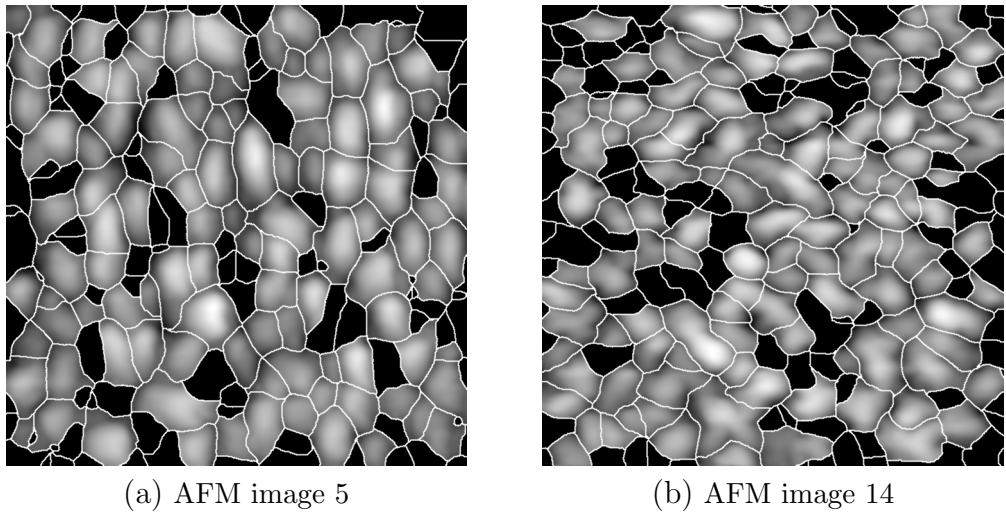
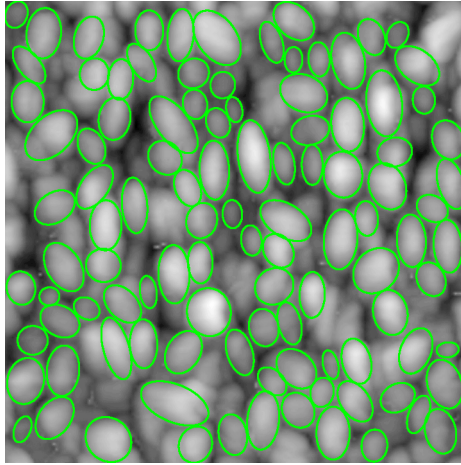


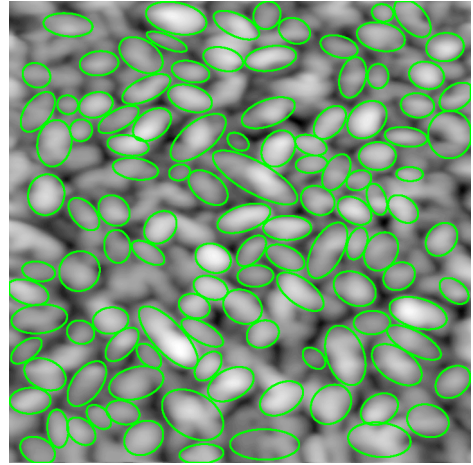
Figure 3.8: Removal of watershed regions with cores on watershed lines or image borders (see Figure 3.7).

proximating ellipses is similar to the concept of reference ellipses. Major axis $2a$ and minor axis $2b$ of the approximating ellipse estimates the length and width, respectively, of the particle.

To compensate for segmentation errors, the method recomputes parameters of the approximating ellipses. The computed shape of the approximating ellipse may be distorted by parts of the region lying outside the real shape of the particle. The method recomputes parameters of the approximating ellipse on the intersection of the region and the ellipse itself. This can be repeated to further increase precision of the estimated shape.

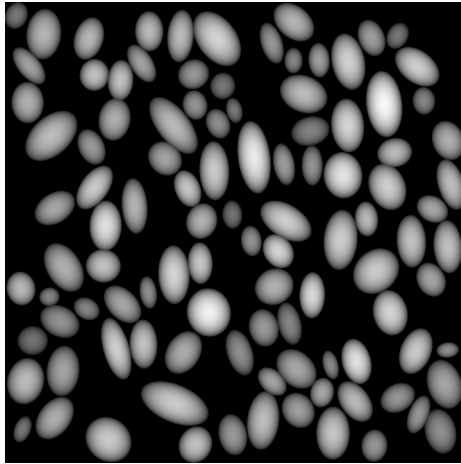


(a) AFM image 5

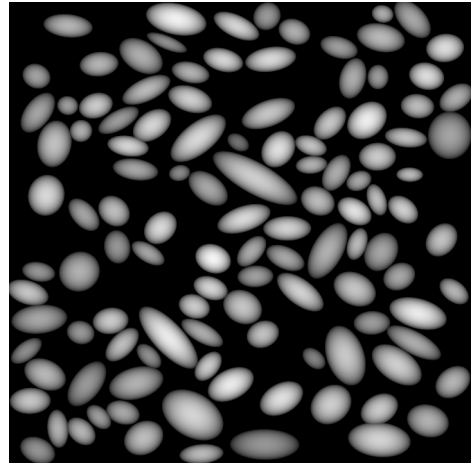


(b) AFM image 14

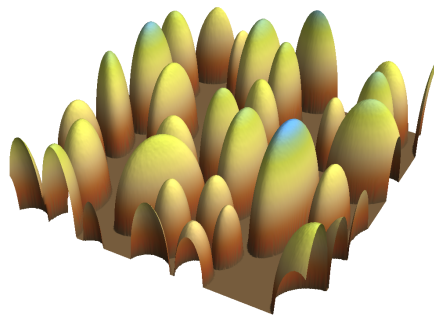
Figure 3.9: Approximation of particles in remaining watershed regions by ellipses.



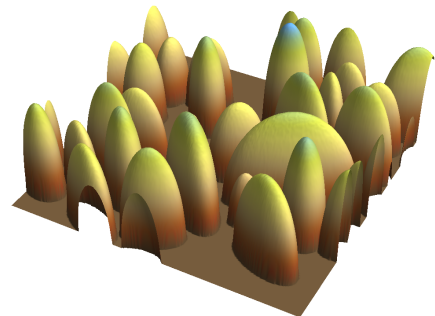
(a) AFM image 5



(b) AFM image 14

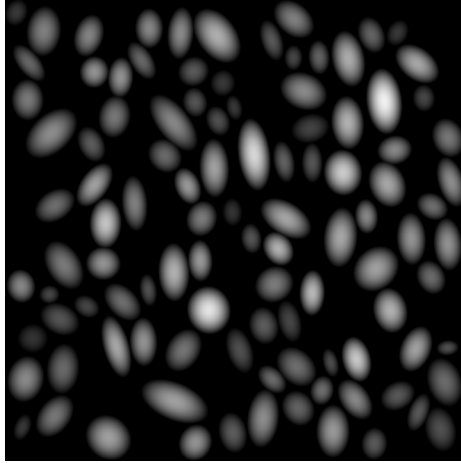


(c) AFM image 5 (detail)

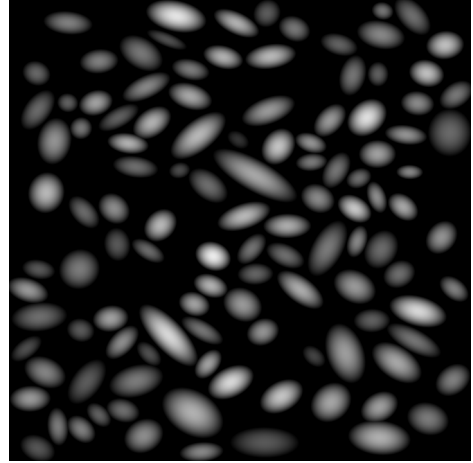


(d) AFM image 14 (detail)

Figure 3.10: Approximation of particles topography within detected ellipses (see Figure 3.9) by ellipsoids.

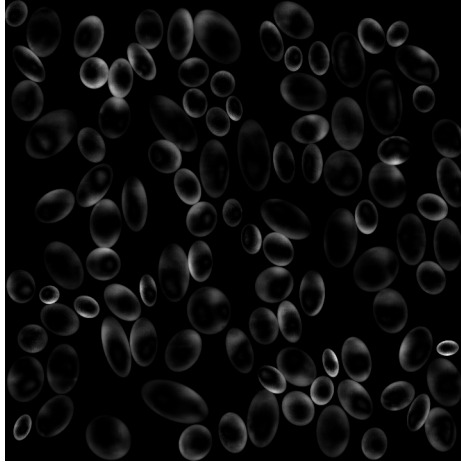


(a) AFM image 5

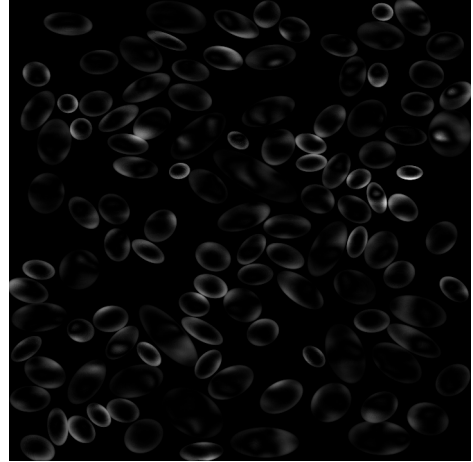


(b) AFM image 14

Figure 3.11: Weights z^2 for computation of approximation error ϵ (see Eq. (3.3.1)).



(a) AFM image 5



(b) AFM image 14

Figure 3.12: Approximation error ϵ (see Eq. (3.3.1)) between the measured topography (see Figure 3.9) and the approximating ellipsoids (see Figure 3.10).

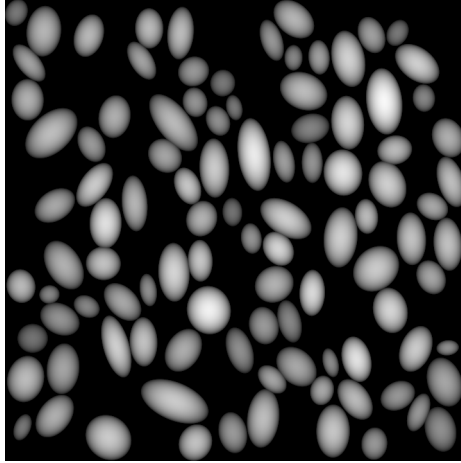
3.3.4 Approximation by ellipsoids

In order to detect ellipsoidal particles, the method approximates the surface of particles by ellipsoids (see Figure 3.10) and computes the approximation error. The method constructs the upper half of ellipsoid

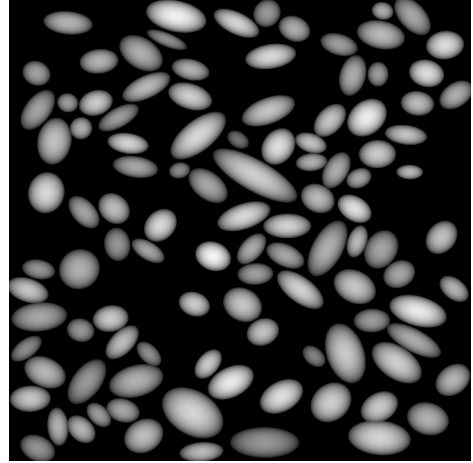
$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1, \quad a, b, c > 0 \quad \& \quad z \geq 0$$

above each approximating ellipse. Semiaxes a and b lie in the plane corresponding to zero brightness and equal the major and minor semiaxes, respectively, of the approximating ellipse. The method estimates the length of semiaxis c perpendicular to the image plane as

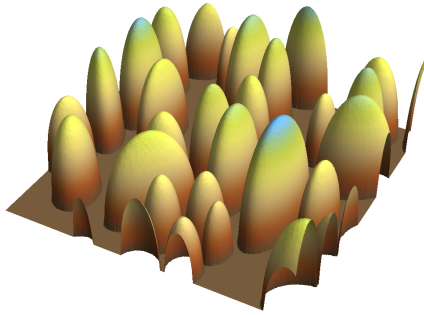
$$c = f(x_t, y_t),$$



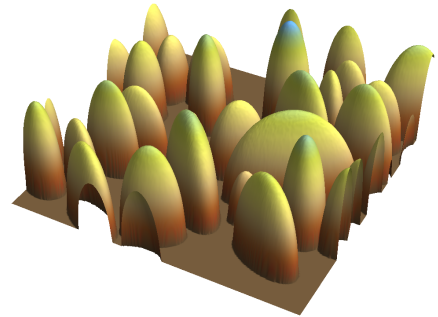
(a) AFM image 5



(b) AFM image 14



(c) AFM image 5 (detail)



(d) AFM image 14 (detail)

Figure 3.13: Ellipsoids approximating particles topography with approximation error $\epsilon \leq 0.015$ (see Eq. (3.3.1))

where (x_t, y_t) is the centroid of the approximating ellipse. The height of the approximating ellipsoid

$$z(x, y) = \begin{cases} c \left(1 - \left(\frac{x^2}{a^2} + \frac{y^2}{b^2}\right)\right)^{\frac{1}{2}}, & \frac{x^2}{a^2} + \frac{y^2}{b^2} \leq 1, \\ 0, & \text{otherwise} \end{cases}$$

above the zero-brightness plane approximates the brightness function of the denoised image within the approximating ellipse (see Figure 3.10).

In order to measure dissimilarity between the surface of the particle and the approximating ellipsoid, the method computes the approximation error. It defines the error as

$$\epsilon = \frac{1}{\iint_D (z(x, y))^2 dx dy} \iint_D (z(x, y))^2 (f(x, y) - z(x, y))^2 dx dy, \quad (3.3.1)$$

where f is the brightness function of the denoised image and D is the shape of the watershed region. Weights $(z(x, y))^2$ suppress the influence of outer parts of the approximation ellipse, which are often distorted by the convolution of the cantilever tip with neighboring particles (see Figure 3.11). The approximation

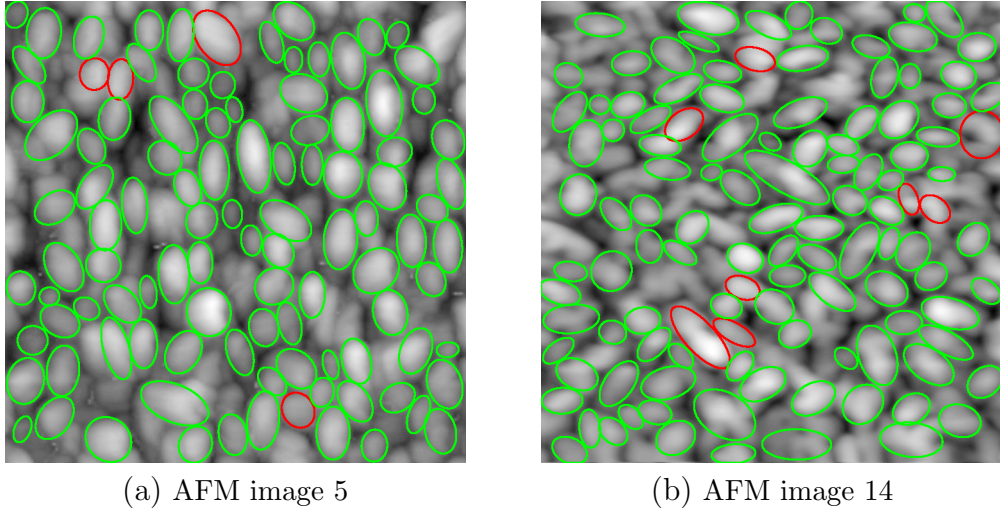


Figure 3.14: Approximating ellipses corresponding to ellipsoids with high (red) and low (green) approximation errors.

error (see Figure 3.12) is a measure of the difference between the surface of the particle and its approximating ellipsoid: the lower the approximation error, the closer the surface of the particle matches the approximating ellipsoid. The error is high for regions that contain overlapping or tilted particles or disrupting AFM noise artifacts.

3.3.5 Measurement

The method then measures the mean length and width of salient particles in the AFM image. It measures only particles with approximation error

$$\epsilon \leq \epsilon_0$$

(see Figure 3.14), where ϵ_0 is a convenient approximation error threshold. The value of ϵ_0 can be set adaptively. The method estimates the length and width of each measured particle (see Figures 3.20a, 3.20b, 3.22a, 3.22b) by the length of major axis $2a$ and minor axis $2b$, respectively, of the corresponding approximating ellipse (see Figures 3.15). From these values, it computes the average length and width.

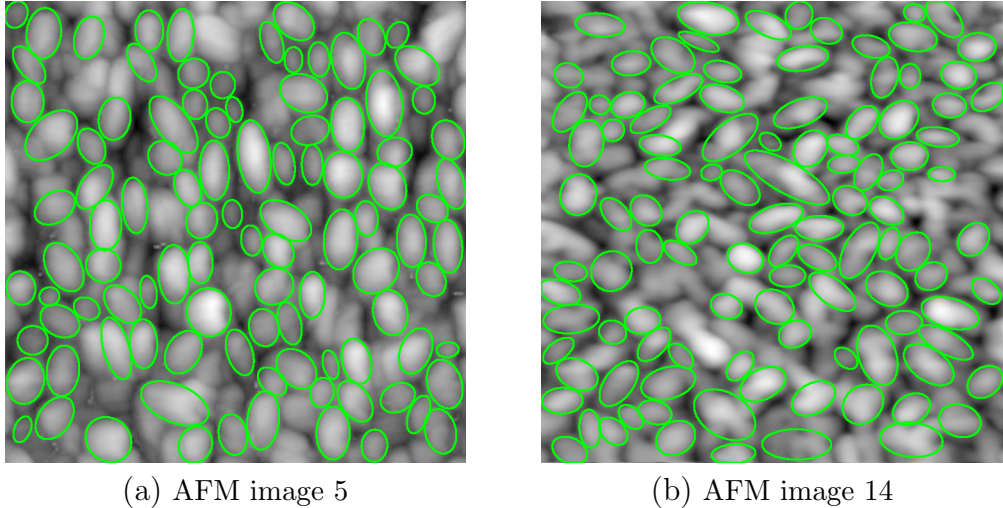


Figure 3.15: Automatically detected shapes of particles convenient for measurements.

3.4 Results

Performance of the proposed method was tested on a set of 11 AFM images of PPDP samples (see Section 3.1). The samples¹ were scanned by the Ntegra Prima apparatus (NT-MDT, Russian Federation) [33]; in order to avoid damage of the soft organic material, the semi-contact (tapping) mode [31] was used. The acquired images and the corresponding color scale of 256 height levels (see Figure 3.16) were saved in a 24bit RGB color format. Dimensions of most images were 719×719 pixels².

As a preprocessing step, the color images were transformed to grayscale (see Figure 3.1) with brightness values linearly corresponding to the measured height of the scanned surface. The acquired images were colored by the AFM imaging software by a color scale of 256 levels; color scales are used by physicists for easier visual evaluation of AFM images. AFM images contain information only about the height of surface, however. For purposes of automatic evaluation, the color images were transformed to a grayscale by mapping the color scale to a gray scale; pixel values were stored in $[0, 1]$ interval.

Automatic measurement of particles was performed in the grayscale AFM images by the method described in Section 3.3 (see Figure 3.3). In each image, the method detected salient particles, and estimated their average lengths and widths in pixels. It suppressed high-frequency noise by convolution with a Gaussian kernel with standard deviation σ ranging from 1 to 3 pixels, depending on the specific image. The segmentation was computed by the watershed by immersion algorithm described in [61] (see Subsection 3.2.1). Global threshold t_0 for com-

¹ The PPDP compound was prepared by reaction of pyrrolinone ester with corresponding nitrile [63]. Thin films of PPDP were prepared by a vacuum evaporation method. The deposition of the active PPDP layer was carried out in a vacuum coating facility with an ultimate pressure of 1×10^{-4} Pa pumped by a diffusion oil pump. Thin films of thickness 100 nm were deposited on selected substrates [39].

² Dimensions of eight images were 719×719 pixels; dimensions of the other images were 909×604 , 723×719 and 715×719 pixels.

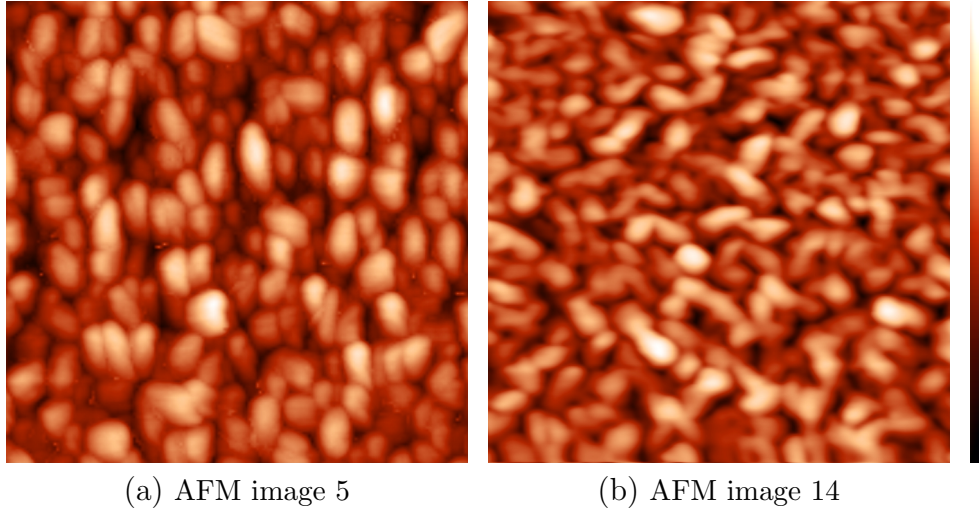


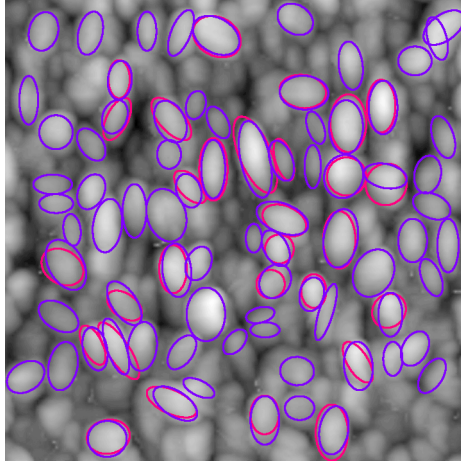
Figure 3.16: Atomic force microscopy (AFM) images of phenylpyridyldiketopyrrolopyrrole (PPDP) samples with color scale (right) of the measured height.

putation of cores in watershed regions was set to $t_0 = 0.97$. Cores on watershed lines and image borders were detected by standard morphological operations, namely binary dilation with a 3×3 structural element. The approximation error threshold was set to $\epsilon_0 = 0.015$.

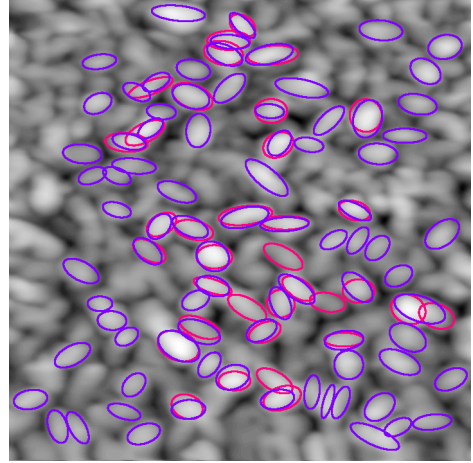
For purposes of comparison, four of the testing AFM images were measured manually by a physicist. Based on his knowledge and experience, he selected a number of representative particles in each image and approximated their shapes by ellipses (see Figures 3.19b and 3.21b); the major and minor axes corresponded to the lengths and widths of the particles. In each image he selected and measured only a small subset of particles that he considered sufficient for estimation of the average length and width of all particles in the image with little error. He measured two of these images again with a time delay, this time in two phases: in the first phase, he again selected a small subset of salient particles for measurement (see Figures 3.19c and 3.21c); in the second phase, he added other salient particles that he also considered convenient for measurement (see Figures 3.19d and 3.21d).

Results of the automatic and manual measurements were compared by the mean values, standard deviations (see Table 3.1), and distributions of lengths and widths (see histograms in Figures 3.20, 3.22) of measured particles in each image.

The results were further compared by two statistics. For the purpose of comparison, two detected elliptical shapes were attributed to the same particle if their centers lied within each other's shape. The first statistic measured the consistency in detection, i.e. the ratio of the number of particles selected by both automatic and manual measurements to the number of manually selected particles (see Table 3.2). This described the relative number of particles selected by the expert as appropriate for measurement that were selected also by the automatic method.

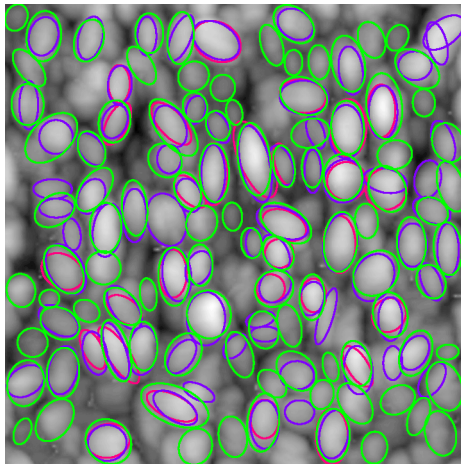


(a) AFM image 5

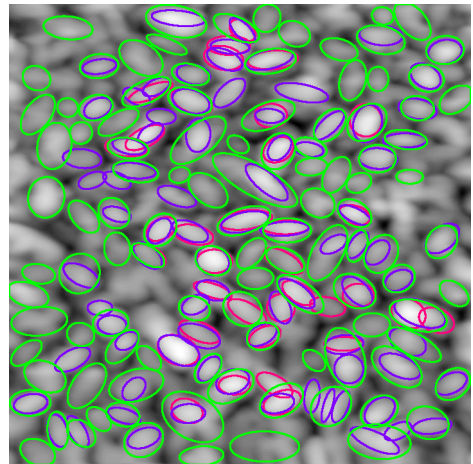


(b) AFM image 14

Figure 3.17: Comparison of manually measured particles (magenta: first measurement session; violet: second measurement session, with additionally measured particles).



(a) AFM image 5



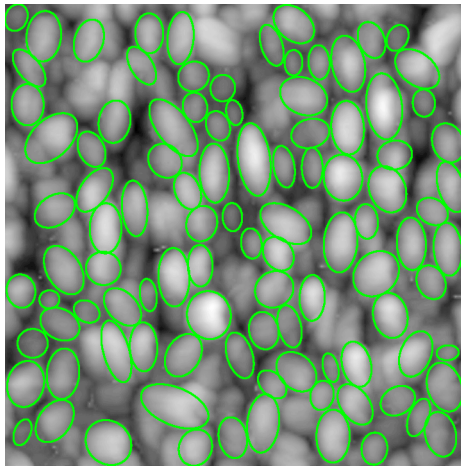
(b) AFM image 14

Figure 3.18: Comparison of automatically (green) and manually (magenta: first measurement session; violet: second measurement session, with additionally measured particles) measured particles.

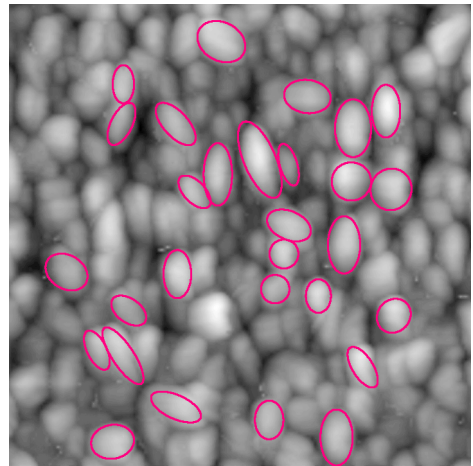
The second statistic measured the Dice similarity coefficient (DSC) [8, 10]

$$\text{DSC} = \sum_{i=1}^n \frac{2|A_i \cap M_i|}{|A_i| + |M_i|}, \quad (3.4.1)$$

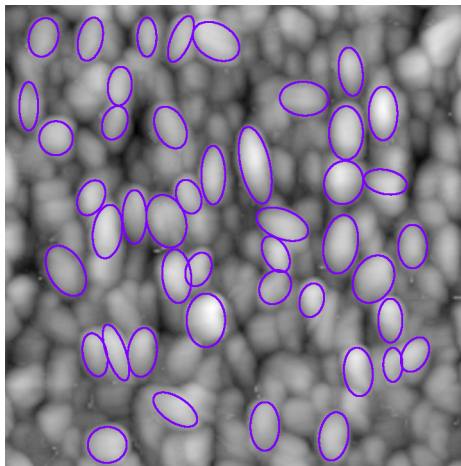
where n denotes the number of particles measured by both methods, A denotes the automatically computed shape of particle i , M the manually measured shape of the same particle, and $|\cdot|$ denotes the number of pixels (see Table 3.2). DSC ranges from zero to one, corresponding to no and full overlap, respectively. The DSC measure describes the difference between automatically and manually estimated shapes of particles that were measured both automatically and manually.



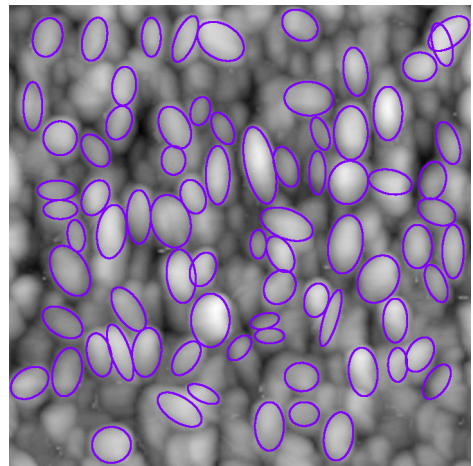
(a) automatic



(b) manual 1

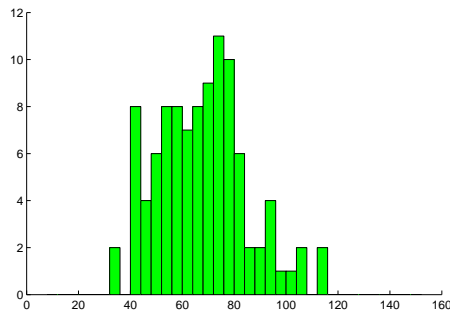


(c) manual 2

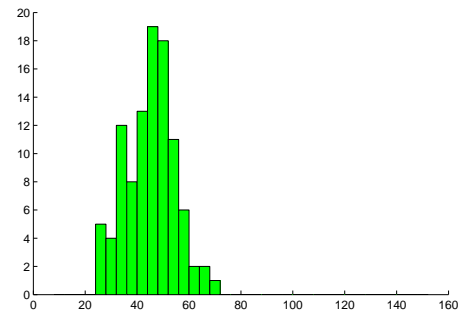


(d) manual 2+

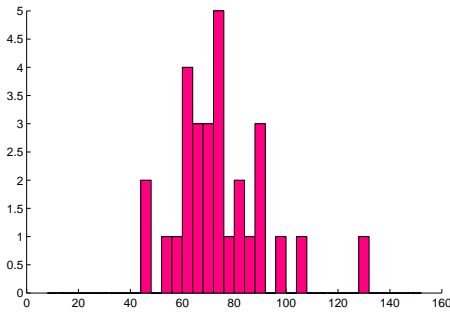
Figure 3.19: Particles measured in AFM image 5: automatic measurement (a), first manual measurement (b), second manual measurement (c), second manual measurement with additionally measured particles (d).



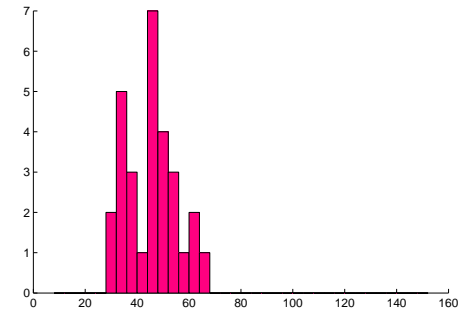
(a) length: automatic



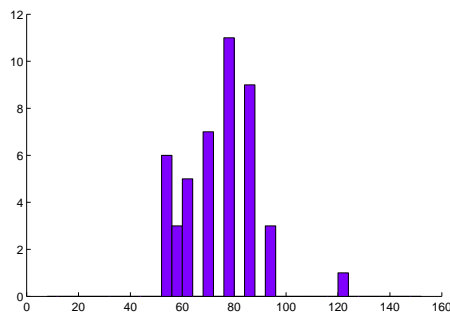
(b) width: automatic



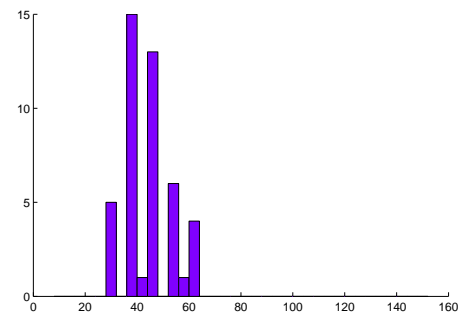
(c) length: manual 1



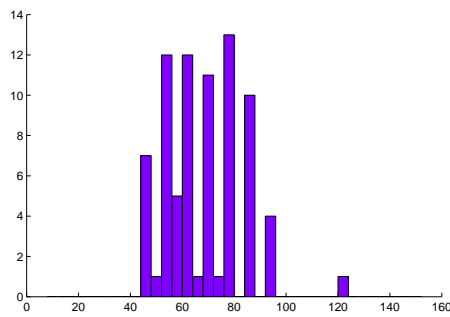
(d) width: manual 1



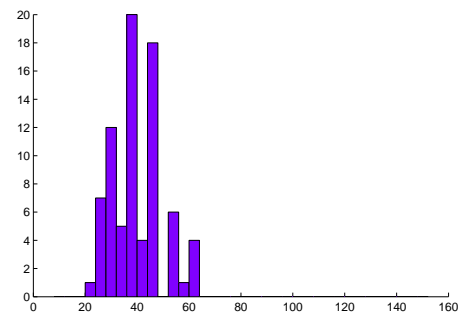
(e) length: manual 2



(f) width: manual 2

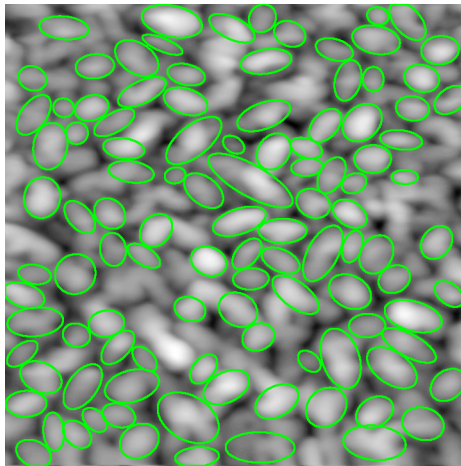


(g) length: manual 2+

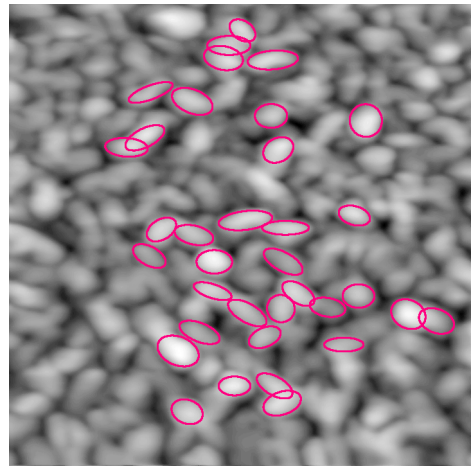


(h) width: manual 2+

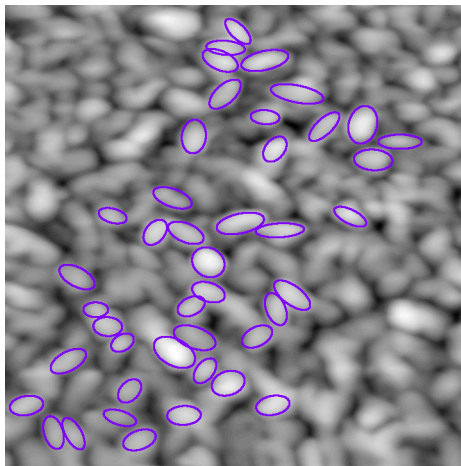
Figure 3.20: Histograms of lengths (left) and widths (right) of automatically measured particles in AFM image 5: automatic measurement (a, b), first manual measurement (c, d), second manual measurement (e, f), second manual measurement with additionally measured particles (g, h).



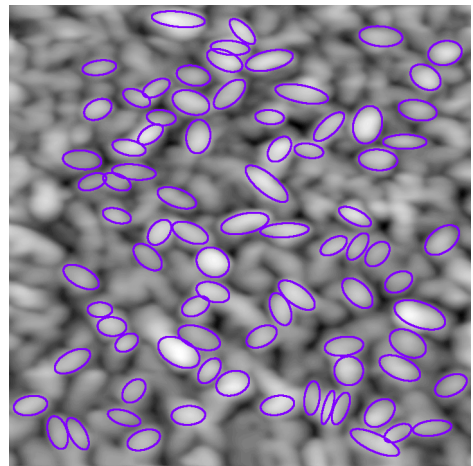
(a) automatic



(b) manual 1

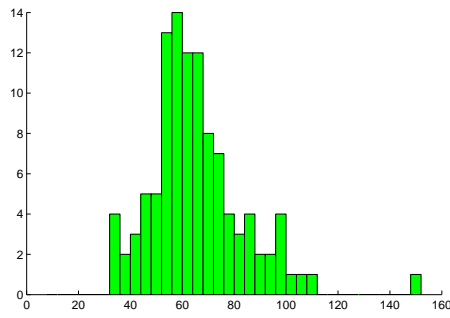


(c) manual 2

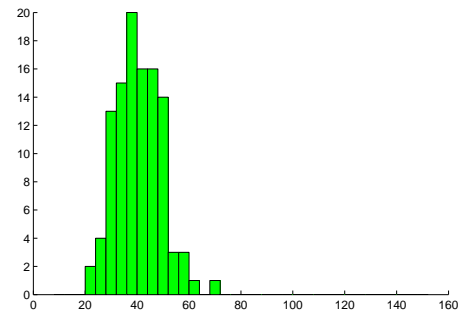


(d) manual 2+

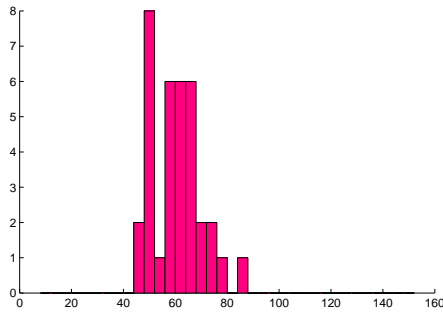
Figure 3.21: Particles measured in AFM image 14: automatic measurement (a), first manual measurement (b), second manual measurement (c), second manual measurement with additionally measured particles (d).



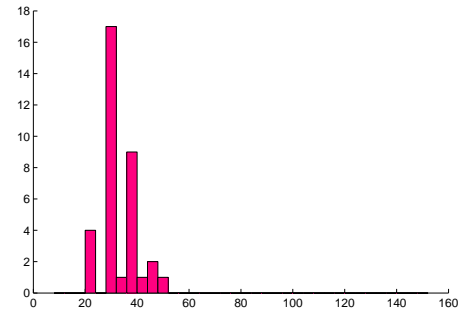
(a) length: automatic



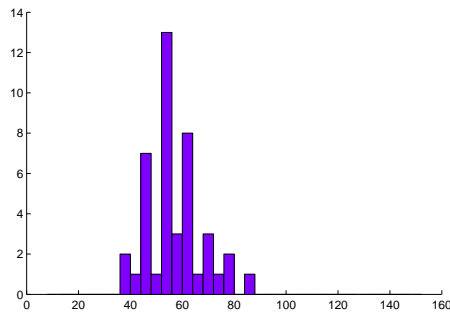
(b) width: automatic



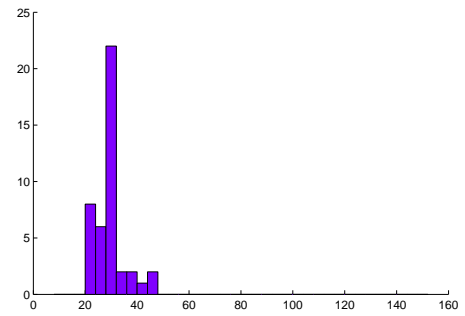
(c) length: manual 1



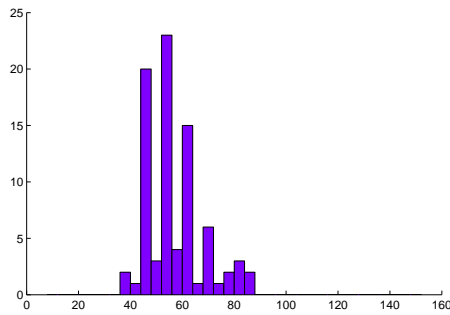
(d) width: manual 1



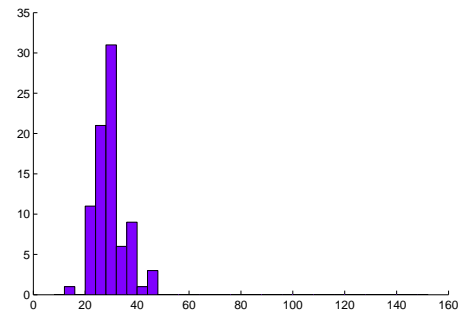
(e) length: manual 2



(f) width: manual 2



(g) length: manual 2+



(h) width: manual 2+

Figure 3.22: Histograms of lengths (left) and widths (right) of automatically measured particles in AFM image 14: automatic measurement (a, b), first manual measurement (c, d), second manual measurement (e, f), second manual measurement with additionally measured particles (g, h).

AFM image	measurement method	number of measured particles	average length	average width
0	automatic all	340	36 ± 10	23 ± 5
0	automatic filtered	287	36 ± 10	22 ± 5
1	automatic all	252	45 ± 13	27 ± 6
1	automatic filtered	218	44 ± 13	27 ± 6
2	automatic all	685	24 ± 6	17 ± 3
2	automatic filtered	674	24 ± 6	17 ± 3
3	automatic all	369	35 ± 8	24 ± 5
3	automatic filtered	368	35 ± 8	24 ± 5
4	automatic all	168	50 ± 13	30 ± 7
4	automatic filtered	144	50 ± 13	31 ± 7
5	automatic all	105	68 ± 17	45 ± 9
5	automatic filtered	101	68 ± 17	45 ± 9
5	manual 1	29	74 ± 18	45 ± 10
5	manual 2	45	74 ± 14	44 ± 9
5	manual 2+	78	68 ± 15	40 ± 9
6	automatic all	172	51 ± 14	30 ± 7
6	automatic filtered	146	50 ± 13	30 ± 7
6	manual 1	31	56 ± 13	27 ± 7
12	automatic all	111	67 ± 17	42 ± 11
12	automatic filtered	107	67 ± 17	42 ± 11
12	manual 1	14	80 ± 13	48 ± 8
14	automatic all	116	66 ± 19	40 ± 9
14	automatic filtered	108	65 ± 18	40 ± 9
14	manual 1	35	60 ± 10	33 ± 7
14	manual 2	43	57 ± 10	30 ± 6
14	manual 2+	83	57 ± 11	30 ± 6
16	automatic all	171	48 ± 14	32 ± 8
16	automatic filtered	164	48 ± 14	32 ± 8
17	automatic all	142	54 ± 17	34 ± 8
17	automatic filtered	100	54 ± 17	35 ± 7

Table 3.1: Results of automatic and manual measurements of the average length and width of phenylpyridyldiketopyrrolopyrrole (PPDP) particles in 11 atomic force microscopy images (AFM). Method “automatic filtered” denotes automatic measurement of particles with approximation error $\epsilon \leq 0.015$ (see Eq. (3.3.1)), while method “automatic all” denotes automatic measurement of all detected particles irrespective of the approximation error; methods “manual1” and “manual2” denote the first and second manual measurement sessions, respectively, and method “manual2+” denotes second session with additionally measured particles.

AFM image	measurement A	measurement B	$ A \& B / B $	DSC
5	automatic filtered	manual 1	90%	0.83
5	automatic filtered	manual 2	89%	0.82
5	automatic filtered	manual 2+	79%	0.81
5	automatic all	manual 1	97%	0.83
5	automatic all	manual 2	93%	0.82
5	automatic all	manual 2+	83%	0.82
5	manual 2	manual 1	93%	0.84
5	manual 2+	manual 1	100%	0.84
6	automatic filtered	manual 1	58%	0.76
6	automatic all	manual 1	71%	0.73
12	automatic filtered	manual 1	100%	0.84
12	automatic all	manual 1	100%	0.84
14	automatic filtered	manual 1	66%	0.76
14	automatic filtered	manual 2	72%	0.71
14	automatic filtered	manual 2+	73%	0.72
14	automatic all	manual 1	80%	0.75
14	automatic all	manual 2	81%	0.72
14	automatic all	manual 2+	80%	0.72
14	manual 2	manual 1	63%	0.83
14	manual 2+	manual 1	89%	0.80

Table 3.2: Comparison of automatic and manual measurements. Method “automatic filtered” denotes automatic measurement of particles with approximation error $\epsilon \leq 0.015$ (see Eq. (3.3.1)), while method “automatic all” denotes automatic measurement of all detected particles irrespective of the approximation error; methods “manual1” and “manual2” denote the first and second manual measurement sessions, respectively, and method “manual2+” denotes second session with additionally measured particles. Column $|A \& B|/|B|$ contains the ratio of the number of particles selected by both measurements to the number of particles selected by measurement B (see Table 3.2). Column DSC contains the Dice similarity coefficient (see Eq. (3.4.1)).

3.5 Discussion

In order to evaluate performance of the automatic measurement method, its results were compared with the results of manual measurements. The results could not be compared with ground truth because real lengths and widths of particles cannot be measured by AFM due to limitations of the scanning technique, namely convolution of the scanned surface with the contacting tip (see Section 3.1). Automatically and manually measured values are both just estimates of the real lengths and widths of particles. Therefore only consistency between the automatic and manual measurements could be evaluated. Note, however, that even a perfect match would not imply accuracy because the results are not compared with ground truth.

For each image, the average lengths and widths of particles from automatic and manual measurements were within the standard deviations (see Table 3.1). Similarities in distributions of lengths and widths of measured particles (see Figures 3.20, 3.22) also indicate that both automatic and manual methods measured particles with similar sizes. The two comparative statistics (see Table 3.2) show that while the automatic methods did not measure some particles selected for manual measurements, the overlap was satisfactory, and comparable to overlap for the two manual measurements. These results indicate that the automatic method could be used in place of manual measurements.

Differences in the computed statistics (see Table 3.2) between the automatic and manual measurements were caused by three main factors. The measurement methods differed in:

- The criteria for selection of particles appropriate for measurement. Whereas the automatic method selected and measured salient particles, i.e. particles with approximation errors $\epsilon \leq \epsilon_0$, the physicist, using his knowledge and experience, selected and measured even particles that were partially occluded or slightly distorted by AFM noise; the automatic method excluded such particles from measurements because it could not estimate their real shape with sufficient accuracy.
- The spatial distribution and the number of selected particles. Whereas the automatic method analyzed particles homogeneously / irrespective of their position in the image, the physicist tended to select particles in clusters (compare Figure 3.15b with Figures 3.21b and 3.21c) in order to facilitate the laborious and time-consuming manual measurement process. As a result, he did not select some particles convenient for measurement.
- The computed lengths and widths of particles that were measured both automatically and manually (see Table 3.2). The automatic measurements were affected by sensitivity of the watershed segmentation to noise and by inaccuracies of the fitted model. The manual measurements were affected by errors in visual estimation of particle shapes caused mainly by tediousness of the process. Note, however, that even manual measurements of the same particles in the same images performed by the same physicist resulted in different measured values (see Figure 3.17).

Accuracy and sensitivity of the automatic method with respect to detection and measurement of particles depend on three main factors:

- Standard deviation σ of the Gaussian low-pass filter affects the rate of false detections of AFM noise artifacts or multiple particles as one particle. On one hand, if σ is too small, the convolution may not suppress high-frequency AFM noise sufficiently and the method may evaluate remaining AFM noise artifacts as particles and include them in measurements. On the other hand, if σ is too large, the convolution may blend neighboring particles, and the method may thus incorrectly identify them as a single particle and include it in measurements. The value of σ that minimizes the number of falsely detected particles depends on the resolution of the input AFM image.
- The watershed transform algorithm and core threshold t_0 may affect the number of falsely detected or falsely removed particles. The watershed algorithm affects not only the shape of regions but also the number of regions incorrectly splitting single particles or containing AFM noise artifacts or multiple particles. Watershed transform by immersion proved convenient for the analyzed AFM images. The value of threshold t_0 influences the number of removed regions. If t_0 is too high, the method identifies more noise artifacts and partly occluded particles as particles convenient for measurement. If t_0 is too low, the method removes even regions with salient particles convenient for measurements.
- Approximation error threshold ϵ_0 affects the number of measured particles. If ϵ_0 is too high, the method measures even falsely detected particles; if ϵ_0 is too low, it does not measure even the most salient particle.

The method could be potentially modified to achieve higher accuracy and better results. Possible modifications include:

- Regions containing cores that only narrowly border watershed lines or image borders may correspond to particles that are not significantly occluded but just lie near AFM noise artifacts or other particles. Such particles could be measured with little error.
- Particles incorrectly divided into separate regions could be merged again and included in measurements. Such regions could be detected by inspecting the continuity of cores across watershed lines.
- The shape of particles could be estimated by a different shape model. Various shape models could be used to detect and measure different types of particles. The scanned surface could be modeled by a mixture of functions corresponding to separated particles. Figure 3.23 shows approximation of particles by Gaussians; ellipsoids used in the proposed method better model the compact support of particles, though.
- The dimensions of approximating ellipses can be modified by coefficients for the major and minor axis to closely correspond to the manually detected ellipses. In this project, however, there was not enough data from manual detections for such supervised learning.

Note, however, that accuracy of the method is strongly limited by the acquisition technique, namely by convolution of the measured surface with an unknown shape of the contacting tip.

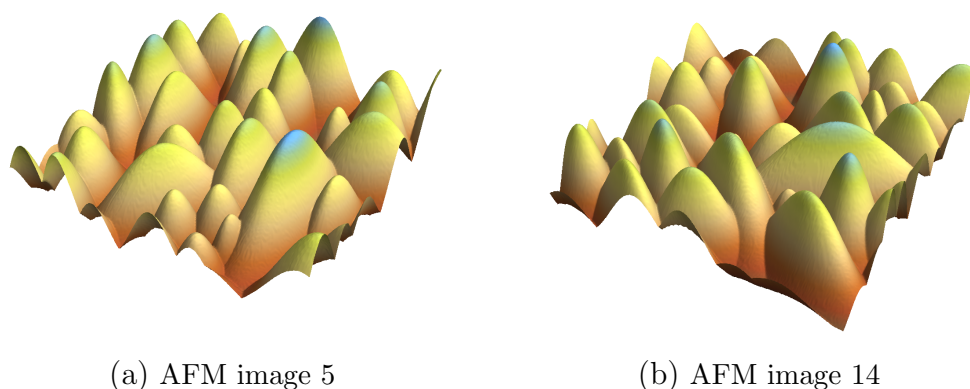


Figure 3.23: Approximation of topography by Gaussians proved less convenient for PPDP particles than approximation by ellipsoids (see Figures 3.10c, 3.10d).

Although the method was primarily developed for measurements in physics, it can be also applied to similarly shaped data in biology and medicine. The method could be used, for example, for measurement of elliptical cells [7] or ellipsoidal microorganisms.

3.6 Conclusion

This project has introduced a new method for detection and measurement of particles in atomic force microscopy (AFM) images. It was developed for purposes of automatic measurement of the average length and width of particles in AFM images of phenylpyridyldiketopyrrolopyrrole (PPDP), a semiconducting material. The method assumes that each AFM image contains similarly sized elliptical particles. It is based on segmentation by a watershed transform and approximation of the shape of particles by ellipses computed by image moments; the length and width of each measured particle are estimated by the major and minor axes, respectively, of the corresponding approximating ellipse. Its performance was successfully tested on eleven AFM images of PPDP samples. Results of the automatic method were compared with results of manual measurements performed by a physicist. The comparison showed consistency between the measurements: differences between automatically and manually computed average lengths and widths were within corresponding standard deviations. The results indicated that the automatic method could be used in practice in place of laborious, time-consuming manual measurements. The main advantage of the proposed method is its robustness to high-frequency noise in AFM images. Possible future improvements include further research on the approximation of particle shapes and on the corresponding approximation error. The method was published in 2011 in *Proceedings of the ICASSP 2011: IEEE International Conference on Acoustics, Speech, and Signal Processing* [45].

Acknowledgment

This work was partly supported by the Ministry of Education of the Czech Republic under the project No. 1M0572 (Research Center DAR), by the Czech Science Foundation under the project No. GACR 203/08/1594, and by the Grant Agency of the Academy of Sciences of the Czech Republic under the project No. KAN401770651. The PPDP samples were kindly provided by and their AFM images acquired in the Institute of Physics, Academy of Sciences of the Czech Republic, Prague, Czech Republic, with special acknowledgement to Jaromír Kopeček, Tatiana Todorciuc and Irena Kratochvílová.

4. Evaluation of vocal fold vibration parameters in videokymographic images

Abstract:

Videokymography is a novel video recording technique used in laryngology and phoniatrics for examination of vocal fold vibrations. Videokymographic cameras repeatedly scan a single line with frame rate *c.* 8000 Hz; they can thus capture fast vibrations of vocal folds along the line of interest. The resulting videokymographic images (videokymograms) consist of successively acquired frames of the scanned line. Visual evaluation of videokymograms is difficult and time-consuming, so development of methods for computer-aided diagnostics is of great interest. The objective of this project was to propose methods for automatic detection of diagnostically important vibration features in videokymograms and computation of corresponding vibration parameters. Performance of the developed methods was tested on a set of videokymograms with a wide range of vibratory patterns; results of the automatic evaluation were comparable with results of visual measurements by clinicians.

Keywords: diagnostics, graph cuts, image analysis, videokymography, vocal fold vibrations

4.1 Introduction

Videokymography is a novel high-speed imaging technique convenient for observation of vocal fold vibrations. Laryngology and phoniatrics examine videokymographic images (videokymograms) for diagnostically important vibration parameters. Currently used visual evaluation of videokymograms is, however, difficult, and its results are inconsistent. The objective of this project was to develop automatic methods for detection of important features in videokymograms and computation of corresponding vibration parameters.

Vibrations of vocal folds are critically important for voice production. Vocal folds are a pair of elastic muscles in the glottal part of larynx (see Figure 4.3a); the right and left vocal folds are mutually symmetric according to the glottal axis. Their vibrations produce phonation. Vibrating folds gradually close and open the space between them (*rima glottidis*) with frequency *c.* 80-1000 Hz [57]. Parameters of vocal fold vibrations determine the quality of phonation. Laryngology and phoniatrics examine these parameters for diagnosis and treatment of voice disorders.

Observation of vocal fold vibrations demands special imaging techniques. Standard video cameras with frame rates 25-30 Hz cannot capture the high-speed vibrations of vocal folds (see Figure 4.1D). In current practice, clinicians examine vocal fold vibrations by strobolaryngoscopy, high-speed videolaryngoscopy or videokymography; all three techniques observe vocal folds from above via a mirror or video camera on a laryngoscope.

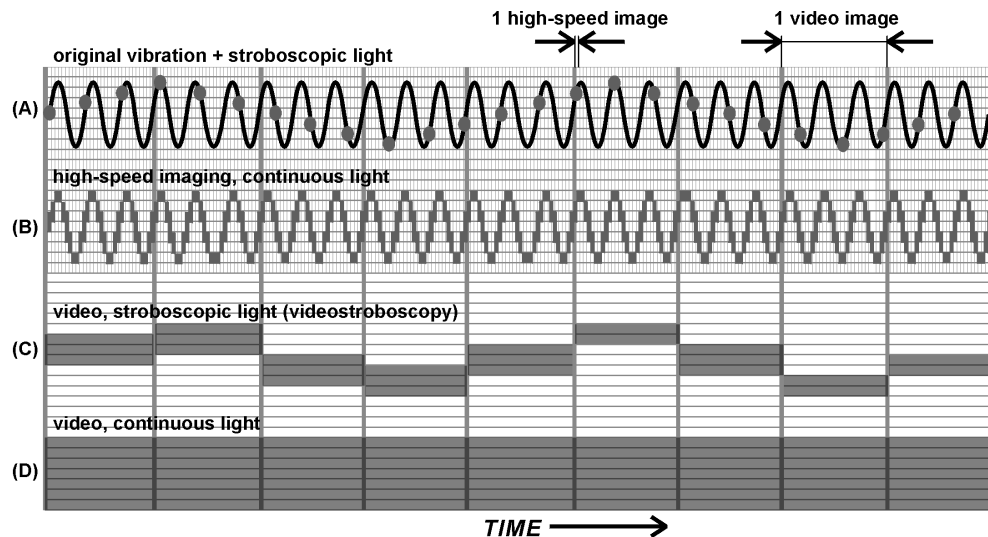


Figure 4.1: Differences in acquisition of vocal fold vibrations by (D) standard video, (A, C) stroboscopy and videostroboscopy, and (B) high-speed imaging. (Reprinted from [53] with permission.)

Strobolaryngoscopy illuminates vocal folds by very short light pulses (see Figure 4.1A). The stroboscopic pulses sample vocal fold vibrations with frequencies that allow observation of subsampled vibrations by a standard video camera (see Figure 4.1C), or directly via a mirror on the laryngoscope. The method is accessible and commonly used in clinical practice. It is, however, inappropriate for examination of irregularly vibrating vocal folds since irregularities in vibrations create distortions in the observed, temporally subsampled signal.

High-speed videolaryngoscopy captures gradual movements of vibrating folds by a high-speed video camera with frame rate *c.* 10000 Hz [9] (see Figure 4.1B). The technique thus allows observation of gradual movements of vocal folds irrespective of their regularity. However, high price and abundance of output data limit use of these systems in clinical practice. Recently developed image processing methods [9, 27, 28] reduce the amount of data by transformation of the high-speed video recordings into 2D images (phonovibragrams) that visualize important vibration features.

Videokymography captures gradual movements of vibrating vocal folds along a line perpendicular to glottal axis by a videokymographic camera that records a single row with frame rate *c.* 8000 Hz (see Figure 4.2). In comparison to standard video cameras, videokymographic cameras increase temporal resolution by decreasing spatial resolution. The resulting videokymograms are real time spatiotemporal images composed of consecutive scans of the recorded row (see Figure 4.3b); the number of unique rows¹ in a videokymogram equals the length of the recorded interval times the frame rate of the videokymographic camera. Videokymography was developed

¹ Modern videokymographic cameras store each acquired frame of the recorded line twice in the videokymogram, i.e. in two consecutive rows.

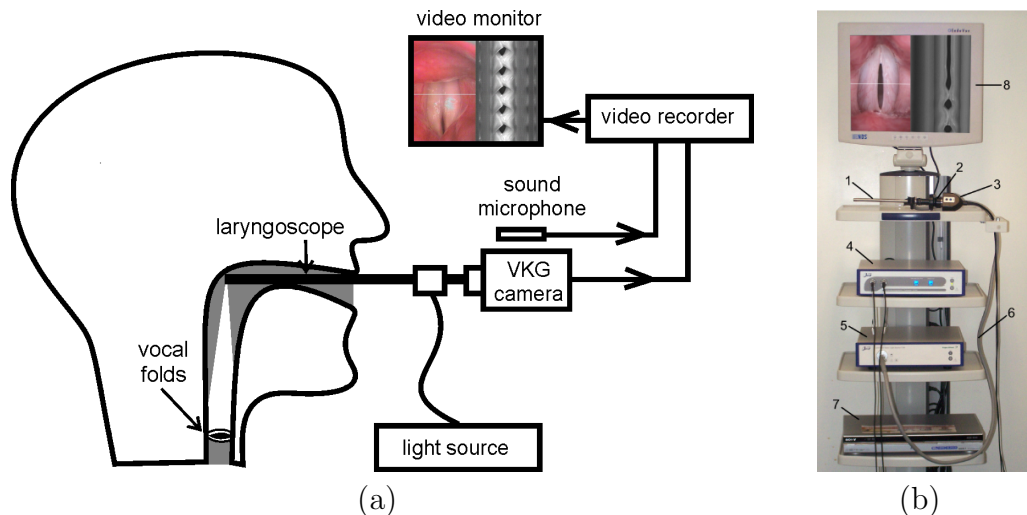


Figure 4.2: (a) Examination of vocal fold vibrations by videokymography. (Reprinted from [57] with permission.) (b) The videokymographic equipment: (1) laryngoscope, (2) an objective with a C-mount, (3) VKG camera head, (4) VKG camera unit, (5) endoscopic high-intensity continuous light source, (6) light cable, (7) digital video recorder, and (8) video monitor. (Photo courtesy of Cymo, B.V., Groningen, The Netherlands. Reprinted from [55] with permission.)

in 1994 [58]. In contrast to high-speed videolaryngoscopy, it significantly reduces the amount of output data as well as the price of the recording equipment. Modern videokymographic cameras allow simultaneous observation in standard and videokymographic modes (see Figure 4.3), so the examining clinician can adaptively position the camera to record the line of interest. Although videokymograms efficiently capture information about a number of important parameters of vocal fold vibrations along the recorded line, medical interpretation requires additional knowledge about the position of the line and the type of phonation. Clinical practice thus often combines videokymography with strobolaryngoscopy. In contrast to stroboscopic techniques, videokymography correctly captures even temporally irregular vibratory patterns; moreover, it allows analysis of vibrations in a still image.

Videokymograms contain a number of features that define diagnostically important vibration parameters. For example, features describing the shape of rima glottidis, e.g. the opening and closing points and lateral and medial peaks (see Figure 4.8), define most of the clinically examined vibration parameters, including cycle-to-cycle variability, relative duration of glottal closure, left-right asymmetries, and relation between opening and closing phases. In another example, features describing trajectories of mucosal waves, i.e. vibratory waves on the surface of vocal folds, define the type and extent of mucosal waves. Laryngology and phoniatrics evaluate vibration parameters to examine properties of vocal folds and characteristics of their vibrations; accurate evaluation is critically important for correct diagnosis of voice disorders and their effective treatment.

In 2007 Švec *et al.* [59] developed a protocol (VKG evaluation sheet) for systematic evaluation of vocal fold vibrations in videokymograms. The evaluation

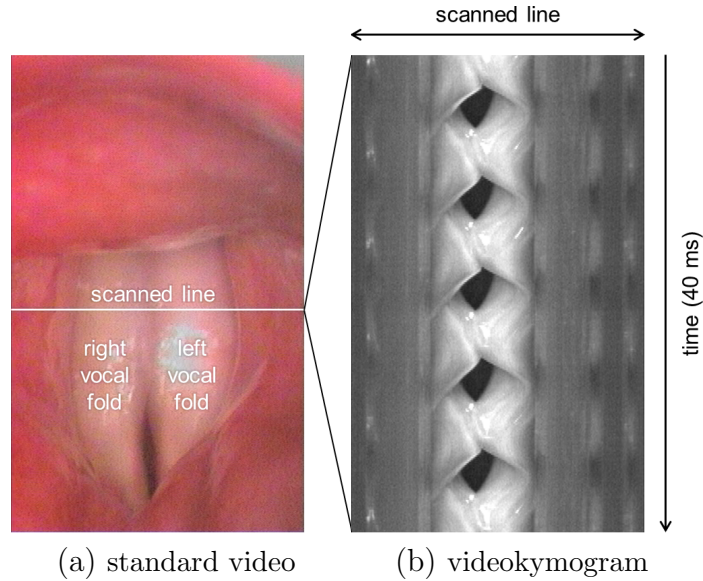


Figure 4.3: Two modes of videokymographic cameras: (a) standard and (b) videokymographic. The videokymogram is composed of successively acquired frames of the scanned line.

sheet defines vibration parameters [56] by categories described by pictograms (see Figure 4.4) and thus provides a convenient tool for visual evaluations. It also enables comparative studies between different evaluations [15].

The current aim is development of methods for computer-aided evaluation of vibration parameters in videokymograms. In current practice, clinicians evaluate videokymograms visually. The visual evaluation process is, however, time-consuming and inconsistent. Its results depend on the evaluators' knowledge and experience, and may even differ between evaluation sessions [15]. Automation of the evaluation process would, in particular, increase the precision and consistency of results. In combination with visual verification, automatic evaluation could be effectively used in clinical practice.

The problem of automatic evaluation of vibration parameters in videokymograms has not been satisfactorily solved yet. Low contrast and high level of noise, as well as irregularities in periodicity of vocal fold vibrations complicate automatic processing of videokymograms. In 2003 Qiu *et al.* [36] introduced an algorithm for detection of glottal contour, i.e. the boundary of rima glottidis, by Otsu's bi-thresholding [34] and active contour fitting, and also for quantification of corresponding vibration parameters by measurement of periodical structures in the signal. The algorithm is, however, applicable only to regular vibratory patterns. In 2006 Manfredi *et al.* [29] developed an algorithm for approximation of glottal contour by active contours. The algorithm is, however, sensitive to low contrast between rima glottidis and vocal folds in videokymograms. In 2008 Jiang *et al.* [23] proposed a curve-fitting algorithm for detection of trajectories of the upper and lower vocal fold lips, which determine the glottal contour. Again, the algorithm is applicable only to regular vibratory patterns. In 2010 Zhang *et al.* [67] introduced a method for segmentation of rima glottidis by Lagrange interpolation, differentiation, and Canny's edge detection. The method was, however,

VKG Evaluation Sheet: Closure

Specify the measured value or select category below

CQ = closure / cycle duration

Duration of closure






					NA
no closure	1-20%	20-40%	40-60%	>60%	NA
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Figure 4.4: Protocol for evaluation of videokymographic images: parameter Duration of closure (see Table 4.4 and Figure 4.17). (Reprinted from [59] with permission.)

designed for high-contrast videokymograms of extracted canine vocal folds.

Since 2010 the problem has been researched also at our department². In 2010 Hauzar [17] developed an algorithm for hierarchical detection of glottal features in videokymograms. The algorithm is, however, sensitive to local contrast between rima glottidis and vocal folds, and may detect some features incorrectly in low-contrast videokymograms with pathologic vibratory patterns. In 2011 Zita [68] developed a correlation-based algorithm for tracking of mucosal waves in videokymograms. The algorithm was, however, designed specifically for videokymograms with physiologic vibratory patterns.

The objective of this project was to develop methods for automatic detection of important vibration features in videokymograms and evaluation of corresponding vibration parameters in the VKG evaluation sheet. The methods should be robust to pathologic vibratory patterns and low contrast in videokymograms. Results of the automatic evaluation should be comparable with results of visual evaluations.

In 2012 we developed new methods for detection of reflections, rima glottidis, mucosal waves, and glottal features, and introduced a method for automatic evaluation of corresponding vibration parameters [44]. One method computes the shape of rima glottidis by thresholding based on graph cuts, and from the segmented shape extracts corresponding vibration features using the hierarchical algorithm developed by Hauzar [17]. Another method detects directions of mucosal waves by Fourier transform; the method can be used in combination with the tracking algorithm proposed by Zita [68] to determine trajectories of mucosal waves with higher accuracy. The developed methods are relatively robust to low contrast and pathologic vibratory patterns. The automatic evaluation method computes from the detected features numerical values of 11 vibration parameters and corresponding categories in the VKG evaluation sheet. Section

² Department of Image Processing, Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic, Prague.

4.3 describes the developed methods in detail. Their performance was tested on videokymograms with a wide range of vibratory patterns. Results of the automatic evaluation were comparable with results of visual evaluations.

The rest of this chapter is organized as follows: Section 4.2 describes the image thresholding algorithm based on graph cuts that was used for segmentation of rima glottidis; Section 4.3 describes in detail the developed methods; Section 4.4 presents and compares results of the automatic evaluation with results of visual evaluations; Section 4.5 discusses results and properties of the developed methods, and outlines possible future enhancements; and Section 4.6 summarizes this chapter.

4.2 Applied image processing tools

The proposed method for automatic segmentation of rima glottidis (see Subsection 4.3.2) uses an algorithm for computation of optimal threshold by normalized graph cuts. Subsection 4.2.1 describes the algorithm in detail.

4.2.1 Thresholding by normalized graph cuts

This subsection describes a thresholding algorithm developed by Tao and Jin in 2008 [50]. It computes the optimal value of threshold by minimization of the normalized graph cut measure proposed by Shi and Malik in 2000 [47].

Graph cuts are an effective tool for optimal bipartitioning of data according to selected criteria. Let $G = (V, E, w)$ denote a weighted complete undirected graph, where V is the set of vertices, $E \subseteq V \times V$ is the set of edges, and $w : E \rightarrow \mathbb{R}_0^+$ is a weight function describing similarity between vertices. Let G be partitioned into two disjoint complementary sets A and $B = V \setminus A$. Graph cut

$$\text{cut}(A, B) = \sum_{\substack{u \in A \\ v \in B}} w(u, v)$$

is a measure that describes the similarity between vertices from different components of the cut. Bipartitioning that minimizes the cut

$$\operatorname{argmin}_{A \subseteq V} \text{cut}(A, B), \quad B = V \setminus A$$

maximizes dissimilarity between the components. However, minimization of the graph cut measure often leads to isolation of a small component. Normalized graph cut measure

$$\text{Ncut}(A, B) = \frac{\text{cut}(A, B)}{\text{asso}(A, V)} + \frac{\text{cut}(A, B)}{\text{asso}(B, V)},$$

where

$$\text{asso}(A, V) = \sum_{\substack{u \in A \\ v \in V}} w(u, v) = \sum_{\substack{u \in A \\ v \in B}} w(u, v) + \sum_{\substack{u \in A \\ v \in A}} w(u, v) = \text{cut}(A, B) + \text{asso}(A, A),$$

solves this problem.

Normalized graph cut minimization can be used for computation of threshold value optimal for image segmentation according to selected criteria. Let

$$f : \{1, \dots, m\} \times \{1, \dots, n\} \rightarrow \{0, \dots, l\}, \quad m, n, l \in \mathbb{N}^+$$

denote a digital gray-scale image. The algorithm constructs complete weighted undirected graph $G = (V, E, w)$, where vertices

$$V = \{(i, j) \mid i \in \{1, \dots, m\}, j \in \{1, \dots, n\}\}$$

correspond to pixels in the image, and weights

$$w(u, v) = \begin{cases} e^{-\frac{|f(u)-f(v)|^2}{c_f} + \frac{\|u-v\|_2^2}{c_d}}, & \|u-v\|_2 \leq r, \\ 0, & \text{otherwise,} \end{cases} \quad (4.2.1)$$

where $\|\cdot\|_2$ denotes ℓ_2 norm, describe both the gray-level similarity and spatial proximity of pixels. Parameters $c_f > 0$ and $c_d > 0$ are gray-level and spatial scaling factors, respectively, and $r \geq 0$ is the radius of pixel neighborhood. Let $L = \{0, \dots, l\}$ denote the set of gray levels in f , and $V_k = \{(i, j) \mid i \in \{1, \dots, m\}, j \in \{1, \dots, n\}, f(i, j) = k\}$ the set of pixels with gray level $k \in L$.

Each gray level $t \in L$ defines bipartitioning of pixels into sets $A_t = \bigcup_{k=0}^t V_k$ and

$B_t = V \setminus A_t = \bigcup_{k=t+1}^l V_k$, and also the corresponding normalized graph cut

$$\text{Ncut}(A_t, B_t) = \frac{\text{cut}(A_t, B_t)}{\text{asso}(A_t, A_t) + \text{cut}(A_t, B_t)} + \frac{\text{cut}(A_t, B_t)}{\text{asso}(B_t, B_t) + \text{cut}(A_t, B_t)},$$

where

$$\text{cut}(A_t, B_t) = \sum_{i=0}^t \sum_{j=t+1}^l \text{cut}(V_i, V_j) = \sum_{\substack{f(u) \leq t \\ f(v) > t \\ u, v \in V}} w(u, v),$$

$$\text{asso}(A_t, A_t) = \sum_{i=0}^t \sum_{j=i}^t \text{cut}(V_i, V_j),$$

$$\text{asso}(B_t, B_t) = \sum_{i=t+1}^l \sum_{j=i}^l \text{cut}(V_i, V_j).$$

Minimization of the normalized graph cut measure

$$t_{Ncut} = \underset{t \in L}{\text{argmin}} \text{Ncut}(A_t, B_t)$$

defines threshold t_{Ncut} that maximizes dissimilarity between complementary objects A_t and B_t according to both spatial and gray-level relations of their pixels. The algorithm thus allows segmentation by optimal thresholding without initiation.

4.3 Methods

This section describes the developed methods for automatic evaluation of videokymograms. The methods described in Subsection 4.3.1 segment and remove specular reflections; the methods in Subsection 4.3.2 segment the shape of rima glottidis and detect corresponding features; the methods in Subsection 4.3.3 detect the direction of mucosal waves and track their trajectories; the methods in Subsection 4.3.4 evaluate vibration parameters from the detected features; and the methods in Subsection 4.3.5 statistically compare results automatic and visual evaluations.

Let $f : D \rightarrow \mathbb{R}$ denote the image function of a videokymogram, where $D = \{1, \dots, x_{max}\} \times \{1, \dots, y_{max}\}$ denotes its domain, and x_{max} and y_{max} the number of columns and rows of the videokymogram, respectively.

4.3.1 Reflections

This subsection describes developed methods for segmentation of specular reflections and reconstruction of corresponding areas in videokymograms. The segmentation method is based on thresholding and region growing, and the reconstruction method is based on inpainting by diffusion.

In laryngoscopy, specular reflections occur on moist surfaces. Liquids on surfaces of vocal folds and surrounding tissues specularly reflect the laryngoscopic illumination. Moist surfaces perpendicular to the axis plane between the illumination source and the camera sensor thus create specular reflections in the recorded data.

In videokymography, specular reflections appear as very bright areas, notably brighter than the rest of the videokymogram (see Figure 4.5a). The videokymograms thus lack information about real appearance of the corresponding tissues. Specular reflections both facilitate and complicate evaluation of videokymograms.

On one hand, specular reflections can be used for detection of vocal fold properties. For example, specular reflections indicate presence of moisture, which is important for correct function of vocal folds. In another example, specular reflections on mucosal waves indicate their trajectory. Specular reflections can be effectively segmented by thresholding because they are generally brighter than the rest of the videokymogram.

On the other hand, specular reflections may also complicate evaluation of vocal fold properties. Specular reflections change the appearance of corresponding tissues in videokymograms, and also create artificial edges and plateaus. Such artifacts may impair results of detection algorithms and thus, for example, lead to detection of false features. In such cases, specular reflections should be first removed from the videokymogram. They can be effectively removed, e.g. by interpolation of image function from their neighborhood.

Segmentation of specular reflections

The proposed method for segmentation of specular reflections in videokymograms consists of three steps. It first normalizes brightness in the videokymogram, then segments the shape of specular reflections by thresholding, and finally refines the segmented shape by region growing.

The brightness normalization step linearly stretches the range of the videokymographic image function f to $[0, 1]$ interval

$$f_n(x, y) = \frac{f(x, y) - \min_{(x,y) \in D} f(x, y)}{\max_{(x,y) \in D} f(x, y) - \min_{(x,y) \in D} f(x, y)}. \quad (4.3.1)$$

This normalization allows use of the same segmentation threshold in all videokymograms.

The thresholding step computes regions corresponding to strong specular reflections by thresholding with convenient high threshold $t_1 \in [0, 1]$

$$g_{t_1}(x, y) = \begin{cases} 1, & f_n(x, y) \geq t_1, \\ 0, & f_n(x, y) < t_1 \end{cases} \quad (4.3.2)$$

(see Figure 4.5b).

The refinement step adjusts distortions in boundaries of the segmented reflections. The method detects regions brighter than convenient low threshold $t_2 \in [0, t_1]$

$$g_{t_2}(x, y) = \begin{cases} 1, & f_n(x, y) \geq t_2, \\ 0, & f_n(x, y) < t_2 \end{cases} \quad (4.3.3)$$

(see Figure 4.5c) and selects regions in g_{t_2} that intersect regions in g_{t_1} . The result is a binary image g_r of segmented specular reflections (see Figure 4.5d).

Removal of specular reflections

The proposed method for removal of specular reflections in videokymograms consists of two steps. It first computes outer boundaries of regions with specular reflections and then reconstructs image function within the regions by diffusion from the outer boundaries.

The first step computes the outer boundaries of specular reflections by binary morphology [46]. It enlarges regions in binary image g_r , i.e. regions with specular reflections computed by the segmentation method described above, by

$$g_d = g_r \oplus \begin{pmatrix} 0 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 0 \end{pmatrix},$$

where \oplus denotes binary dilation with a structuring element, and then subtracts the original regions from the enlarged regions

$$g_b = g_d - g_r.$$

The result is a binary image of the outer boundaries of specular reflections.

The second step reconstructs image function f_n within regions with specular reflections g_r by inpainting from their outer boundaries g_b . The inpainting can be performed by a number of methods [2, 3]. The aim of the reconstruction is to suppress sharp edges created by specular reflections. Diffusion [52], for example, is simple but effective for this purpose; it reconstructs the image function smoothly and does not introduce significant edges. The result is a videokymogram with specular reflections removed by interpolation from their neighborhood (see Figure 4.5f).

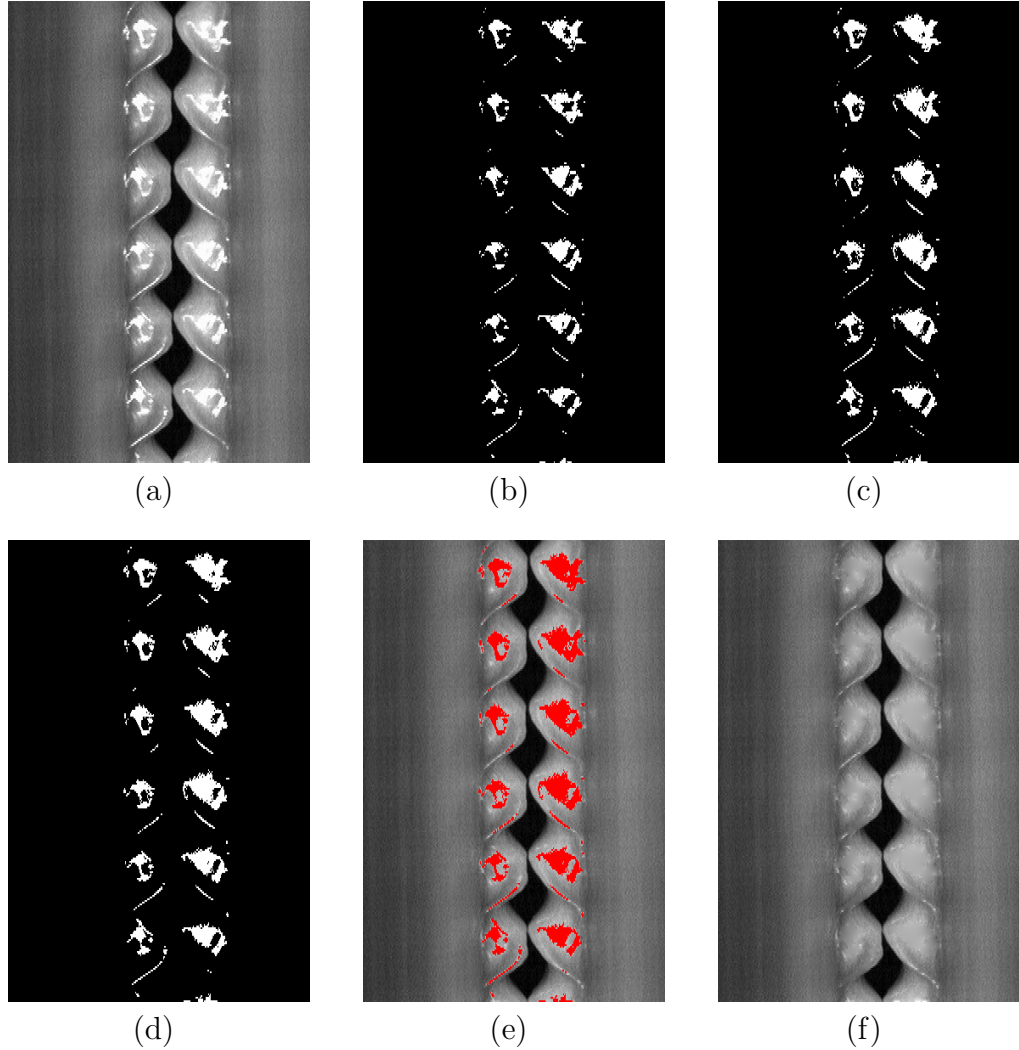


Figure 4.5: Detection and removal of specular reflections: (a) videokymogram with specular reflections, (b) thresholding with high threshold t_1 (see Eq. (4.3.2)), (c) thresholding with low threshold t_2 (see Eq. (4.3.3)), (d) growing of regions in (b) within regions from (c), (e) segmented specular reflections (red), (f) removal of the reflections by diffusion from outside boundaries.

4.3.2 Rima glottidis

This subsection describes developed methods for segmentation of rima glottidis and detection of corresponding vibration features in videokymograms. The segmentation method is based on thresholding by graph cuts (see Subsection 4.2.1) and morphological operations; the feature detection method is based on the hierarchical algorithm developed by Hauzar [17].

Rima glottidis (or glottal area) is the space between vocal folds. The boundary between vocal folds and rima glottidis is called glottal contour. Lateral movements of vibrating vocal folds change the shape of rima glottidis in periods corresponding to vibration cycles. The shape gradually widens and narrows during the opening and closing phases, respectively, and completely disappears during closure (see Figure 4.8).

In videokymograms, rima glottidis appears as a dark area between vocal folds. Closures, if present, divide its shape into separate regions corresponding to open phases of vibration cycles. Rima glottidis is mostly darker than the rest of the videokymogram (see Figure 4.6a). It may, however, contain bright areas created by reflections on underlying tissues. Moreover, some areas of vocal folds or surrounding tissues may, due to insufficient illumination, also appear dark. For these reasons, rima glottidis cannot be generally segmented only by thresholding.

Segmentation of rima glottidis

The proposed segmentation method consists of three main steps. The first step roughly estimates maximal lateral extent of rima glottidis and crops the videokymogram accordingly; the second step segments the shape of rima glottidis by thresholding; and the third step removes incorrectly segmented regions corresponding to dark areas of vocal folds and surrounding tissues and also refines the shape of regions corresponding to rima glottidis.

In order to estimate the extent of rima glottidis, the method first computes the centerline of vibrations (see Figure 4.6b) by the algorithm developed by Hauzar [17]. The centerline is a vertical line approximately bisecting the shape of rima glottidis in the videokymogram. The algorithm first computes sums of pixel values in columns of the videokymogram, and orders the columns accordingly from minimum. It estimates the centerline x_c by the column with the lowest sum that has vibrations in its neighborhood.

The method then normalizes brightness of the videokymogram f to $[0, 1]$ (see Eq. (4.3.1)) and suppresses noise by median filtering (see Figure 4.6c). Then it computes minima of columns

$$c_{min}(x) = \min_y f(x, y),$$

and thresholds the vector

$$c_{t_r}(x) = \begin{cases} 1, & c_{min}(x) \geq t_r, \\ 0, & \text{otherwise} \end{cases} \quad (4.3.4)$$

by threshold t_r corresponding to an estimate of maximal brightness of rima glottidis in videokymograms. The thresholding divides the vector into regions (see Figure 4.6d). The method then selects the region that contains the centerline x_c and denotes its boundaries as x_1 and x_2 , i.e. $x_c \in \{x_1, \dots, x_2\}$ and

$$c_{t_r}(x) = \begin{cases} 1, & x \in \{x_1, \dots, x_2\}, \\ 0, & x = x_1 - 1 \vee x = x_2 + 1. \end{cases}$$

The interval encloses the maximal lateral extent of rima glottidis in the videokymogram. The method now crops the videokymogram to columns in interval $\{x_1, \dots, x_2\}$ (see Figure 4.6e).

In the second step, the method segments the cropped videokymogram f_c by thresholding based on normalized graph cuts (see Subsection 4.2.1). For this purpose, it discretizes the range of the videokymogram to $\{0, \dots, 255\}$

$$f_d(x, y) = [255f_c(x, y)],$$

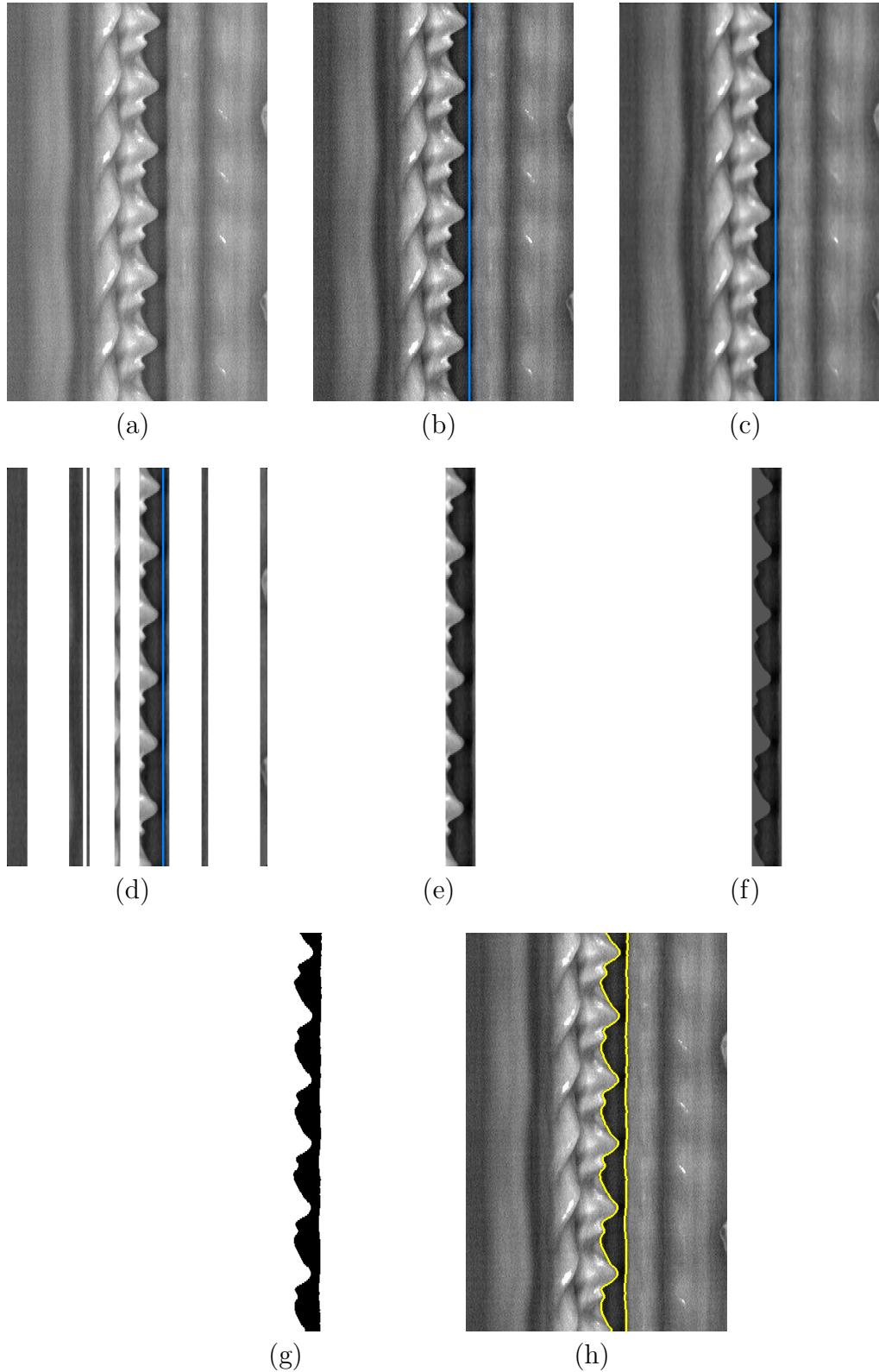


Figure 4.6: Segmentation of the shape of rima glottidis: (a) videokymogram, (b) centerline (blue) computed by the algorithm developed by Hauzar [17], (c) median filtering, (d) thresholding by column minima with threshold t_r (see Eq. (4.3.4)), (e) region corresponding to the centerline, (f) semi-thresholding of bright areas with threshold t_s (see Eq. (4.3.5)), (g) thresholding with threshold t_{GC} computed by graph cuts (see Eq. (4.3.6)), (h) segmented boundary of rima glottidis (yellow).

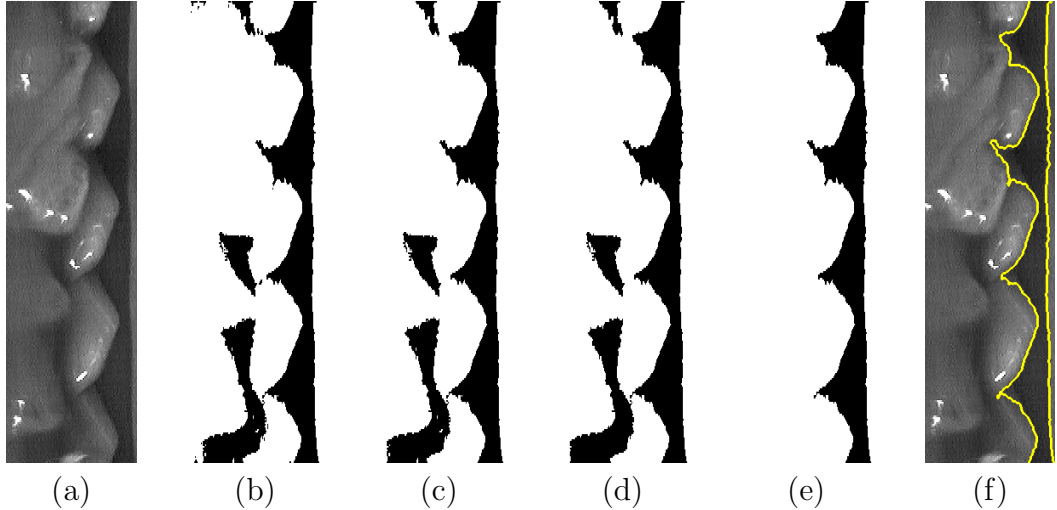


Figure 4.7: Refinement of the segmented shape of rima glottidis by morphological operations (cropped): (a) videokymogram with low contrast and pathologic vibratory pattern, (b) thresholding by graph cuts, (c) removal of small objects, (d) filling of holes, (e) removal of outer objects, (f) refined shape of rima glottidis (yellow).

where $[\cdot]$ denotes rounding. The method then removes bright values by semi-thresholding

$$f_{t_s}(x, y) = \begin{cases} t_s, & f_d(x, y) \geq t_s, \\ f_d(x, y), & \text{otherwise} \end{cases} \quad (4.3.5)$$

to reduce influence of very bright areas on the value of computed threshold (see Figure 4.6f). Then it constructs graph G for f_{t_s} according to the graph cut based thresholding algorithm described in Subsection 4.2.1, and minimizes the normalized graph cut measure

$$t_{GC} = \underset{t \in \{0, \dots, t_s\}}{\operatorname{argmin}} \operatorname{Ncut}(A_t, B_t), \quad A_t \cup B_t = D_c, \quad (4.3.6)$$

where $D_c = \{x_1, \dots, x_2\} \times \{1, \dots, y_{max}\}$ is the domain of f_{t_s} . The method then thresholds the digital image by the computed optimal threshold

$$f_{GC}(x, y) = \begin{cases} 1, & f_{t_s}(x, y) \leq t_{GC}, \\ 0, & \text{otherwise} \end{cases}$$

to regions defined by components $A_{t_{GC}}$ and $B_{t_{GC}}$, corresponding to the area rima of rima glottidis and its surroundings, respectively.

The resulting binary image f_{GC} may contain—apart from regions corresponding to rima glottidis—also incorrectly detected regions corresponding to dark areas of vocal folds and surrounding tissues, or holes in the regions of rima glottidis, corresponding to bright areas of reflections from tissues under rima glottidis. The refinement method gradually detects such regions and holes and removes them by standard morphological operations [46]. It namely removes small regions and regions outside the centerline, and detects the holes by labeling and fills them in

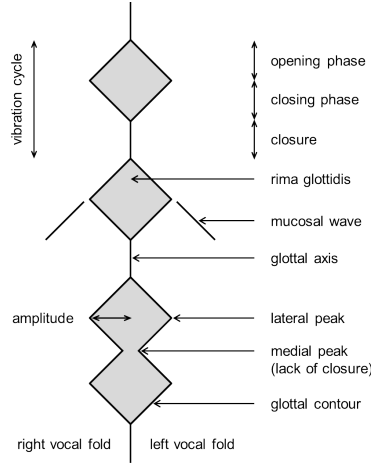


Figure 4.8: Vibration features in videokymograms.

(see Figure 4.7). Finally, the method positions the refined binary image within the domain of the original videokymogram f . The result is a binary image f_{rg} of the shape of rima glottidis (see Figure 4.6g). For visualization purposes, the method smooths the detected shape by morphological closing (see Figure 4.6h).

Detection of glottal features

This part describes a method for detection of vibration features defined by the shape of rima glottidis (glottal features). The method detects basic glottal features from the segmented shape of rima glottidis and from them derives other glottal features.

The shape of rima glottidis defines a number of vibration features in videokymograms. Clinically important glottal features include opening and closing points, lateral and medial peaks, vibration cycles and their opening, closing, open and closed phases, and glottal and vocal fold amplitudes (see Figure 4.8). The opening and closing points, and the lateral and medial peaks (basic glottal features) determine the other aforementioned features (derived glottal features).

Hauzar [17] developed an algorithm for extraction of basic glottal features directly from videokymograms. The algorithm computes the shape of rima glottidis locally for each feature by thresholding and region growing; the segmentation step of the algorithm is, however, sensitive to low contrast between rima glottidis and vocal folds. The algorithm may thus compute some features incorrectly, namely in poorly illuminated videokymograms of pathologic vibratory patterns.

feature	notation
opening points	O_i
closing points	C_i
lateral peaks	A_i^R, A_i^L
medial peaks	M_i^R, M_i^L

Table 4.1: Basic glottal features in videokymograms (see Figure 4.8); upper indices R and L denote the right and left vocal folds, respectively, and lower index i denotes the number of the vibration cycle in the videokymogram.

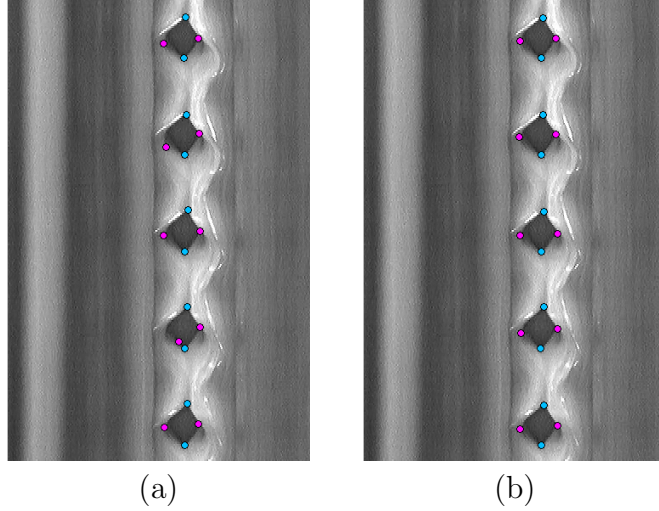


Figure 4.9: Detection of basic glottal features in a videokymogram with irregular vibratory pattern and low contrast by the algorithm developed by Hauzar: (a) directly from the videokymogram [17], (b) from the shape of rima glottidis segmented by the proposed method.

feature	notation and definition
generalized opening points	$\tilde{O}_i^j = \{O_i, M_i^j\}$
generalized closing points	$\tilde{C}_i^j = \{C_i, M_i^j\}$
opening phase duration	$t_i^{oj} = A_i^j(y) - \tilde{O}_i^j(y)$
closing phase duration	$t_i^{cj} = \tilde{C}_i^j(y) - A_i^j(y)$
open phase duration	$T_i^{oj} = t_i^{oj} + t_i^{cj} = \tilde{C}_i^j(y) - \tilde{O}_i^j(y)$
closed phase duration	$T_i^{cj} = \tilde{O}_{i+1}^j(y) - \tilde{C}_i^j(y)$
vibration cycle duration	$T_i^j = T_i^{oj} + T_i^{cj} = t_i^{oj} + t_i^{cj} + T_i^{cj} = \tilde{O}_{i+1}^j(y) - \tilde{O}_i^j(y)$
vocal fold amplitudes	$a_i^j = \max(A_i^j(x) - \tilde{O}_i^j(x) , A_i^j(x) - \tilde{C}_i^j(x))$
glottal amplitudes	$a_i = A_i^L(x) - A_i^R(x)$

Table 4.2: Derived glottal features in videokymograms (see Figure 4.8); upper index $j \in \{R, L\}$ denotes the right and left vocal folds, respectively, and lower index i denotes the number of the vibration cycle in the videokymogram.

The proposed feature detection method combines this algorithm with the method for global segmentation of rima glottidis by graph cuts. It first computes the shape of rima glottidis by the graph-cut-based segmentation method described above. The feature detection method then applies the algorithm developed by Hauzar to the inversion of the binary image of segmented rima glottidis³. The algorithm computes temporal (top-down, y) and lateral (left-right, x) coordinates of opening and closing points and lateral peaks, and only temporal coordinates of medial peaks. The proposed method then computes lateral coordinates of medial peaks from the intersection of their temporal coordinates with the segmented shape of rima glottidis.

The method computes derived glottal features from the detected basic glottal features. It first denotes the basic glottal features according to Table 4.1; upper

³ That is, rima glottidis black, vocal folds and surrounding tissues white.

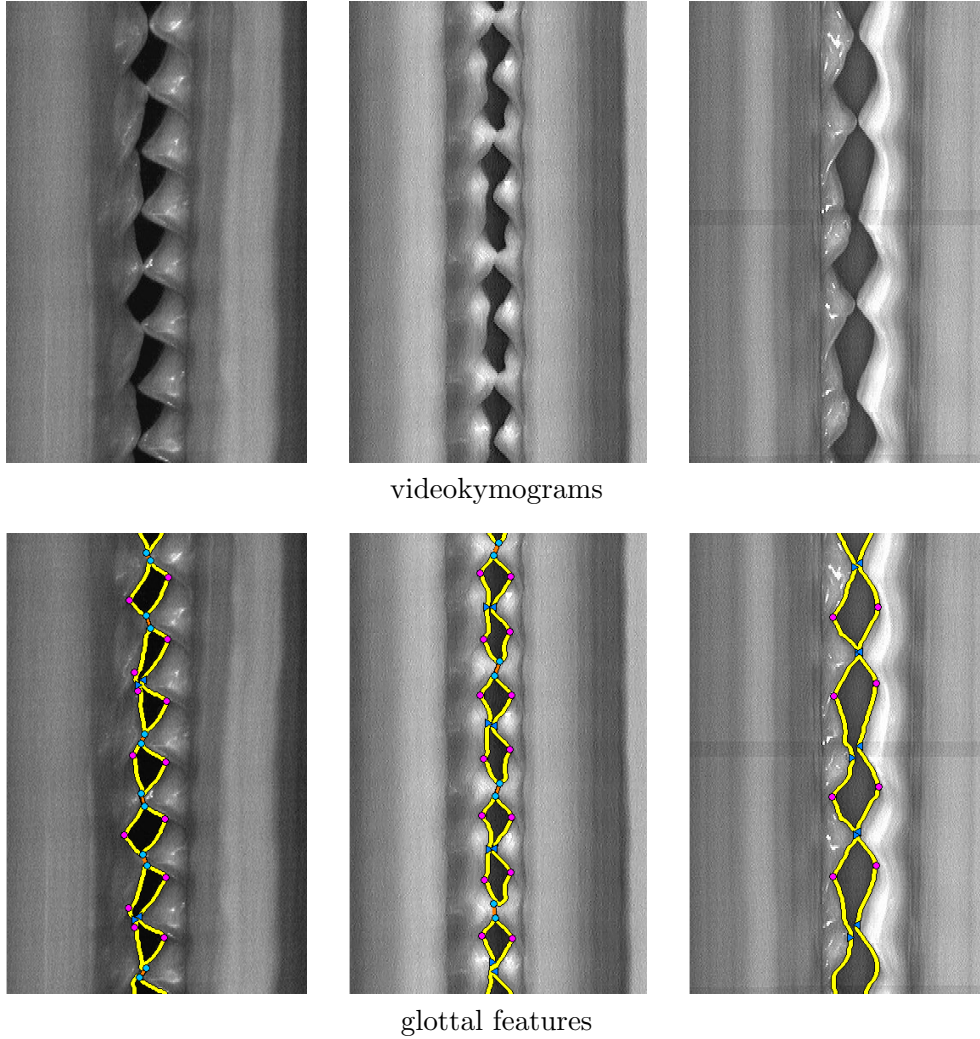


Figure 4.10: The shape of rima glottidis and basic glottal features (magenta: lateral peaks, light blue: opening and closing points, dark blue: medial peaks) detected by the proposed method in videokymograms with irregular vibratory patterns.

index R or L denotes correspondence of a feature to the right or left vocal fold, respectively, and lower index $i \in \{1, \dots, n\}$ denotes the number of the corresponding vibration cycle, where n is the number of cycles in the videokymogram. The method then computes the derived glottal features according to definitions in Table 4.2. It denotes the union of opening points and medial peaks as generalized opening points, and the union of closing points and medial peaks as generalized closing points. The generalized opening and closing points enclose open phases, while the generalized opening points separate vibration cycles⁴.

The detected glottal features determine a number of vocal fold vibration parameters evaluated in the VKG evaluation sheet. Subsection 4.3.4 defines these parameters and sets correspondences between their numerical values and categories in the evaluation sheet.

⁴ Depending on definition, vibration cycles can be separated either by generalized opening points or by lateral peaks or by generalized closing points.

4.3.3 Mucosal waves

This subsection describes the developed method for detection of mucosal wave directions in videokymograms. It detects the direction of each mucosal wave by Fourier transform. The method uses positions of lateral peaks computed by the method described in Subsection 4.3.2. The method can be used in combination with tracking methods, e.g. the correlation based algorithm proposed by Zita [68], as a consistency check of the computed trajectory direction.

Mucosal waves are vibratory waves on the upper surface of vocal folds. Transition of vocal fold movement from lateral to medial between the opening and closing phases raises a wave on the upper surface. The wave propagates laterally across the surface until it disappears or reaches the lateral border of the vocal fold. Presence of mucosal waves and their extent indicate how pliable a vocal fold is. The top or slope of mucosal waves are sometimes highlighted by specular reflections. Good illumination is critical for observation of mucosal waves.

In videokymograms, mucosal waves appear as diagonal paths on vocal folds from lateral peaks in the direction of the opening movement. The wave may gradually slow down, in which case the trajectory bends slightly downward. Specular reflections, if present, may either highlight the top of the wave and thus the mucosal wave path itself, or create a close parallel trajectory corresponding to a slope of the wave. Mucosal wave paths are not generally brighter or darker than the rest of the vocal fold; they rather appear as edges in direction of their propagation. Inappropriate illumination often complicates detection of mucosal waves in videokymograms.

Detection of mucosal wave direction

The proposed method detects mucosal wave direction by Fourier transform. The method first removes duplicate rows. Modern videokymographic cameras store each frame of the scanned line twice in the videokymogram, i.e. in successive odd and even rows

$$f_2(x, 2y - 1) = f_2(x, 2y), \quad y \in \{1, \dots, \frac{y_{max}}{2}\},$$

where f_2 denotes the recorded videokymogram and y_{max} the number of its rows. Duplicate rows would, however, double the frequency pattern in Fourier spectrum (see Figure 4.11c), and thus complicate detection of maximal directional energy. The method thus removes the duplicate rows. In videokymograms compressed by JPEG lossy compression, the successive odd and even row may slightly differ due to quantization. To compensate for the quantization noise, the method does not simply select either odd or even rows but rather averages them

$$f(x, y) = \frac{f_2(x, 2y - 1) + f_2(x, 2y)}{2}, \quad y \in \{1, \dots, \frac{y_{max}}{2}\}.$$

The result is a videokymogram without duplicities and with half the original number of rows (see Figure 4.11d).

The method detects the direction of individual waves in windows. For each lateral peak, except the first and the last one, it creates a rectangular window D vertically from the previous to the successive lateral peak and horizontally from the lateral peak in lateral direction to a convenient distance (see Figures 4.11d,

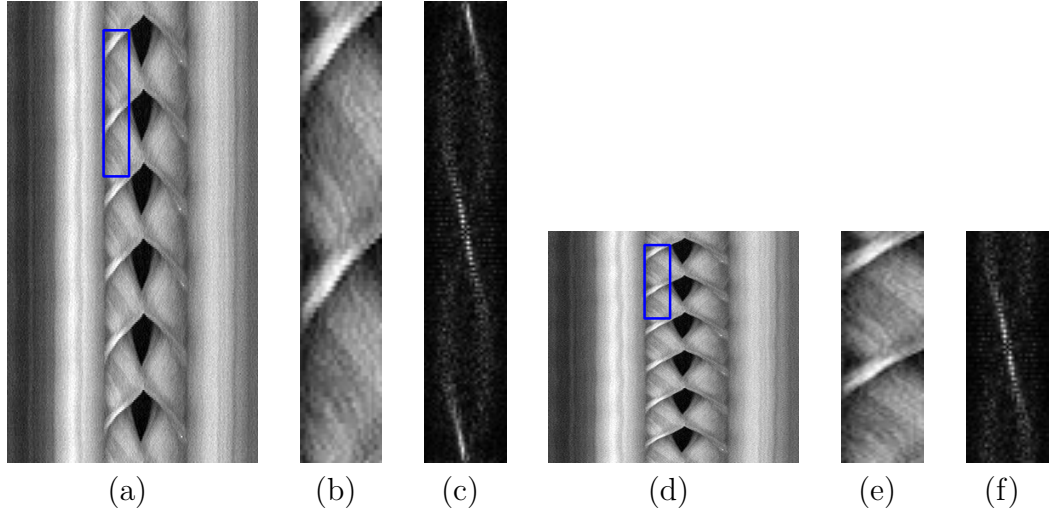


Figure 4.11: Impact of duplicate rows in videokymograms on pattern in Fourier spectrum: (a) videokymogram with duplicate rows, (b) window for detection of mucosal wave direction, (c) double pattern in Fourier spectrum, (d) videokymogram with removed duplicate rows, (e) window for detection of mucosal wave direction, (f) pattern in Fourier spectrum.

4.11e, and 4.12a), so that the width of the window is sufficient for directional analysis of frequency spectrum in discrete domain. For simplicity, the following description considers videokymographic image function f an absolutely integrable real function $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ of two variables $\{x, y\} \in \mathbb{R}^2$ with compact support D corresponding to the window; conversion to discrete domain is straightforward.

The method computes Fourier spectrum of function f

$$F(u, v) = \iint_D f(x, y) e^{-2\pi i(ux+vy)} dx dy,$$

and extracts its amplitude

$$|F(u, v)| = \iint_D f(x, y) \cos(2\pi(ux + vy)) dx dy.$$

Amplitude spectrum $|F(u, v)|$ usually contains horizontal stripes (see Figure 4.12b) caused by discontinuities between the left and right borders of the window. In order to suppress the stripes, the method removes the central value

$$F(0, 0) = \iint_D f(x, y) dx dy,$$

which is typically significantly larger than other values, from the amplitude spectrum

$$|F_0(u, v)| = \begin{cases} 0, & u = v = 0, \\ |F(u, v)|, & \text{otherwise} \end{cases}$$

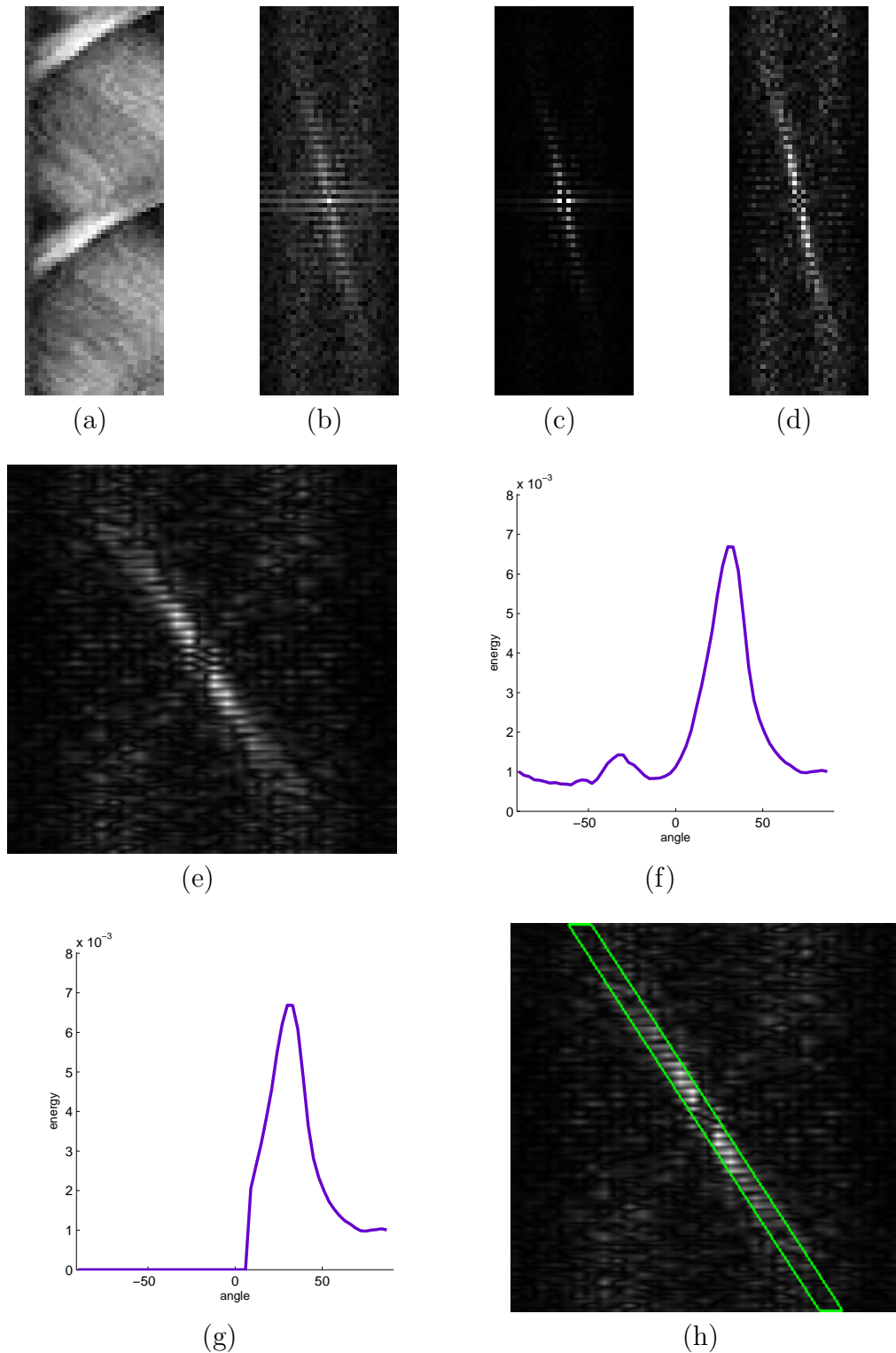


Figure 4.12: Detection of mucosal wave direction by Fourier transform in videokymograms with removed duplicities: (a) window, (b) amplitude of Fourier spectrum (logarithmic scale), (c) amplitude spectrum with removed center, (d) spectrum with normalized rows, (e) spectrum stretched to square shape, (f) directional energy of the spectrum, (g) restriction to angles of possible mucosal wave directions, (h) direction with maximal energy (green) indicates mucosal waves in perpendicular direction (compare with (a)).

(see Figure 4.12c) and divides the spectrum by the integral of horizontal direction⁵

$$|F_1(u, v)| = \frac{|F_0(u, v)|}{\int_D |F_0(u, v)| du}$$

(see Figure 4.12d).

The method measures directional energy in spectrum $|F_1(u, v)|$. It stretches the computed spectrum to square shape (see Figure 4.12e) so that its distribution of angles corresponds to distribution of angles in image domain. The method then computes directional energy of the spectrum

$$E(\psi) = \iint_{S_\psi} |F_1(u, v)|^2 dudv, \quad \psi \in (-\frac{\pi}{2}, \frac{\pi}{2})$$

(see Figure 4.12f). To increase robustness to noise, it measures the energy in bands

$$S_\psi = \{(u, v) \mid |u \cos \psi - v \sin \psi| \leq r\}, \quad \psi \in (-\frac{\pi}{2}, \frac{\pi}{2}), \quad r \geq 0 \quad (4.3.7)$$

of convenient width $2r$ centered in and rotated around zero. It computes the angle with maximal directional energy as

$$\psi_{max}^j = \operatorname{argmax}_{\psi \in \Psi^j} (E(\psi)), \quad j \in \{R, L\}$$

(see Figure 4.12h), where

$$\Psi^j = \begin{cases} (-\frac{\pi}{2}, -\varepsilon), & j = R, \\ (\varepsilon, \frac{\pi}{2}), & j = L \end{cases} \quad (4.3.8)$$

are intervals of possible mucosal wave directions for the right and left vocal folds, respectively (see Figure 4.12g). The method excludes angles from interval near zero

$$[-\varepsilon, \varepsilon], \quad \varepsilon \geq 0$$

because the corresponding bands are often disrupted by frequency peaks caused by discontinuities on vertical borders of window D in its periodic composition; these angles correspond to vertical directions in image domain. The discontinuities on borders could be suppressed by smooth fading of the image function near vertical borders to zero, e.g. by multiplication with a Gaussian. In discrete case, however, such fading would further reduce amount of information in Fourier spectrum. Because the pixel-wise extent of mucosal waves is narrow and because the direction of mucosal waves is not vertical, the method just excludes angles within an interval near zero from search of the angle with maximal directional energy.

Because Fourier transform projects edges from image domain to peaks in perpendicular direction in Fourier domain, the angle with maximal directional energy in Fourier domain corresponds to perpendicular angle

$$\phi_{max}^j = \psi_{max}^j - \frac{\pi}{2}, \quad j \in \{R, L\}$$

⁵ This corresponds to division by row sums in discrete domain.

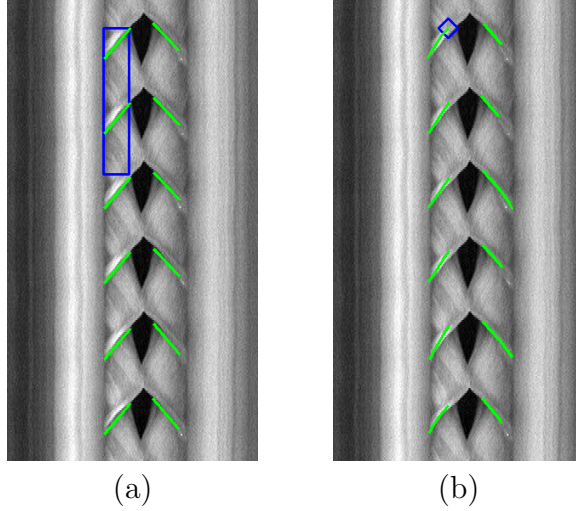


Figure 4.13: Detection of mucosal waves: (a) directions detected by Fourier transform, (b) trajectories detected by the correlation-based algorithm developed by Zita [68].

in the videokymogram. Angle ϕ_{max}^j thus estimates the direction of mucosal waves in the videokymogram within window D . Finally, the method adjusts the detected angle to correspond to the size of the original videokymogram with duplicate rows

$$\phi_{2max}^j = \arctan(2 \tan(\phi_{max}^j)), \quad j \in \{R, L\}.$$

Tracking of mucosal waves

The tracking algorithm proposed by Zita [68] detects the path of mucosal waves by cross-correlation. The algorithm estimates the direction of a mucosal wave as the direction between the corresponding generalized opening point and lateral peak (see Subsection 4.3.2). It then creates a correlation window (see Figure 4.13b) above the wave and tracks its path from the lateral peak in the estimated direction. It estimates the extent of the wave by thresholding of correlation values.

The directions between a generalized opening point and the corresponding medial peak may, however, differ from the direction of the mucosal wave (see Figure 4.14). The direction computed by the Fourier-based method does not depend on the position of generalized opening points and can thus be used as a consistency check.

The combined algorithm estimates direction of the mucosal wave

$$\alpha^j = \phi_{2max}^j$$

by Fourier transform, and also by the direction of glottal opening β^j , where

$$\beta^L = \arctan\left(\frac{\tilde{O}^L(x) - A^L(x)}{A^L(y) - \tilde{O}^L(y)}\right),$$

$$\beta^R = \beta^L - \frac{\pi}{2},$$

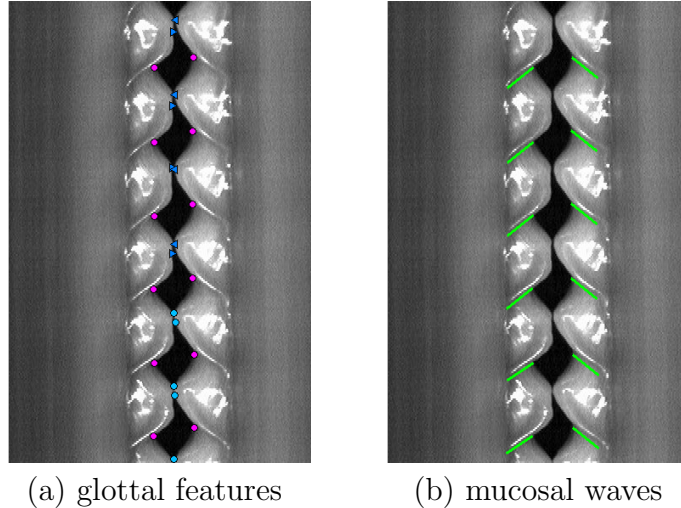


Figure 4.14: Directions between generalized opening points and lateral peaks may differ from directions of mucosal waves: (a) generalized opening points (light blue: opening points, dark blue: medial peaks) and lateral peaks (magenta) detected by the proposed graph-cut-based method using the feature extraction algorithm developed by Hauzar [17] (see Subsection 4.3.2); (b) directions of mucosal waves (green) detected by the proposed Fourier-based method.

are computed from the corresponding generalized opening point \tilde{O}^j and lateral peak A^j ; $j \in \{R, L\}$, where R and L denote the right and left vocal folds. Significant difference between the estimated mucosal wave direction α^j and the glottal opening direction β^j

$$|\alpha^j - \beta^j| > \delta, \quad \delta > 0,$$

where δ is a convenient maximal allowed difference, indicates inconsistency. The inconsistency can be caused by high curvature of the glottal contour during the opening phase or by a weak or distorted mucosal wave pattern. In such a case, the trajectory computed by tracking in either direction should be visually verified.

4.3.4 Parameters of vocal fold vibrations

This subsection quantifies vibration parameters determined by the computed glottal features (see Subsection 4.3.2), and sets correspondence between their numerical values and categories in the VKG evaluation sheet [59]. It numbers the parameters in accordance with notation in [15] and quantifies them in accordance with [54]; the intervals for classification into categories in [54] were estimated manually from the corresponding pictograms in the evaluation sheet.

For computation of vibration parameters, the evaluation method has to perform a consistency check of the detected features. The employed methods may not detect all features, especially in videokymograms with poor illumination or pathologic vibratory pattern. The evaluation method detects inconsistencies by inspection of the detected features. It first hierarchically divides the features into vibration cycles of the right and left vocal folds, and then compares the

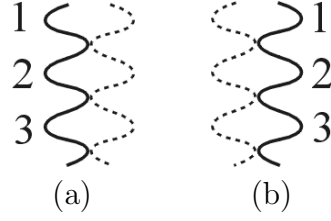


Figure 4.15: (8)-(9) Number of cycles: (a) right vocal fold, (b) left vocal fold. Correspondence between pictograms and categories in the VKG evaluation sheet. (Reprinted from [15, 59] with permission.)

position and length of corresponding cycles. The method accounts for possible inconsistencies in computation of the parameters.

Number of cycles

$$(8) \text{NumberOfCyclesR} = \frac{y_{max}}{\bar{T}^R}$$

$$(9) \text{NumberOfCyclesL} = \frac{y_{max}}{\bar{T}^L}$$

The Number of cycles parameter (see Figure 4.15) is defined by the duration of the recorded videokymogram y_{max} and of the average length of vibration cycle

$$\bar{T}^j = \frac{1}{n_j} \sum_{i=1}^{n_j} T_i^j, \quad j = R, L,$$

where $n_R = |\tilde{O}^R| - 1$ and $n_L = |\tilde{O}^L| - 1$ are the number of full cycles of the right and left vocal fold, respectively, in the videokymogram.

Cycle-to-cycle variability

$$(10) \text{VariabilityR} = \text{median}_{i=1, \dots, n-1} API(i, R)$$

$$(11) \text{VariabilityL} = \text{median}_{i=1, \dots, n-1} API(i, L)$$

The Cycle-to-cycle variability parameter (see Figure 4.16 and Table 4.3) is defined by the Amplitude Periodicity Index (API) [36]

$$API(i, j) = \frac{\min\{a_i^j, a_{i+1}^j\}}{\max\{a_i^j, a_{i+1}^j\}}, \quad i = 1, \dots, n-1, \quad j = R, L.$$

Duration of closure

$$(13) \text{ClosureDuration} = \text{median}_{i=1, \dots, n} CQ(i)$$

The Duration of closure parameter (see Figure 4.17 and Table 4.4) is defined by the Closed Quotient (CQ)

$$CQ(i) = \frac{T_i^c}{T_i}, \quad i = 1, \dots, n.$$

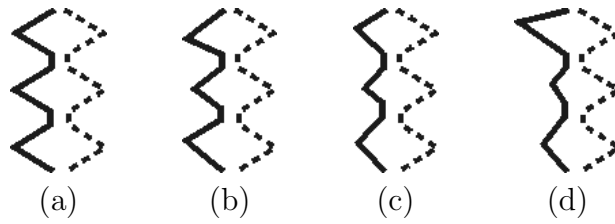


Figure 4.16: (10)-(11) Cycle-to-cycle variability: (a) negligible, (b) small, (c) medium, (d) large. Correspondence between pictograms and categories in the VKG evaluation sheet. (Reprinted from [15, 59] with permission.)

category	description	VariabilityR, VariabilityL
1	negligible	(0.85, 1]
2	small	(0.61, 0.85]
3	medium	(0.5, 0.61]
4	large	[0, 0.5]

Table 4.3: (10)-(11) Cycle-to-cycle variability: correspondence between numerical values and categories of the parameter in the VKG evaluation sheet.

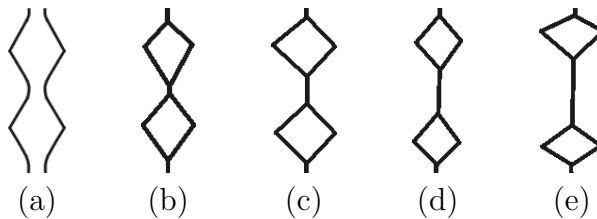


Figure 4.17: (13) Duration of closure: (a) no closure, (b) 1-20%, (c) 20-40%, (d) 40-60%, (e) >60%. Correspondence between pictograms and categories in the VKG evaluation sheet. (Reprinted from [15, 59] with permission.)

category	description	ClosureDuration
1	no closure	[0, 0.01]
2	1-20%	(0.01, 0.2]
3	20-40%	(0.2, 0.4]
4	40-60%	(0.4, 0.6]
5	>60%	(0.6, 1]

Table 4.4: (13) Duration of closure: correspondence between numerical values and categories of the parameter in the VKG evaluation sheet.

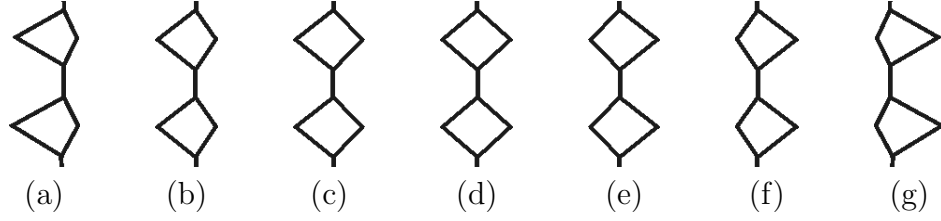


Figure 4.18: (14) Amplitude differences: (a) R much larger, (b) R larger, (c) R slightly larger, (d) R \sim L, (e) L slightly larger, (f) L larger, (g) L much larger. Correspondence between pictograms and categories in the VKG evaluation sheet. (Reprinted from [15, 59] with permission.)

category	description	AmplitudeDifferences
1	R much larger	$[-1, -0.6)$
2	R larger	$[-0.6, -0.31)$
3	R slightly larger	$[-0.31, -0.1)$
4	R \sim L	$[-0.1, 0.1]$
5	L slightly larger	$(0.1, 0.31]$
6	L larger	$(0.31, 0.6]$
7	L much larger	$(0.6, 1]$

Table 4.5: (14) Amplitude differences: correspondence between numerical values and categories of the parameter in the VKG evaluation sheet.

Amplitude differences

$$(14) \text{AmplitudeDifferences} = \text{median}_{i=1, \dots, n} ASI(i)$$

The Amplitude differences parameter (see Figure 4.18 and Table 4.5) is defined by the Amplitude Symmetry Index (ASI) [36]

$$ASI(i) = \frac{a_i^L - a_i^R}{a_i^L + a_i^R}, \quad i = 1, \dots, n.$$

Frequency differences

$$(15) \text{FrequencyDifferences} = \frac{\text{NumberOfCyclesL}}{\text{NumberOfCyclesR}}$$

The Frequency differences parameter (see Figure 4.19 and Table 4.6) can be defined by the number of cycles (see parameters (8)-(9)).

Phase differences

$$(16) \text{PhaseDifferences} = \text{median}_{i=1, \dots, n} PSI(i)$$

The Phase differences parameter (see Figure 4.20 and Table 4.7) is defined by the Phase Symmetry Index (PSI) [36]

$$PSI(i) = \frac{A_i^L(y) - A_i^R(y)}{T_i}, \quad i = 1, \dots, n.$$

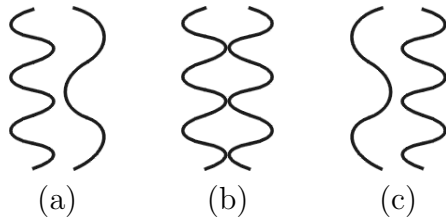


Figure 4.19: (15) Frequency differences: (a) R faster than L, (b) L and R equal, (c) L faster than R. Correspondence between pictograms and categories in the VKG evaluation sheet. (Reprinted from [15, 59] with permission.)

category	description	FrequencyDifferences
1	R faster than L	$(0, 0.91)$
2	L and R equal	$[0.91, 1.1)$
3	L faster than R	$[1.1, \infty)$

Table 4.6: (15) Frequency differences: correspondence between numerical values and categories of the parameter in the VKG evaluation sheet; the evaluation sheet numbers the “R faster than L” category by 2, and the “L and R equal” category by 1.

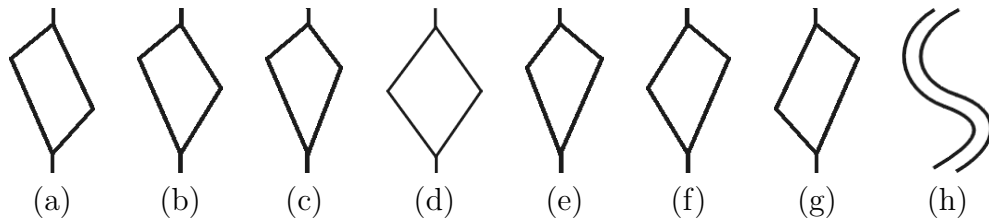


Figure 4.20: (16) Phase differences: (a) R ahead of L: large, (b) medium, (c) small, (d) negligible, (e) L ahead of R: small, (f) medium, (g) large, (h) lambda. Correspondence between pictograms and categories in the VKG evaluation sheet. (Reprinted from [15, 59] with permission.)

category	description	PhaseDifferences
1	R ahead of L: large	$(0.3, 1]$
2	R ahead of L: medium	$(0.15, 0.3]$
3	R ahead of L: small	$(0.05, 0.15]$
4	negligible	$[-0.05, 0.05]$
5	L ahead of R: small	$[-0.15, -0.05)$
6	L ahead of R: medium	$[-0.3, -0.15)$
7	L ahead of R: large	$[-1, -0.3)$
14	lambda	yet to be quantified

Table 4.7: (16) Phase differences: correspondence between numerical values and categories of the parameter in the VKG evaluation sheet; the evaluation sheet numbers the “lambda” category by 8.

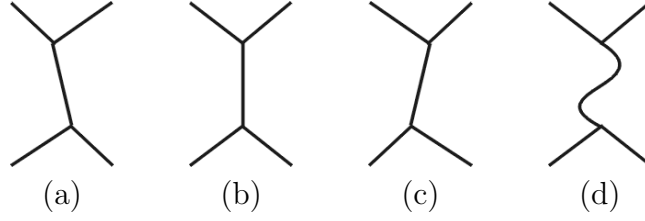


Figure 4.21: (17) Axis shift: (a) R → L, (b) negligible, (c) L → R, (d) complex. Correspondence between pictograms and categories in the VKG evaluation sheet. (Reprinted from [15, 59] with permission.)

category	description	AxisShift
1	R → L	$(0.1, \infty)$
2	negligible	$[-0.1, 0.1]$
3	L → R	$(-\infty, -0.1)$
6	complex	yet to be quantified

Table 4.8: (17) Axis shift: correspondence between numerical values and categories of the parameter in the VKG evaluation sheet; the evaluation sheet numbers the “R → L” category by 2, the “negligible” category by 1, and the “complex” category by 4.

Axis shift

$$(17) \text{ AxisShift} = \text{median}_{i=1, \dots, n-1} AS(i)$$

The Axis shift parameter (see Figure 4.21 and Table 4.8) is defined by the Axis Shift (AS) [30]

$$AS(i) = \frac{O_{i+1}(x) - C_i(x)}{a_i}, \quad i = 1, \dots, n.$$

Opening versus closing duration

$$(18) \text{ SkewingR} = \text{median}_{i=1, \dots, n} SI(i, R)$$

$$(19) \text{ SkewingL} = \text{median}_{i=1, \dots, n} SI(i, L)$$

The Opening versus closing duration parameter (see Figure 4.18 and Table 4.9) is defined by the Speed Index (SI) [20]

$$SI(i, j) = \frac{t_i^{oj} - t_i^{cj}}{T_i^o} = \frac{t_i^{oj} - t_i^{cj}}{t_i^o + t_i^c} = \frac{SQ(i, j) - 1}{SQ(i, j) + 1}, \quad i = 1, \dots, n, \quad j = R, L,$$

which can be defined by the Speed Quotient (SQ) [20]

$$SQ(i, j) = \frac{t_i^{oj}}{t_i^{cj}}, \quad i = 1, \dots, n, \quad j = R, L.$$

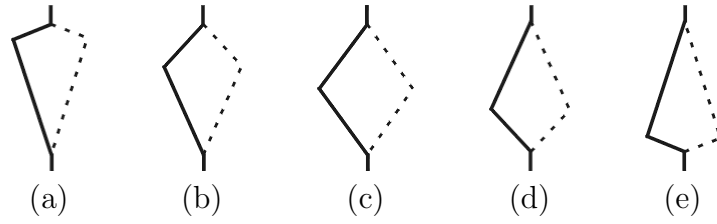


Figure 4.22: (18)-(19) Opening versus closing duration: (a) much shorter, (b) shorter, slightly shorter, (c) equal, (d) slightly longer, longer, (e) much longer. Correspondence between pictograms and categories in the VKG evaluation sheet. (Reprinted from [15, 59] with permission.)

category	description	SkewingR, SkewingL
1	much shorter	$[-1, -0.75]$
2	shorter	$[-0.75, -0.35]$
3	slightly shorter	$[-0.35, -0.05]$
4	equal	$[-0.05, 0.05]$
5	slightly longer	$[0.05, 0.35]$
6	longer	$[0.35, 0.75]$
7	much longer	$[0.75, 1]$

Table 4.9: (18)-(19) Opening versus closing duration: correspondence between numerical values and categories of the parameter in the VKG evaluation sheet.

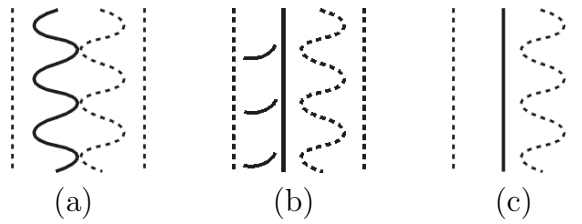


Figure 4.23: (1)-(2) Presence of vibrations: right vocal fold (a) vibrating, (b) vibrating partly, (c) not vibrating. Correspondence between pictograms and categories in the VKG evaluation sheet. (Reprinted from [15, 59] with permission.)

category	description
1	vibrating
2	vibrating partly
3	not vibrating

Table 4.10: Presence of vibrations: categories in the VKG evaluation sheet.

The VKG evaluation sheet evaluates parameters (8)-(9) by integers, and parameters (10)-(11), (13), (14), (15), (16), (17), and (18)-(19) by disjoint categories numbered by positive integers.

The VKG evaluation sheet defines special categories and causalities. All parameters include category “NA” (not applicable), numbered⁶ $-\infty$. The NA category indicates that the parameter is not applicable or that its evaluation is problematic or just that it was not evaluated. Whereas neighboring non-NA categories correspond to similar values of vibration parameters, the NA category is qualitatively different. The evaluation sheet also defines causalities between categories of several parameters (see Table 4.11); the order of parameters in the evaluation sheet follows the causalities.

condition	implication
$(1) = 3 \vee (2) = 3$	$\Rightarrow (14), (15), (16), (17) = -\infty$
$(1) = 3$	$\Rightarrow (8), (10), (18) = -\infty$
$(2) = 3$	$\Rightarrow (9), (11), (19) = -\infty$
$(13) = 1$	$\Rightarrow (17) = -\infty$
$(15) = 1 \vee (15) = 3$	$\Rightarrow (16), (17) = -\infty$

Table 4.11: Causality of parameter categories in the VKG evaluation sheet.

Most parameters depend on categories of the Presence of vibrations (1)-(2) parameter, which indicates whether the right and left vocal folds vibrate (see Figure 4.23 and Table 4.10). The parameter has yet to be quantified, however. The automatic evaluation computes non-NA values of the depending parameters. It can then apply the implications from Table 4.11 based on a visual evaluation of the Presence of vibrations parameter, or the consensus result of multiple visual evaluations of the parameter (see Subsection 4.3.5).

To separate qualitatively different categories and order categories according to their similarity, this chapter renumbered the categories in the evaluation sheet:

- the “NA” category from 0 to $-\infty$,
- the “R faster than L” and “L and R equal” categories of the Frequency differences (15) parameter from 2 to 1 and 1 to 2, respectively,
- the “lambda” category of the Phase differences (16) parameter from 8 to 14,
- and the “negligible”, “R \rightarrow L” and “complex” categories of the Axis shift (17) parameter from 1 to 2, 2 to 1, and 4 to 6, respectively.

4.3.5 Comparison of automatic and visual evaluations

This subsection describes a method for comparison of the performance of automatic and visual evaluations. It compares results of one automatic evaluation with results of one or multiple visual evaluations.

⁶ The evaluation sheet numbers the NA category by 0.

The method introduces the following notation. Let P denote the set of both automatically and visually evaluated parameters, n the number of evaluated videokymograms, and m the number of visual evaluations. Let $E_A(p, i)$ ($p \in P; i = 1, \dots, n$) denote the result of automatic evaluation of parameter p in videokymogram i , and $E_V(p, i, j)$ ($p \in P; i = 1, \dots, n; j = 1, \dots, m$) the result of j^{th} visual evaluation of parameter p in videokymogram i . Let $V^+(p, i)$ denote set of indices of visual evaluations of parameter p in videokymogram i with non-NA results

$$V^+(p, i) = \{j \mid j \in \{1, \dots, m\} \ \& \ E_V(p, i, j) > 0\}, \quad p \in P, \quad i = 1, \dots, n.$$

The method defines the consensus result of m visual evaluations of parameter p in videokymogram i as

$$E_V(p, i) = \begin{cases} \text{mode}_{j \in V^+(p, i)} E_V(p, i, j), & |V^+(p, i)| > \frac{m}{2}, \\ -\infty, & \text{otherwise,} \end{cases} \quad p \in P, \quad i = 1, \dots, n.$$

That is, the consensus result is NA if and only if the result of at least half of corresponding visual evaluations was NA; otherwise, the definition estimates it by the most frequent non-NA result.

For each parameter $p \in P$ and videokymogram $i \in \{1, \dots, n\}$, the proposed method compares the consensus result of visual evaluations $E_V(p, i)$ with the result of automatic evaluation $E_A(p, i)$ (automatic–visual match) as well as with the results of visual evaluations $E_V(p, i, j)$ ($j = 1, \dots, m$) (visual–visual match). It considers two evaluations of parameter $p \in P$ in a videokymogram matching if their respective results v_1, v_2 fall into the same category or into neighboring non-NA categories

$$\text{match}(v_1, v_2, d_p) = \begin{cases} 1, & v_1 = v_2 = -\infty \vee |v_1 - v_2| \leq d_p, \\ 0, & \text{otherwise,} \end{cases}$$

where $d_p \geq 0$ denotes tolerated distance between similar categories of parameter p .

The comparison method defines the automatic–visual match for parameter $p \in P$ with tolerance d_p as

$$\text{match}_{AV}(p, d_p) = \frac{1}{n} \sum_{i=1}^n \text{match}(E_A(p, i), E_V(p, i), d_p),$$

and the visual–visual match for parameter p as

$$\text{match}_{VV}(p, d_p) = \frac{1}{n} \sum_{i=1}^n \frac{1}{m} \sum_{j=1}^m \text{match}(E_V(p, i, j), E_V(p, i), d_p).$$

The automatic–visual and visual–visual matches indicate the success rate of the automatic and visual evaluations, respectively, of parameter $p \in P$ in all n videokymograms with respect to the consensus results of visual evaluations and tolerance d_p . For each parameter, the automatic–visual and visual–visual matches

relate performance of the automatic evaluation to performance of the visual evaluations. Error rate

$$\text{error}_{\text{AV}}(p, d_p) = 100\% - \text{match}_{\text{AV}}(p, d_p)$$

indicates how often a verifier would have to correct the automatically computed categories of parameter p to achieve a 100% match with the consensus category of visual evaluations with tolerance d_p .

4.4 Results

The developed methods were tested on two sets of videokymographic images. The first set consisted of 50 videokymograms⁷ with a representative range of both physiologic and pathologic vibratory patterns. The second set consisted of 6 videokymograms⁸ with physiologic vibratory patterns. The images were acquired by a videokymographic camera⁹ with scanning frequency 7200 rows per second; the camera stored each frame of the scanned line twice in the videokymograms.

The 50 images from the first set were visually evaluated according to the VKG evaluation sheet in 18 evaluations [15]. They were analyzed by 10 evaluators with different level of experience. Eight of them analyzed the images twice, in a randomized order and with a time delay so as to minimize dependence between the first and the second evaluations.

The proposed methods (see Section 4.3) automatically evaluated all 56 videokymograms. As a preprocessing step, the methods first eliminated duplicate rows¹⁰; the detected features were then recomputed to match the original size of the videokymograms. The method described in Subsection 4.3.1 detected and removed specular reflections from the videokymograms; the high and low thresholds were set to $t_1 = 0.85$ and $t_2 = 0.75$, respectively (see Eqs. 4.3.2 and 4.3.3). The method described in Subsection 4.3.2 segmented the shape of rima glottidis and computed basic and derived glottal features (see Tables 4.1 and 4.2); the maximal brightness of rima glottidis in normalized videokymograms was experimentally estimated by $t_r = 0.25$ (see Eq. (4.3.4)), and the semi-threshold suppressing the impact of very bright areas was set to $t_s = \frac{1}{3}255 = 85$ (see Eq. (4.3.5)); parameters of the graph cut algorithm were set to $c_f = 625$, $c_d = 4$, $r = 2$ (see Eq. (4.2.1)). The method described in Subsection 4.3.3 then estimated directions of mucosal waves; the width of bands for computation of directional energy was set to $2r = 0.05w_D$, where w_D denotes the width of the amplitude spectrum stretched to square (see Eq. (4.3.7)), and the interval of angles near zero to $[-\epsilon, \epsilon] = [-\frac{\pi}{30}, \frac{\pi}{30}]$ (see Eq. (4.3.8)). The method described in Subsection 4.3.4 computed from the detected glottal features numerical values and VKG evaluation sheet categories of corresponding vibration parameters.

⁷ Dimensions of the videokymograms ranged from 211×544 to 360×544 pixels, corresponding to intervals of 40 ms.

⁸ Dimensions of the videokymograms were 352×574 pixels, also corresponding to intervals of 40 ms.

⁹ The videokymographic camera stored the images in a 24bit JPEG lossy compression format. For further processing, the images were converted to an 8bit grayscale PNG lossless compression format.

¹⁰ Although the elimination step is important only for detection of mucosal wave directions by Fourier transform, it reduces the amount of processed data.

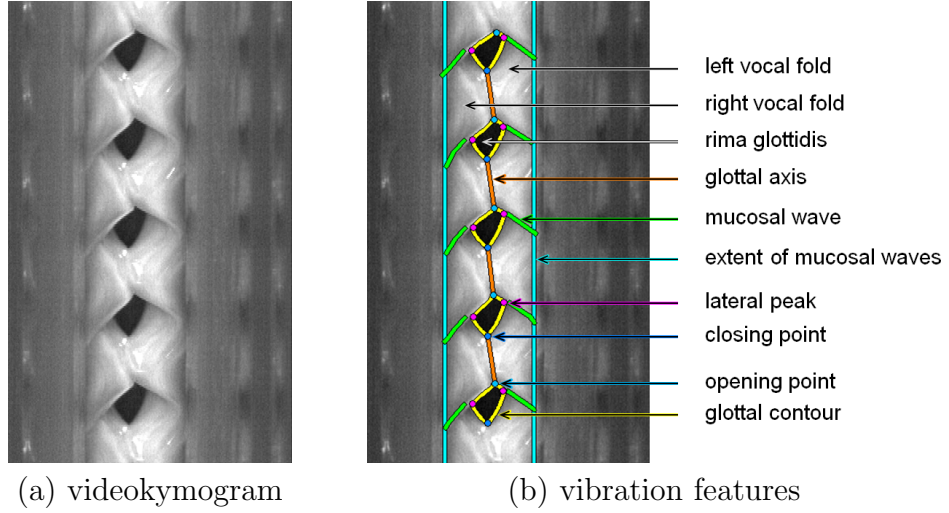


Figure 4.24: (a) Videokymogram of vocal fold vibrations. (b) Vibration features detected automatically by the proposed method with use of the feature detection algorithm developed by Hauzar [17] and the mucosal wave tracking algorithm developed by Zita [68].

parameter	automatic–visual match	visual–visual match
(8) NumberOfCyclesR	98%	95%
(9) NumberOfCyclesL	98%	96%
(10) VariabilityR	92%	93%
(11) VariabilityL	88%	93%
(13) ClosureDuration	98%	93%
(14) AmplitudeDifferences	100%	88%
(15) FrequencyDifferences	100%	95%
(16) PhaseDifferences	88%	85%
(17) AxisShift	88%	79%
(18) SkewingR	86%	87%
(19) SkewingL	90%	85%

Table 4.12: Comparison of results of automatic and visual evaluations on a set of 50 videokymograms by the automatic–visual and visual–visual match with tolerance between closely neighboring classes ($d_p = 1$, see Subsection 4.3.5). The similarity in comparative statistics for each parameter indicate that the performance of the automatic evaluation is comparable with visual evaluations.

For the 50 videokymograms from the first set—and for each automatically evaluated parameter separately—the comparison method described in Subsection 4.3.5 also computed consensus categories of the 18 visual evaluations, the automatic–visual and visual–visual matches (see Table 4.12), and contingency tables between the automatic and visual evaluations (see Table 4.13); the tolerance of the match was set to $d_p = 1$ for all parameters.

A \ V	$-\infty$	1	2	3	4	5	\sum	rate ($d_p = 1$)	rate ($d_p = 0$)
$-\infty$	0	0	0	0	0	0	0	–	–
1	18	217	34	1	0	0	270	<i>93%</i>	80%
2	6	11	90	1	0	0	108	<i>94%</i>	83%
3	14	6	142	76	14	0	252	<i>92%</i>	30%
4	7	1	30	78	90	28	234	<i>84%</i>	38%
5	8	0	0	0	3	25	36	<i>78%</i>	69%

Table 4.13: Contingency table comparing results of the automatic evaluation (left) with results of 18 visual evaluations (top) of the Closure duration (13) parameter in 50 videokymograms. The tables shows the rates of match with tolerance between closely neighboring categories ($d_p = 1$), and, for illustration, also the absolute match ($d_p = 0$).

4.5 Discussion

In order to evaluate performance of the developed methods, results of the automatic evaluation were compared with results of visual evaluations (see Subsection 4.3.5). Visual evaluations are, however, just approximations of ground truth, since even experienced clinicians may differ in evaluation of videokymograms, i.e. even a perfect match between multiple evaluations does not necessarily imply accuracy. The project thus computed just the consistency between automatic and visual evaluations. To evaluate performance of the visualizations, it compared their results with the consensus results of visual evaluations, which represent the best available approximation of ground truth. The comparison method tolerated classification into closely neighboring categories because the VKG evaluation sheet divides continuously distributed parameters into disjoint categories and also defines the categories ambiguously by labels and pictograms.

Similarities in results of the automatic and visual evaluations indicate that the developed methods could be used for computer-aided evaluation in clinical practice. The automatic–visual and visual–visual statistics were comparable for all automatically evaluated parameters (see Table 4.12). These result indicate similar performance of automatic and visual evaluations. The error rates of the automatic evaluation imply that a verifier would have to correct only a small percentage of automatically computed parameter values to match the corresponding consensus categories with the defined tolerance between directly neighboring categories. Because the automatic–visual and visual–visual comparative statistics do not account for differences in probability of categories, the project compared results of the automatic evaluation with results of all visual evaluations by contingency tables (see Table 4.13). The match rates indicate good performance of the developed automatic evaluation with respect to visual evaluations.

Differences in results of the automatic and visual evaluations were caused by two main factors. Firstly, the feature detection methods were not errorless. They did not detect some features correctly, namely in videokymograms with poor illumination or pathologic vibratory patterns. This was caused by errors either in detection of the rima glottidis shape or in the extraction of glottal features. The employed feature detection algorithm developed by Hauzar [17] suppresses false positives, i.e., it does not include features detected with low confidence in

its results. Although the evaluation method performed consistency check on the computed features, the errors in detection decreased accuracy of the computed parameter values, namely for irregular vibratory patterns. Secondly, the exact correspondence between numerical values and categories of parameters in the VKG evaluation sheet was unknown to the evaluators. They selected the categories according to corresponding descriptions and pictograms. The ambiguity in categorization implies necessity for tolerance between similar categories. It also projected into differences between results of individual visual evaluations.

Accuracy and precision of the automatic evaluation depend on three main factors. Firstly, high variance of vibratory patterns and various artifacts in videokymograms limit the success rate of the automatic methods. High variance—typical of medical data—limits applicability of automatic methods because they cannot account for all possible variances of vibratory patterns. Secondly, the values of parameters of the automatic methods affect their sensitivity to various factors, e.g. illumination and type of vibratory pattern. Different parameter values could lead to different results of the automatic evaluation; the thresholds for segmentation of specular reflections, for example, directly affect the size of segmented regions (see Subsection 4.3.1). The parameters of the developed methods were set experimentally. Thirdly, correspondences between numerical values and categories of parameters in the evaluation sheet were estimated manually from the pictograms [54]. In contrast to visual evaluation, however, the automatic methods evaluate numerical values of parameters, not just categories. This means that the automatic methods allow evaluation with higher precision.

The proposed methods could be modified in two main ways to increase accuracy of the automatic evaluation. Firstly, the parameters of the methods could be computed by supervised learning. This would require a large representative training set, however, to account for high variability of medical data. The learning could increase accuracy and applicability of the methods in practice. Secondly, the proposed methods could be combined with other existing methods for evaluation of vibration parameters to increase robustness of the evaluation to high variability of videokymographic data.

The developed methods will be integrated into a software application for computer-aided evaluation of vocal fold vibrations in clinical practice. The methods for growth reconstruction (see Chapter 2) and particle measurement (see Chapter 3) developed in the previous two projects were designed for academic research purposes; they were therefore implemented in Matlab with use of the Image Processing Toolbox [21] to allow their further development with easy modification and testing. The methods for evaluation of videokymograms are, however, aimed for use in clinical practice. Performance of the methods described in this chapter was successfully tested in Matlab. After development and proper testing of methods for evaluation of other vibration parameters, e.g. vibrations of ventricular folds, developed by colleagues at our department, all these methods will be implemented in a standalone application. The application will automatically detect and display vibration features in the analyzed videokymogram and compute the corresponding parameter categories in the evaluation sheet. The clinician will thus be able to verify the computed parameters by quick visual inspection of the detected features. The software application could speed up the evaluation process and increase accuracy of its results.

4.6 Conclusion

This chapter has introduced new methods for detection of vocal fold vibration features in videokymograms and for automatic evaluation of corresponding parameters in the videokymographic evaluation sheet [59]. The methods were developed for purposes of computer-aided evaluation of vibratory patterns in videokymograms in laryngology and phoniatrics. The developed methods detect specular reflections, the shape of rima glottidis and directions of mucosal waves. The methods combine various image processing algorithms, e.g. optimal thresholding by normalized graph cuts, and incorporate or modify existing algorithms for processing of videokymograms, namely the algorithm for extraction of glottal features developed by Hauzar [17] and the algorithm for tracking of mucosal waves developed by Zita [68]. The methods also detect vibration features defined by the shape of rima glottidis. The proposed evaluation method computes from the detected features numerical values of 11 vibration parameters as well as corresponding categories in the evaluation sheet. Performance of the automatic evaluation was tested on a set of 50 videokymograms with a wide range of vibratory patterns. The comparison showed high consistency in results between the automatic evaluation and 18 visual evaluations; such consistency indicates that the developed methods could be used, in combination with visual verification, for computer-aided evaluation of videokymograms in clinical practice. The automatic methods could speed up the evaluation process and increase accuracy of its results. We plan to develop a software tool that would allow computer-aided evaluation of videokymograms according to the VKG evaluation sheet in clinical practice. The tool will include the methods described in this chapter as well as other methods currently developed in our department. The main contribution of this chapter was development of new methods for automatic evaluation of a relatively high number of vibration parameters. The chapter also provides a consistent survey on quantification of the automatically evaluated parameters and their correspondence with categories in the evaluation sheet. The proposed methods were presented at the *BioImage Informatics 2012* conference in Dresden, Germany [44].

Acknowledgment

The videokymographic images in the first and second sets were acquired from studies [24] and [19], respectively, with special acknowledgment to Jan G. Švec from the Department of Biophysics, Faculty of Science, Palacký University, Olomouc, Czech Republic. The algorithm for computation of optimal threshold by graph cuts was implemented and kindly provided by Adam Novozámský; the implementations of algorithms for detection of basic glottal features [17] and tracking of mucosal waves [68] were kindly provided by David Hauzar and Aleš Zita, respectively.

5. Conclusion

The Thesis has introduced new methods for automatic processing of image data in biology, physics and medicine. It focused on three specific image processing problems: reconstruction of light microscopy images showing the growth of microorganisms during intervals between observations, measurement of particles in atomic force microscopy images, and evaluation of vocal fold vibration parameters in videokymographic images. The developed methods could facilitate documentation, analysis and evaluation of the image data in practice. Their performance was tested on real data; the automatically computed results were comparable with ground truth or results of visual evaluations.

The growth reconstruction method addresses the problem of temporally sparse documentation of growing microorganisms. Phytopathogenic specimens which cannot be monitored continuously are typically observed in sessions with relatively long intervals. The developed method reconstructs light microscopy images of settled filamentous microorganisms corresponding to intervals between observation sessions. The method thus allows visualization of gradual development of the specimens during the intervals. In contrast to most model-based growth simulation methods, the proposed method reconstructs the missing images directly from the observed ones, which contributes to authentic appearance of the reconstructed images.

The particle measurement method addresses the problem of measurement of spatially separated ellipsoidal particles in atomic force microscopy (AFM) images. In contrast to other AFM image processing methods, the proposed method uses prior knowledge about ellipticity of the measured particles, which reduces sensitivity of the method to noise and increases accuracy of its results. The method also reconstructs the topography of particles distorted by convolution of the surface with the AFM scanning tip. The developed method could be used instead of time-consuming manual measurements.

The proposed videokymographic evaluation methods address the problem of measurement of vocal fold vibration parameters in videokymograms. In comparison to existing methods for detection of vibration features, the developed methods are relatively robust to low contrast and pathologies in the vibratory pattern. The methods evaluate a number of diagnostically important vibration parameters not only by numerical values but also by categories in the videokymographic evaluation sheet, which allows direct comparison of the automatically computed parameters with results of visual evaluations. The developed methods could be used, in combination with visual verification, for computer-aided evaluation in clinical practice; the methods will be integrated, along with other methods currently developed in our department, in a software tool for automatic evaluation of videokymograms according to the evaluation sheet.

Application of the proposed methods is not limited to these three problems, however. The methods could be used generally for any data with similar properties. For instance, the method developed for measurement of ellipsoidal particles in AFM images could be used also for measurement of elliptical cells or microorganisms in biology and medicine.

Development of the presented methods helped me gain experience in biomed-

ical image processing and analysis. I will continue my research in a post doc position at Flinders University in Adelaide, South Australia, with focus on segmentation in medical images. I am looking forward to using knowledge acquired during my work on the Thesis, as well as some of the developed methods, for solution of new problems.

Bibliography

- [1] H. Aboul-Ella and M. Nakajima. Image metamorphosis transformation of facial images based on elastic body splines. *Signal Processing*, 70(2):129–137, 1998.
- [2] M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester. Image inpainting. In *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, SIGGRAPH '00, pages 417–424, New York, NY, USA, 2000. ACM Press/Addison-Wesley Publishing Co.
- [3] M. Bertalmio, L. Vese, G. Sapiro, and S. Osher. Simultaneous structure and texture image inpainting. *Trans. Img. Proc.*, 12(8):882–889, August 2003.
- [4] G. Binnig, C. F. Quate, and C. Gerber. Atomic force microscope. *Phys. Rev. Lett.*, 56:930–933, 1986.
- [5] H. D. Buf and M. M. Bayer, editors. *Automatic Diatom Identification*, volume 51 of *Machine Perception and Artificial Intelligence*. World Scientific, 2002.
- [6] J. Chelkowski and A. Visconti, editors. *Alternaria: Biology, Plant Diseases and Metabolites*, volume 3 of *Topics in Secondary Metabolism*. Elsevier Science Publishers B.V., 1992.
- [7] F. Cloppet and A. Boucher. Segmentation of complex nucleus configurations in biological images. *Pattern Recogn. Lett.*, 31(8):755–761, 2010.
- [8] W. R. Crum, O. Camara, and D. L. G. Hill. Generalized overlap measures for evaluation and validation in medical image analysis. *IEEE Trans Med Imaging*, 25(11):1451–61, 2006.
- [9] D. D. Deliyski, P. P. Petrushev, H. S. Bonilha, T. T. Gerlach, B. Martin-Harris, and R. E. Hillman. Clinical implementation of laryngeal high-speed videoendoscopy: Challenges and evolution. *Folia Phoniatr Logop*, 60(1):33–44, 2008.
- [10] L. R. Dice. Measures of the amount of ecologic association between species. *Ecology*, 26:297–302, 1945.
- [11] L. Fekete, K. Kůsová, V. Petrák, and I. Kratochvílová. AFM topographies of densely packed nanoparticles: a quick way to determine the lateral size distribution by autocorrelation function analysis. *Journal of Nanoparticle Research*, 14:1–10, 2012.
- [12] J. Flusser, T. Suk, and B. Zitová. *Moments and Moment Invariants in Pattern Recognition*. John Wiley & Sons Ltd, Chichester, 2009.
- [13] K. Fujimura and M. Makarov. Foldover-free image warping. *Graphical Models and Image Processing*, 60(2):100–111, 1998.

- [14] Z. Guo and R. W. Hall. Parallel thinning with two-subiteration algorithms. *Communications of the ACM*, 32(3):359–373, 1989.
- [15] V. Hampala. Vizuální hodnocení videokymografických snímků u hlasových poruch [Visual evaluation of videokymographic features in voice disorders]. Master’s thesis, Palacký University, Olomouc, 2011.
- [16] R. M. Haralick and L. G. Shapiro. *Computer and Robot Vision*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1st edition, 1992.
- [17] D. Hauzar. Zpracování digitálních snímků videokymografických záznamů jako podpůrný nástroj pro diagnostiku hlasivek [Analysis of digital videokymographic images as a tool for vocal folds diagnostics]. Master’s thesis, Charles University, Prague, 2010.
- [18] P. S. Heckbert. Fundamentals of texture mapping and image warping. Technical report, 1989.
- [19] C. T. Herbst, Q. Qiu, H. K. Schutte, and J. G. Švec. Membranous and cartilaginous vocal fold adduction in singing. *J Acoust Soc Am*, 129(4):2253–2262, 2011.
- [20] M. Hirano. *Clinical examination of voice*. Springer-Verlag, Wien, Austria, 1981.
- [21] Image Processing Toolbox. <http://www.mathworks.com/products/image/>, MATLAB.
- [22] J. Jan. *Medical Image Processing, Reconstruction and Restoration - Concepts and Methods*. CRC Press, Taylor and Francis Group, Boca Raton, FL, USA, signal processing and comm. edition, 2006.
- [23] J. Jiang, Y. Zhang, M. Kelly, E. Bieging, and M. Hoffman. An automatic method to quantify mucosal waves via videokymography. *Laryngoscope*, 2008.
- [24] T. A. Van Kalkerren, H. K. Schutte, Q. Qiu, J. G. Švec, and H. F. Mahieu. First clinical experiences of second generation videokymography. In *Proceedings AQL 2006: Advances in Quantitative Laryngology, Voice and Speech Research, October 6-7, 2006, Groningen, the Netherlands*, pages 1–8. Groningen Voice Research Lab, University of Groningen, Groningen, the Netherlands, 2006.
- [25] T. Y. Kong and A. Rosenfeld, editors. *Topological Algorithms for Digital Image Processing*. Elsevier Science Inc., New York, NY, USA, 1996.
- [26] L. Lam, S.-W. Lee, and C. Y. Suen. Thinning methodologies—a comprehensive survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(9):869–885, 1992.
- [27] J. Lohscheller, U. Eysholdt, H. Toy, and M. Dollinger. Phonovibrography: mapping high-speed movies of vocal fold vibrations into 2-D diagrams for visualizing and analyzing the underlying laryngeal dynamics. *IEEE Trans Med Imaging*, 27(3):300–309, 2008.

- [28] J. Lohscheller, H. Toy, F. Rosanowski, U. Eysholdt, and M. Döllinger. Clinically evaluated procedure for the reconstruction of vocal fold vibrations from endoscopic digital high-speed videos. *Med Image Anal*, 2007.
- [29] C. Manfredi, L. Bocchi, S. Bianchi, N. Migali, and G. Cantarella. Objective vocal fold vibration assessment from videokymographic images. *Biomedical Signal Processing and Control*, 1(2):129 – 136, 2006. Voice Models and Analysis for Biomedical Applications.
- [30] D. D. Mehta, D. D. Deliyski, T. F. Quatieri, and R. E. Hillman. Automated measurement of vocal fold vibratory asymmetry from high-speed videendoscopy recordings. *J.Speech Lang.Hear.Res.*, 54(1):47–54, 2011.
- [31] V. L. Mironov. *Fundamentals of Scanning Probe Microscopy*. Russian Academy of Sciences, Institute for Physics of Microstructures, Nizhniy Novgorod, 2004.
- [32] J. Mizuguchi, T. Imoda, H. Takahashi, and H. Yamamaki. Polymorph of 1,4-diketo-3,6-bis-(4'-dipyridyl)-pyrrolo-[3,4-c]pyrrole and their hydrogen bond network: A material for H₂ gas sensor. *Dyes and Pigments*, 68(1):47–52, 2006.
- [33] NT-MDT. <http://www.ntmdt.com/>.
- [34] N. Otsu. A thresholding selection method from gray-level histograms. *IEEE Trans Systems Man Cybernetics*, 9:62–66, 1979.
- [35] P. Prusinkiewicz, M. Hammel, J. Hanan, and R. Měch. L-systems: from the theory to visual models of plants. In M. T. Michalewicz, editor, *Plants to Ecosystems: Advances in Computational Life Sciences I*, pages 1–27. CSIRO Publishing, 1996.
- [36] Q. Qiu, H. K. Schutte, L. Gu, and Q. Yu. An automatic method to quantify the vibration properties of human vocal folds via videokymography. *Folia Phoniatr Logop*, 55(3):128–136, 2003.
- [37] J. B. T. M. Roerdink and A. Meijster. The watershed transform: Definitions, algorithms and parallelization strategies. *Fundamenta Informaticae*, 41:187–228, 2000.
- [38] K. Rohr. *Landmark-Based Image Analysis: Using Geometric and Intensity Models*, volume 21 of *Computational imaging and vision*. Dordrecht: Kluwer Academic Publishers, 2001.
- [39] O. Salyk, J. Vyňuchal, I. Kratochvílova, T. Todorciuc, J. Pavluch, and P. Toman. Study of phenylpyridyldiketopyrrolopyrrole interaction with hydrogen in gas and in acids. *Dyes and Pigments*, 207(10):2327–2333, 2010.
- [40] J. Schlecht, K. Barnard, E. Spriggs, and B. Pryor. Inferring grammar-based structure models from 3D microscopy data. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 1–8, 2007.

- [41] J. Sedlář. Digital processing of microscopy images in biology. Master's thesis, Charles University, Prague, 2004.
- [42] J. Sedlář, J. Flusser, and M. Sedlářová. Tracking the growth of filamentous fungi by means of the morphological skeleton. In I. Provazník J. Jan, J. Kozumplík, editor, *Analysis of Biomedical Signals and Images. BIOSIGNAL 2006*, pages 308–311, Brno, Czech Republic, 2006.
- [43] J. Sedlář, J. Flusser, and M. Sedlářová. Photorealistic modeling of the growth of filamentous specimens. *EURASIP Journal on Advances in Signal Processing*, 2008:1–9, 2008. Article ID 520972.
- [44] J. Sedlář, A. Novozámský, J. G. Švec, B. Zitová, and J. Flusser. Measurement of vocal fold features in videokymography images. In *Proc. BioImage Informatics 2012*, page 53, Max Plank Inst. MCG, Dresden, Sep 16-19 2012.
- [45] J. Sedlář, B. Zitová, J. Kopeček, T. Todorciuc, and I. Kratochvílová. Detection of elliptical particles in atomic force microscopy images. In *ICASSP 2011: IEEE International Conference on Acoustics, Speech, and Signal Processing*, pages 1233–1236. IEEE, 2011.
- [46] J. Serra. *Image Analysis and Mathematical Morphology*. Academic Press, London, 1988.
- [47] J. Shi and J. Malik. Normalized cuts and image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(8):888–905, August 2000.
- [48] L. Streit, P. Federl, and M. C. Sousa. Modelling plant variation through growth. In *Computer Graphics Forum*, volume 24, pages 497–506, 2005.
- [49] Frost & Sullivan. 2004 Healthcare storage report, 2004.
- [50] W. Tao, H. Jin, Y. Zhang, L. Liu, and D. Wang. Image thresholding using graph cuts. *Trans. Sys. Man Cyber. Part A*, 38(5):1181–1195, September 2008.
- [51] M. Toman. Analýza STM snímků pořízených elektronovým mikroskopem [Analysis of STM images acquired by electron microscope]. Master's thesis, Charles University, Prague, 2001.
- [52] D. Tschumperlé and R. Deriche. Vector-valued image regularization with PDEs: a common framework for different applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(4):506–517, 2005.
- [53] J. G. Švec. *On vibration properties of human vocal folds: voice registers, bifurcations, resonance characteristics, development and application of videokymography*. Doctoral dissertation, University of Groningen, Groningen, the Netherlands, 2000.
- [54] J. G. Švec. *Vocal Fold Protocol Quantification*, 2012.
- [55] J. G. Švec and F. Šram. Videokymographic examination of voice. In E. P. M. Ma and E. M. L. Yiu, editors, *Handbook of Voice Assessments*, pages 129–146. San Diego, CA: Plural Publishing, 3rd edition, 2011.

- [56] J. G. Švec, F. Šram, and H. K. Schutte. Videokymography in voice disorders: what to look for? *Ann Otol Rhinol Laryngol*, 116(3):172–80, 2007.
- [57] J. G. Švec, F. Šram, and H. K. Schutte. Videokymography. In M. P. Fried and A. Ferlito, editors, *The Larynx*. San Diego, CA: Plural Publishing, 3rd edition, 2009.
- [58] J. G. Švec and H. K. Schutte. Videokymography: high-speed line scanning of vocal fold vibration. *J Voice*, 10(2):201–205, 1996.
- [59] J. G. Švec, H. Švecová, C. Herbst, and H. K. Schutte. Evaluation protocol for videokymographic images. MsAccess software application, 2007. Groningen Voice Research Lab, University of Groningen, Groningen, the Netherlands.
- [60] J. S. Villarrubia. Algorithms for scanned probe microscope image simulation, surface reconstruction, and tip estimation. *J Res Natl Inst Stand Technol*, 102:102–425, 1997.
- [61] L. Vincent and P. Soille. Watersheds in digital spaces: An efficient algorithm based on immersion simulations. *IEEE Transactions on Pattern Recognition and Machine Intelligence*, 13(6):583–598, 1991.
- [62] F. Šroubek and J. Flusser. Fusion of blurred images. In R. S. Blum and Z. Liu, editors, *Multi-Sensor Image Fusion and Its Applications*, volume 26 of *Signal Processing and Communications Series*, pages 405–430, San Francisco, 2005. CRC Press.
- [63] J. Vyňuchal, S. Luňák Jr., A. Hatlapatková, R. Hrdina, A. Lyčka, L. Havel, K. Vyňuchalová, and R. Jirásko. The synthesis, absorption, fluorescence and photoisomerisation of 2-aryl-4-arylmethylidene-pyrroline-5-ones. *Dyes and Pigments*, 77(2):266–276, 2008.
- [64] G. Wolberg. *Digital Image Warping*. IEEE Computer Society Press, 1990.
- [65] C. Xu and J. L. Prince. Generalized gradient vector flow external forces for active contours. *Signal Processing*, 71(2):131–139, 1998.
- [66] M. Zemánková and A. Lebeda. Fusarium species, their taxonomy, variability and significance in plant pathology. *Plant Protection Science*, 37:25–42, 2001.
- [67] Y. Zhang, E. Bieging, H. Tsui, and J. J. Jiang. Efficient and effective extraction of vocal fold vibratory patterns from high-speed digital imaging. *J Voice*, 24(1), 2010.
- [68] A. Zita. Zpracování videokymografických záznamů [Analysis of videokymographic images]. Bachelor’s thesis, Charles University, Prague, 2011.
- [69] B. Zitová and J. Flusser. Image registration methods: a survey. *Image and Vision Computing*, 21(11):977–1000, 2003.

List of Abbreviations

AFM	...	atomic force microscopy	(see Section 3.1)
CPs	...	control points	(see Section 2.3)
PPDP	...	phenylpyridyldiketopyrrolopyrrole	(see Section 3.1)
RT	...	room temperature	(see Section 2.1)
VKG	...	videokymography	(see Section 4.1)