

BREAST CANCER HISTOPATHOLOGY IMAGE ANALYSIS

MITKO VETA

BREAST CANCER HISTOPATHOLOGY IMAGE ANALYSIS
PhD thesis, Utrecht University, The Netherlands

This thesis was typeset by the author using L^AT_EX 2_ε
Cover design: Mitko Veta and Darko Krstevski
Printed by: Proefschriftmaken.nl || Uitgeverij BOXPress
ISBN: 978-90-393-6209-9

Copyright © 2014 Mitko Veta

All rights reserved. No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any other information storage and retrieval system, without the written permission from the copyright owner.

BREAST CANCER HISTOPATHOLOGY IMAGE ANALYSIS

BEELDANALYSE VAN BORSTKANKER HISTOPATHOLOGIE
(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht op gezag van de rector magnificus, prof.dr. G.J. van der Zwaan, ingevolge het besluit van het college voor promoties in het openbaar te verdedigen op dinsdag 23 september 2014 des middags te 14.30 uur

door

MITKO VETA

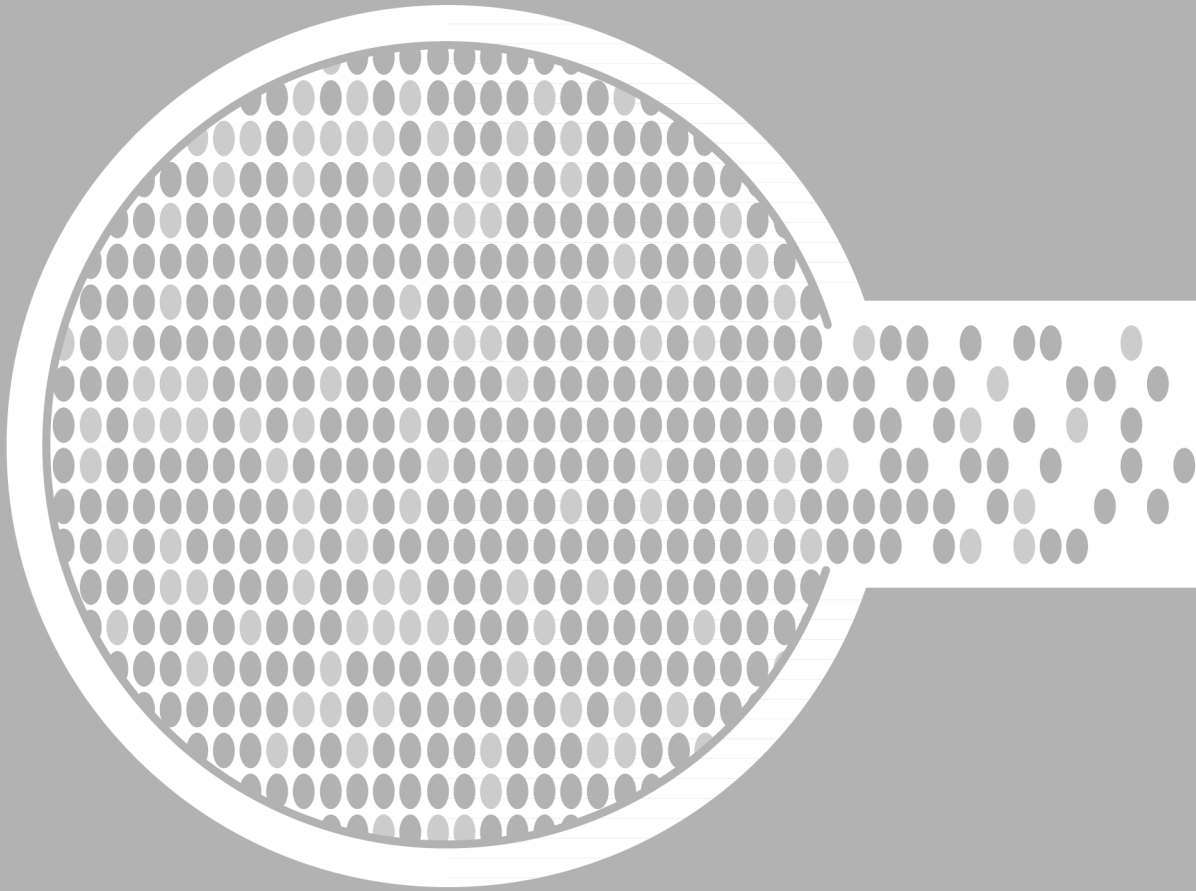
geboren op 20 augustus 1984 te Strumica, Macedonië

Promotoren: prof. dr. ir. M. A. Viergever
prof. dr. P. J. van Diest
prof. dr. J. P. W. Pluim

Contents

1	General introduction	3
2	Breast cancer histopathology image analysis: a review	7
2.1	Introduction	9
2.2	Tissue preparation and imaging	9
2.3	Staining analysis	11
2.4	Quantification of immunohistochemistry	12
2.5	Object detection and segmentation	14
2.6	Discussion and conclusions	23
3	Automatic nuclei segmentation	27
3.1	Introduction	29
3.2	Materials and methods	31
3.3	Results	40
3.4	Discussion and conclusions	42
4	Prognostic value of nuclear morphometric features	49
4.1	Introduction	51
4.2	Materials and methods	52
4.3	Results	55
4.4	Discussion and conclusions	56
5	Assessment of algorithms for mitosis detection	61
5.1	Introduction	63
5.2	Materials	65
5.3	Methods	71
5.4	Evaluation	79
5.5	Results	81
5.6	Discussion	83
5.7	Conclusions	88
6	Automatic detection of mitotic figures	91

6.1	Introduction	93
6.2	Materials and methods	93
6.3	Results	100
6.4	Discussion and conclusions	102
7	General summary	105
	Bibliography	109
	Nederlandse samenvatting	119
	Acknowledgements	123
	List of publications	127
	Curriculum vitae	131



CHAPTER

1

General introduction

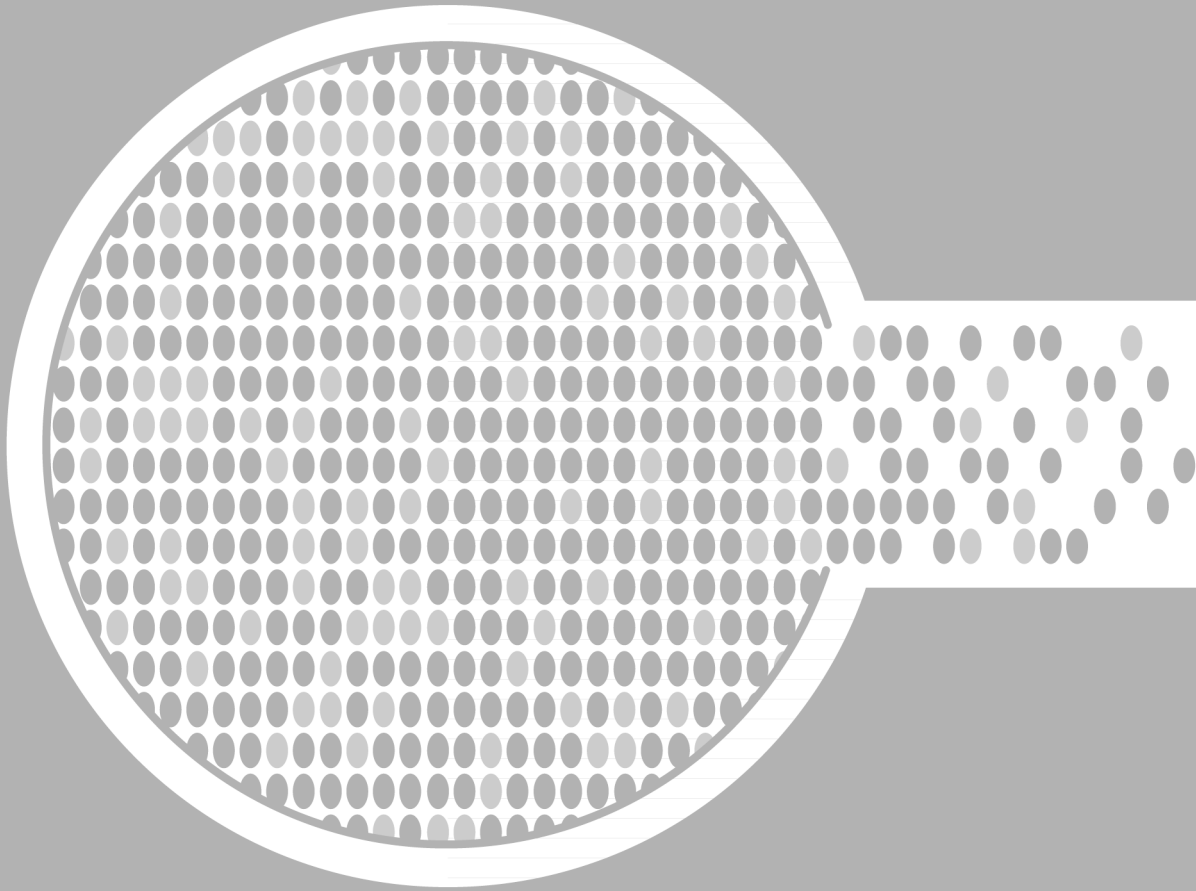
The process of digitization of pathology imaging has been ongoing for several decades. As early as in 1986, telepathology was made possible after the introduction of video cameras mounted on microscopes, enabling live images to be shared with experts at remote locations (Al-Janabi et al. 2011; Pantanowitz et al. 2011). This enabled live teleconsultation and remote diagnosis of frozen sections (Baak et al. 2000), although at a relatively low resolution. In the last two decades, high quality digital cameras became available, allowing capturing of still digital images from a microscope at a high resolution. This was followed by the introduction of whole-slide imaging (WSI) scanners that are gradually making their way into pathology labs as the digital age alternative to the conventional microscope. State-of-the-art WSI scanners are table-top devices that take glass slides as input and produce whole-slide digital images (sometimes referred to as digital or virtual slides) as output, automating all intermediate steps such as localization of the tissue and focus plane selection. The goal of whole-slide imaging, coupled with whole-slide image viewers, is to simulate slide viewing by a conventional microscope on a computer screen.

Pathology labs are currently undergoing a transformation towards a fully digital workflow (Stathonikos et al. 2013). In addition to the digital management of tissue samples, pathology orders and reports, this includes the digitization of histopathology slides and use of computer monitors for viewing them, which aims to replace the optical microscope as the primary tool used by pathologists. This transformation has only recently been enabled by the introduction of cost and time efficient WSI scanners. This process of adoption of WSI is somewhat analogous to the digitization of radiological imaging. However, a full analogy is difficult to establish because in pathology the primary object of analysis is the tissue rather than the image (Hipp et al. 2011). Adoption of a fully digital clinical workflow in radiology benefitted from the ability to acquire radiological images digitally. By contrast, the barrier to entry is higher for histology, which is a distinctly analog modality that must always undergo an analog-to-digital conversion before computerized analysis. Nevertheless, there are numerous advantages to WSI that provide incentives for acceptance as a primary diagnostic modality. Many advantages stem from the fact that compared with glass slides, whole-slide images are more portable entities that can be easily retrieved from a digital archive. They are accessible through a computer network from remote locations and allow simultaneous access by multiple people. Whole-slide image viewers can offer an enriched user experience, such as showing an overview image along with the high-power view to enable better orientation and navigation, or simultaneously displaying of two or more slides side by side. One of the main benefits of digital slides compared with conventional glass slides is that they enable

seamless integration of quantitative automatic image analysis methods into the workflow. These methods have the potential to tackle the problems that stem from the subjective interpretation by pathologists and, at the same time, reduce their workload (Meijer et al. 1997).

A relatively large percentage of the samples that are analyzed in pathology labs are from breast cancer patients, since this disease is the most prevalent form of cancer among women (Bray et al. 2004). Analysis methods that are routinely performed by pathologists, such as determination of the histological grade and the hormone receptor status by immunohistochemistry, can be tedious and are hampered by observer variability (Meyer et al. 2005; Perez et al. 2006). The histological tumor grade is commonly determined according to the modified Bloom-Richardson system, which consists of semi-quantitative assessment of nuclear atypia, tubule formation and mitotic activity (Elston et al. 1991; Rakha et al. 2010) in hematoxylin and eosin (H&E) stained sections. The analysis of immunohistochemically stained slides mainly involves the estimation of the number of cells that are positive for a particular antigen and the degree of positivity (staining intensity) (Hammond et al. 2010; Wolff et al. 2007).

The focus of this thesis is on automatic image analysis of H&E stained breast cancer histopathology images. Chapter 2 gives a detailed review of the literature on the topic of analysis of breast cancer histopathology images. The tissue preparation and imaging processes are also covered and particular attention is given to techniques for detection and segmentation of various objects, such as nuclei, tubules and mitotic figures, as well as computer-aided diagnosis and prognosis methods. In Chapter 3, the development and evaluation of a method for automatic segmentation of nuclei in H&E stained breast cancer histopathology images is presented. The proposed nuclei segmentation method is then used in Chapter 4 to extract and evaluate the prognostic value of nuclear morphometric features in a cohort of 101 male breast cancer patients. Chapter 5 presents the results from a challenge workshop on the difficult problem of detection of mitotic figures in H&E stained breast cancer histopathology images. The challenge was based on a data set consisting of 12 training and 11 testing subjects, with more than one thousand annotated mitotic figures by multiple observers. The development and evaluation of a mitotic figures detection method using the same data set is described in Chapter 6. Chapter 7 provides a general discussion of the results of this thesis and wraps up the results once more.



**Breast cancer histopathology
image analysis: a review**

Summary

In this chapter, an overview of methods that have been proposed for analysis of breast cancer histopathology images is presented. This research area has become particularly relevant with the advent of whole slide imaging (WSI) scanners, which can perform cost-effective and high-throughput histopathology slide digitization, and which aim at replacing the optical microscope as the primary tool used by pathologist.

Breast cancer is the most prevalent form of cancers among women, and image analysis methods that target this disease have a huge potential to reduce the workload in a typical pathology lab and to improve the quality of the interpretation. The chapter starts with an overview of the tissue preparation, staining and slide digitization processes followed by a discussion of the different image processing techniques and applications, ranging from analysis of tissue staining to computer-aided diagnosis and prognosis of breast cancer patients.

2.1 Introduction

In this chapter, we give an overview of image analysis methods that have been proposed for breast cancer histopathology images. We focus on automatic image analysis of histopathology tissue preparations imaged by bright-field microscopy, since this covers the bulk of the work that is performed by pathologists for this disease. However, some techniques developed for other tissue types or microscopy modalities that are relevant for the scope of this review are mentioned throughout the text when appropriate. For a broader overview of digital pathology and the use of automatic methods for analysis of histopathology slides, we refer the reader to the recent reviews in Di Cataldo et al. (2012), Fuchs et al. (2011), Ghaznavi et al. (2013), Gurcan et al. (2009) and Al-Janabi et al. (2011).

2.2 Tissue preparation and imaging

Before we proceed to discuss the different image analysis algorithms and applications, we give an overview of the tissue preparation and staining processes and digitization of histological slides. In the typical hospital workflow, breast tumor excisions or biopsies are performed in the operating room after which the material is sent for analysis to the pathology lab. The first step of the tissue preparation process is formalin fixation and embedding in paraffin. From the paraffin blocks, sections with a thickness of 3-5 μm are cut using a microtome (a high precision cutting instrument) and mounted on glass slides. The structures of interest in the tissue, in most instances the nuclei and cytoplasm, are not readily visible on the mounted sections. They therefore need to be dyed with stains that highlight them. The standard staining protocol uses hematoxylin and eosin (H&E, Figure 2.1(a)). In spite of the fact that this staining protocol has been in use for around a century, the diagnostic and prognostic procedure for all patients still almost always starts by staining the sections with H&E. Hematoxylin binds to DNA and thereby dyes the nuclei blue/purple, and eosin binds to proteins and dyes other structures (cytoplasm, stroma, etc.) pink.

Immunohistochemistry (IHC) is a more advanced staining technique, which makes use of antibodies to highlight specific antigens in the tissue (Figure 2.1(b)- 2.1(c)). In breast cancer, IHC is commonly used to highlight the presence of estrogen (ER), progesterone (PR) and human epidermal growth factor 2 (HER2) receptors, as well as to assess the proliferation of the tumor, for example, by highlighting the Ki-67 protein, which is associated with cell proliferation (Fitzgibbons et al. 2000; Hammond et al. 2010; Wolff et al. 2007). When performing IHC, the tissue is usually counterstained with hematoxylin to identify the nuclei and to visualize the tissue architecture (in

case of nuclear antigens — visualize the nuclei in which the target antigen is absent). When quantification is of primary interest, such as with the determination of the ER, PR and HER2 receptor status for breast cancer patients by IHC, the staining protocols need to be standardized and quality-controlled in order to obtain reproducible results, which are comparable across different patients (Taylor et al. 2006; Walker 2006).

Very often in breast cancer research, many different markers highlighted by IHC from hundreds or thousands of patients need to be considered. In order to achieve high-throughput analysis, researchers resort to using tissue microarrays (TMAs) (Kononen et al. 1998; Skacel et al. 2002). TMAs are constructed by punching small core biopsies (usually with a core diameter of 0.6 mm) from selected regions of the paraffin blocks containing the tissue to be analyzed, and transferring them to a recipient paraffin block in a regular pattern. The recipient paraffin block, which now contains tissue from many different subjects, is then cut and stained in a standardized manner. Because with TMAs tissues from different patients are stained under the same conditions, the resulting staining variability is significantly lower than with routinely prepared histopathology slides, which makes them more suitable for image analysis.

Currently, the typical pathology lab workflow is concluded by staining and coverslipping of the glass slides, after which they are sent to the pathologist for analysis. As digital pathology becomes more commonplace, slide digitization is added as an additional stage to this workflow (Stathonikos et al. 2013). The early slide digitization systems were digital cameras mounted on standard microscopes, which could capture still images. Present day whole slide imaging scanners, which enable high throughput slide digitization at relatively low cost, handle the entire scanning procedure automatically. This includes loading of the slides on the scanning platform, detection of the relevant tissue regions and focus point selection, image acquisition, compression, storing and registration on a laboratory information system. Most of the WSI scanners that are currently in use perform slide scanning at $\times 20$ or $\times 40$ magnification with a spatial resolution in the order of $0.5 \mu\text{m}/\text{pixel}$ and $0.25 \mu\text{m}/\text{pixel}$, respectively. Because of the large size, the captured RGB image is compressed most commonly with JPEG or JPEG 2000. For faster navigation, the images can be stored in a pyramid structure with increasing magnification at each level of the pyramid. This also facilitates multi-scale image analysis.

The tissue preparation, staining and slide digitization processes can have a significant impact on the tissue/image appearance, and insight into them may lead to a better design of image analysis algorithms. The impact can be manifested in several ways. For example, improper fixation can lead to changes in tissue morphology and thus induce incorrect tissue morphometry

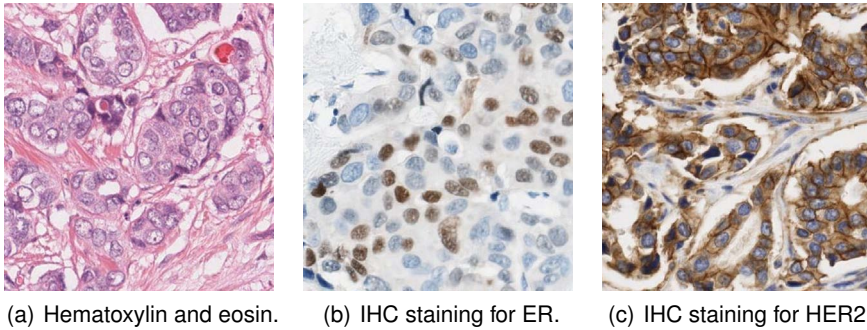


Figure 2.1: Example histological stains. For the IHC stained sections, the antibody is visualized with the DAB chromogen (3, 3'-Diaminobenzidine) and the tissue is counterstained with hematoxylin.

by automatic image analysis. If sections are not properly stained (i.e., over- or under-stained) and mounted, this may result in out-of-focus regions and/or missing parts. Even small variations of the staining conditions can lead to variations in tissue appearance and cause problems for automatic image analysis algorithms. This issue of staining/appearance variability will be considered in more detail in the following sections.

2.3 Staining analysis

When performing image analysis of histopathology images, it is of interest to separate the histological stains that dye different tissue components. For instance, if the application is nuclei detection in H&E stained sections, it can be beneficial to obtain a hematoxylin only image since the eosin-only stained components can contribute towards false positives. One approach is to perform clustering or supervised classification of the RGB pixel values in order to obtain binary or probability maps for the different stains. These methods require labeled data or identification of the cluster that corresponds to the stain of interest. Another approach for staining separations is based on the fact that the image formation process in brightfield microscopy can be modeled according to the Lambert-Beer law for light absorption. According to this law, the optical density (the logarithm of the intensity) is proportional to the concentration of stain in the tissue. Given that with RGB image sensors there are three detection channels, linear decomposition can be performed to determine the concentration of up to three stains for each pixel location (Ruifrok et al. 2001). Once the stain concentrations are determined, single-stain images can be derived by an inverse approach. These techniques require definition of characteristic absorption spectra for the stains that need to be separated.

However, there are blind techniques that do not have this limitation (Gavrilovic et al. 2013).

It should be noted here that the commonly used DAB chromogen (3, 3'-Diaminobenzidine) is not a true light absorber but it exhibits light scattering behavior. Thus, the behavior of DAB cannot be perfectly modelled by the Lambert-Beer law. However, in practice, good unmixing results can still be achieved (Taylor et al. 2006).

One of the major difficulties in breast cancer histopathology image analysis, particularly of H&E stained sections, is appearance variability. In part, this can be explained by the heterogeneity of the disease, but a large portion is a result of the tissue preparation and staining processes. The color appearance can significantly vary between different labs owing to differences in fixation and in staining protocols and reagents. Color appearance may also vary between samples produced in the same lab as a result of preanalytic features as fixation delays and inconsistencies in the staining conditions. Finally, variability in appearance can be due to slide digitization conditions, including notably differences in optics, light detectors or light sources used in the scanners.

Automatic image analysis methods can be significantly hampered by the variability of the tissue appearance. In addition, methods that are developed and tested on data from a single center often must be reevaluated and adjusted when used with slides from external labs. It is desirable that histopathology image analysis methods are designed in a way that is robust to appearance variability, such as in Monaco et al. (2012). Alternatively, the image appearance can be standardized prior to further processing. One simple approach is to determine the concentrations of the individual stains for each pixel with some of the staining separation techniques mentioned before, normalize the staining concentrations and then digitally mix the stains with common characteristic absorption coefficients to obtain a standardized image (Bilgin et al. 2012; Macenko et al. 2009). In (Basavanhally et al. 2013a), a method for color standardization was proposed based on unsupervised segmentation into tissue components. The tissue was divided into four components: nuclei, stroma, epithelium and background and the RGB histograms for each component were aligned to a template image.

2.4 Quantification of immunohistochemistry

In H&E stained slides, the features of the nuclei that are of interest to pathologists are relatively complex - their size, shape and texture, their spatial arrangement and organization into tubules, interaction with the stroma, etc. In contrast with this, most of the information that is of interest in IHC stained

sections is contained in the color and the intensity of the staining, which makes IHC stained samples more open to design and implementation of image processing algorithms. For example, a useful and readily obtained characteristic of IHC digital slides is the determination of the percentage of pixels that are positively stained for a particular antigen.

In breast cancer patients, the ER, PR and HER2 receptor statuses can have a major influence on the planning of adjuvant systemic treatment. Currently, the standard method of scoring IHC stained slides is by visual examination under a microscope. This procedure is prone to variability among pathologists even when strict guidelines are followed. Recent recommendations issued by the American Society of Clinical Oncology and the College of American Pathologists for testing of the ER, PR and HER2 receptor status include encouragement of the use of quantitative image analysis techniques with the goal of improving the consistency of the interpretation (Hammond et al. 2010; Wolff et al. 2007).

The ER and PR receptor statuses are customarily determined by counting the percentage of positively stained nuclei. If this percentage is above a pre-defined threshold (10% in Europe and 1% in the USA) the tissue is defined positive. The automatic quantification of the ER and PR status thus usually involves the use of an automated nuclei detection or segmentation algorithm (Rexhepaj et al. 2008) (this topic will be covered in detail in the following section). As an alternative to this, the percentage of positively stained nuclear area can be determined (Tuominen et al. 2010). In Amaral et al. (2013), a method which computes features that reflect the area of positively stained nuclei and the nuclear intensity was proposed. These features were then mapped to an ordinal scale that is used by pathologists.

In comparison with the ER and PR receptors, which are expressed in the cell nuclei, the HER2 receptor is expressed on the cell membranes. The tumors are scored positive when more than 30% of the cell membranes show complete, uniform and intensive staining. Cases with complete but non-uniform or weak staining in more than 10% of the cells are equivocal, and cases with no staining or incomplete staining are defined to be negative. Equivocal cases are further evaluated by other methods (Moelans et al. 2009).

The largest challenge in HER2 staining quantification lies in correct membrane segmentation, which can be particularly challenging in negative cases and cases with incomplete and faint staining. A method for automated assessment of HER2 immunohistochemistry is presented in Masmoudi et al. (2009). In the first stage of this approach, image pixels are classified as belonging to epithelial nuclei or cell membranes. The nuclear regions are further segmented into individual nuclei by watershed segmentation and the cell membranes are determined by adaptive ellipse fitting. Slides are then classified

into one of the three scoring groups based on features describing the membrane staining intensity and completeness. In Ficarra et al. (2011), following nuclei segmentation, approximate membrane contours were determined by Voronoi tessellation. The approximate contours were then refined based on the membrane staining intensity. Hall et al. (2008) proposed a method where features for HER2 quantification were extracted based on positive controls, thus eliminating the effect of the variability in the staining between different slides.

Most of the commercially available image analysis tools include algorithms for positive nuclei counting and membrane staining quantification. A review of imaging solutions for quantitative immunohistochemistry can be found in Rojo et al. (2009). Two publicly available web applications for ER/PR and HER2 quantifications are described in Tuominen et al. (2012) and Tuominen et al. (2010). In various recent studies, automatic scoring has shown high agreement with expert scoring and other methods (such as FISH for HER2 scoring) (Bolton et al. 2010; Brüggmann et al. 2012; Gavrielides et al. 2011; Laurinaviciene et al. 2011; Lloyd et al. 2010; Minot et al. 2012; Mohammed et al. 2012; Nassar et al. 2011; Turashvili et al. 2009).

2.5 Object detection and segmentation

2.5.1 Tissue and tissue components segmentation

The typical histopathology slide contains a tissue area of approximately 15×15 mm. At the resolutions at which digital slides are captured, this will result in images with a size of up to several gigapixels. Because processing of these very large images might result in computational problems, it is common practice to identify the regions of the slides that are of interest prior to performing more detailed image analysis.

Generally, large portions of the slides are empty, i.e. they do not contain tissue. Most WSI scanners have the ability to identify empty tiles in the slide during the scanning process and avoid scanning them, which results in reduction of scanning time and lowers the required amount of storage. One such approach for supervised tissue localization was proposed in Alomari et al. (2009).

When analysis of TMAs is performed, the individual TMA cores need to be identified and segmented and their coordinates on the TMA grid need to be assigned so they can be matched to the donor paraffin block (Foran et al. 2011; Lahrman et al. 2010; Wang et al. 2011b).

In general, large areas of the tissue are not relevant for the problem at hand. For example, for computer-aided diagnosis of breast cancer (classification into the classes benign or malignant) only the epithelial regions of

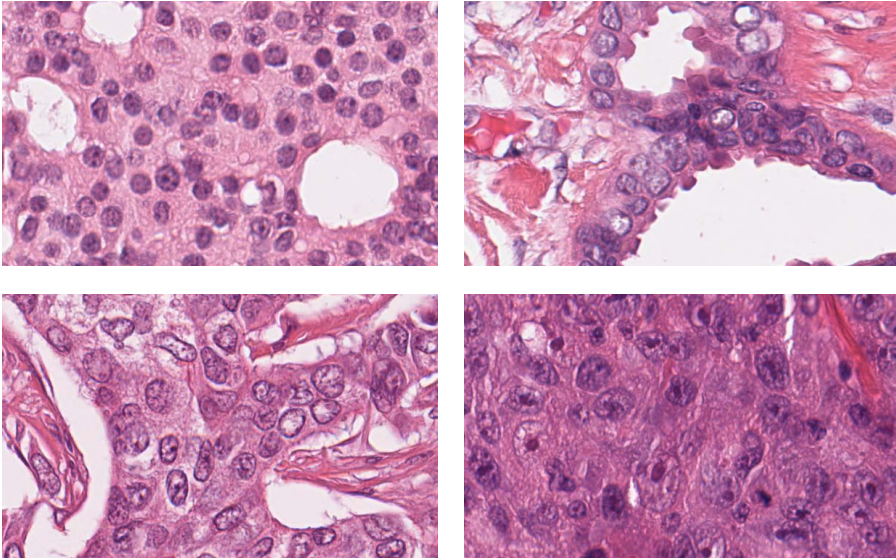


Figure 2.2: Different nuclei appearances in breast cancer histopathology images (from top-left to bottom-right): small and round with relatively uniform chromatin, organized into tubules, marginalized chromatin, and prominent nucleoli.

the tissue are relevant. When quantification of IHC or histological grading needs to be performed, only the tumor tissue is of interest and non-tumor regions need to be excluded from the analysis. Furthermore, epithelial and stromal regions of the tumor have different significance for diagnosis and prognosis. Segmentation of the tumor into these components is often used as a pre-processing step when performing automatic morphometry and histological grading, or for guidance when constructing TMAs (Karaçali et al. 2007). In the majority of the published studies on tissue segmentation, this kind of compartmentalization is achieved by supervised pixel-wise classification of small rectangular image regions based on color and texture features (Bahlmann et al. 2012; Linder et al. 2012; McKenna et al. 2013; Peikari et al. 2013; Wang et al. 2011a), although unsupervised methods have been proposed (Khan et al. 2013).

2.5.2 Nuclei detection and segmentation

The segmentation of nuclei in breast cancer histopathology images can be considered a basic functional block in many different applications. Quantification of IHC nuclear staining has already been mentioned in the previous section. Other applications include extraction of prognostically relevant morphometric

features (size, shape, chromatin texture), automatic nuclear pleomorphism grading as part of a computer-aided prognosis system, detection of lymphocytic infiltration, and detection of malignancy and tubule formation in a bottom-up manner.

Nuclei segmentation remains a very challenging problem, particularly for routinely stained H&E sections, despite the numerous attempts to solve it. Many of the challenges arise from the variability of the tissue appearance, which is in part due to imperfections in the staining process. Furthermore, there is an inherent diversity of the appearance of epithelial cancerous nuclei, which may vary from almost normal-like round to highly irregularly shaped and enlarged nuclei with coarse and marginalized chromatin and prominent nucleoli (small round structures inside the nuclei). This is illustrated in Figure 2.2. Different nucleus types, such as elongated fibroblasts and lymphocyte nuclei, often appear together with epithelial nuclei, which can decrease the specificity when only epithelial nuclei need to be detected or segmented. Additionally, nuclei may be overlapping, clustered or tightly clumped, which makes them difficult to separate. Lastly, hematoxylin-stained “junk” particles, which tend to appear in high grade tumors, can hamper the nuclei segmentation.

A large variety of approaches for segmentation of nuclei in breast cancer histopathology images have been proposed. They vary not only in the core segmentation methods, but also in the pre- and postprocessing steps that aim to improve segmentation performance.

The most difficult aspect of nuclei segmentation in breast cancer histopathology images is the detection of individual nuclei, especially when they are clustered closely together and overlap. Some proposed methods try and identify the individual nuclei prior to performing the segmentation procedure (marker extraction), whereas others first segment clumps of nuclei from the rest of the tissue and then proceed to separate those into individual nuclei. Methods such as the Hough transform (Cosatto et al. 2008) and voting along the direction of the image gradient to infer the center of the object (Qi et al. 2012; Veta et al. 2013b) have been used for both identification of nucleus markers and separation of segmented clumps. Another popular approach to separate clumps of nuclei is to split them along points of high concavity (Fatakdawala et al. 2010; Wienert et al. 2012). In Jung et al. (2010), separation of clustered nuclei is achieved by unsupervised Bayesian classification. Supervised methods for nuclei detection have also been proposed, with good performance (Vink et al. 2013).

With regards to the core segmentation technique, active contours are among the more popular methods (Ali et al. 2012; Cosatto et al. 2008; Qi et al. 2012). Their objective is to find a minimum energy fit of moving con-

tours to the image, the energy being defined in such a way that the contours are attracted to the boundaries of the objects of interest. The contours need to be seeded close to the target nuclei locations, which is why the approach is usually paired with a nuclei detection method. In Qi et al. (2012) the level set active contours segmentation includes a repulsion term to prevent the contours of adjacent cells from overlapping. The active contour model for nuclei and lymphocyte nuclei segmentation proposed in Ali et al. (2012) incorporates boundary, region and shape prior terms, and performs simultaneous segmentation of multiple objects in the image.

Several methods have been proposed that work by initially defining a large number of candidate regions and then selecting the ones that are likely to represent a correctly segmented object. In Arteta et al. (2012), candidate regions are identified by the maximally stable extremal regions (MSER) detector. This detector produces a large number of potentially overlapping regions. Each candidate region is evaluated with a statistical model, and dynamic programming is used to select a set of non-overlapping regions that best fit the model. The authors of Wienert et al. (2012) proposed generating a very large number of candidate objects by identifying all possible closed contours within the image by contour tracing. A set of non-overlapping contours is obtained by defining the objects that are most “fit” based on the contour gradient strength. In Veta et al. (2013b), candidate regions are initially identified by a marker-controlled watershed approach at multiple scales and using multiple marker types, which yields a large number of overlapping contours. Regions unlikely to represent valid nuclei are removed based on size, shape, boundary and chromatin distribution features. Local concurrences are resolved by greedy selection of the contours most likely to represent nuclei, using the solidity of the object as a fitness value.

Other recently proposed and promising methods for nuclei segmentation in breast cancer histopathology are based on dictionaries of discriminative image patches (Karsnas et al. 2012) and marked point processes (Avenel et al. 2013; Kulikova et al. 2012).

Detection and segmentation of lymphocyte nuclei and detection of lymphocytic infiltrations can be considered special cases of nuclei detection/segmentation. In Basavanahally et al. (2010), region growing with high sensitivity and low specificity is used to initially segment lymphocyte nuclei and other objects. Then, maximum a posteriori (MAP) estimation that incorporates size, luminance and spatial proximity information is used to improve the specificity of the detector. Finally, the results from the lymphocyte nuclei detection are input to a classifier that discriminates between the lymphocyte infiltration phenomenon and the baseline level of lymphocytes. In Fatak-dawala et al. (2010), output from a Gaussian mixture clustering algorithm is

used to initialize geodesic active contour segmentation. The overlapping objects are resolved by splitting them along high concavity points. Lymphocyte nuclei are distinguished from other objects by texture based clustering.

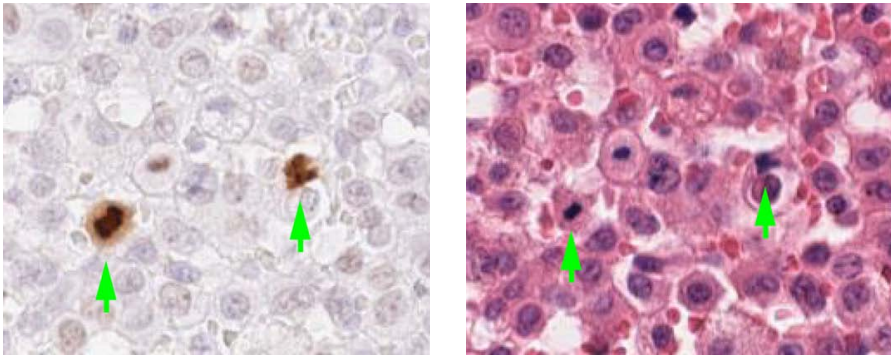
The choice of the segmentation method is closely related to the intended application and the available time and computational resources. For example, active contours may yield superior nuclear segmentation, but at the cost of increased computational complexity, which limits their use for high throughput applications. When such high throughput processing is needed, a simple thresholding of the hematoxylin channel followed by morphological operations and splitting of clusters along high concavity points might be used, but at the cost of less accurate segmentation results.

So far we have only listed methods that specifically aim at dealing with nuclei segmentation in breast cancer histopathology sections imaged with brightfield microscopy as an application. However, methods developed for other tissue types (e.g., prostate) or microscopy modalities (e.g., fluorescence microscopy) can also be applied to this problem (Adiga et al. 2006; Cong et al. 2000; Gudla et al. 2008; Al-Kofahi et al. 2010; Kong et al. 2011; Li et al. 2008; Malpica et al. 1997; Naik et al. 2008; Wahlby et al. 2004). In order to use these methods, some adaptation of some of the image processing steps will likely be needed. For example, modification of the way a nuclear staining image is obtained.

2.5.3 Tubules segmentation

Along with nuclear pleomorphism, the degree of structural differentiation of the tissue is one of the earliest prognostic factors for breast cancer patients that have been identified by pathologists. Cancer disrupts the ability of the nuclei to communicate with each other and organize themselves into structures such as tubules, making the lack of tubule formation an indicator for advanced malignancy. The tubules are generally round or oval structures consisting of lumen surrounded by a layer of epithelial cells. The major challenge in tubule segmentation is the similar appearance of other structures, such as adipose tissue or tears formed during the tissue preparation process, which only lack the outside layer of well-arranged epithelial nuclei.

A color gradient-based geodesic active contour model for segmentation of the tubular lumen areas was proposed in Xu et al. (2010b). The segmentation is initialized by weighted mean shift clustering and normalized cuts, and performs more favorably than the Chan-Vese region-based active contour model. This work was extended in Basavanhally et al. (2011) by incorporating domain knowledge to distinguish between tubules and other lumen-like areas. The authors show that the segmentation result can be used to infer the degree of tubule formation as defined in the Bloom-Richardson grading system.



(a) Region from a PPH3 labeled section. This staining method is specific to cells in the M phase which are clearly visible as dark brown objects (marked with green arrows).

(b) The same section stained with H&E. The corresponding mitotic figures are visible as hyperchromatic objects. However, many other similar objects appear which complicates mitosis counting.

Figure 2.3: Mitotic figures in breast cancer histopathology images.

Although the literature for tubule segmentation in breast cancer is not extensive, methods developed for segmentation of other related structures are also applicable to this problem. A very closely related application is segmentation of glands in prostate cancer histopathology images (Adiga et al. 2006; Nguyen et al. 2012; Peng et al. 2011; Xu et al. 2010a). The methods developed for this application might be used for tubule segmentation in breast cancer with little adaptation. It should also be mentioned that some of the features that can be derived by performing segmentation of tubules, can also be captured by computing features that describe the overall architecture of the tissue (van Diest et al. 1992a; Doyle et al. 2008).

2.5.4 Mitotic figures detection and assessment of proliferation

Of the three components that are part of histological grading of breast cancer, the assessment of tumor proliferation is probably the most important and prognostically significant one (van Diest et al. 2004). The oldest and still most widely used form of assessment of tumor proliferation is counting of mitotic figures in a predefined tissue area (usually 2 mm^2). The nuclei of the cells that are in the M phase of the cell cycle have distinctive morphological appearance in the H&E sections. Most commonly, mitotic figures manifest themselves as hyperchromatic objects without a clear nuclear membrane, with “hairy” protrusions around the edges and basophilia instead of eosinophilia in the surrounding cytoplasm. However, these are more instructive than defining features, and the bulk of the training of pathologists consists of looking at

specific examples of mitotic figures designated as such by experts. The task of identifying mitotic figures is notoriously time-consuming and difficult, due to the fact that many other objects such as apoptotic and necrotic nuclei may have similar appearance, which renders it difficult even for trained experts to make a distinction. Lymphocyte nuclei, compressed nuclei, “junk” particles and other artifacts from the tissue preparation process can also have a hyperchromatic appearance.

In addition to mitosis counting in H&E sections, there are also IHC techniques that can be used for assessment of proliferation of breast cancer tumors. However, they are not routinely used owing to increased cost and additional time delay. The most widely used IHC technique is labeling of the Ki-67 antigen that is associated with cell proliferation and expressed in all cell cycle phases except G0 (Dowsett et al. 2011). Because Ki-67 labeling is unspecific to the M phase, the assessment of the number of mitotic figures is usually higher than by counting in H&E sections and it might not have the same significance. A proposed alternative to Ki-67 is phosphohistone H3 (PPH3) that has the advantage of targeting only nuclei in the M phase, and has been shown to have prognostic significance (Skaland et al. 2007). An example region labeled for PPH3 is shown in Figure 2.3(a). With both of these staining methods, the previously discussed techniques for quantification of positively stained nuclei can be used.

The standard approach of assessing tumor proliferation in pathology labs, however, still remains mitosis counting. Given that this is the most tedious part of the Bloom-Richardson grading system, there is a large incentive to develop an automatic mitosis detection algorithm that works with the routinely prepared H&E sections.

The earliest proposed approaches were unavoidably hampered by the limited image acquisition quality and computational power (Kaman et al. 1984) and the need of specialized staining such as Feulgen to better highlight the chromatin (Beliën et al. 1997; ten Kate et al. 1993). Recently, two publicly available datasets of H&E stained breast cancer histopathology images with annotated mitotic figures were made available (Roux et al. 2013; Veta et al. 2014a), which sparked further development of different mitosis detection approaches (Ciresan et al. 2013; Huang et al. 2012; Irshad 2013; Irshad et al. 2013; Malon et al. 2013; Malon et al. 2012; Rajpoot et al. 2013; Sommer et al. 2012; Tek 2013; Veta et al. 2013a).

The majority of the proposed approaches work by first identifying candidate objects or locations that are then classified as mitotic figures or other objects. By far, the most distinctive feature of the mitotic figures is their hyperchromicity. In most cases, the intensity of the staining of the mitotic figures is noticeably darker than normal epithelial nuclei and only comparable to ap-

optotic, necrotic or compressed (artifact from the tissue preparation) nuclei and lymphocyte nuclei. This is illustrated in Figure 2.3(b). The candidate extraction phase often may make use of this distinctiveness by performing thresholding, local intensity minima detection or pixel-wise classification followed by refining of the detected regions by morphological operations and/or active contours segmentation. In the second stage, more specialized features designed to capture the specific morphology of mitotic figures are used to train a classification model. However, convolutional neural networks that operate on raw RGB image patches appear to exhibit the best performance for this detection task (Ciresan et al. 2013; Malon et al. 2013; Malon et al. 2012). The approach that was proposed in (Ciresan et al. 2013) is unique in the sense that it uses deep convolutional neural networks and does not perform candidate detection as an initial stage. Instead, it performs classification at every pixel location, which achieves excellent results.

A limiting factor of automatic mitosis detection is that whole slide images are typically scanned at a single focal plane. During mitosis counting, pathologists rely on “fine tuning” of the focus — something that is missing in whole slide images at a single focal plane. Digital slide scanners that perform image acquisition at multiple focal planes are now becoming more common; however, this feature is not widely used because of the increased storage demands. We anticipate that in the future, as storage costs go down and new image compression techniques become available, this limitation will be removed.

2.5.5 Computer-aided diagnosis and prognosis

The objective of image analysis of digitized histopathology slides is to facilitate, and preferably automate, computer-aided diagnosis and prognosis (CAD and CAP) in pathology labs. CAD is defined as the detection of cancer within the examined tissue, whereas CAP addresses the more complex problem of predicting the outcome for the patient based on the available data. There are numerous challenges in achieving this objective, including the large image sizes and the lack of representative datasets with high quality annotations by multiple observers, and with patient follow-up. On the other hand, the incentives for developing CAD and CAP systems are overwhelming. Patients that are suspected of having breast cancer undergo a biopsy that is examined by a pathologist. Large numbers of cases are found to be benign (Bulte et al. 2013), often easily distinguishable from cancer, which implies that a CAD system operating at high sensitivity can significantly reduce the workload of the pathologist, even if the specificity is moderate. This kind of system can also be used for quality control and assurance, for example, to identify positive cases that have been missed during the routine examination. Furthermore, the extraction of quantitative parameters from tumor regions can go a long

way towards reducing the inter- and intra-observer variability of breast cancer grading, which has been well documented in the literature.

Some of the proposed breast cancer detection techniques operate by examining nuclear features, as cancer nuclei have a distinct morphology — large size, coarse chromatin texture and irregular shape. In Cosatto et al. (2008) ROIs from breast cancer histopathology slides are classified as benign or malignant based on two features related to the nuclear size: the median nuclear area and the number of large well-formed nuclei in the region. Nuclear morphometric features are also used in Chekkoury et al. (2012) in combination with texture and topology features for malignancy detection in breast cancer histopathology. The extraction of nuclear morphometric features, in most cases, relies on a nuclei segmentation procedure. The influence of the nuclei segmentation accuracy on the subsequent feature extraction for classification into the classes benign and malignant is examined in Boucheron et al. (2010). The conclusion of the authors was that perfect segmentation accuracy is not needed as it does not necessarily guarantee optimal performance. In Doyle et al. (2008), cancer tissue is distinguished from non-cancer tissue based on Gabor texture features, without relying on a nuclei segmentation algorithm.

It should be noted here that breast cancer grades can be seen as estimates of patient outcome based on an expert opinion. In that sense, CAP systems that aim at predicting the histological tumor grade use this as an intermediate end-point for the prediction of survival. The use of reliable intermediate end-points is crucial in a situation where the actual patient survival is not known or it is difficult to obtain. Besides the histological grade, other intermediate end-points can be used, for example, the risk scores from gene expression assays. Such work was presented in Basavanhally et al. (2012) and Madabhushi et al. (2011), where tissue architecture features were used to predict the Oncotype DX recurrence score with high accuracy, which suggests that image analysis methods have the potential to be used as a cheap alternative or supplement to gene expression profiling.

When patient cohorts with known survival outcome are available, systems can be built that directly predict patient survival based on the available data, both from imaging and from other sources, thus avoiding the use of possibly unreliable intermediate end-points. One such example can be found in Veta et al. (2012), where the mean nuclear area calculated by automatic nuclei segmentation was shown to be prognostically relevant for male breast cancer patients in addition to other histological and clinical features. The fractal dimension of the tissue, extracted from invasive breast cancer TMAs stained with pan-cytokeratin (specific to epithelial components), which was previously associated with tumor grade (Tambasco et al. 2008), was also shown

to have prognostic value (Tambasco et al. 2010). In Beck et al. (2011), the authors proposed a system that successfully predicts patient outcome based on automatically extracted quantitative features from histopathology images. In addition, they identified three novel and previously unrecognized stromal features that are significantly associated with patient survival.

An exciting and promising area of research is the integration of imaging biomarkers from histopathology images with genomic data. Structural information about the tissue is lost when preparing the molecular assays, hence imaging biomarkers may be complementary to genomic data. In a recent study (Yuan et al. 2012), the authors used an image analysis approach to derive knowledge about the tumor cellular composition (percentage of cancer, stromal and lymphocytic nuclei), which they used to correct copy number data and more accurately estimate the HER2 amplification. In addition, they combined image features with genomic information to train a predictor of survival of ER-negative breast cancer patients. This predictor had a better performance in comparison with using image features or genomic information only. In another study (Wang et al. 2013), the authors developed a workflow for image analysis of histopathology images and integration of morphological features with genomic data for biomarker discovery. Four of the morphological features were identified as biomarkers that can separate patients into groups with different outcomes.

Many of the proposed methods for automatic breast cancer detection, grading and prognosis have been trained and evaluated only on relatively small regions from the image slides, either digitized TMA slides or manually selected regions from whole-slide images. Although these methods are useful in a semi-automatic setting, it is necessary for a high throughput and automatic application to either produce an output for the entire slide (or even, from a set of slides originating from one specimen), as in Basavanhally et al. (2013b), where this is achieved by using a multi-field-of-view framework, or to perform extraction of relevant ROIs for the problem at hand and limit the analysis solely to those regions (Bahlmann et al. 2012; Huang et al. 2011; Peikari et al. 2013).

2.6 Discussion and conclusions

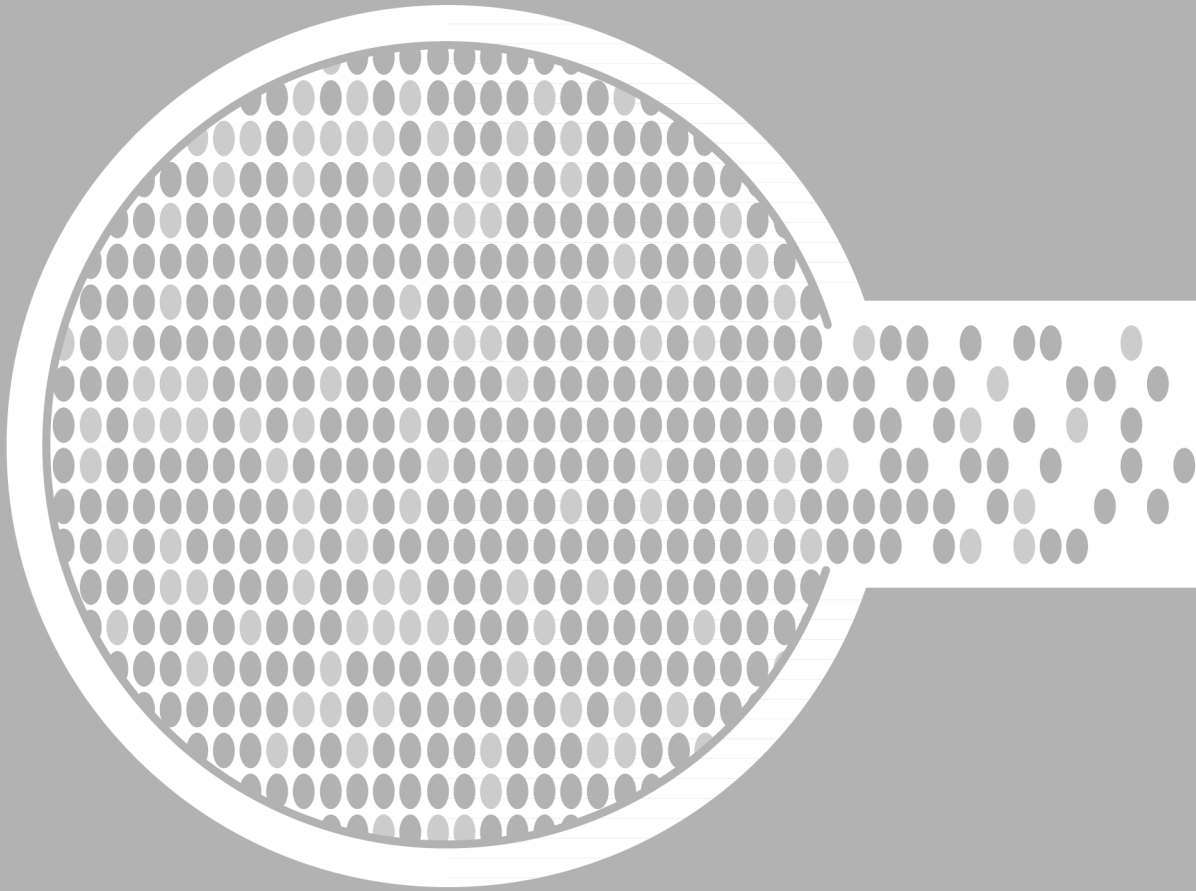
In the past few years the interest in analysis of histopathology images has been steadily increasing, prompted by the introduction of whole slide imaging scanners into pathology labs and the imminent acceptance of digital slides as a primary diagnostic modality.

The use of quantitative techniques is viewed as a solution to the problem of observer variability of the interpretation of histopathology slides, both by

pathology professionals and image analysis experts. In the case of breast cancer, several applications for quantification of immunohistochemically stained tissue have already gained approval from the United States Food and Drug Administration (FDA). Methods that work with routinely prepared H&E stained slides have great potential to make an impact on the pathology workflow. However, ensuring robustness is challenging owing to the complexity of the tissue characteristics that need to be analyzed. Nevertheless, progress has been made both in the development of basic image analysis tools for H&E stained sections (i.e., object detection and segmentation methods) and in the development of systems that predict patient outcome either directly or by use of intermediate endpoints such as the histological grade as assigned by pathologists. Further improvements are needed in order to produce methods that have performance levels that are suitable for clinical application.

At multiple points throughout the text we have mentioned that the steps of tissue preparation, staining and slide digitization that precede image analysis can influence the results. The more quantitative analysis of histopathology image data will become part of routine pathology practice, the more optimization of tissue preparation, staining and slide digitization will be needed. However, some variation in the appearance will still be present even under closely monitored conditions, so analysis methods will have to be developed in a way that is robust to such variation.

Perhaps the main obstacle in the development of new histopathology image analysis methods lies in the lack of large publicly available annotated datasets. While the advent of WSI scanners has produced vast quantities of image data, it is difficult to obtain ground truth annotations in a form that can readily be used for development and testing of image analysis methods, even when this data is tied to the pathology reports. For example, the tumor grade or the mitotic activity index that can be found in pathology reports are often based on the analysis of a particular region from a single slide viewed under a microscope, while the slide itself has been selected from a set of slides originating from the same sample. Information about the selected slide and the region within that slide where analysis was performed is not recorded and thus correspondence between the routine annotation and the image data is difficult to establish. In addition, owing to the large observer variability, annotation by multiple observers is needed to produce high quality ground truths, which is both time-consuming and expensive, particularly for large datasets. Making annotated image datasets publicly available will provide a breeding ground for the development of new image analysis algorithms and will enhance the objectivity of method comparison and improve the quality of computer-aided diagnosis and prognosis.



CHAPTER

3

Automatic nuclei segmentation

Summary

The introduction of fast digital slide scanners that provide whole slide images has led to a revival of interest in image analysis applications in pathology. Segmentation of cells and nuclei is an important first step towards automatic analysis of digitized microscopy images. We therefore developed an automated nuclei segmentation method that works with hematoxylin and eosin (H&E) stained breast cancer histopathology images, which represent regions of whole digital slides.

The procedure can be divided into four main steps: 1) pre-processing with color unmixing and morphological operators, 2) marker-controlled watershed segmentation at multiple scales and with different markers, 3) post-processing for rejection of false regions and 4) merging of the results from multiple scales. The procedure was developed on a set of 21 breast cancer cases (subset A) and tested on a separate validation set of 18 cases (subset B). The evaluation was done in terms of both detection accuracy (sensitivity and positive predictive value) and segmentation accuracy (Dice coefficient).

The mean estimated sensitivity for subset A was 0.875 (± 0.092) and for subset B 0.853 (± 0.077). The mean estimated positive predictive value was 0.904 (± 0.075) and 0.886 (± 0.069) for subsets A and B, respectively. For both subsets, the distribution of the Dice coefficients had a high peak around 0.9, with the vast majority of segmentations having values larger than 0.8.

Based on: Mitko Veta et al. (2013b). „Automatic nuclei segmentation in H&E stained breast cancer histopathology images”. In: *PLoS ONE* 8, e70221

3.1 Introduction

Assessment of breast cancer prognosis from excision biopsy slides relies largely on the Bloom-Richardson grading system. It is based on semiquantitative scoring of the degree of tubule formation, nuclear pleomorphism, and mitotic rate, which has proven to be prognostically strong (Elston et al. 1991). However, the scoring is done traditionally by visual examination through the microscope which has suboptimal reproducibility (Robbins et al. 1995). The use of automatic image analysis methods, which can provide reproducible quantitative parameters that describe the tumor tissue, has been suggested as a way to overcome this drawback (Meijer et al. 1997).

Traditional image analysis of conventional glass slides was hampered by the selective approach due to limitations of the scanning equipment and the need for special stains (Beliën et al. 1997). The introduction of fast digital slide scanners that provide whole slide images has led to a revival of interest in image analysis applications in pathology. Optimal integration of such applications in pathology workflow necessitates using hematoxylin and eosin (H&E) stained slides since this is the standard staining protocol (the diagnostic process for each case always starts with staining the specimen with these dyes). Given the complexity and the diversity of the tissue appearance, the automatic analysis of H&E stained images can be very challenging.

Segmentation of cells and nuclei is an important first step towards automatic analysis of digitized microscopy images. Most of the developed cell and nuclei segmentation techniques revolve around active contours, watershed segmentation, pixel-wise clustering/classification or a combination of the above, supplemented by different pre-processing and post-processing steps and detection/localization schemes. Bamford et al. (1998) used a dual active contour model for the task of segmenting cell nuclei from cytoplasm in conventional Papanicolaou stained cervical cell images. Cosatto et al. (2008) detected candidate nuclei locations in breast histopathology images using the Hough transform and evolved an active contour around each point, rejecting malformed outlines with a trained classifier. They used the segmentation output for predicting nuclear pleomorphism scores, however, the segmentation method by itself was not rigorously evaluated. Fatakdawala et al. (2010) presented an expectation-maximization driven geodesic active contour with overlap resolution for segmentation of lymphocytes in breast cancer histopathology images. Ali et al. (2012) presented an active contour model that integrates region, boundary and shape information, and showed that it can be used for nuclei, lymphocytes and gland segmentation in prostate and breast cancer biopsy images. Wienert et al. (2012) proposed a method for nuclei detection and segmentation based on contour tracing and subsequent pruning of contours to retain the most probable ones. They evaluated the detection

Table 3.1: Nuclei segmentation dataset summary.

	Number of slides	Pleomorphism grade distribution (I, II and III)	Total number of manually segmented nuclei	Average number of manually segmented nuclei per slide
Subset A	21	8; 8; 5	2191	104.3 (± 12.2)
Subset B	18	1; 10; 7	2073	115.2 (± 12.2)

Representative regions from subset A were used for tuning of parameters during the development of the segmentation procedure. Representative regions from subset B were used for an independent validation of the chosen parameters. From each slide, approximately 100 representative nuclei were manually segmented with systematic random sampling.

performance of the algorithm in a set of breast, liver, gastric mucosa and bone marrow images.

Watershed segmentation is a method particularly suited for cell and nuclei segmentation (Beucher et al. 1993; Mousses et al. 2003). The results of the classical watershed segmentation can be significantly improved by modifying the segmentation function (topographical relief) to contain regional minima only at specific locations that mark the objects of interest and the background. These markers can be obtained in a variety of ways and the process is usually application-dependent. Malpica et al. (1997) examined the use of this technique in bone marrow and peripheral blood microscopy images. Marker-controlled watershed for segmentation and subsequent tracking of cells in time lapse microscopy was proposed by Yang et al. (2006). Huang et al. (2010) described a method for segmentation of nuclei in hepatocellular carcinoma biopsy images based on marker-controlled watershed segmentation of initial contours followed by refinement with a snake model. Marker-controlled watershed, with markers produced by template matching, was also used by Kachouie et al. (2010) for segmentation of mammalian cells in microscopy images.

Although many nuclei/cell segmentation methods exist in the literature, they are usually closely related to the microscopy technique, tissue type, staining and target nuclei/cell types. Thus, they are not directly applicable to an arbitrary type of image. In this chapter we present a marker-controlled watershed based technique for segmentation of cancer nuclei in H&E stained breast cancer histopathology images. In addition to the combination of the different processing steps, the novelty of the method lies in the multiscale approach to the pre-processing of the images and the marker extraction for the watershed

segmentation, the use of multiple marker types and the relatively simple but effective merging of the segmentations produced at different scales and from multiple markers. This multiscale and multimarker approach yields much better results than simply performing segmentation at a single scale and with a single marker type. The method was evaluated with regards to both the detection and the segmentation accuracy on a set of breast cancer images of diverse tissue appearance, and showed excellent results. In addition to the evaluation on our dataset, we evaluated our method on the dataset used in Wienert et al. (2012) and achieved comparable results.

3.2 Materials and methods

3.2.1 Breast cancer cases

For this study a total of 39 slides from 38 patients from breast cancer excision biopsies were used. The slides were routinely prepared with the standard procedure consisting of formalin fixation and paraffin embedding of the tissue, followed by cutting of 3-5 μm thick sections and staining with H&E. The digitization of the complete slides was done using a ScanScope XT whole slide scanner (Aperio, Vista, CA, USA) at a magnification of $\times 40$ (0.75 NA) and a resolution of 0.25 $\mu\text{m}/\text{pixel}$. JPEG2000 compression with a quality factor of at least 80 was used to reduce the storage requirements. With this compression type and quality, no visible compression artifacts were present in the digital slides. From each digital slide a representative region of approximately 1×1 mm was selected and marked by an experienced pathologist (PJvD) and graded for nuclear pleomorphism according to the Bloom-Richardson grading system (grade I, II or III ranging from good to poor prognosis). The regions of interest were selected using predefined guidelines that are also used when performing grading by pathologists. More precisely, only areas with high epithelial cellularity and preferably on the periphery of the tumor were selected. Regions with severe lymphocytic infiltration and necrosis were avoided, as well as regions with scanning artifacts and out-of-focus problems.

The regions were divided into two subsets. Subset A consisted of 21 slides and was used during the development of the segmentation procedure. These slides were selected by an experienced pathologist (PJvD) to represent the diversity in tissue appearance and to have an approximately balanced distribution of pleomorphism grades. Subset B consisted of 18 slides of consecutive patients collected from our Pathology Department archive based solely on the availability. The segmentation procedure was developed on subset A and validation was performed on subset B. All the experiments in this study were performed on the selected representative regions from the digital slides.

3.2.2 Ground truth segmentation

To set the gold standard, manual segmentation was performed in the marked regions on all 39 slides. Since each region contains many thousands of nuclei, manual segmentation of all nuclei was impractical and a systematic random sampling approach was followed (Fleege et al. 1990). This involved overlaying a grid of measurement frames over the marked region and segmenting one nucleus within each measurement frame (Figure 3.1). The grid was overlaid starting from an arbitrary location according to a distribution rule. The distribution rule depended on the area of the measurement frame and of the region, on the desired number of segmentations and on the estimated tumor area within the region (for more details see Fleege et al. (1990)). Each measurement frame was subdivided into five rows. Scanning the rows from left to right, the first unscathed epithelial breast cancer nucleus with identifiable contours whose center of mass lied within the row was chosen for manual segmentation (Figure 3.1(a)). Measurement frames of size $50 \times 50 \mu\text{m}$ and a target of 100 nuclei per region were used. An expert (RK) performed one manual segmentation per measurement frame.

A summary of the dataset is presented in Table 3.1. We point out that in some cases the target number of 100 nuclei was not reached when too many of the sampling frames fell into non-tumor tissue, while in other cases this number was overreached. The sample size of 100 nuclei was chosen because it has been shown that this number of segmentations is sufficient to reliably estimate certain morphometric features such as the mean nuclear area (Janink et al. 1995). At the resolution at which the digital slides were scanned, the average area of the manually segmented nuclei was approximately 900 pixels.

3.2.3 Overview of the method

A block-diagram with an overview of the proposed method is presented in Figure 3.2. This is an extension and improvement of our previously published nuclei segmentation method (Veta et al. 2011). The entire procedure can be divided into four main steps: 1) pre-processing, 2) marker-controlled watershed segmentation, 3) post-processing and 4) merging of the results from multiple scales. The aim of the pre-processing is to remove irrelevant content while preserving the boundaries of the nuclei. The pre-processing starts with color unmixing for separation of the hematoxylin stain from the RGB image (the nuclei are dyed by this stain; Figure 3.3(b)). The grayscale version of the hematoxylin image is then processed with a series of morphological operations in order to remove irrelevant structures (Figure 3.3(c)). The core part of the procedure is the marker-controlled watershed segmentation. Two

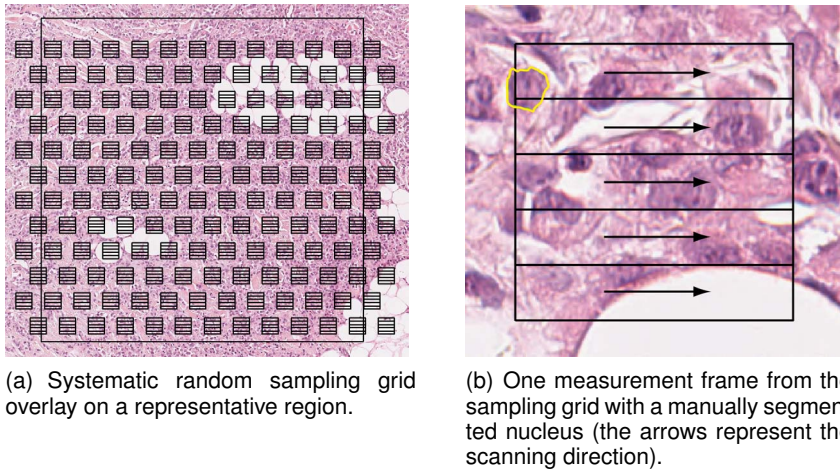


Figure 3.1: Systematic random sampling method used for manual nuclei segmentation.

types of nuclear markers are used: markers extracted using an image transform that highlights structures of high radial symmetry (Figure 3.3(d)- 3.3(f)) and regional minima of the pre-processed image (Figure 3.3(g)-3.3(h)). In the post-processing step, regions unlikely to represent nuclei are removed and the contours of the remaining regions are parameterized as ellipses. By varying the size of the structuring element in the pre-processing step, the segmentation procedure can be tuned to look for nuclei at different scales, allowing multiscale analysis. The segmentation results from the multiple scales and two marker types are then merged by resolving concurrent regions to give the final segmentation.

3.2.4 Color unmixing

The first step is separation of the H&E stains with the color unmixing technique suggested in Ruifrok et al. (2001), which is a special case of true spectral unmixing techniques that work with multispectral cameras (Garini et al. 2006). The technique uses the fact that the image formation process in bright field microscopy can be modeled by the Lambert-Beer law. Given that the images are captured by three detection channels (R, G and B) with known optical densities and the stain-specific absorption coefficients can be experimentally determined from single stain images, the concentrations of the two stains can be determined for each pixel location. These in turn can be used to obtain single stain images. Since the nuclei are stained with hematoxylin, the gray-scale version of the hematoxylin single stain image is used in all subsequent processing. An example of color unmixing is presented in Figure 3.3(b).

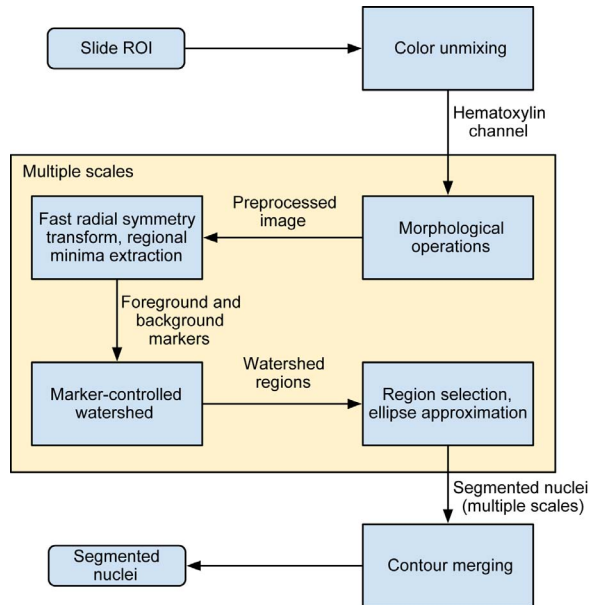
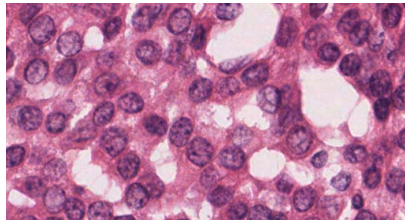


Figure 3.2: Schematic overview of the different steps in the automated image analysis method for nuclei segmentation.

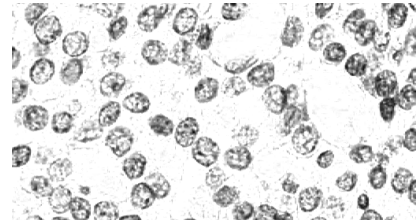
3.2.5 Morphological operations

The now separated hematoxylin image still contains spurious structures within the nuclei. These present obstacles for the marker extraction and segmentation and can be filtered out with a series of operations based on morphological grayscale reconstruction (Vincent 1993). Opening by reconstruction removes unconnected bright objects that are smaller than the structuring element (SE). Similarly, closing by reconstruction removes unconnected dark objects smaller than the SE. Applying these two operators in sequence produces “flat” images and the amount of detail present can be controlled by the size of the SE. In the hematoxylin images, best results were obtained by first applying opening and then closing by reconstruction (both with a disk-shaped SE with radius n). The size of the SE, as defined by the radius n , should be chosen according to the size of the spurious structures which in turn is related to the size of the nuclei and the resolution of the image.

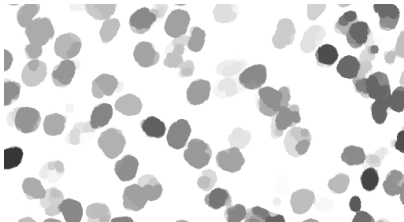
After application of these two operations the main contours of the nuclei often have an irregular shape and protrusions emanating from the edges hampering the segmentation result. To remedy this problem, additional morphological closing with a small SE is applied. This simplifies the shape of the object, eliminates small protrusions, disconnects “loosely” connected objects and does not significantly affect the location of the main contours. The SE for this operation is chosen to be a disk with half the radius of the one used for the opening and closing by reconstruction operators. An example of prepro-



(a) Original image.



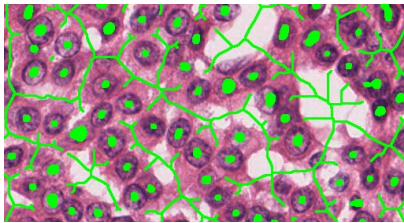
(b) Hematoxylin channel.



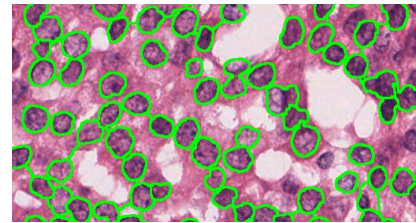
(c) Pre-processed image (hematoxylin channel processed with series of morphological operations).



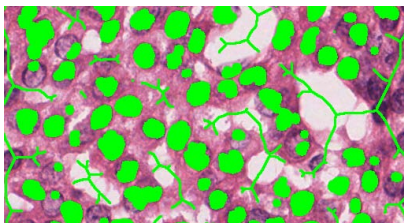
(d) Fast radial symmetry transform (FRST).



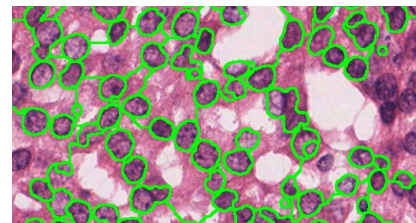
(e) FRST foreground and background markers.



(f) Watershed segmentation with FRST markers.



(g) Regional minima foreground and background markers.



(h) Watershed segmentation with regional minima markers.

Figure 3.3: Marker imposition and watershed segmentation of nuclei. Prior to applying the FRST the image is preprocessed with color unmixing and morphological operations ($n = 10$). The set of radii for the FRST is $R = \{10, 11, \dots, 20\}$. Note: the markers and watershed ridges (given in green in the figure) were dilated by one pixel for better visualization.

cessing with the series of morphological operations is shown in Figure 3.3(c).

It is difficult to set one parameter n that will work well across all images in our data set, or, in many instances, across different nuclei within one image. The optimal simplification factor is closely related to the size of the undesired structures that need to be removed (as all unconnected objects smaller than the SE will be removed). Employing a large SE oversimplifies the image, while using too small a SE does not always produce desirable results as many of the substructures within the large nuclei remain, affecting segmentation performance. This is why a multiscale approach was chosen — each image is preprocessed with SEs of different sizes and segmentation is performed at each scale. For the problem at hand, the range of SE radii is set to be $n \in \{10, 11, \dots, 18\}$ pixels, which corresponds to the approximately expected range of minor semi-axes in breast cancer nuclei imaged at this magnification.

3.2.6 Fast radial symmetry transform

The fast radial symmetry transform (FRST) (Loy et al. 2003) is a computationally efficient, non-iterative procedure that operates along the direction of the image gradient to infer centers of radial symmetry. This transform was originally developed for face detection tasks in computer vision, but was recently used in automatic analysis of follicular lymphoma (Kong et al. 2011; Sertel et al. 2010) and bears similarity to other operators specifically designed for cell and nuclei segmentation (Schmitt et al. 2008). A generalized version of this transform was used in Chekkoury et al. (2012) for segmentation of nuclei in breast cancer biopsy images.

The nuclear contours, in most cases, exhibited high radial symmetry making this transform suitable for their localization. To produce candidate nuclei locations, we use the orientation-based version of the transform, which discards gradient magnitude information and relies only on the orientation. This can be beneficial in the case of low contrast between the nuclei and the background. The FRST is computed for a set of radii R that reflects the size of the symmetric features that need to be detected. An example of the FRST applied to a morphologically pre-processed image is given in Figure 3.3(d).

3.2.7 Marker imposition and segmentation

Given an input image preprocessed with the morphological operators at scale n , two marker-controlled watershed segmentations, each targeting a specific type of nuclei, are performed — one using FRST markers and one using regional minima markers. The FRST S is computed for the set of radii $R \in \{n, n + 1, \dots, 2n\}$ pixels. This set of radii reflects the size of the nuclei that are reconstructed well in the preprocessed image. The FRST nuclei mark-

ers are extracted as the extended regional minima of S , with an empirically set height parameter $h = 0.4$. The extended regional minima of S are calculated as the regional minima of the h -minima transformation of S .

For successful watershed segmentation the background also has to be marked. To achieve this, a naïve assumption that each detected foreground marker corresponds to a nucleus with maximal size (the largest radius in the set R) is made. In this way, provisional foreground (nuclei) and background maps can be formed. The morphological skeleton of the background map is used as a background marker.

After foreground and background markers have been obtained, the Sobel gradient magnitude image of the pre-processed image, which is used as a segmentation function for the watershed, is modified by imposing regional minima on the locations of the markers. In this way, only one watershed region per marker is obtained.

Although the FRST markers are very successful in marking nuclei even in more complex situations like clustered nuclei, sometimes a proper marker is not produced in situations when the symmetry assumption is violated or in case of overly elongated nuclei. To address these situations, at each scale, an additional watershed segmentation is produced using the regional minima of the pre-processed image as markers as in Huang et al. (2010). The background markers are defined in the same way as for the FRST case. Figure 3.3 gives an example of marker-controlled watershed segmentation with FRST and regional minima markers. Figure 3.3(e) and 3.3(g) give the foreground and background markers from the FRST and the regional minima respectively, and corresponding results from the segmentation are given in Figure 3.3(f) and 3.3(h).

3.2.8 Post-processing

Many of the resulting watershed regions do not correspond to nuclei or represent erroneous segmentations (severe over- or under-segmentation, regions spilled into the background etc.). In the post-processing step we aim to remove those regions based on the following extracted features:

Solidity (s): The ratio of the area of the object and of the convex hull of the object (the convex polygon with smallest area that contains the object). This value should be high for the nuclei regions since they are rarely concave. In our previous work (Veta et al. 2011) we have shown that this feature can be highly discriminative between correct and incorrect segmentations produced by marker-controlled watershed.

Boundary saliency (l): The difference between the intensity level of the outside boundary and the intensity level of the inside boundary of the nucleus. The outside intensity level is taken as the median of the intensity values in a tight band around the segmented region. The inside intensity level is defined in an analogous way.

Mass displacement (d): The distance between the centroid and the weighted centroid of the region (the pixel locations are weighted by the inverse intensity values) normalized by the smaller axis of the region. Low values of this feature imply near symmetric distribution of the intensity inside the nucleus region. In certain situations regions that do not correspond to correct segmentations have high mass displacement (regions spilled into the background, over-segmentations, under-segmentations etc.).

Although the problem of identifying the non-nuclei regions can be posed as a one- or two-class statistical classification task, we found that a simple rule-based rejection scheme is a much better and flexible solution. For each of the defined features a range of probable values is defined. If for a given region one of the features is outside of the probable range, the region is discarded. Additionally, regions that are too small (area $< n^2\pi$) or too large (area $> 4n^2\pi$) for the scale at which they are segmented (as defined by n) are removed. Since the coarseness of the extracted contours depends on the scale at which they were extracted (smaller scales result in contours with finer details and vice versa), all the contours are standardized by approximating them with ellipses.

The ranges for the features were empirically determined and are as follows: $s \in (0.875, 1)$, $l \in (20, 255)$, $d \in (0, 0.08)$. Qualitative analysis of the influence of the selected feature ranges can be found in the supplementary material of Veta et al. (2013b). Most of the segmentations outside of the excluded range correspond to false objects, and this effect is robust with respect to difference in tissue appearance.

3.2.9 Merging results from multiple scales

The outputs from the multiple scales and the two types of markers often produce overlapping regions. For example, a nucleus might be properly segmented at a certain scale, but a substructure within the nucleus might be segmented at a higher scale, and/or oversegmentation containing another nucleus might be produced at a lower scale. Much more commonly, almost identical segmentations are produced at neighboring scales and/or with the two types of markers. These situations are resolved by identifying all overlaps and selecting the most probable regions according to a fitness value. For all pairs of

regions (X_i, X_j) segmented in a given image we define the following overlap measure:

$$OV(X_i, X_j) = \frac{|X_i \cap X_j|}{\min(|X_i|, |X_j|)} \quad (3.1)$$

This measure has a maximum value of 1 when one of the regions is completely contained in the other one and a minimum value of 0 when the two regions do not intersect. Given this measure, the following adjacency matrix is defined:

$$A(i, j) = \begin{cases} 1 & \text{if } OV(X_i, X_j) > T_h \\ 0 & \text{otherwise} \end{cases} \quad (3.2)$$

The threshold T_h defines when two regions are considered to be overlapping. All pairs of regions with a non-zero overlap measure smaller than this value are considered to be only “touching”. Each region is also assigned a fitness value f that is used for comparing concurrent regions and selecting the one that is most likely to represent a nucleus. The region overlaps are then resolved according to the following simple algorithm:

1. Find the region r with the maximum fitness value f (see below);
2. Mark r as accepted and reject all regions that are adjacent to it;
3. Repeat steps 1. and 2. for the remaining regions until all are accepted or rejected.

The threshold T_h was chosen to be 0.2. This value allows small overlap of touching nuclei. Simply using the solidity of the region as a fitness value proved to give good results, although a linear combination of other features might be an alternative to consider.

3.2.10 Evaluation

The Dice coefficient is a measure of overlap between two regions, commonly used for evaluation of segmentation techniques. It is defined as:

$$D(X, Y) = 2 \frac{|X \cap Y|}{|X| + |Y|} \quad (3.3)$$

The automatic segmentations were compared with the manual segmentations obtained with systematic random sampling in the following way: if a manual segmentation was not intersected by an automatic segmentation with

a Dice coefficient of at least 0.2, it was counted as a false negative (FN). Otherwise, it was counted as a true positive (TP). The Dice coefficient was also taken as a measure of quality of the segmentation.

The reasoning behind a cut-off value of 0.2 was to avoid unsegmented nuclei that are “touched” by a neighboring segmentation to be counted as true positives. The value of 0.2 is arbitrary, but it should be pointed out that in case of a lower value, more nuclei will be counted as TP at the cost of having more segmentations with very poor quality and vice versa.

To estimate the positive predictive value a subset of 100 automatically segmented nuclei from each slide was randomly generated. An expert (RJK) labeled all segmentations that did not correspond to epithelial nuclei, such as stroma, lymphocytes, “junk” particles etc.

For each representative region the sensitivity, positive predictive value and the median Dice coefficient were estimated. Because of the asymmetric left-skewed distribution, the median of the Dice coefficient is a better measure of central tendency than the mean. We refer to the sensitivity, positive predictive value and median Dice coefficient measures as estimates because they are based on an annotated subset of the entire population of nuclei in the images.

In addition to the evaluation on our dataset, we evaluated the proposed method on a publicly available dataset used in a recently published paper on nuclei detection and segmentation (Wienert et al. 2012). This dataset contains 36 histopathology images of breast, liver, gastric mucosa and bone marrow imaged at $\times 20$ magnification. The ground truth is provided as manually annotated nuclei centroids. We evaluated the detection performance on this data set in the same way as in (Wienert et al. 2012), i.e. in terms of overall positive predictive value, sensitivity and conglomerate score (a score of the ability of the method to successfully separate conglomerates). For this experiment, no parameter values were adapted, except for the adjustment of the expected range of nuclei semi-axes, to account for the smaller magnification ($n = \{5, 6, \dots, 9\}$).

3.3 Results

Segmentation results for a few regions from our data set are given in Figure 3.4 for qualitative evaluation, along with the intermediate results prior to rejection of spurious contours and prior to the merging of concurrent regions. The four examples are chosen to represent tissue types with different appearance: large and small nuclei, nuclei organized into tubules, highly marginalized chromatin etc. In the same figure, the intermediate results prior to the rejection of false contours and merging of the contours from multiple scales are also shown. The visual examination shows overall good perform-

ance with a limited number of severe over- or under-segmentations. Also, it is apparent that a segmentation is produced for most of the nuclei in the image, with few contours corresponding to non-epithelial nuclei objects. The results from all the regions in our data set are available for download from: <http://www.isi.uu.nl/People/Mitko/segmentation.html>.

The sensitivity, positive predictive value and median Dice coefficient for each case in subsets A and B are summarized in Figure 3.5. Note that subset A was used during the development of the algorithm and subset B is used as an independent validation set. The sensitivity was estimated as the percentage of manual segmentations that were matched to an automatic segmentation, as explained in the previous section. The positive predictive value was estimated as the percentage of the annotated automatic segmentations (100 per slide) marked as corresponding to an epithelial nucleus. The mean estimated sensitivity for subset A was 0.875 (± 0.092) and for subset B 0.853 (± 0.077). The mean estimated positive predictive value was 0.904 (± 0.075) and 0.886 (± 0.069) for subsets A and B, respectively. For both subsets, the distribution of the estimated Dice coefficients had a high peak around 0.9, with the vast majority of segmentations having values larger than 0.8. The one outlier in terms of sensitivity in the first subset was due to the tissue being over-stained with eosin, which negatively affected the color unmixing procedure. The cases with low sensitivity in the second subset had a large proportion of nuclei that were not segmented due to their very small size (comparable to the size of lymphocytes). The outlying cases with low positive predictive value were either high grade cancer and/or had a large proportion of relatively large fibroblasts. In the high grade cancer cases, there were often many junk particles, usually of small size, that were picked up by the segmentation procedure. Although the scales for the segmentation were chosen so that most of the lymphocytes were not segmented, some were still included in the segmentation and they affected the positive predictive value negatively. Most of the segmentations had a high value of the Dice coefficient. The tail in the distribution of the Dice coefficients represents severe over- or under-segmentations (two or more nuclei segmented as one or a segmented sub-structure of a nucleus).

Specifying wider ranges of probable feature values during the post-processing will result in higher sensitivity but at the cost of decreasing the positive predictive value, and vice versa. The solidity feature is the most discriminative between true and false segmentations. This is because highly convex segmented regions are unlikely to occur by chance, and the convex regions that do occur correspond to correctly segmented nuclei in the vast majority of cases. This motivated the use of this feature as a fitness value during the region merging process.

A comparison of our multiscale method to the same method on only a single scale and with a single marker ($n = 12$ and FRST markers were chosen as best performing) showed that the sensitivity of the multiscale method on the validation set was significantly higher (0.853 compared to 0.579 on average). This exemplifies the added value of our multiscale approach. Our method generalized well when used to detect and segment nuclei in a diverse set of histopathology images, including breast, liver, gastric mucosa and bone marrow tissues. We achieved an overall positive predictive value of 0.904, sensitivity of 0.833 and a conglomerate score of 0.989 which is comparable to the results of the method presented in (Wienert et al. 2012) (0.908, 0.859 and 0.958 respectively).

One of the potential uses of an automatic nuclei segmentation method is to extract prognostically meaningful morphometric parameters. As an example, we show that the proposed nuclei segmentation technique can be used to reliably estimate the mean nuclear area (MNA) from the representative regions. The area of all segmented nuclei was calculated and then averaged for each representative region to produce the MNA. We trained a linear regression on the training set to correct for the systematic underestimation of the MNA. We observed that the main reason for this systematic underestimation is that the “junk particles” that are segmented are typically several times smaller than that of the large epithelial nuclei. In addition, undersegmentation of large nuclei is more common than oversegmentation of small nuclei. The learned linear regression was used to correct the MNA estimates of the cases in the validation sets. The results are presented in the form of a scatter plot in Figure 3.6. It can be observed that there is good correspondence between the two measurements and that there is no noticeable systematic bias.

3.4 Discussion and conclusions

This study set out to develop a segmentation method for breast cancer nuclei that works on H&E stained breast cancer histopathology images. The evaluation revealed that the proposed method has good performance in both detection and segmentation accuracy. The evaluation was done on two subsets of images, one of which was used for parameter tuning and the other for validation. The segmentation results were slightly worse for the validation subset, probably due to the fact that this data set contained more cases with high grade cancer that are generally more difficult to segment. Nevertheless, the results on this validation set provide a good idea of the performance of the algorithm in real life scenarios.

We did not perform standardization of the tissue appearance (Kayser et al. 2008), inasmuch as the techniques we used aim for robustness with respect

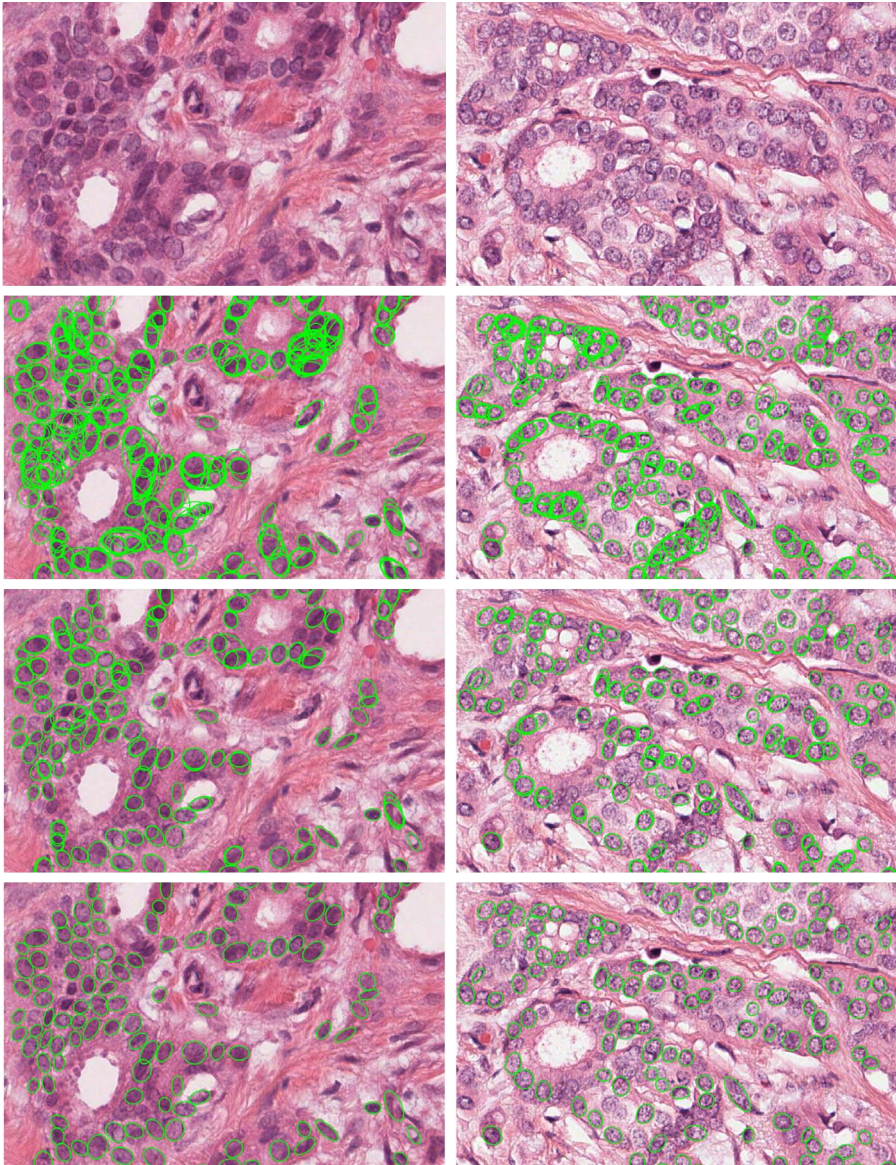


Figure 3.4: Examples of automated nuclei segmentation in breast cancer sections. All images are shown at the same scale. First row: original images. Second row: intermediate results prior to the rejection of spurious regions based on solidity, boundary saliency and mass displacement. Third row: intermediate results prior to the merging of contours from multiple scales. Fourth row: final segmentation results.

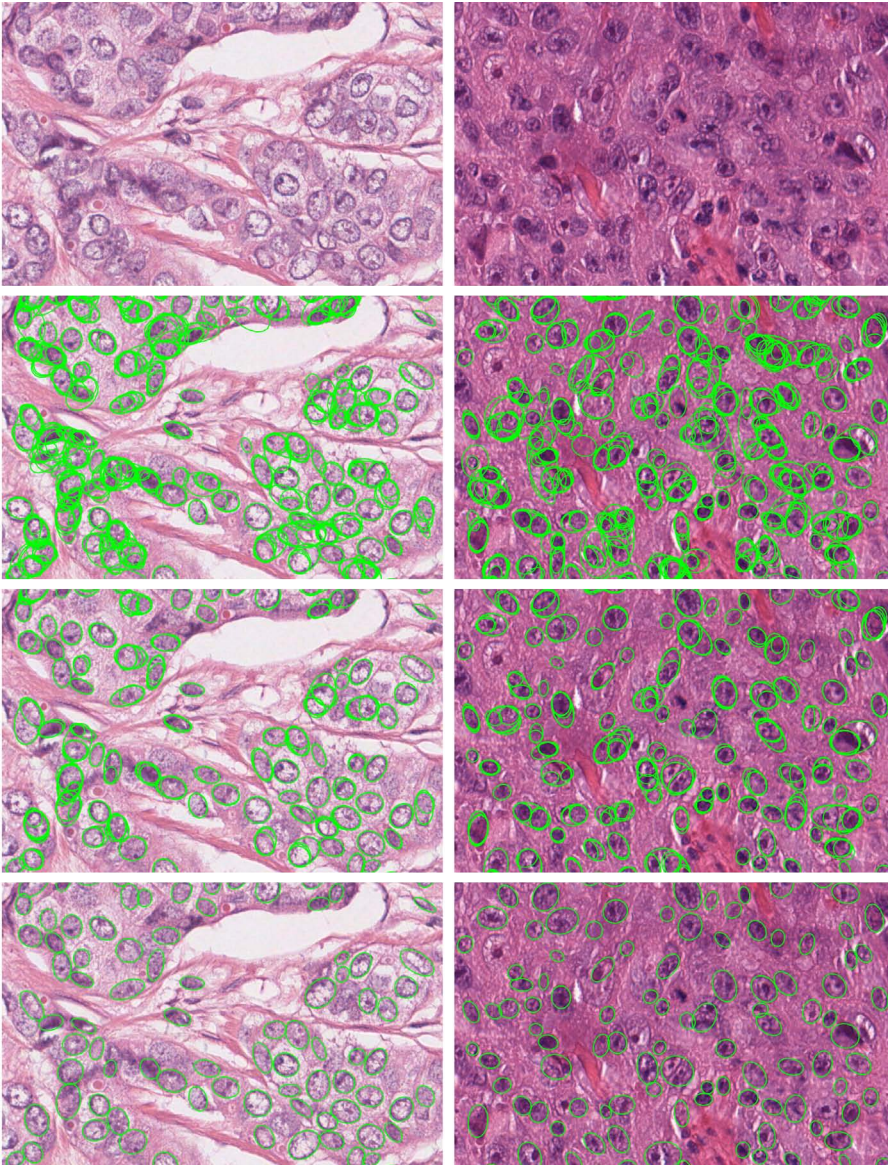


Figure 3.4: *Continued.* Examples of automated nuclei segmentation in breast cancer sections. All images are shown at the same scale. First row: original images. Second row: intermediate results prior to the rejection of spurious regions based on solidity, boundary salience and mass displacement. Third row: intermediate results prior to the merging of contours from multiple scales. Fourth row: final segmentation results.

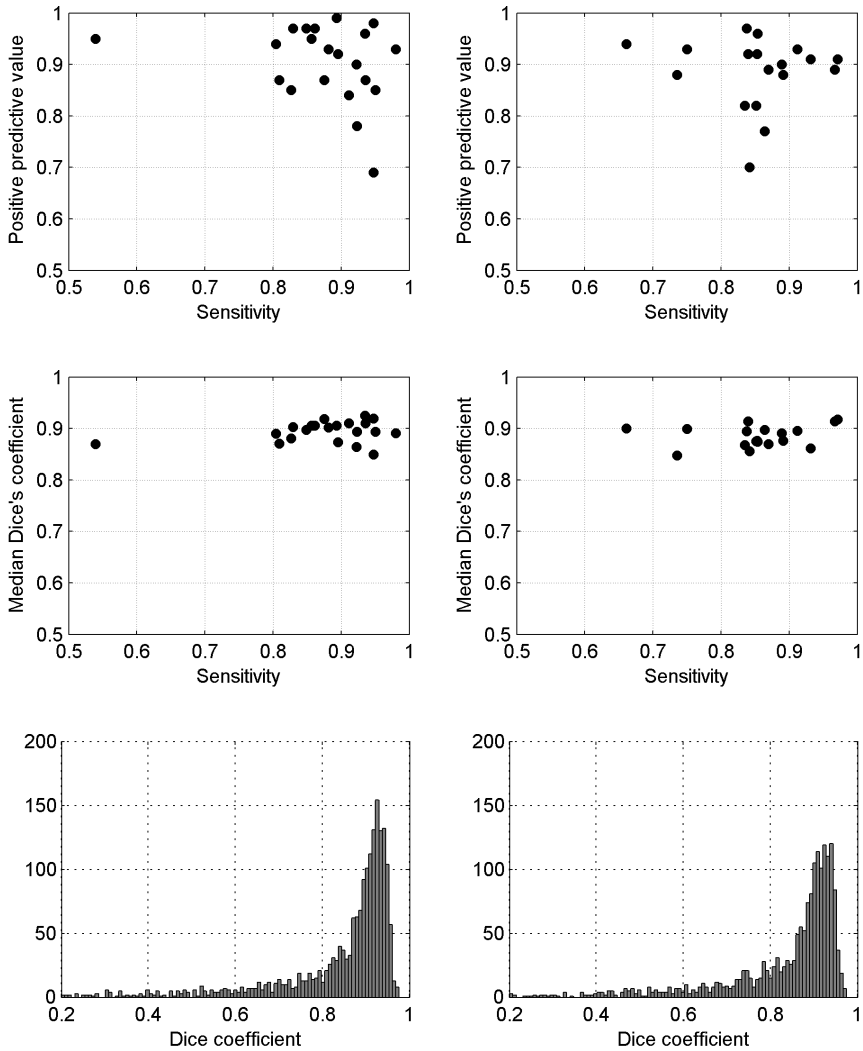


Figure 3.5: Plot of the performance measures. The performance measures in the first column refer to subset A and in the second row to subset B.

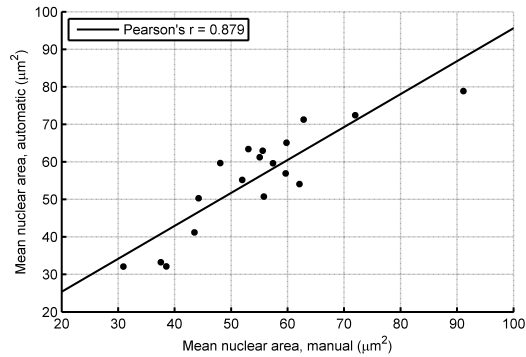


Figure 3.6: Scatter plot of the mean nuclear area as calculated by manual and automatic segmentation of nuclei.

to variation in the preparation of the samples that is within the “nominal range”. However, it should be noted that very poor sample preparation (such as very thick sections, overstraining, poorly fixed tissue etc.) or poor digitization (failed autofocusing, stitching artifacts etc.) can adversely affect the segmentation technique. Still, these problems rarely occur and can be remedied with a stricter quality control during the tissue preparation and slide scanning.

One possible point of improvement of our segmentation technique may be the inclusion of a pre-segmentation step that divides the tissue into epithelial and stromal regions. This would help to eliminate some of the false positives that arise in the stromal areas. Another improvement would be the use of a dedicated lymphocyte segmentation/detection procedure, as presented in Fatakdwala et al. (2010).

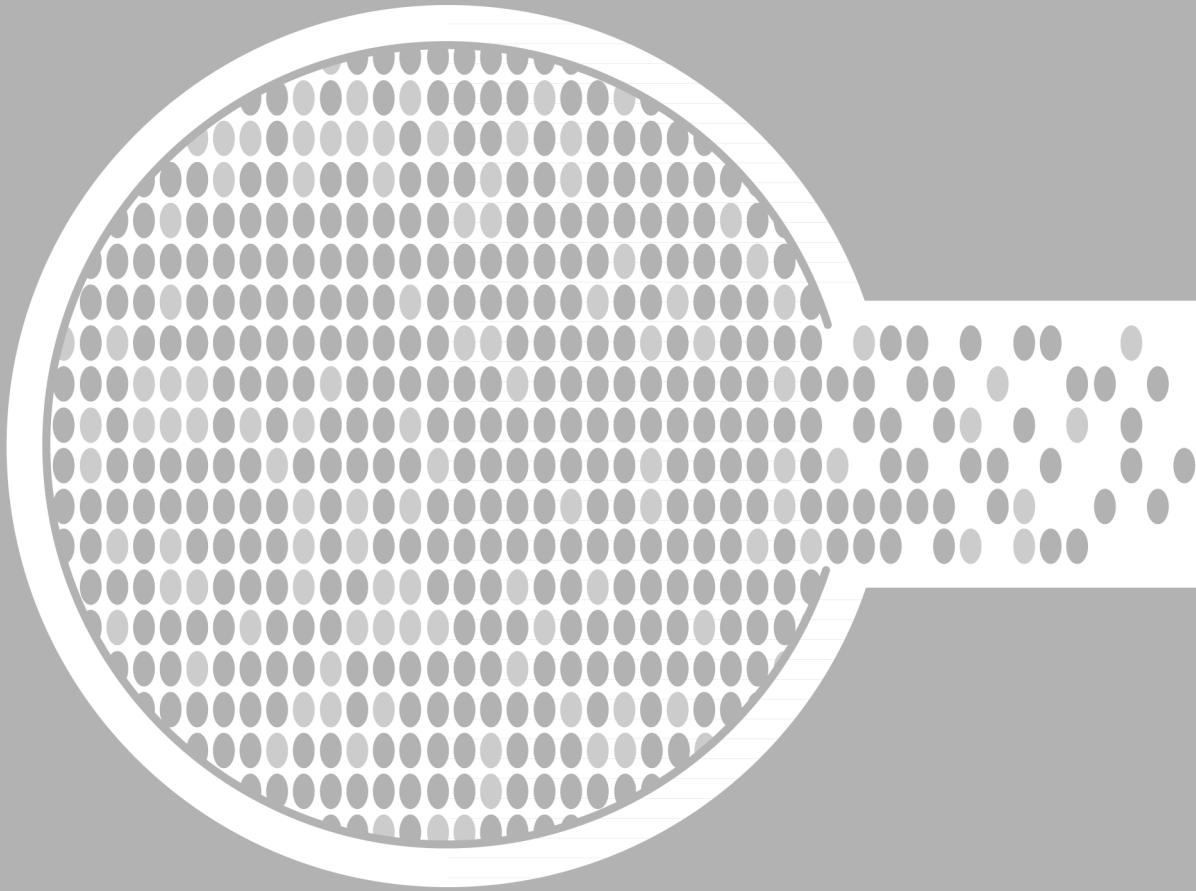
In our current work, we decided to concentrate on nuclear size features and nuclear architecture, because these are more robust with respect to the tissue preparation and staining processes compared with nuclear shape and chromatin texture features. For this purpose, elliptical approximations of the contours were sufficient. However, this approximation is a drawback when certain morphometric shape features need to be calculated. If computation of shape features is required, our segmentation algorithm can be extended to include an additional step of refining the contours.

The implementation of the method was done in MATLAB. The segmentation procedure for one image of size 1000×1000 pixels takes approximately 90 seconds on a PC with an Intel Core2Quad Q9500 processor. We note that this is only an experimental implementation, with processing times too slow for full slide segmentation, but further speed improvements are possible. In addition to this, tissue sampling methods (Belhomme et al. 2011; Kayser et al. 2009), and/or supervised extraction of relevant regions of interest (Gutiérrez et al. 2011; Romo et al. 2011) can be used in order to reduce the number of regions from the full slide that need to be processed, while still providing a

relevant result.

In another recent study (Veta et al. 2012) we have shown that the mean nuclear area (MNA) measurement extracted with the method presented in this chapter is a relevant prognostic marker in a cohort of 101 male breast cancer patients, outperforming the traditional nuclear pleomorphism score. Development of other prognostic markers, derived for example from analysis of the nuclear texture or architecture of the tissue is also a possibility. This analysis can potentially be done on whole slide images, which opens the possibility for integration into the workflow of routine pathology practice. Segmentation of nuclei can also be used, in a bottom-up manner, to locate the tumor regions within the slide or to assess the degree of tubule formation.

In conclusion, we have presented an accurate technique for automated segmentation of nuclei in images derived from digital slides of H&E stained breast cancer sections. The technique was evaluated on a number of representative regions and showed good performance in terms of detection and segmentation accuracy. This technique can be used to estimate prognostically relevant quantitative features such as MNA for breast cancer grading.



Prognostic value of nuclear morphometric features

Summary

Numerous studies have shown the prognostic significance of nuclear morphometry in breast cancer patients. Wide acceptance of morphometric methods has however been hampered by the tedious and time consuming nature of the manual segmentation of nuclei and the lack of equipment for high throughput digitization of slides. Recently, whole slide imaging became more affordable and widely available, making fully digital pathology archives feasible. In this study, we employ an automatic nuclei segmentation algorithm to extract nuclear morphometry features related to size and we analyze their prognostic value in male breast cancer.

The study population comprised 101 male breast cancer patients for whom survival data was available (median follow-up of 5.7 years). Automatic segmentation was performed on digitized tissue microarray slides and for each patient the mean nuclear area and the standard deviation of the nuclear area were calculated. In univariate survival analysis, a significant difference was found between patients with low and high mean nuclear area ($p = 0.022$), while nuclear atypia score did not provide prognostic value. In Cox regression, mean nuclear area had independent additional prognostic value ($p = 0.032$) to tumor size and tubule formation

In conclusion, we present an automatic method for nuclear morphometry and its application in male breast cancer prognosis. The automatically extracted mean nuclear area proved to be a significant prognostic indicator. With the increasing availability of slide scanning equipment in pathology labs, these kinds of quantitative approaches can be easily integrated in the workflow of routine pathology practice.

Based on: Mitko Veta et al. (2012). „Prognostic value of automatically extracted nuclear morphometric features in whole slide images of male breast cancer”. In: *Modern Pathology* 25, 1559–1565

4.1 Introduction

Histologic grading of breast cancer is liable to inter- and intra-observer variability and it has suboptimal reproducibility (Robbins et al. 1995) because of the subjective nature of the three components that constitute the grading system: nuclear pleomorphism, tubule formation and mitotic count. The usefulness of nuclear morphometry by image analysis in providing more objective and reproducible prognosis for breast cancer patients has been recognized for a long time (Baak et al. 1985; Chiusa et al. 2000; Cui et al. 2007; Kronqvist et al. 1998; Mommers et al. 2001; Pienta et al. 1991; Tan et al. 2001). Prognostically important features express the nuclear size and shape (and in some cases the nuclear texture and the architecture of the tissue (van Diest et al. 1992a)) in a quantitative manner. Incorporating nuclear morphometry features into grading would therefore make sense, but wider acceptance of such an approach to histological grading has been hampered by the tedious and time consuming nature of the manual segmentation of nuclei. Another contributing factor has been the lack of technology for high throughput digitization of histological slides. Recently, whole slide imaging (Al-Janabi et al. 2011; Pantanowitz et al. 2011) has become more affordable and thus more accepted into pathology labs, with the scanning and processing time constantly decreasing. Thus, fully digital pathology archives are already feasible (Huisman et al. 2010). This development of scanning equipment has in turn prompted the development of automatic image analysis methods of histopathology images that aim at reducing or completely eliminating the manual input to the quantitative analysis of the tissue (Gurcan et al. 2009).

In this study, we demonstrate the usefulness of automatic image analysis in breast cancer grading by employing a segmentation method to extract prognostically relevant morphometric features related to size from cancer nuclei in male breast cancer patients. This relatively rare type of cancer represents less than 1% of all breast cancers. Despite this, the mortality and morbidity of the disease are significant. Owing to the rare occurrence, large series are lacking and most of the knowledge is generalized from breast cancer in females. One previous study on 50 male breast cancer patients revealed that nuclear morphometry features from manually produced segmentations are predictive for the survival of the patients (Chiusa et al. 2000). In the work presented here, we extract two morphometric features, the mean nuclear area and the standard deviation of nuclear area, using a fully automatic segmentation method on whole slide images, and we analyze their prognostic value.

4.2 Materials and methods

4.2.1 Patients

The study population comprised 101 invasive male breast cancer patients. These are consecutive cases from the years 1986 to 2010 from four pathology labs in The Netherlands: St. Antonius hospital Nieuwegein, Diakonessenhuis Utrecht, University Medical Center Utrecht, and Laboratory for Pathology East Netherlands. This group of patients was previously used to analyze the molecular sub-typing, fibrotic focus and hypoxia in male breast cancer (Kornegoor et al. 2011; Kornegoor et al. 2012). The patients for whom survival data was not available were excluded in the current study. Age, tumor size and lymph node status were extracted from the pathology reports. Cases with isolated tumor cells were considered as lymph node positive. Hematoxylin and eosin (H&E) slides were reviewed by three experienced observers (PJvD, RK, AHJV-M) to confirm the diagnosis and to characterize the tumor. Histological type (WHO), tubule formation, nuclear grade, mitotic activity index (MAI) (van Diest et al. 1992b), and histological grade according to the modified Bloom and Richardson (Elston et al. 1991) score were recorded. The grades were assigned by consensus of the three observers in one microscope session. Prognostic information was obtained from the Integral Cancer Centre Netherlands. A summary of the clinicopathological data is given in 4.3.

4.2.2 Tissue preparation and slide scanning

Hematoxylin and eosin stained slides were used to identify representative tumor areas. From these areas three 0.6 mm punch biopsies from formalin-fixed and paraffin-embedded tissue blocks were obtained and embedded in a recipient paraffin block to produce a tissue microarray as described previously (Kornegoor et al. 2011). Sections of 4 μm were cut and stained with H&E. The tissue microarray slides were digitized using a ScanScope XT whole slide scanner (Aperio, Vista, CA, USA). The digitization was done at $\times 40$ magnification with a resolution of 0.25 $\mu\text{m}/\text{pixel}$. In order to reduce the storage requirements, JPEG2000 compression with high quality factor was used. The tissue microarray cores were manually extracted and stored as separate image files. Entire cores, or parts of them, were removed from the analysis in certain cases like folded tissue, non-tumor tissue, poor fixation and poor digitization due to failed autofocusing. The latter two negatively affect the automatic segmentation method. In total, 16 cores were completely removed and 16 cores were partially removed. This includes two patients for whom all three cores were completely removed.

Table 4.1: Summary of clinicopathological features of 101 male breast cancer patients.

Total number of patients	101
Age	
Median	67
Range	32-88
Tumor type	
ductal	90
lobular	3
mixed type (ductal/lobular)	2
invasive cribriform	1
papillary	1
mucinous	2
invasive micropapillary	1
adenoid cystic carcinoma	1
Tumor size	
< 2.0 cm	47
≥ 2.0 cm	51
unknown	3
Lymph node metastases	
absent	42
present	47
unknown	12
Nuclear atypia	
I (mild)	7
II (moderate)	60
III (severe)	34
Tubule formation	
I (> 75%)	9
II (10 – 75%)	41
II (< 10%)	51
Mitotic activity index (MAI)	
≤ 8	46
> 8	55
Bloom and Richardson grade	
I	21
II	44
III	36

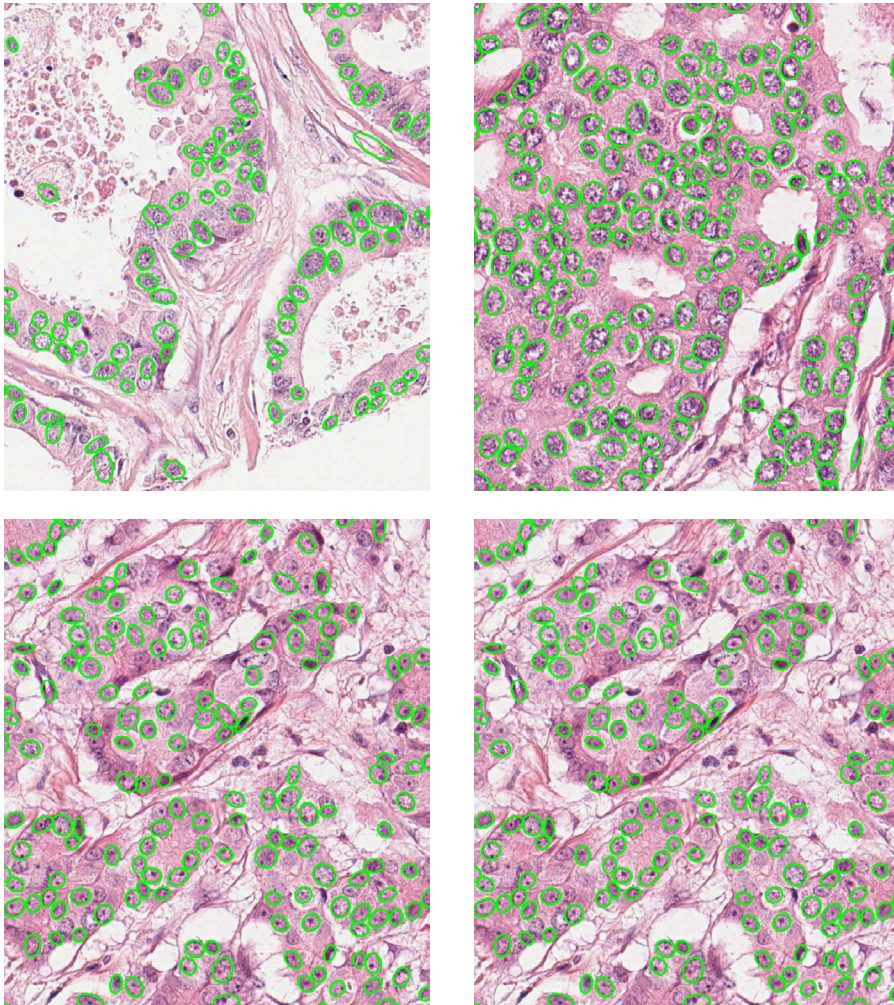


Figure 4.1: Example automatic nuclei segmentation results in male breast cancer histopathology images (all images are reproduced at the same scale).

4.2.3 Image analysis

For the task of automatic nuclei segmentation of the epithelial nuclei the method presented in Chapter 3 was used. Automatic nuclei segmentation was performed on all tissue microarray cores after which the mean and standard deviation of the nuclear area were calculated for all patients. The values were corrected to compensate for the systematic over- and under-estimation error with linear regression trained on the 39 regions mentioned before. Some example segmentation results are given in Figure 4.1.

4.2.4 Statistical analysis

A result was considered statistically significant if $p < 0.05$. Because the number of patients with score I for nuclear atypia, tubule formation and Bloom and Richardson grading system was very low, scores I and II were pooled. The mean and standard deviation of the nuclear area were correlated with clinicopathological features using the independent sample *t*-test.

The median follow-up time for the patients was 5.7 years, so all survival analysis was based on the 5-year survival rates. For the univariate survival analysis, patients were divided into groups of high and low mean and standard deviation of the nuclear area. The low group included the patients in the first two tertiles, whereas the high group included the last tertile. The rationale behind this kind of dichotomization was to establish an analogy with the nuclear and Bloom and Richardson grades, for which approximately one third of the patients were assigned a high grade. Tumor size and MAI were dichotomized using previously defined cut-off values (Kornegoor et al. 2012). Univariate survival analysis was done by plotting the Kaplan-Meier survival curves and performing the logrank test.

Features that proved significant in univariate analysis were entered in multivariate survival analysis using Cox's proportional hazards model (forward stepwise selection). Due to the relatively high median age of the patients, age was also taken as a covariate in multivariate survival analysis.

4.3 Results

The comparison of the mean and standard deviation of the nuclear area between patients grouped by clinicopathological features is summarized in Table 4.1. Significant differences were found for the mean nuclear area between patients with low and high nuclear atypia ($p = 0.032$), low and high MAI ($p = 0.011$) and Bloom and Richardson grades I and II versus III ($p = 0.007$). For the standard deviation of the nuclear area, significant differences were observed between patients with low and high MAI ($p = 0.014$) and Bloom and Richardson grades I and II versus III ($p = 0.047$).

Results from the univariate survival analysis are given in Table 4.3 and the survival curves according to the mean and standard deviation of the nuclear area are presented in Figure 4.2. Large tumor size ($p = 0.036$), low tubule formation ($p = 0.019$), high MAI ($p = 0.015$), high Bloom and Richardson grade ($p = 0.027$) and high mean nuclear area ($p = 0.022$) were associated with poor survival. Five-year survival rates for low and high mean nuclear area were 77% and 52%. The standard deviation of the nuclear area was not a significant predictor of outcome.

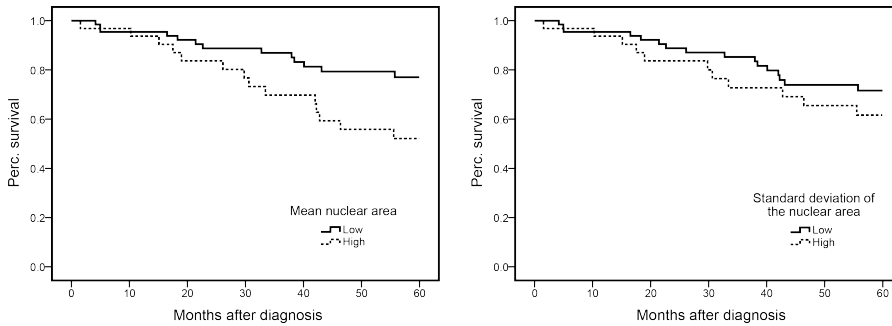


Figure 4.2: Kaplan-Meier survival curves for male breast cancer patients grouped by automatically extracted nuclear morphometry features.

In Cox proportional regression, tumor size ($p = 0.017$), tubule formation ($p = 0.035$) and mean nuclear area ($p = 0.032$) were retained as independent prognostic factors. The coefficients for the model are given in Table 4.4.

4.4 Discussion and conclusions

The goal of this study was to analyze the prognostic significance of automatically extracted nuclear morphometric features from whole slide images in male breast cancer patients. From the two examined features, only the mean nuclear area provided significant prognostic value. In contrast to others studies, the lymph node status was not a univariate prognostic predictor of survival. It should be pointed out that this remains the case even if the patients with isolated tumor cells ($n = 4$) are regarded as node negative. A significant difference in mean nuclear area was found between patients with low to moderate and high nuclear atypia as graded by an expert pathologist. This was expected since subjective assessment of the size of the nuclei is a major part of the nuclear atypia scoring. Interestingly, nuclear atypia score on its own was not a significant prognostic factor, which is in agreement with previous studies in female breast cancer (Abdel-Fatah et al. 2010). In Cox regression, automatically extracted mean nuclear area was an independent prognostic factor to tumor size and tubule formation. These results are in agreement with multiple studies that have shown the prognostic significance of nuclear morphometry in female and male breast cancer patients (Baak et al. 1985; Chiusa et al. 2000; Cui et al. 2007; Kronqvist et al. 1998; Mommers et al. 2001; Pienta et al. 1991; Tan et al. 2001). Altogether, this suggest that prognostication in male breast cancer could benefit from replacing classical nuclear atypia scoring according to the Bloom and Richardson grading system by automated nuclear morphometry of whole slide images.

Table 4.2: Association between automatically assessed mean and standard deviation of the nuclear area on whole slide images with classical clinicopathological features in male breast cancer patients (*t*-test).

Feature	Mean nuclear area (μm^2)	<i>p</i> -value	Standard deviation of the nuclear area (μm^2)	<i>p</i> -value
All patients	42.80 (8.98)		17.03 (4.34)	
Tumor size				
< 2.0 cm	43.02	0.884	16.84	0.802
\geq 2.0 cm	42.75		17.06	
Lymph node metastases				
absent	42.85	0.976	16.47	0.321
present	42.79		17.42	
Nuclear atypia				
I and II	41.26	0.032 ^a	16.32	0.065
III	45.37		18.02	
Tubule formation				
I and II	42.56	0.938	16.94	0.909
III	42.70		16.84	
Mitotic activity index				
\leq 8	40.18	0.011 ^a	15.75	0.014 ^a
$>$ 8	44.76		17.87	
Bloom and Richardson grade				
I and II	40.85	0.007 ^a	16.25	0.047 ^a
III	45.90		18.05	

^a Significant at the 5% level.

Table 4.3: Univariate survival analysis results of male breast cancer patients according to classic clinicopathological features and automatically extracted mean and standard deviation of nuclear area on whole slide images.

Feature	N	5-year survival	s.e.	<i>p</i> -value
Tumor size				
< 2.0 cm	47	0.79	0.07	0.036 ^a
≥ 2.0 cm	51	0.59	0.07	
Lymph node metastases				
absent	42	0.82	0.06	0.132
present	47	0.66	0.08	
Nuclear atypia				
I and II	67	0.71	0.06	0.430
III	34	0.62	0.09	
Tubule formation				
I and II	50	0.82	0.06	0.019 ^a
III	51	0.55	0.08	
Mitotic activity index				
≤ 8	46	0.81	0.06	0.015 ^a
> 8	55	0.57	0.07	
Bloom and Richardson grade				
I and II	65	0.77	0.06	0.027 ^a
III	36	0.53	0.09	
Mean nuclear area				
< 46.19 μm ²	66	0.77	0.06	0.022 ^a
≥ 46.19 μm ²	33	0.52	0.09	
Standard deviation of the mean nuclear area				
< 18.3 μm ²	66	0.72	0.06	0.328
≥ 18.31 μm ²	33	0.61	0.09	

^a Significant at the 5% level.

Table 4.4: Coefficients for the Cox proportional hazards model. Variables were encoded with 0 for the group with better prognosis and 1 with the group with worse prognosis. Age, MAI and Bloom and Richardson grade were not included in the model.

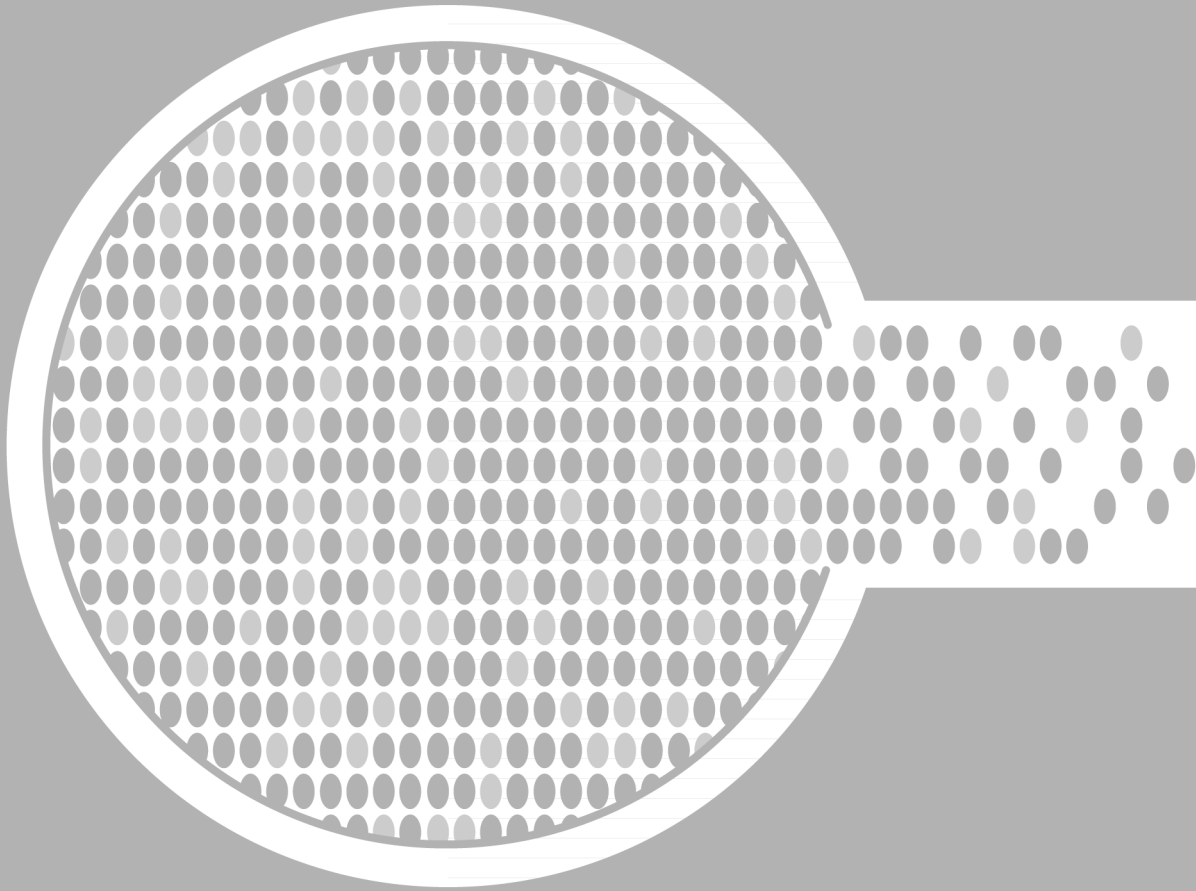
Feature	N	s.e.	<i>p</i> -value
Tumor size	1.060	0.445	0.017
Tubule formation	0.942	0.447	0.035
Mean nuclear area	0.858	0.399	0.032

Careful examination of the results revealed that many of the cases that have nuclei of small or intermediate size were given grade III for nuclear atypia due to the irregular chromatin texture and presence of large nucleoli and vice versa. Chromatin texture and presence of nucleoli are more easily evaluated by visual examination than nuclear size and thus, arguably, they are more influential in the grading process. In this study we did not consider automatically extracted nuclear texture features or nucleolar size, but this presents an interesting topic for future work.

As mentioned before, the recent advancements in slide scanning equipment and automatic image analysis methods of histopathology images may increase the use of quantitative methods in pathology. Software applications for analysis of immunohistochemically stained slides are already available from commercial vendors, some of which have approval by the USA Food and Drug Administration (Rojo et al. 2009). Such applications should be considered additional decision support tools for the pathologists; not overruling them (Hipp et al. 2011). One application example would be presenting the automated mean nuclear area to the pathologist at the time of grading a particular case, together with mean nuclear area values of reference cases and their nuclear atypia score. This additional quantitative input would help “steer” the decision for the nuclear atypia grading. These “hybrid” approaches, however, are yet to be examined in an experimental setting.

It should be pointed out that the automatic estimation has some drawbacks. Poorly fixed or poorly stained tissue, inclusion of regions with severe necrosis or lymphocytic infiltration, and failed autofocusing during the digitization may negatively affect segmentation performance, which in turn affects the estimation of the prognostic features. These situations, however, can be easily identified in a revision step by an experienced observer.

In conclusion, we here present an automatic method for nuclear morphology in male breast cancer grading. This approach using whole digital slides offers all the benefits of a quantitative method while eliminating the tediousness of the previous interactive methods. With the increasing availability of slide scanning equipment in pathology labs, this kind of quantitative approach can be easily integrated in the workflow of routine pathology practice.



CHAPTER

5

Assessment of algorithms for mitosis detection

Summary

The proliferative activity of breast tumors, which is routinely estimated by counting of mitotic figures in hematoxylin and eosin stained histology sections, is considered to be one of the most important prognostic markers. However, mitosis counting is laborious, subjective and may suffer from low inter-observer agreement. With the wider acceptance of whole slide images in pathology labs, automatic image analysis has been proposed as a potential solution for these issues.

In this chapter, the results from the Assessment of Mitosis Detection Algorithms 2013 (AMIDA13) challenge are described. The challenge was based on a data set consisting of 12 training and 11 testing subjects, with more than one thousand annotated mitotic figures by multiple observers. Short descriptions and results from the evaluation of eleven methods are presented. The top performing method has an error rate that is comparable the inter-observer agreement among pathologists.

Based on: Mitko Veta et al. (2014a). „Assessment of algorithms for mitosis detection in breast cancer histopathology images”. Submitted for review

5.1 Introduction

Breast cancer patients can considerably benefit from adjuvant therapy. However, aggressive adjuvant therapies are costly, can lead to potentially serious side effects and thus are only given to patients that are at a high risk. Assessing the patient risk requires use of good prognostic factors. In this regard, prognostic factors related to tumor proliferation have proven to be among the most powerful ones (van Diest et al. 2004).

The proliferation of cells occurs through a process that can be divided into several phases: resting phase (G0), first gap phase (G1), synthesis phase (S), second gap phase (G2) and mitotic phase (M). After the M-phase, the cells either enter the G0-phase or the G1-phase repeating the process. The cells that are in the M-phase can be visually determined under a microscope by their characteristic morphology. In hematoxylin and eosin (H&E) stained breast cancer sections, mitoses are discernible as hyperchromatic objects that lack a clear nuclear membrane and have therefore specific shape properties. Counting of mitotic figures in H&E stained sections is the oldest and still most widely used form of assessment of proliferation of breast cancer tumors by pathologists.

The proliferative activity of the tumor is estimated as the number of mitoses in an area of 2 mm², which corresponds to 8-10 microscope high power fields (HPFs, refers to the area that is visible using the microscope under very high magnification, usually $\times 40$) depending on the microscope model. This number is referred to as the mitotic activity index (MAI). Mitosis counting is routinely performed in pathology labs all over the world and is widely used as a prognostic factor. Although MAI assessment can be well reproducible if a strict protocol is followed after rigorous training (van Diest et al. 1992b), it is a subjective procedure that is liable to intra-observer variation. Several factors contribute to this. First of all, the task of identifying mitotic figures in H&E sections is not trivial. They can display a number of different appearances, with their hyperchromacy being the most salient feature. Moreover, many other objects can have a similar hyperchromatic appearance, such as apoptotic or necrotic nuclei, compressed nuclei, “junk” particles and other artifacts from the tissue preparation process. This makes the identification of mitoses difficult. Furthermore, the assessment of the proliferative state by counting mitotic figures is performed only in a small area of the tumor selected to be at the tumor periphery and to have high cellularity. The choice of the area is also a matter of subjective interpretation and one of the potential sources of low reproducibility.

In addition to being subjective, mitosis counting is a laborious task, compared with the assessment of other prognostic factors for breast tumors, such as nuclear pleomorphism and tubule formation. For a typical case, it takes

5-10 min. for a pathologist to perform mitosis counting, and the process must sometimes be repeated in different areas or different tumor slides for borderline cases.

In the last decade, pathology labs have started to move towards a fully digital workflow, with the use of digital slides being the main component of this process (Stathonikos et al. 2013). This was made possible by the introduction of scanners for whole slide imaging (WSI) that enable cost-effective production of digital representations of glass slides. In addition to many benefits in terms of storage and browsing capacities of the image data, one of the advantages of digital slides is that they enable the use of image analysis techniques that aim to produce quantitative features to help pathologists in their work. An automatic mitosis detection method with good performance could alleviate both the subjectivity and the tediousness of manual mitosis counting, for example, by independently producing a mitotic activity score or guiding the pathologist to the region within the tissue with highest mitotic activity.

This chapter gives an overview of the Assessment of Mitosis Detection Algorithms 2013 (AMIDA13) challenge that we recently launched. The main goal of the challenge was to evaluate and compare the performance of different (semi-)automatic mitosis detection methods that work on regions extracted from whole slide images on a large common data set. Since only the number of mitoses present in the tissue is of importance, i.e. their size and shape is not of interest, the challenge was defined as a detection problem.

5.1.1 Challenge format

The challenge was opened on March 28th, 2013, at which point interested groups or individuals could register on the challenge website and download the training data set that they could use to develop their methods. The training data set consisted of image data accompanied by ground truth mitosis annotations. Approximately two months after the release of the training data set a testing data set of similar size was released. The testing data set consisted only of image data, i.e. the ground truth annotations were withheld by the challenge organizers in order to ensure independent evaluation. After the release of the testing data set, the participants were able to run their methods and upload results to the challenge website for evaluation. The number of submissions from each registered participant was limited to three in order to avoid overfitting of the method to the testing data. Each submission had to be accompanied by a short method description, or, in the case it was the second or third submission of the participating team, a description of how the method differed from previous submissions. The submitted results were evaluated by the challenge organizers and the evaluation result was made available to the participants.

The first part of the AMIDA13 challenge was concluded with the workshop held in conjunction with the Medical Image Computing and Computer Assisted Interventions (MICCAI) 2013 conference on September 22nd, 2013 in Nagoya, Japan. The deadline for submissions that were to be presented at the challenge workshop and included in this overview was September 8th, 2013. After the conclusion of the workshop, the challenge website was reopened for new submissions.

Prior to the challenge workshop, over 110 teams or individuals from more than 30 countries registered to download the data set. Fourteen teams submitted results for evaluation. All submissions were automatic methods. Although the option was provided, none of the submissions was a semi-automatic method. This overview includes eight methods proposed by teams that submitted results for evaluation and attended the challenge workshop. Additionally, three methods by teams that submitted results for evaluation and achieved good performance but could not attend the challenge workshop are also included. The full list of results along with any new submissions after the challenge workshop is available on the challenge website.

5.1.2 Previous work

The earliest methods for automatic mitosis detection in breast cancer tissue date back to more than two decades ago (Beliën et al. 1997; Kaman et al. 1984; ten Kate et al. 1993). However, those approaches were inevitably constrained in their performance and potential use by the limited slide digitization technology and available computational resources at that time. The recent interest in this problem (Ciresan et al. 2013; Malon et al. 2012; Roux et al. 2013; Veta et al. 2013a) was ignited by the increased availability of WSI scanners. This resulted in the organization of the MITOS challenge in 2012 on the same topic, which was well attended and helped advance the state of the art for this problem. However, the challenge was based on a data set of relatively small size (5 slides in total, 10 annotated HPFs per slide), and it did not account for the inter-subject variability in tissue appearance and staining as regions of the same slides were included in both the training and testing data sets. We aim to address these issues with the data set used in this challenge.

For a wider overview of histopathology image analysis techniques for breast cancer and other tissue types, we refer the reader to the recent reviews in Gurcan et al. (2009) and Veta et al. (2014b).

5.2 Materials

In this section, we describe the process by which the challenge data set was compiled and annotated.

5.2.1 Patient, slide and region selection

The histology slides that were used for the creation of the challenge data set were produced at the Pathology Department of the University Medical Center Utrecht, Utrecht, The Netherlands. This is a pathology laboratory of medium size handling more than 144,000 surgical pathology slides and 12,000 cytology slides each year. From the archives of the department, slides from 23 consecutive invasive breast carcinoma patients, admitted between July 2009 and January 2010, were retrieved based solely on their availability (other selection criteria were not employed). All slides were prepared according to the standard laboratory protocol that consists of formalin fixation and paraffin embedding of the tissue, followed by cutting of 3-5 μm thick sections and staining with H&E.

One of the most difficult challenges in histopathology image analysis is the variability of tissue appearance, which is mostly the result of the variability in the conditions of the tissue preparation and staining processes. The challenge data set consists of slides that were routinely prepared at different time points during a longer period of time. In this way, it can be expected that appearance variability, which can be avoided when the tissues are processed in a single batch, will be reflected in the data set and the challenge will thus provide realistic performance estimates.

Data of a single patient typically consist of multiple slides. In clinical practice, the pathologist selects the slide and the region within the slide that is most suitable for the analysis at hand. Automating these selection steps is interesting by itself, and is certainly needed for a fully automatic workflow. However, for the challenge data set we decided to perform these steps manually and focus the challenge on the problem of mitosis detection.

An expert pathologist selected one representative slide per patient and marked a large region of the tumor on the glass slides in which mitosis annotation was to be performed. For the larger tumors, the marked areas within the slides were selected to encompass the most invasive part of the tumor, to be located at the periphery and to have high cellularity, which are the standard guidelines for performing mitosis counting. Smaller tumors were included in their entirety. The size of the outlined area varied from 7 mm^2 to 58 mm^2 with a median of 26 mm^2 . It should be noted here that this diverges from the routine pathology practice of estimating the mitotic activity as the number of mitotic figures within an area of 2 mm^2 . The choice of marking a larger area for mitosis annotation was made in order to ensure that a larger number of mitoses could be identified, which would result in data set of a size that is sufficient for training and evaluation of an automatic mitosis detection method.

5.2.2 Image acquisition

The representative regions were digitized with a ScanScope XT whole slide scanner (Aperio, Vista, CA, USA). This scanner model can perform the steps of tissue selection, patch focus point selection, calibration, image acquisition and compression in a fully automatic manner. The digitization of the candidate regions was done at $\times 40$ magnification with a spatial resolution of $0.25 \mu\text{m}/\text{pixel}$. The automatic focus points were manually reviewed before single focus plane scanning to reduce the possibility of blurry patches. High quality JPEG2000 compression (quality factor 85) was used to store the images, almost completely eliminating any visible compression artifacts, which was confirmed with side-by-side comparison of compressed and uncompressed versions of several regions. The slide digitization parameters were optimized to ensure the highest image quality possible and differs from the standard practice of digital slide archiving at the UMC Utrecht (Huisman et al. 2010).

5.2.3 Ground truth annotation

The ground truth for the challenge was assigned based on the annotations by multiple observers, to reduce the influence of observer variability. We used the following protocol to establish the ground truth mitosis annotations.

1. Two pathologists independently traversed the indicated regions of interest and marked the locations of mitoses by drawing an ellipse encompassing the object in the whole slide image viewer;
2. The objects on which both pathologists agreed were directly taken as ground truth mitosis objects;
3. The discordant objects (annotated as mitoses by only one of the observers) were presented to a panel of two additional pathologists to make the final decision. Note that the panel did not traverse the slides but only examined the discordant objects.

With this annotation protocol, all objects that were accepted as ground truth mitoses have been agreed upon by at least two experts.

The first set of annotations was done by a pathologist at the UMC Utrecht using the ImageScope whole slide viewer (Aperio, Vista, CA, USA). The second set of annotations was done by an external pathologist using pathoconsult.nl — an online digital slide viewing and collaboration platform maintained by the UMC Utrecht. The second observer was blinded to the results from the first observer. The observers did not receive a standardized definition for mitotic figures, but were instructed to perform the mitosis counting using the criteria they employ in daily practice.

The total number of annotations made by the first and second observers was 1088 and 1599, respectively. The number of locations upon which they agreed was 649, which left 1389 annotations to be resolved by the panel of two additional observers. The panel revisited all discordant objects and after discussion, decided together which objects to accept as ground truth mitoses. The total number of remaining objects after the consensus annotation was 1157 (this number also includes the concordant objects that were directly accepted as ground truth).

We note here that although the difference in the absolute counts between the two initial observers is quite large (1088 compared to 1599), a large portion of it can be traced back to only a few subjects (Table 5.1). Later investigation into this showed that this discrepancy can be largely attributed to the difference in the interpretation of objects that are difficult to interpret without fine focusing ability. It is also reasonable to assume that some of the difference can be attributed to the inter-institutional differences in mitosis counting.

5.2.4 Image data distribution

For simplicity, the image data for the challenge was not distributed as whole slide images, but instead, each whole slide image was divided into a number of smaller TIFF images that could be easily read by most image analysis platforms. We chose the size of the TIFF images to be 2000×2000 pixels, representing an area of 0.25 mm^2 , which is in the order of one microscope high power field (the exact area of one HPF is different for different microscope models). We refer to the TIFF images as HPFs. HPFs that were outside the marked region of interest on the slide or that intersected the black marker annotation were excluded (Figure 5.1). Since for some subjects the total number of HPFs was very high (in the order of several hundreds), only the HPFs that contained at least one mitosis were included as part of the data set. For the subjects that had fewer than 10 HPFs with mitosis occurrence, additional empty HPFs were included to extend the total number to 10 in order to include sufficient background image data, necessary for good training and evaluation. Note that some mitotic figures fell into a HPF that intersects the marker annotation and are thus not included in the data set. This was the case for around 6% of the annotations.

The set of 23 subjects was split into two subsets — one intended for training of the methods and one used as an independent testing data set. The division into training and testing data sets was done in such a way that the number of HPFs and ground truth mitoses was approximately balanced. A summary of the two data sets is given in Table 5.1. The ground truth was provided to the participants only for the training data set, in the form of coordinates of mitoses (the centres of the elliptical annotations) for each HPF.

Table 5.1: Data set summary.

ID	Outlined area (mm ²)	Number of annotated objects by observer 1	Number of annotated objects by observer 2	Number of HPFs in the challenge data set	Number of ground truth mitotic figures
Training dataset					
1	20	60	188	39	73
2	36	51	32	28	37
3	41	13	46	16	18
4	24	178	338	61	224
5	20	5	8	10	6
6	43	104	89	61	96
7	22	55	140	43	68
8	21	2	5	10	3
9	58	0	12	10	2
10	26	0	4	10	0
11	20	5	25	13	15
12	41	9	5	10	8
Testing dataset					
1	7	2	4	10	3
2	36	1	37	15	16
3	40	69	73	49	66
4	21	11	10	10	9
5	27	8	9	10	6
6	26	277	247	67	212
7	7	0	2	10	2
8	26	0	5	10	0
9	31	111	155	44	115
10	34	93	133	48	72
11	23	34	32	22	32

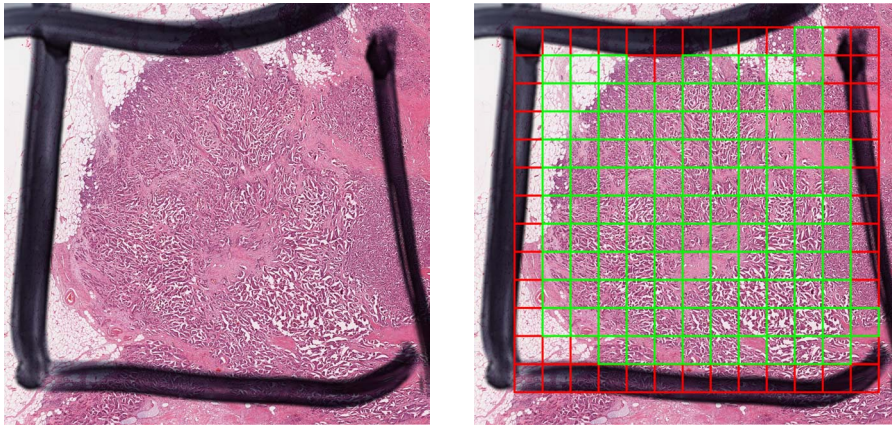
5.2.5 Object-level observer agreement

Given the notorious difficulty of the mitosis detection task even for expert pathologists, the performance of the automatic mitosis detection methods should be evaluated within the context of the inter-observer agreement.

The similarity of two sets of annotations can be expressed using the Dice overlap coefficients, which is computed as:

$$D(A, B) = \frac{2|A \cap B|}{|A| + |B|} \quad (5.1)$$

where A and B are the two sets of annotations and $|\cdot|$ indicates the number of elements in the sets.



(a) Example slide with the area for annotation indicated with a black marker.

(b) Each rectangle is one HPF. HPFs that intersect the black marker lines (given in red) are not included in the data set.

Figure 5.1: Separation into high power fields (HPFs).

On the entire data set of 23 subjects, prior to the separation and selection of HPFs, the Dice overlap coefficient was 0.483 (i.e., there were 2687 annotations by the two observers in total and they agreed for 649 objects). For the HPFs in the training data set, the Dice overlap coefficient between the sets of annotations by the two independent observers was 0.527, and for the testing data set 0.566. These numbers are higher¹ compared to the value for the entire data set due to the selection of non-empty HPFs (prior to separation into HPFs and rejection) based on the consensus annotation.

The Dice overlap coefficients between the individual annotations and the consensus annotation are 0.749 and 0.763 for the first and second observer respectively for the selected HPFs in the training data set, and 0.796 and 0.773 for the selected HPFs in the testing data set. It should be noted however that this is a biased measure since the consensus annotation is partly based on the two independent annotations.

¹Note that the annotation procedure led to rejection of some HPFs that had only discordant annotations by the two initial observers that were then in turn all rejected by the consensus annotation. This produces higher values for the Dice overlap when only the selected HPFs are considered.

5.3 Methods

5.3.1 CCIPD/MINDLAB²

Preprocessing: In addition to the three color channels from the RGB input images, four additional color channels were computed and used for candidate segmentation and feature extraction: L from the LAB color space, V and L from the LUV color space and the blue ratio image (ratio of the blue color channel and the sum of the other two channels in the RGB color space).

Candidate detection and segmentation: For each HPF, a set of candidate regions was defined by thresholding the blue ratio image.

Feature extraction and classification: This method fuses two classification strategies: a feature learning method based on Convolutional Neural Networks (CNN), and a set of handcrafted features combined with a random forests (RF) classifier. For each candidate region, both learned features and handcrafted features were extracted independently, and then classified using the corresponding classifier. The CNN model has 4-layer architecture, including two consecutive convolution-pooling layers, a fully-connected layer and a softmax classification layer. It operates on 80×80 pixel patches in the YUV color space centered at the candidate regions. The first 3 layers comprise 64, 128, and 256 neurons, respectively. For the convolution-pooling layers, fixed 8×8 convolutional kernel and 2×2 pooling kernel were applied. The CNN was trained using stochastic gradient descent. In addition, morphological, statistical and texture features were extracted for each of the seven color channels. Principal component analysis (PCA) was applied to reduce the dimension of the extracted features by retaining 98.5% of the principal components. Using this reduced representation, a cascade of two RF classifiers with 100 random trees was trained. To balance the numbers of mitosis and non-mitosis objects, the number of non-mitosis nuclei was reduced by eliminating overlapping objects and oversampling the positive class. The final prediction score is a weighted average of the outputs of the two classifiers. More details about this method can be found in (Wang et al. 2014).

5.3.2 DTU³

Preprocessing: Candidate detection was performed on the blue ratio image, calculated as the ratio of the blue color channel and the sum of the red and green channels.

²CCIPD at Case Western Reserve University, USA and MindLab at National University of Colombia

³Technical University of Denmark

Candidate detection and segmentation: Candidate detection was performed by thresholding of the Gaussian of Laplacian blob detector applied to the blue ratio image, followed by connected component labelling.

Feature extraction and classification: For each detected candidate object, a 100×100 pixel patch was extracted and each RGB color channel was independently normalized to have values in the range $[0, 1]$. From the color normalized patches, three types of image features were extracted:

1. Color intensity histograms, one for each color channel;
2. Gradient orientation histograms. The orientations of the gradients are calculated relative to a vector from the cell center to the location of the gradient;
3. Shape index histograms (Larsen et al. 2014). The shape index captures second-order differential structure from the local Hessian eigenvalues. The two eigenvalues are mapped to a continuous interval providing a smooth and intuitive transition between the second-order shapes (cup, rut, saddle, ridge and cap).

Each image feature was computed for different concentric donut-like spatial pooling regions centered on the candidate object. The donut-shaped pooling regions vary in radius and width such that they capture image structure in different parts of the candidate object. Note that the features are rotationally invariant because both the image features and the spatial decomposition are rotationally invariant. The extracted image features were used to train a support vector machine (SVM) classifier with radial basis function (RBF) kernel.

5.3.3 IDSIA ⁴

In this approach, Multi Column Max-Pooling Convolutional Neural Networks (MCMPCNN) are used for supervised pixel classification. MPCNNs alternate convolutional layers with max-pooling layers. A similar technique won the MITOS mitosis detection competition (Ciresan et al. 2013) and recently produced outstanding results in image classification (Ciresan et al. 2012a) and segmentation (Ciresan et al. 2012b).

The inputs to the MPCNN are 63×63 pixel patches directly sampled from the raw RGB images. The output is the probability that the central pixel of the patch is within 20 pixels of the centroid of a mitosis. Three networks with identical 10-layer architecture were trained on 20 million patches extracted

⁴IDSIA, Dalle Molle Institute for Artificial Intelligence, USI-SUPSI, Lugano, Switzerland

from the training images. One tenth of such instances were randomly sampled from mitosis pixels (which represent a tiny fraction of all pixels in the training data set); 40% were randomly sampled from all non-mitosis pixels; the remaining 50% were sampled only from non-mitosis pixels that were found to be similar to mitosis (therefore, more challenging to classify) during a simple preprocessing phase. The resulting training data set was augmented by adding rotated and mirrored instances. Each network was trained for a maximum of five epochs, which required about three days of computation using an optimized GPU implementation. Mitoses in the test images were detected by sliding the neural network detectors over the images by means of an efficient algorithm (Giusti et al. 2013). This resulted in a map where each pixel represented the probability of belonging to the mitosis class. Each test image was processed in eight different rotation/mirroring combinations by each of the three networks. The 24 resulting probability maps were averaged, and then convolved with a 20-pixel radius disk kernel. Nonzero values obtained after performing non-maxima suppression in a 50-pixel radius corresponded to detected centroids of mitotic figures.

5.3.4 ISIK⁵

This method is an extension of previous work on mitosis detection (Tek 2013).

Preprocessing: Prior to candidate detection, image contrast stretching was performed.

Candidate detection and segmentation: Candidate objects were initially segmented by a morphological double threshold operation. The resulting binary image was filtered by an area opening operation (with the minimum area set to 50 pixels), to remove isolated spurious regions. The candidate extraction procedure was finalized with a morphological hole filling step.

Feature extraction and classification: The candidate object classification consists of two separate stages. In the first stage, a set of simple object features were employed to significantly reduce the number of false objects, while keeping the loss of true mitotic figures to a minimum. The following set of features was used in this stage: area, major- and minor-axis lengths, perimeter, equivalent diameter, ratio of the area to the perimeter, eccentricity, extent, mean intensities of the three RGB channels and the ratios of the mean intensities of the three RGB channels to the corresponding means of all candidate components of the same image.

⁵Department of Computer Science and Engineering, Isik University, Istanbul, Turkey

In the second stage, which operates on candidate objects that have not been eliminated in the first stage, a set of windows of increasing width around the candidate object were defined. For each window, a feature vector that is formed of five different groups of features is calculated: color (statistics of the RGB color channels and similarities to the average mitosis and non-mitosis histograms), binary shape, Laplacian, morphological (area granulometry) and gray-level co-occurrence. For both stages of the classification, an ensemble of multi-stage AdaBoost classifiers was used.

5.3.5 MINES⁶

Preprocessing: Separate hematoxylin and eosin channels were obtained with color unmixing (deconvolution) (Ruifrok et al. 2001).

Candidate detection and segmentation: The segmentation of candidate objects was performed entirely using the hematoxylin channel. In order to detect potential nuclei (candidates), a diameter closing operation (Walter et al. 2007) was applied to the median filtered image removing all dark structures with maximal extension smaller than a predefined parameter. By calculating the difference to the median filtered image, these small dark structures could be segmented by simple double thresholding. The connected components of this binary image were considered candidate objects.

Feature extraction and classification: The candidate segmentation procedure failed to identify only 2 mitoses in the training data set. With the aim to reduce the high number of false positives, a supervised classifier was trained. For each candidate object, shape and texture features (basic geometric and gray level features, Haralick features, statistical geometric features, morphological granulometries, convex hull features, etc.), as defined in (Walter et al. 2010), were computed. In addition, Haralick and basic gray level features were calculated for the candidate region in the eosin channel and for a ring around each candidate region in the hematoxylin channel in order to quantify the local environment.

A training data set of three classes was built: non-mitosis, early mitosis (prophase/prometaphase) and late mitosis (metaphase, anaphase). The rationale of transforming the binary classification problem into a three-class problem was that the two mitosis classes are morphologically very different and some preliminary runs showed that this strategy gives better results (for this particular problem). In order to distinguish the three classes, an SVM

⁶Centre for Computational Biology - Mines ParisTech, Institut Curie and U900 INSERM, Paris, France

classifier was trained (RBF kernel, parameters automatically determined by a grid search with 5-fold cross validation). The mitotic cells were taken as the union of the early mitosis and the late mitosis class obtained from this classifier.

5.3.6 NTUST⁷

In this approach, a diverse cascade learning framework (Wang et al. 2010b) with the cwBoost learning algorithm (Wang et al. 2010a) is used for supervised pixel classification. The hierarchical ensemble classifier contains 10 layers of simple mitosis detectors, which evaluate various types of inputs with different models, and can quickly filter out negative areas. A similar technique was used for obscured human head detection in video sequences (Wang et al. 2010b). The inputs to the learning methods were 50×50 pixel patches directly sampled from image data extracted from the red color channel. The red color channel was chosen because in preliminary tests it was found to give the best cross-validation accuracy compared to the other color channels. Ten layers of boosting ensembles were trained on image patches extracted from the training images. Each ensemble contained 10 C4.5 decision tree classifiers (Quinlan 1996). Mitoses in the test images were then detected by sliding the hierarchical boosting detectors over the images. The output from the detector was the probability that the central pixel of the patch is within 25 pixels of the centroid of a mitosis. A confidence weight is generated by computing the number of detections in the same area, and regions with a weight greater than 2 are defined as possible mitosis regions.

5.3.7 PANASONIC⁸

Preprocessing: The RGB images were first transformed into a number of different color spaces that later facilitated the candidate segmentation and feature extraction: L*a*b, HSV, blue ratio image (BR), red ratio image (RR) and blue-red ratio image (BRR). The blue ratio image, which accentuates the nuclear dye, was computed as the ratio of the blue channel and the sum of the other two channels. The red ratio and red-blue ratio image were computed in a similar manne

Candidate detection and segmentation: Candidate mitosis regions were extracted by binary thresholding of the BR image. The threshold was auto-

⁷Graduate Institute of Biomedical Engineering, National Taiwan University of Science and Technology

⁸Panasonic Healthcare Co., Ltd.

matically determined as three times the standard deviation of the BR image. Regions that were smaller than 80 pixels were eliminated.

Feature extraction and classification: The following morphological features were computed for each candidate region: area, major axis length, minor axis length, eccentricity, orientation, convex area, filled area, equivalent diameter, solidity, extent, and perimeter. In addition to this, for each candidate region, a rectangular window was defined, and the following features were extracted within the window:

1. Histogram of local binary patterns (LBP) for the BR, RR, BRR, L^* , H, S, and V images. The LBP features were computed for three radii (1, 3 and 5 pixels), and the histograms were concatenated;
2. Haralik features (contrast, correlation, energy, and homogeneity) for the BR, RR, BRR, L^* , H, S, and V images;

The candidate objects were classified as mitoses or non-mitoses using a random forest classifier.

5.3.8 POLYTECH/UCLAN ⁹

Preprocessing: The candidate detection was performed in the blue corrected images, which were intensity adjusted to calibrate both the image contrast and the average intensity, partially compensating for the differences in tissue appearance.

Candidate detection and segmentation: Candidate objects were detected by thresholding and binary morphological operations. Patches of 128×128 pixels were then extracted around the centroid of each candidate object and subsequently used for feature extraction and classification.

Feature extraction and classification: The features used for classification of the candidate objects were selected to represent both textural and shape information. The first set of 10 features was extracted from the average of the run-length matrices calculated in four directions $0^\circ, 45^\circ, 90^\circ, 135^\circ$ (Irshad et al. 2013). The second set of eight features was extracted from the average of the co-occurrence matrices for the same four directions and includes: energy, entropy, correlation, difference moment, inertia, cluster shade, cluster prominence and Haralick's correlation. To capture spatial information, each patch was divided into seven rings and a central circle, and for each region an

⁹University Nice - Sophia Antipolis, France and University of Central Lancashire, UK

eight bin intensity histogram was calculated giving in total 64 features. The final feature that was calculated is the area of the segmented candidate region by binary thresholding. The OpenCV implementation of random forests was used for classification. Due to the highly imbalanced training data set, additional positive (mitosis) patches were randomly selected within a small neighborhood of the ground truth mitosis location.

5.3.9 SURREY¹⁰

Preprocessing: To compensate for the variability of the tissue staining and preparation, the images were first aligned in color space. This was done using histogram matching, with the mean histogram from the whole training data set as the target, and the histograms of the individual subjects as the input. Each color channel was adjusted independently. The computed histograms excluded pixels that belong to regions that do not contain tissue (i.e., white regions), found by thresholding of the green channel.

Candidate detection and segmentation: Candidate mitosis locations were detected based on color. Each color channel was quantized to 64 levels, and these values were used to address a 3-dimensional (RGB) lookup table that points to the likelihoods of the color being present in a mitotic figure. The color lookup table was defined based on the histograms of 10-pixel circular neighborhoods of ground truth locations.

After obtaining a likelihood map for an input image, it was low-pass filtered and thresholded, and the centers of the connected component regions were taken as candidate locations. Around each candidate location, 70x70 pixel patches were extracted and converted to grayscale. Up to two largest objects within the patch were segmented by a threshold that provides the best combination of high boundary gradient and low variance within the object(s). Objects that had area and contrast with the background below pre-defined thresholds were removed. When a pair of objects was segmented, it was ensured that they had roughly the same area and intensity.

Feature extraction and classification: For each candidate, a set of rotation invariant features reflecting the shape, contrast, edge properties and texture of both the segmented object and the background was calculated. In addition, pairs were characterized by the ratio and average of a subset of parameters from each of the objects. For classification, RBF SVM classifier was used with dominant class subsampling and model averaging, to deal with the class imbalance.

¹⁰Centre for Vision, Speech and Signal Processing (CVSSP), University of Surrey, UK

5.3.10 SHEFFIELD/SURREY ¹¹

This method requires a minimal input in its design and models the space of mitosis images using a low-dimensional manifold. An advantage of manifold learning models is that they enable practitioners to easily visualize the range of mitosis appearances.

The preprocessing and candidate extraction steps of this submission are the same as the ones described in SURREY, but instead of computing a set of predefined features, the normalized gray-level candidate patches were represented as vectors and modeled as observations by a Bayesian Gaussian Process Latent Variable Model (BGPLVM) (Titsias et al. 2010). This method learns a low dimensional latent space that is mapped nonlinearly back to the original space of observations. In addition, it enables the computation of an approximate density function of novel samples given the known image samples. Therefore, one BGPLVM was used for the positive samples (mitotic cells) and another for the negative samples (false-positive candidate objects). The two classes were assumed to be independent and classification was done using maximum likelihood. Visual inspection of the models (by reconstructing samples in different positions of the latent space) showed that images reconstructed from the positive model were sharper than those of the negative samples, i.e, there was a higher appearance variation among negative samples. Furthermore, by navigating through the positive latent space, the reconstructed images showed a smooth transition between different phases of mitosis as well as between different types of mitosis appearances (including tripolar mitosis that occur in cancer cells). However, the classification results were relatively poor, probably because the assumption of independence between the classes does not hold, indicating that a method that jointly models samples and labels can be more promising (e.g. using Manifold Relevance Determination (Damianou et al. 2012)).

5.3.11 WARWICK ¹²

Preprocessing: Staining normalization by non-linear color mapping (Khan et al. 2014) was performed in order to neutralize the inherent variation in the color of the staining.

Candidate detection and segmentation: Candidate objects were extracted by statistical modelling of the pixel intensities of the b-channel from the Lab color space. The pixel intensities from mitosis regions were modeled by a

¹¹Department of Computer Science and Sheffield Institute for Translational Medicine, University of Sheffield, UK and Centre for Vision, Speech and Signal Processing (CVSSP), University of Surrey, UK

¹²University of Warwick, UK

Gamma distribution and those from non-mitosis regions by a Gaussian distribution (Rajpoot et al. 2013). Maximum-likelihood estimation was employed in order to estimate the unknown parameters of the distributions. The posterior probability map was then binarized to identify candidate objects. The threshold value for obtaining the binary map was selected by performing receiver operating characteristic (ROC) curve analysis. All candidate objects with an area between 40 and 500 pixels were considered candidate objects. This range of areas was obtained by examining the size of the ground truth mitoses in the training data set.

Feature extraction and classification: For each candidate object, a set of object level color, shape and texture features was computed. In addition, a small context window around the candidate objects was defined, and used to compute contextual features (first order statistics over a set of textural features). These contextual features were combined with the object features to train a classifier. Since the number of non-mitosis candidate objects was disproportionately higher compared to mitosis candidate objects, the classification was performed with the RUSBoost classifier (Seiffert et al. 2010), which combines under-sampling and boosting to handle the class imbalance problem. Other classifiers (such as SVM, LDA) were also evaluated, but their cross-validation performance on training data proved to be lower than RUSBoost.

5.4 Evaluation

A detected object was considered to be a true positive if the Euclidean distance to a ground truth location is less than $7.5 \mu\text{m}$ (30 pixels). This value corresponds approximately to the average size of mitotic figures in the data set, and provides a reasonable tolerance for misalignment of the ground truth location and the detection. When multiple detections fell within $7.5 \mu\text{m}$ of a single ground truth location (eg., when two components of a single mitotic figure were detected separately), they were counted as one true positive. All detections that were not within $7.5 \mu\text{m}$ of a ground truth location are counted as false positives. All ground truth locations that do not have detected objects within $7.5 \mu\text{m}$ were counted as false negatives.

For each proposed method two types of evaluation measures relating to the detection accuracy were computed. The first evaluation measure was the overall F_1 -score, where all ground truth objects were considered a single data set regardless to which subject they belong. The F_1 -score was defined as the harmonic mean of the precision (Pr; positive predictive value) and the recall (Re; sensitivity):

Table 5.2: Overall F_1 -scores of the proposed methods. To compute the overall F_1 -score, all ground truth objects were considered a single data set regardless to which subject they belong.

IDSIA	0.610	0.612	0.611
DTU	0.427	0.555	0.483
SURREY	0.357	0.332	0.344
ISIK	0.306	0.351	0.327
PANASONIC	0.336	0.310	0.322
CCIPD/MINDLAB	0.353	0.291	0.319
WARWICK	0.171	0.552	0.261
POLYTECH/UCLAN	0.186	0.263	0.218
MINES	0.139	0.490	0.217
SHEFFIELD/SURREY	0.119	0.107	0.113
NTUST	0.011	0.685	0.022

$$F_1 = \frac{2Pr \times Re}{Pr + Re} \quad (5.2)$$

The overall F_1 -score is dominated by the subjects that have a high number of mitotic figures. To complement this measure, individual F_1 -scores for each subject were also calculated. For the submissions that included probability estimates associated with the detections, the precision-recall (PR) curves were also computed.

After a visual inspection of the detection results, it was observed that many of the false positives produced by the top performing methods closely resemble mitotic figures. Indeed, owing to the difficulty of the task it is possible that some mitotic figures were missed during the ground truth annotation, but were then detected by the automatic methods. To further examine this, the false positives from the top two methods, which had notably better performance than the remaining methods, were presented to a panel of two observers for re-annotation, along with the ground truth mitoses as a control. This panel consisted of one of the observers that initially traversed the slides and one of the observers that participated in the panel that resolved the discordant objects. Separate images of size 200×200 pixels, centered at the false positive detections and the ground truth objects, were extracted and presented in random order to the new panel for re-annotation, i.e., labelling as mitosis or non-mitosis.

The mitotic activity of the tumors was ultimately expressed as the density of mitotic figures, i.e., the number of mitoses per tissue area. To evaluate the performance of the methods for this task, the correlation coefficient between the number of detections and the number of ground truth mitoses per HPF for the subjects in the testing set was computed.

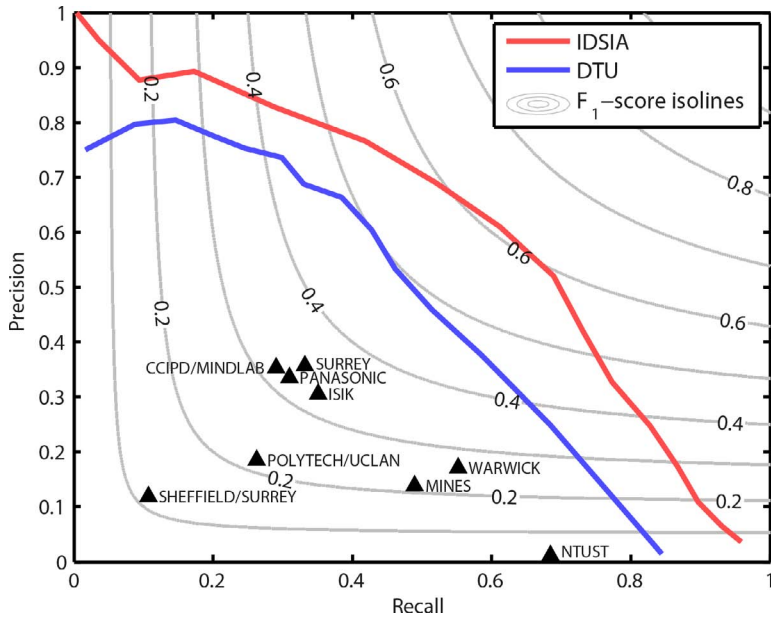


Figure 5.2: Precision-recall curves of the two top ranking methods. The performance of the other methods is plotted for comparison.

5.5 Results

5.5.1 Mitosis detection

The overall F_1 -scores along with the precision and recall of the proposed methods are summarized in Table 5.2. The top ranking method is IDSIA with an overall F_1 -score of 0.611. The individual F_1 -scores for each subject are summarized in Table 5.3, along with the average across all subjects. Note that subject #8 from the testing data set has zero annotated ground truth mitotic figures, thus the F_1 -score is undefined. Instead of the F_1 -score, for this subject, Table 5.3 contains the number of false positive detections. According to this evaluation measure, the top ranking method is again IDSIA with an average F_1 -score of 0.445.

Figure 5.2 gives the PR curves for the IDSIA and DTU methods. The performance of the other methods is also plotted on the same graph for comparison.

Note that these results are based on the original ground truth data and are not influenced by the re-annotation of false positives.

Table 5.3: Individual (per subject) and average F_1 -scores of the proposed methods. To compute the individual F_1 -scores, every subject was considered a separate dataset

Team name	1	2	3	4	5	6	7	8 ^a	9	10	11	Average F_1 -score
IDSIA	.50	.00	.66	.57	.29	.73	.00	0	.46	.60	.64	0.445
DTU	.00	.11	.41	.56	.20	.63	.00	0	.35	.61	.68	0.352
WARWICK	.13	.34	.14	.59	.40	.62	.00	3	.09	.38	.34	0.302
ISIK	.00	.10	.37	.44	.00	.47	.00	1	.16	.26	.46	0.226
PANASONIC	.00	.00	.28	.20	.40	.52	.00	0	.10	.30	.33	0.213
CCIPD/MINDLAB	.00	.15	.36	.38	.00	.42	.00	0	.07	.38	.33	0.208
SURREY	.00	.10	.48	.30	.00	.47	.00	2	.14	.19	.38	0.205
MINES	.00	.14	.11	.40	.16	.49	.00	1	.07	.32	.34	0.203
POLYTECH/UCLAN	.00	.00	.07	.18	.00	.50	.00	0	.04	.30	.39	0.148
SHEFFIELD/SURREY	.03	.07	.35	.15	.00	.05	.00	54	.09	.08	.18	0.099
NTUST	.01	.03	.01	.09	.03	.23	.02	29	.01	.05	.21	0.068

^a Subject #8 has no ground truth mitotic figures, thus the F_1 -score is not defined. For this subject, the number of false positives is given in the table. This subject was excluded when computing the average F_1 -score.

5.5.2 Re-annotation of false positives

The results from the re-annotation experiment are given in Figure 5.3. The proportion of false positives from the IDSIA method re-annotated as mitotic figures was $p = 0.29$, 95% CI [0.23, 0.36]. In other words, out of the 208 false positives from the IDSIA method, 61 were re-annotated as true mitotic figures. The proportion of false positives from the DTU method re-annotated as mitotic figures was $p = 0.16$, 95% CI [0.12, 0.19]. In other words, out of the 397 false positives from the DTU method, 62 were re-annotated as true mitotic figures. This is illustrated in Figure 5.3(a).

The proportion of objects re-annotated as mitotic figures from the entire set of detections produced by the IDSIA method was $p = 0.61$, 95% CI [0.57, 0.65]. This means that out of the 534 detections that this method produced, 326 were re-annotated as true mitotic figures. The corresponding proportion for the DTU method was $p = 0.42$, 95% CI [0.39, 0.46] (293 out of 693 detected objects). For comparison, the proportion of objects re-annotated as mitotic figures from the ground truth dataset was $p = 0.71$, 95% CI [0.67, 0.75] (379 out of 533 objects). This is illustrated in Figure 5.3(b).

5.5.3 Number of mitoses per HPF

The scatter plots for the number of detections and the number of ground truth objects per HPF for the four methods with highest performance according to the overall F_1 -score is given in Figure 5.4. The plots for the other methods are

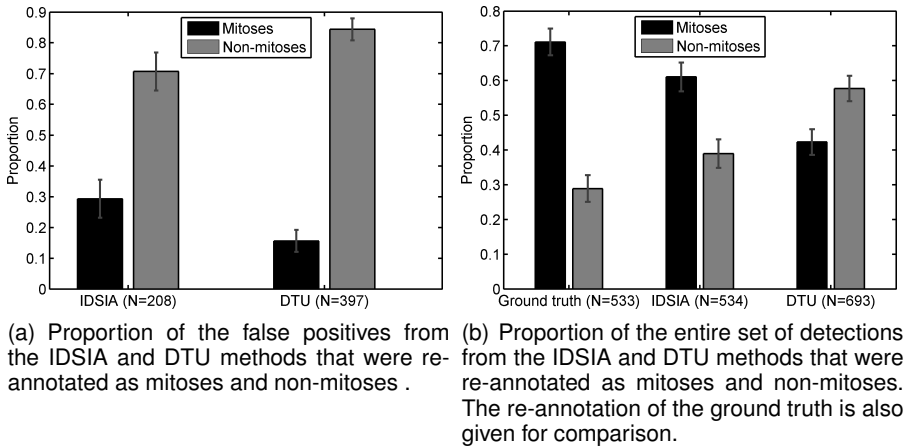


Figure 5.3: Results from the re-annotation of the false positives.

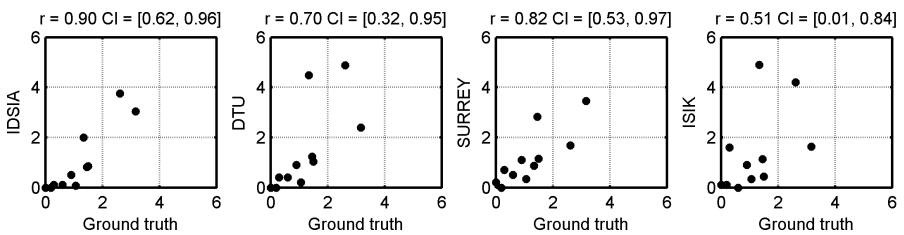


Figure 5.4: Scatter plots for the estimated and ground truth number of mitoses per HPF for the first four methods with highest overall F_1 -score.

omitted for brevity. The best correlation was achieved by the IDSIA method ($r = 0.90$, 95% CI [0.62, 0.96]).

Tables with more detailed results are available for download from the challenge website.

5.6 Discussion

5.6.1 Summary of the proposed methods

The majority of the proposed methods followed a two-step object detection approach. The first step identified candidate objects that were then classified in the second step as mitoses or non-mitoses. Some of the proposed methods prior to the candidate extraction and classification steps performed transformation of the color channels (DTU, PANASONIC, POLYTECH/UCLAN) or staining unmixing (MINES) to obtain a nuclear/hematoxylin channel, thus eliminating eosinophilic structures that can hamper the detection perform-

ance. Three of the proposed methods (SURREY, SHEFFIELD/SURREY and WARWICK) perform explicit staining normalization to tackle the problem of staining variability.

The most popular technique for candidate extraction was thresholding of a grayscale image in combination with linear filtering or morphology operators. This technique worked relatively well because of the hyperchromacity of the mitoses — by selecting only the “darkest” nuclei in the images as candidates, a large number of the non-mitoses can be rejected while achieving high sensitivity for the mitosis class. Three of the proposed methods (SURREY, SHEFFIELD/SURREY and WARWICK) used a supervised method to obtain a mitosis likelihood map, which was then thresholded in order to obtain candidate objects.

In the second step, features were computed for segmented mitosis candidates and/or image patches centered at the detected candidate locations. The size of the image patches varied from 63×63 to 128×128 pixels. These patch sizes were selected to be somewhat larger than the expected size of mitotic figures in order to capture contextual information. Spatial pooling was employed by some methods (DTU, POLYTECH/UCLAN) to capture information about structures in different regions of the candidate objects.

A variety of different generic color, texture and shape features, with emphasis on rotational invariance, was employed by the different methods. The CCIPD/MINDLAB method used a combination of classification based on hand-crafted features and a feature learning method (convolutional neural networks), and SHEFFIELD/SURREY models the set of observations with manifold learning.

For classification, RBF SVMs (DTU, SURREY, MINES) and random forests (CCIPD/MINDLAB, PANASONIC, POLYTEC/UCLAN) were the most commonly used classifiers, with some methods employing different boosting techniques (ISIK, WARWICK, NTUST). The problem of class imbalance was addressed by subsampling of the dominant negative class or oversampling of the class of mitotic figures.

The IDSIA and NTUST methods did not perform candidate extraction, but instead, evaluated the detector for every pixel location. IDSIA used a very efficient implementation of deep convolutional neural networks to obtain a mitosis probability map for each image, from which mitoses were detected by non-maxima suppression.

5.6.2 Performance of the proposed methods

The best performing method according to all evaluation measures was IDSIA. The overall F_1 -score of this method was comparable to the inter-observer agreement among pathologists. The DTU method also achieved solid per-

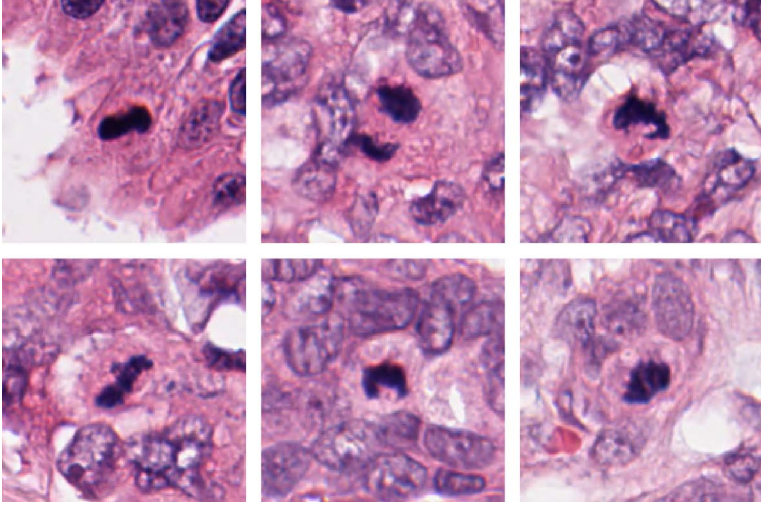
formance according to the overall and average F_1 -scores. The performance of these two methods was notably better than that of the remaining methods. The analysis of the PR curves (Figure 5.2) indicated that this is not related to the choice of the operating point of the detectors (the points indicating the performance of the other methods are in the interior of the areas spanned by the curves).

The majority of the false positives produced by the methods with lower performance were dark objects that lack other characteristics of mitotic figures such as protrusions around the edges. On the other hand, many of the mitotic figures that had less intensive staining were not correctly detected. This can be either explained by the fact that the texture and shape features used lack discriminative ability and do not capture these fine structural details or by the fact that the mitotic figures with less intensive staining were underrepresented in the training set. Examples of the most commonly detected and missed mitotic figures are given in Figure 5.5.

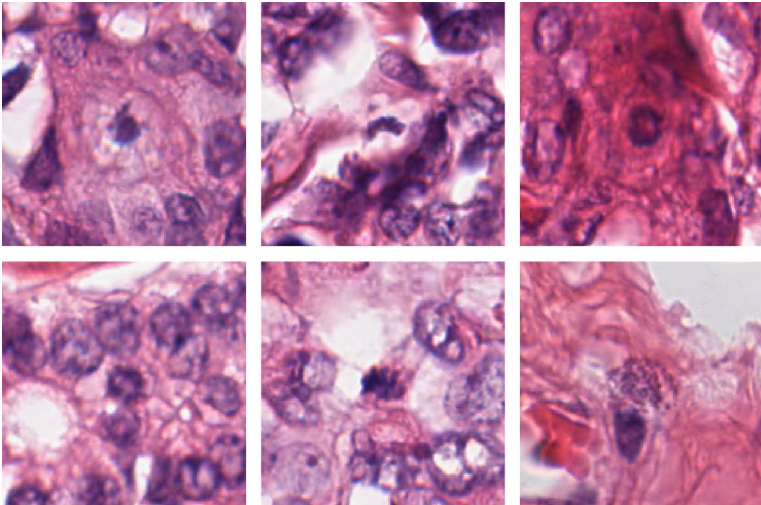
The results from the re-annotation experiment indicate that a large portion of the “false positives” from the IDSIA method can in fact be considered true mitotic figures (Figure 5.3). They may have been missed during the ground truth annotation because of the intricacy of the task and the observer variability. In addition, the distribution of the assigned labels during the re-annotation of the original ground truth set and of the set of detections from this method was very similar.

Good correlation between the ground truth and estimated number of mitoses per HPF was achieved even for some proposed methods that have lower overall and average F_1 -scores (Figure 5.4). This indicates that estimation of the mitotic activity index or the mitotic activity grade might be possible with lower object-level detection performance or even from global image features. This represents an interesting subject for future research. It should be also pointed out that the number of mitoses per HPF does not correspond to the estimated MAI of the outlined regions in the slides because empty HPFs were removed when forming the challenge data set.

Experiments with combining the results from the different methods by majority voting or intersection of the better performing methods did not show improved results over the best individual method. One of the conclusions of the discussion during the challenge workshop was that the variation in the staining appearance is one of the major obstacles for mitosis detection. After the workshops, attempts to improve the IDSIA and DTU methods were made by incorporating explicit staining normalizations that showed promising results. However, due to the preliminary nature of the experiments we chose not to include the results here.



(a) Mitotic figures that were detected by most (at least ten) of the proposed methods.



(b) Mitotic figures that were not detected by any of the proposed methods.

Figure 5.5: Examples of the most commonly detected and missed mitotic figures.

5.6.3 Performance for individual subjects

The worst performance in terms of the F_1 -score was achieved for subjects #1 and #7. They both have very low mitotic activity, and the few ground truth mitoses have atypical appearance (lack of hyperchromacity). The methods that have good overall performance produce a very low number of false positives for these two subjects, suggesting that they would produce a low mitotic activity estimate.

Another subject for which very poor performance was achieved is #2. Note that for this subject there was large discrepancy between the numbers of objects indicated as mitoses by the two independent observers (Table 5.1). This indicates that there is an intrinsic difficulty in identifying mitotic figures in this particular case. Case #9 contained many dark nuclei that are not mitotic figures, thus the methods that extensively rely on the staining intensity as a features had very low performance.

The best performance was achieved for subject #6 that had very high mitotic activity and most of the hyperchromatic objects indeed represented mitotic figures.

5.6.4 Feasibility of mitosis counting on whole slide images

Digital slides are still not widely accepted as primary diagnostic modality pending validation studies, with the main concern being the image quality and lack of fine focusing ability (Al-Janabi et al. 2011). In the context of breast cancer histopathology grading, the image quality of whole slide images is principally sufficient for the scoring of nuclear atypia and tubule formation, which together with mitosis counting constitute the commonly used modified Bloom-Richardson grading system (Elston et al. 1991). However, the task of mitosis counting is a more delicate one. Detailed examination is required to distinguish mitoses from other mitosis-like objects, which requires the use of fine focusing on the conventional microscope. This feature is missing in whole slide images captured with a single focal plane. Although whole slide imaging scanners that support slide digitization at multiple focal planes are becoming increasingly available, the use of this feature is still rather limited due to the increased time and storage requirements. Taking this into account, there is a possibility of discrepancy between the mitotic activity as estimated by light microscopy and on unifocal whole slide images.

This challenge compared the performance of human experts and computer algorithms on digital slides. It should be emphasized that both the automatic algorithms and the expert observers worked with the same input. The patient and slide selection was done in a way that captures the intra-laboratory variability of the tissue preparation. However, it is likely that the inter-laboratory

variability is greater due to suppliers of reagents, different staining protocols and different scanners. One possible extension of this challenge is inclusion of data from multiple centers.

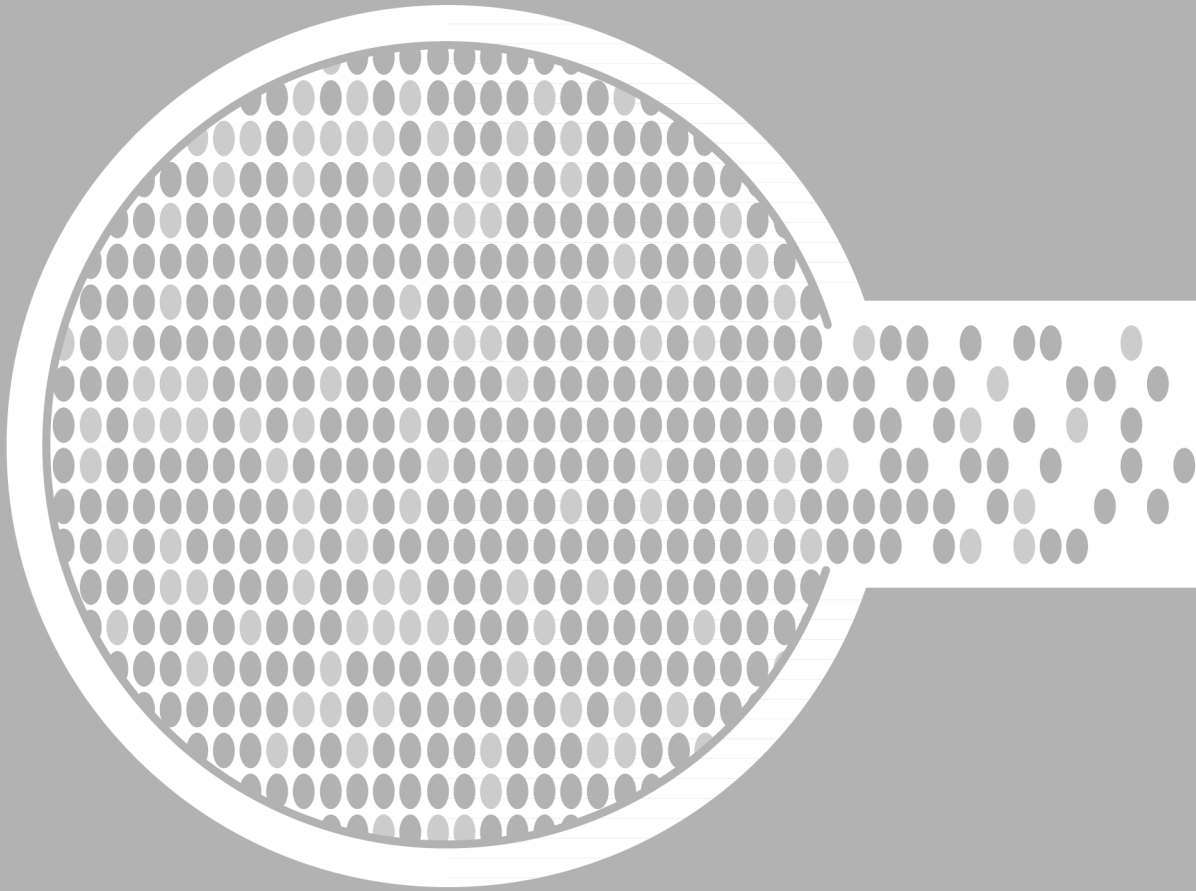
At the time when we started the work on annotating the challenge data set, studies that examined the feasibility of mitosis counting on digital slides were lacking. The qualitative impression of pathologists that we interviewed was that the image quality of digital slides imaged at $\times 40$ magnification and a single focal plane is sufficient for mitosis counting in most cases, but for some instances it can be difficult to make a firm decision if an object represents mitosis without the possibility to fine tune the focus. We performed a small scale internal validation study and concluded that there is generally a good correlation between the assessment of the mitotic activity index by light microscopy and on whole slide images (Stathonikos et al. 2013). In a recent larger study involving multiple observers and a large number of subjects it was found that the scoring of mitotic activity on whole slide images is as reliable as on conventional glass slides viewed under a microscope (Al-Janabi et al. 2013).

As an alternative to expert annotation, we are currently investigating the use of Phosphohistone H3 (PPH3) labelling of mitotic figures to produce ground truth mitotic figure locations in a more objective manner. These ground truth locations can then be registered to conventionally stained H&E slides. The initial results look promising, but we are still optimizing the procedure.

5.7 Conclusions

In this chapter, we summarized the proposed methods and results from a challenge workshop on mitosis detection in breast cancer histopathology images. The challenge data set consisted of 12 subjects for training and 11 for testing, both with more than 500 annotated mitotic figures by multiple observers. In total 14 teams submitted methods for evaluation, 11 of which are described here. The best performing methods achieved an accuracy that is in the order of inter-observer variability.

Our intention is for this challenge to be ongoing with incremental extensions of the training and testing data sets. By keeping the challenge website (<http://amida13.isi.uu.nl>) open for new submissions, we hope to keep a record of the state of the art of mitosis detection in breast cancer histopathology images.



CHAPTER

6

Automatic detection of mitotic figures

Summary

This chapter describes the development and evaluation of an automated method for mitosis detection in breast cancer histopathology images. The proposed mitosis detection method follows a two stage approach. In the first stage, candidate mitosis objects are detected and segmented based on the staining intensity and in the second stage the candidate objects are classified as mitotic figures or non-mitoses using a set of size, shape, color and texture features. A staining normalization procedure is employed to reduce the influence of the staining variability. The method was evaluated on the AMIDA13 dataset and achieved good results compared to other methods on the same dataset. Two baseline models, one omitting the staining normalization and one omitting texture features were evaluated for comparison. They achieved worse results compared to the full model, which illustrates the added value of the staining normalization and the texture features.

Based on: M. Veta et al. (2013a). „Detecting mitotic figures in breast cancer histopathology images”. In: *SPIE Medical Imaging*. Vol. 8676, 867607–867607-7

6.1 Introduction

In this chapter, we present a mitosis detection method that is an extension and improvement of the one published in Veta et al. (2013a). That study was performed prior to the Assessment of Mitosis Detection Algorithms 2013 (AMIDA13) challenge using the same image data but with ground truth from only a single observer. Here we repeat the experiments with the AMIDA13 dataset and make several extensions and improvements of the method. Following the conclusions from AMIDA13, we examine the added value of texture features, compared to using size, shape and intensity/color features only. In addition, we analyse the influence of employing a staining normalization procedure on the detection results.

6.2 Materials and methods

6.2.1 Dataset

The dataset used in this study consists of 23 cases divided into two subsets, one with 12 cases for training and another one with 11 cases for testing of the detection method. The cases are represented with a number of images (at least 10 per case) each corresponding to an area of 0.25 mm^2 or one microscope high power field (HPF). The spatial resolution of the images is $0.25 \mu\text{m}/\text{pixel}$. The ground truth is provided in the form of ground truth mitotic figure locations for each HPF. The annotation was performed by two observers that independently traversed the slides. Discordant objects were resolved by a panel of additional two observers. In total there are 1157 ground truth mitotic figures, 550 of which are in the training set. More details about the patient selection, slide digitization and annotation are given in Chapter 5.

6.2.2 Overview of the method

The proposed mitosis detection method follows a two stage approach. In the first stage, candidate mitosis objects are detected and segmented and in the second stage the candidate objects are classified as mitotic figures or non-mitoses. Prior to the candidate detection, a staining normalization procedure is employed to reduce the influence of the staining variability on the detection.

6.2.3 Staining normalization

One of the main difficulties in automatic image analysis of hematoxylin and eosin stained tissue is the appearance variability. This is mainly the result of variations in the tissue preparation and staining processes that are difficult to control. Consequently, even tissue sections that are prepared in the same

lab but at different time points, as is the case with this dataset, can exhibit substantial variability of the staining intensity and color appearance. One of the ways to tackle this issue is by performing staining normalization prior to the analysis. For this task, we use the method outlined in Macenko et al. (2009). In short, this method consists of the following steps:

1. For each image in the dataset, stain-specific absorption coefficients (staining vectors) are automatically determined;
2. Using the automatically determined absorption coefficients each image is decomposed into two staining concentration maps (one for each constituent stain) by performing staining unmixing (Ruifrok et al. 2001);
3. For each image, the staining concentration distributions are normalized to have the same maximum value across the dataset;
4. The images are recomposed from the normalized concentration maps using common staining absorption coefficients.

The third step of this method normalizes the intensity of the staining across the dataset and the fourth step normalizes the color appearance of the hematoxylin and eosin stains. The stain-specific absorption coefficients for each image are automatically determined based on the distribution of the optical density vectors of the image pixels (ODs). More details can be found in Macenko et al. (2009) and Vink et al. (2013).

6.2.4 Candidate detection and segmentation

The most salient feature of mitotic figures is their hyperchromicity. Mitoses are stained noticeably darker than non-dividing epithelial nuclei, however, other objects such as lymphocytes and necrotic nuclei can also have hyperchromatic appearance. Nevertheless, by identifying the “darkest” objects in the images, most mitoses can be successfully detected along with a limited number of false positives.

We detect candidate objects by performing non-minima suppression in a radius R on the staining-normalized red color channel. Prior to performing non-minima suppression, the image is filtered with a Gaussian low-pass filter with standard deviation σ to eliminate small structures that can be regarded as noise. Only local minima below a threshold T are considered to be candidate objects. By varying this threshold value, different combinations of true and false positive rates can be obtained, which can be used to plot the free-response receiver operating characteristic (FROC) curve.

Following the detection, candidate objects are segmented by the level-set method proposed in Chan et al. (2001). This method fits a piecewise constant model to the image, and can segment objects whose boundaries are not

necessarily defined by a strong gradient or are discontinuous. It works by minimizing the following energy functional:

$$F(C, c_1, c_2) = \mu \cdot \text{Length}(C) + \lambda_1 \cdot \int_{\text{inside}(C)} |I(x, y) - c_1|^2 dx dy + \lambda_2 \cdot \int_{\text{outside}(C)} |I(x, y) - c_2|^2 dx dy \quad (6.1)$$

In the previous equation, I is the image function, C denotes the segmentation curve and c_1 and c_2 are the average intensity of I inside and outside C . The parameters $\mu > 0$, $\lambda_1 > 0$ and $\lambda_2 > 0$ are fixed and defined by the user. The parameter μ controls the scale of the segmented objects (smaller μ means that smaller objects will be segmented and vice versa) and λ_1 , λ_2 are the fit penalties for the inside and outside of the curve. For each candidate location, the level-set segmentation was performed in a local rectangular support region centered at the local minimum.

6.2.5 Feature extraction and classification

The candidate classification component aims at classifying all extracted candidates as being a mitotic figure (or part of a mitotic figure) or a false object. To achieve this we trained a statistical classifier with a set of size, shape, color and texture features. Mitotic figures often exhibit specific structures at the border regions, usually in the form of protrusions. In addition, the immediate context of mitotic figures is free from other hyper-chromatic objects in most cases, and can contain other potentially useful information. To capture this information, for each segmented candidate region, two additional regions are defined in which features are extracted: boundary region and neighborhood region (Figure 6.1).

Following the candidate detection and segmentation, the locations of the detected objects are compared to the ground truth to produce labels. In concordance with the evaluation criteria of the AMIDA13 challenge, all objects that are within 30 pixels of a ground truth location are labelled as belonging to the positive class. Note that multiple objects can be matched to a single ground truth location.

For each candidate object, the following features were extracted:

1. *Size and shape features*: Area, perimeter, compactness, eccentricity, solidity and sphericity of the segmented candidate region.
2. *Color features*: Mean, standard deviation, skewness, kurtosis, median, minimum and maximum value of the three color channels from the RGB color space of the segmented candidate region and the two additional regions.

3. *Texture features*: The texture features were extracted by recording the responses from a set of Gaussian, Laplacian of Gaussian, edge and bar filters (first and second order Gaussian derivatives), similar to the MR8 filter bank described in Varma et al. (2005) (Figure 6.1). The images were processed at six scales for each filter type. The scales for the Gaussian and Laplacian filters were chosen to be $\sigma = \{1, \sqrt{2}, 2\sqrt{2}, 4, 4\sqrt{2}\}$. The scales for the edge and bar filters, which are anisotropic, were chosen to be $(\sigma_x, \sigma_y) = \{(1, 3), (\sqrt{2}, 3\sqrt{2}), \dots, (4\sqrt{2}, 12\sqrt{2})\}$. Since the Gaussian derivative filters are oriented, to achieve orientation invariance the filtering is performed at six orientations uniformly spread between 0 rad and π rad, with the maximum response for each pixel taken as the filter output. The mean, standard deviation, skewness, kurtosis, median, minimum and maximum value of the responses from all filters for each of the three color channels, and for each of the three regions are used as texture features.

The total number of computed features is 1581. Many of the texture features, however, are highly correlated, and thus contain redundant information. To reduce the dimensionality of the set of texture features, principal component analysis (trained on the training set) is performed prior to the classification, retaining only the principal components that account for 99% of the variance.

For classification, we used the L_2 -regularized linear support vector machine classifier as implemented in the LIBLIENAR library (Fan et al. 2008). This implementation supports class weighting, which can be used to address the class imbalance problem (there are many more objects from the non-mitosis class than mitoses). In initial cross-validation experiments on the training set, this classifier proved to give the best performance among several examined classification models including k -NN and non-linear support vector machines.

6.2.6 Parameter selection and training

We used an H&E stained slide that was not part of the mitosis detection dataset to define reference maximum concentrations and staining absorption coefficients for the staining normalization method. Prior to performing candidate detection, segmentation and feature extraction, all images in the dataset were normalized. A staining normalization example using this method is given in Figure 6.2. It can be observed that after the normalization, both images have similar intensities and color, which can also be seen from the OD distributions.

The parameters R and σ for the candidate detection were optimized on the training set by performing exhaustive search for the ranges $R =$

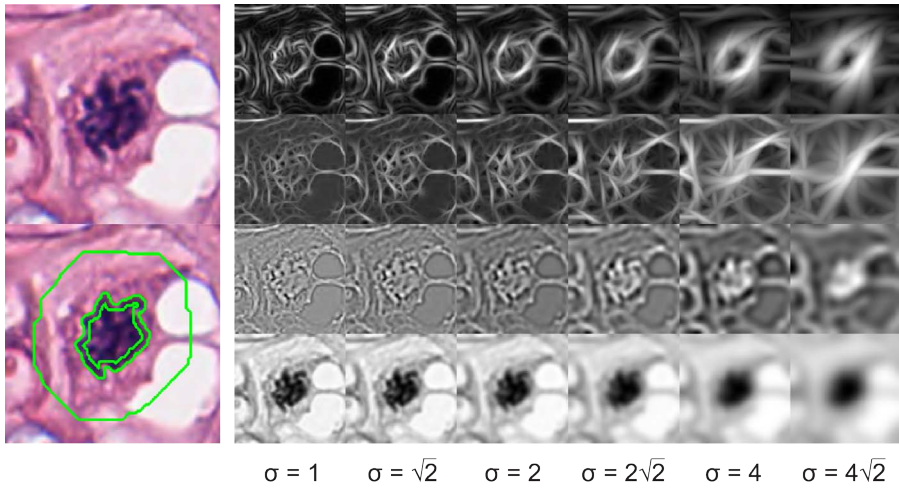
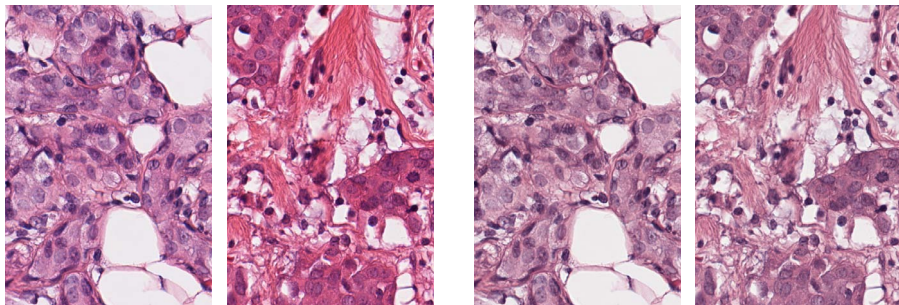
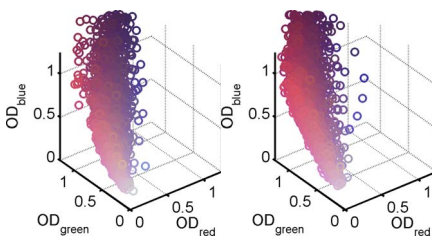


Figure 6.1: Example texture feature extraction. Left: A segmented mitotic figure with two additionally defined regions. Right: Filter responses for different scales. From top to bottom, the filter responses correspond to the edge, bar, Laplacian of Gaussian and Gaussian filters.

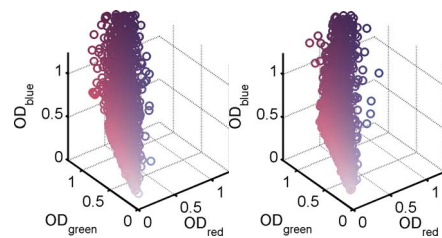


(a) Original images.

(b) Staining-normalized images.



(c) Distribution of the OD values of the original images.



(d) Distribution of the OD values of the staining-normalized images.

Figure 6.2: Staining normalization example.

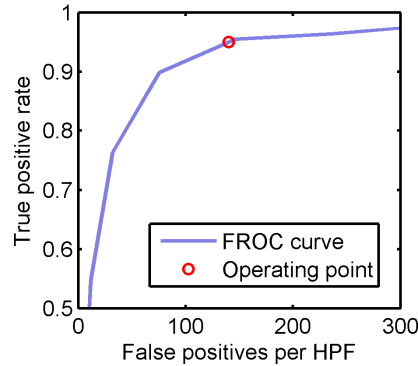


Figure 6.3: Candidate detection FROC curve for $(R, \sigma) = (45, 2)$. The threshold for the selected operating point is $T = 70$.

$\{7.5, 15, 30, 45, 60\}$ pixels and $\sigma = \{1, 2, 4, 8, 10, 12\}$ pixels. For each combination of parameters (R, σ) , the free-response ROC curve was plotted by varying the threshold T . The optimal set of parameters was chosen to be the one that produced the lowest number of false positives for a target true positive rate of 95%. This was determined to be $(R, \sigma) = (45, 2)$ pixels, which produces approximately 150 false candidates per HPF at 95% TP rate for a threshold $T = 70$ (Figure 6.3).

Note that by using this set of parameters and threshold value, an estimated 5% loss of mitotic figures after the candidate detection stage is introduced in the system. This, however, provides a good compromise between the number of detected true objects and false positives that are input to the segmentation and candidate classification stages.

Experiments using the hematoxylin channel obtained by staining unmixing for candidate detection instead of the red color channel were also performed but this approach did not provide superior results.

For each candidate, the level-set segmentation was performed in a local rectangular support region of 120×120 pixels ($30 \times 30 \mu\text{m}$) around the detected location. The segmentation was initialized with a circular region with a diameter of 15 pixels ($3.75 \mu\text{m}$). The parameters μ , λ_1 and λ_2 were empirically set to be 0.2, 5, and 1, respectively. The segmentation results proved to be robust with respect to the choice of parameters and initialization, as very similar results were obtained for a wide range of values. Example segmentation results using the chosen parameters are given in Figure 6.4.

The boundary region that aids the feature extraction was defined as a band with a width of 3 pixels ($0.75 \mu\text{m}$) on the interior boundary of the segmented candidate region. Similarly, the neighborhood region was defined as a band with a width of 20 pixels ($5 \mu\text{m}$) outside of the segmented candidate region. The size of these additional regions was empirically set based on the expected size of mitotic figures and their distance to neighboring objects.

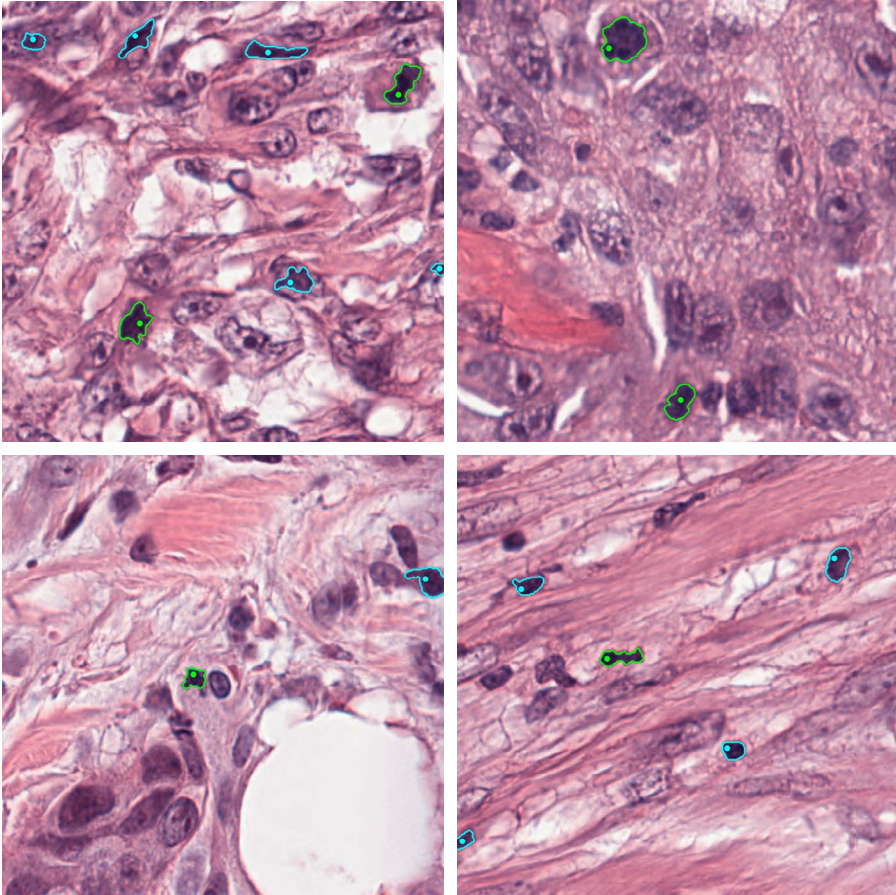


Figure 6.4: Example segmentation results. The dots indicate the detected local minima. Ground truth mitotic figures have green outlines, the other objects are false positives. All images are shown at the same scale and are staining-normalized.

The candidate detection and segmentation generated 47955 non-mitoses and 529 mitosis candidates that were used to train the classifier. The regularization parameter C for the L_2 -regularized linear support vector machine classifier was optimized by leave-one-patient-out cross-validation on the training set using the area under the receiver operating characteristic curve (AUC) as classification performance measure. The dimensionality reduction of the texture features was performed within each fold separately. To address the imbalance in the dataset, the classes were weighted by the inverse of their frequencies. The optimal regularization parameter was used to train a classifier using all cases in the training set. The optimal operating point of the classifier (probability cut-off) was selected to be the one that maximizes the

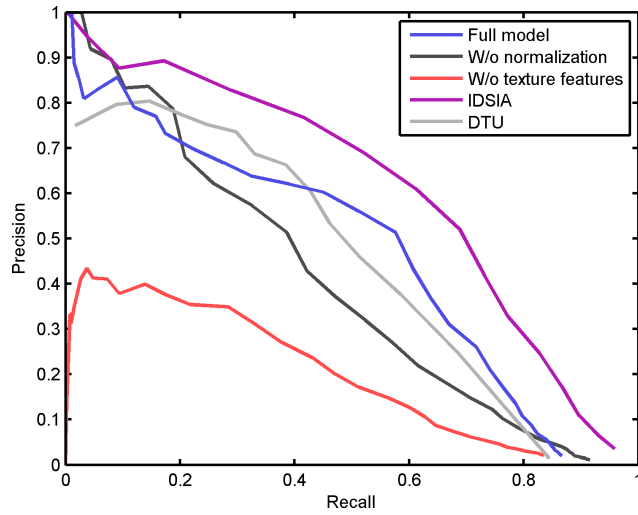


Figure 6.5: Precision-recall curves.

F_1 -score in the leave-one-out experiment.

In addition to the full mitosis detection model, two additional models were trained for comparison. The first one omits the staining normalization procedure prior to the candidate extraction, segmentation and feature extraction. The second one omits the texture feature extraction, and performs classification using size, shape and color features only. In both cases, complete re-training of the detection system was performed, including optimization of the candidate detection and classifier. These baseline models illustrate the influence of different components of the proposed method on the detection results.

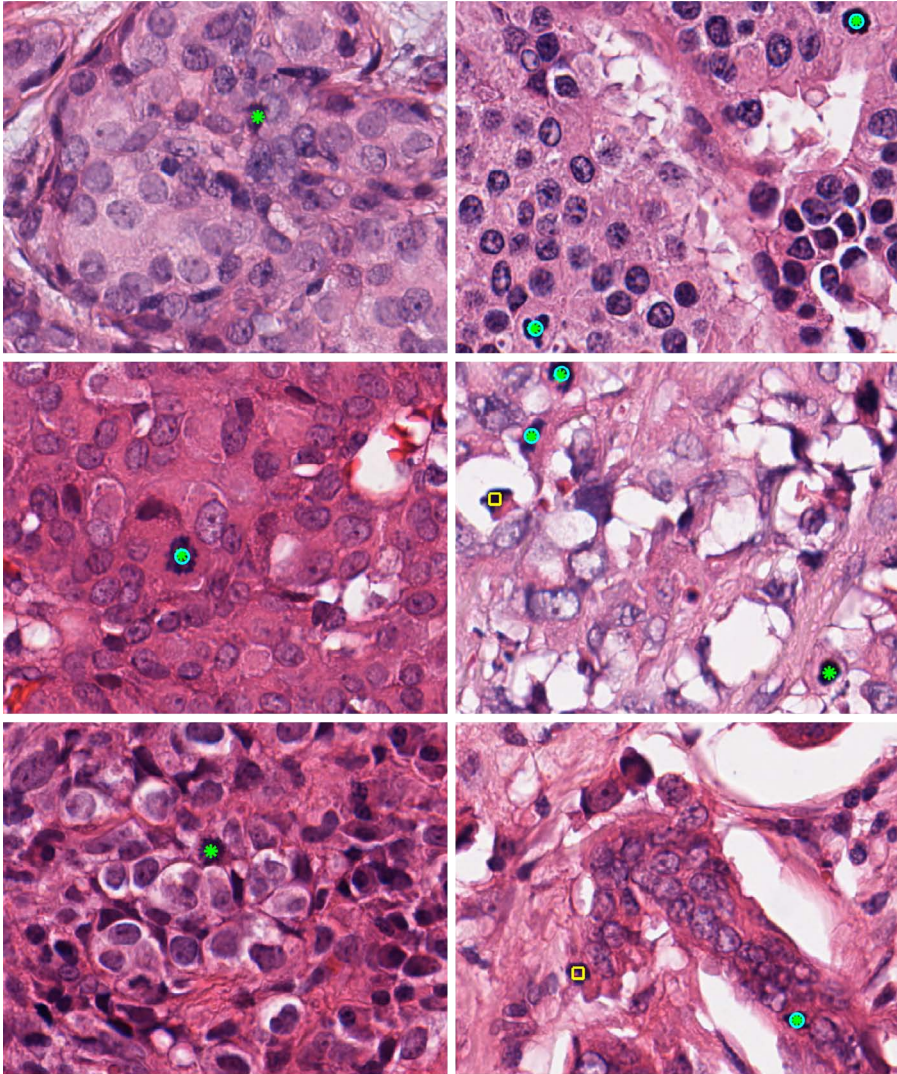
6.3 Results

The evaluation of the detection results is performed according to the AMIDA13 rules, i.e., a detection is considered to be a true positive if the Euclidean distance to a ground truth location is less than $7.5 \mu\text{m}$ or 30 pixels.

The precision-recall (PR) curves of the proposed method along with the two baseline models (without staining normalization and without texture features) are given in Figure 6.5. In addition, the PR curves for the two top performing methods from the AMIDA13 challenge described in the previous chapter are also shown for comparison.

The detection results for the optimal operating point of the classifier are given in Table 6.1. The overall precision, recall and F_1 -score are 0.63, 0.33 and 0.43, respectively. Example detection results for visual evaluation are shown in Figure 6.6.

The correlation coefficient between the estimated and ground truth number of mitotic figures per HPF was $r = 0.54$ $\text{CI} = [-0.34, 0.97]$.



* Ground truth ○ True positive □ False positive

Figure 6.6: Example detection results. From top left to bottom right, the examples originate from cases #1, #3, #4, #6, #9 and #10 in the testing set. The images are shown at the same scale and are not staining-normalized.

Table 6.1: Performance of the mitosis detection method for the individual cases in the testing set.

Case #	True positives	False positives	False negatives	Precision	Recall	F_1 -Score
1	0	5	3	0.00	0.00	0.00
2	0	0	16	0.00	0.00	0.00
3	31	3	35	0.91	0.47	0.62
4	2	0	7	1.00	0.22	0.36
5	0	0	6	0.00	0.00	0.00
6	102	59	110	0.63	0.48	0.55
7	0	7	2	0.00	0.00	0.00
8	-	7	0	-	-	-
9	6	6	109	0.50	0.05	0.09
10	18	4	54	0.82	0.25	0.38
11	14	9	18	0.61	0.44	0.51

6.4 Discussion and conclusions

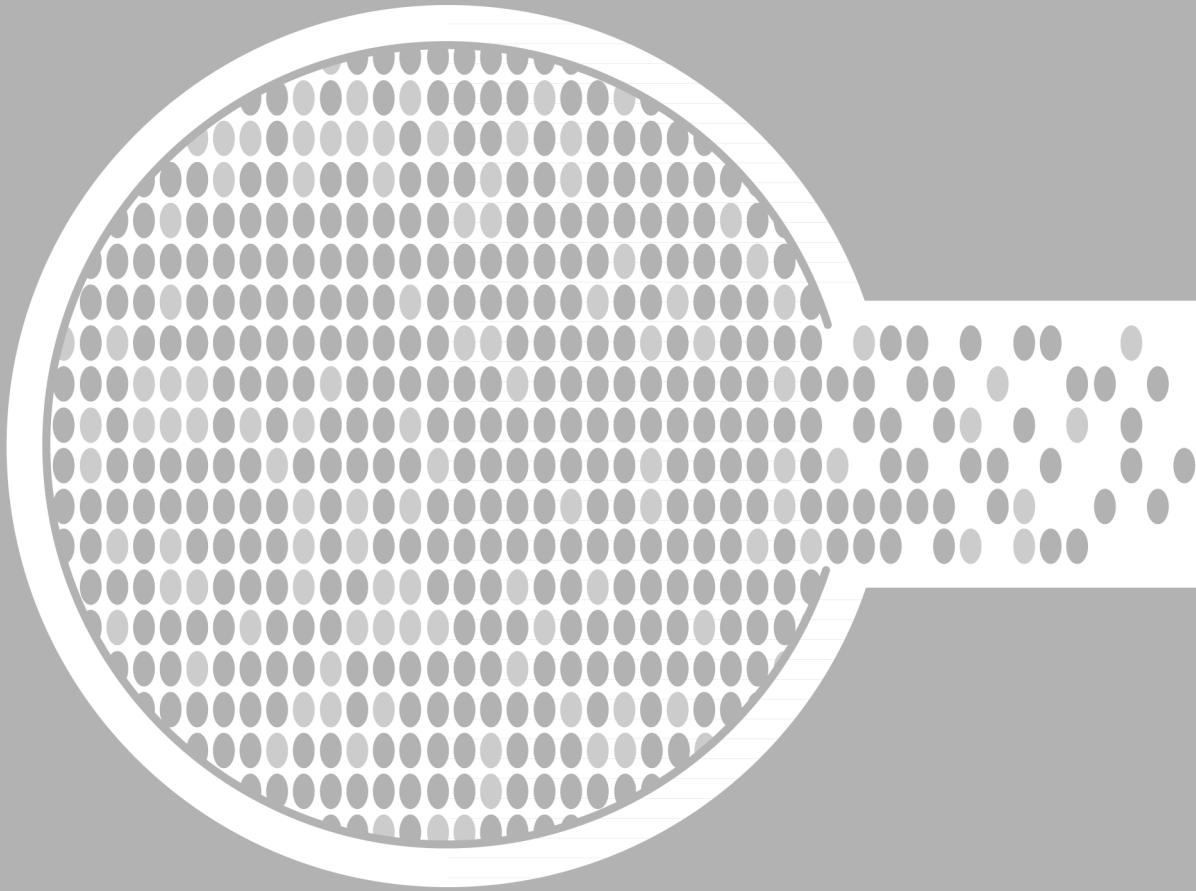
In this chapter, we proposed a two-stage method for the challenging task of mitosis detection in breast cancer histopathology images. In the first stage, candidate objects are identified and segmented, and in the second stage, the candidate objects are classified as mitoses or non-mitoses. With the first stage, a large number of locations that are distinctly not mitotic figures are quickly eliminated, which reduces the required computational resources and enables the second stage to be focused on the more challenging instances. Two baseline methods, produced by omission of different components of the processing pipeline, were also analysed for comparison.

The overall detection results that were achieved by the proposed method are relatively good. From Figure 6.5 it can be concluded that the overall performance of the proposed mitosis detection method is comparable to the DTU method from the AMIDA13 challenge, but somewhat below the IDSIA method. The same conclusion can be drawn from the overall precision, recall and F_1 -score. The omission of the texture features from the system significantly reduced the detection performance. This signifies that the texture features add important information in addition to the size, shape and color features. The inclusion of the staining normalization prior to the candidate detection and feature extraction had a moderate positive effect on the detection performance. However, the method performs poor for some of the individual cases in the testing dataset, as can be seen from Table 6.1. This is also reflected in the correlation between the estimated and ground truth number of mitotic figures per HPF, which is relatively low compared to the top performing methods of the AMIDA13 challenge. Relative to the IDSIA and DTU methods, the

worst performance is achieved for case #9. This case represents a unique challenge because it both contains many non-mitotic objects with appreciably strong staining intensity, and mitotic figures with weak staining intensity. This can not be successfully addressed by the staining normalization procedure. In fact, this procedure has a negative effect for this case as the staining concentrations are normalized with respect to the maximum concentration within each HPF.

In this study we examined only one staining/image normalization technique. Other approaches may provide more robust results and a good candidate for future analysis.

A larger dataset, preferably consisting of hundreds of cases from multiple centers, is needed in order to better capture the variability of the tissue and mitotic figure appearance. However, this kind of dataset is difficult to obtain because of the time-consuming nature of the annotation process. An alternative to slide annotation by expert observers is to use a specialized stain that is specific to mitotic figures, such as PPH3, to obtain ground truth locations that can then be registered to an H&E stained image. This would not only enable “mass production” of ground truth data, but would also eliminate any observer variability and disagreement. We are currently investigating the feasibility of such an approach.



CHAPTER **7**

General summary

Wide acceptance of whole-slide imaging (WSI) in pathology labs now seems imminent. The proliferation of WSI is driven by the reduced cost of implementation as well as the many advantages that whole-slide images can offer to the end users. This includes easy access to the slides even from remote locations, improved slide annotation, ability to simultaneously view and compare multiple slides and use of automated image analysis tools. This chapter presents a short summary of the research and results in this thesis that is focused on breast cancer histopathology image analysis.

In Chapter 2, a review of the literature on breast cancer histopathology image analysis is presented. The greatest progress has been made in automatic analysis of immunohistochemistry (IHC), which has resulted in regulatory approval of image analysis tools for routine use. This is not surprising since IHC stained sections are more amenable to image analysis. Progress has also been made in automatic analysis of hematoxylin and eosin (H&E) stained sections, although at a slower pace. For this staining, the features that are of interest are more complex, thus ensuring robustness is more challenging. A large number of object detection/segmentation techniques and computer-aided diagnosis and prognosis methods have been proposed. However, further improvements are needed in order to reach performance levels that are suitable for routine practice. Two major issues that need to be addressed in order to speed up the development of histopathology image analysis algorithms were identified. First, the tissue preparation process needs to become more standardized in order to reduce the variability of the appearance that is a major debilitating factor for automated image analysis. Second, there is an increasing need for large and well-annotated datasets that can be used for training and evaluation of new methods.

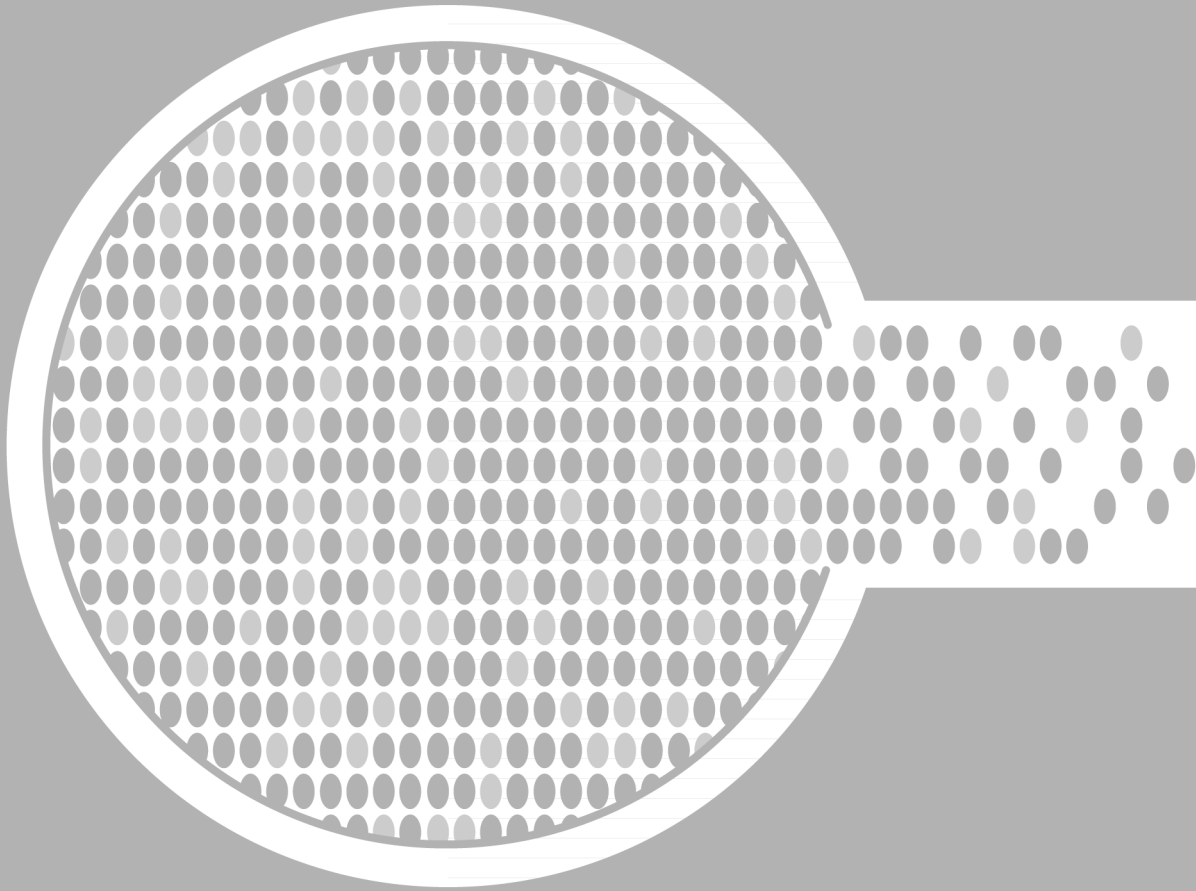
Automatic nuclei segmentation can be an important first step in the analysis of histopathology images. Chapter 3 describes a nuclei segmentation method that works with images from routinely prepared breast cancer H&E sections. The proposed method is based on marker-controlled watershed segmentation that operates on multiple scales and uses multiple marker types. The multi-scale approach enables nuclei with a wide range of sizes to be successfully segmented. This is very important as the size of the nuclei can have prognostic significance for the patients. The technique was evaluated on a dataset of 18 slides with more than 2000 manually segmented nuclei. It showed very good performance in terms of nuclei detection and segmentation accuracy as well in the ability to estimate the mean nuclear area, which is an established morphometric parameter. In Chapter 4, this technique was used to extract and evaluate nuclear morphometric features related to size in a cohort of 101 male breast cancer patients. In univariate survival analysis, a significant difference was found between patients with low and high

mean nuclear area as estimated by automatic nuclei segmentation. The mean nuclear area was retained as an independent prognostic factor in multivariate survival analysis, in addition to tumor size and tubule formation score.

Chapters 5 and 6 deal with automatic detection of mitotic figures. Identification of mitotic figures in H&E stained breast cancer histology sections is challenging even for expert pathologists. It is the most tedious and time-consuming component of histological grading of breast cancer and suffers extensively from observer variability. Regardless of these issues, the mitotic activity index, which is expressed as the number of mitoses in a tissue area of 2 mm^2 is a very powerful proven prognostic factor. Automatic detection of mitotic figures can reduce the workload and the impact of the subjective interpretation.

Recently we organized a challenge workshop on this difficult problem, titled Assessment of Mitosis Detection Algorithms 2013 (AMIDA13). The main goal of the challenge was to evaluate and compare the performance of different automatic mitosis detection methods on a common dataset. The challenge dataset consists of 12 cases for training and 11 for testing with more than 1000 annotated mitotic figures by multiple observers. Chapter 5 summarizes the achieved results from this workshop. The detection error of the best scoring method is on a par with the inter-observer variability. After a visual examination of the results, it was observed that many of the false positives closely resemble mitotic figures and might actually represent objects that have been missed during the annotation owing to the difficulty of the task. In a reannotation experiment, close to 30% of the false positives of the best performing method were annotated as true mitotic figures. This result suggests that the state-of-the-art methods for mitosis detection can already be used in a semi-automatic setting with the goal of reduced variability and faster annotation.

In Chapter 6, the development and evaluation of a two-stage mitosis detection method was described. In the first stage, candidate objects are detected and segmented based on the staining intensity. Since hyperchromicity is the most salient feature of mitotic figures, a large number of locations that are distinctly not mitotic figures can be quickly eliminated. In the second stage the candidate objects are classified as mitoses or non-mitoses using a statistical classifier trained with a number of size, shape, color and texture features. Prior to the candidate detection and classification, a staining normalization procedure is employed in order to reduce the impact of the staining variability. The achieved detection results are comparable to the second best method from AMIDA13.



Bibliography

- Abdel-Fatah, T. M. A. et al. (2010). „Proposal for a modified grading system based on mitotic index and Bcl2 provides objective determination of clinical outcome for patients with breast cancer”. In: *The Journal of Pathology* 222, 388–399.
- Adiga, U. et al. (2006). „High-throughput analysis of multispectral images of breast cancer tissue”. In: *IEEE Transactions on Image Processing* 15, 2259–2268.
- Ali, S. and A. Madabhushi (2012). „An integrated region-, boundary-, shape-based active contour for multiple object overlap resolution in histological imagery”. In: *IEEE Transactions on Medical Imaging* 31, 1448–1460.
- Alomari, R. S. et al. (2009). „Localization of tissues in high-resolution digital anatomic pathology images”. In: *SPIE Medical Imaging*. Vol. 7260, 726016–726016-10.
- Amaral, T. et al. (2013). „Classification and immunohistochemical scoring of breast tissue microarray spots”. In: *IEEE Transactions on Biomedical Engineering* 60, 2806–2814.
- Arteta, C. et al. (2012). „Learning to detect cells using non-overlapping extremal regions”. In: *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 348–356.
- Avenel, C. and M. S. Kulikova (2013). „Marked point processes with simple and complex shape objects for cell nuclei extraction from breast cancer H&E images”. In: *SPIE Medical Imaging*. Vol. 8676, 86760Z–86760Z-7.
- Baak, J. P., P. J. van Diest and G. A. Meijer (2000). „Experience with a dynamic inexpensive video-conferencing system for frozen section telepathology”. In: *Analytical Cellular Pathology* 21, 169–175.
- Baak, J. P. A. et al. (1985). „The value of morphometry to classic prognosticators in breast cancer”. In: *Cancer* 56, 374–382.
- Bahlmann, C. et al. (2012). „Automated detection of diagnostically relevant regions in H&E stained digital pathology slides”. In: *SPIE Medical Imaging*. Vol. 8315, 831504–831504-8.
- Bamford, P. and B. Lovell (1998). „Unsupervised cell nucleus segmentation with active contours”. In: *Signal Processing* 71, 203–213.
- Basavanhally, A. and A. Madabhushi (2013a). „EM-based segmentation-driven color standardization of digitized histopathology”. In: *SPIE Medical Imaging*. Vol. 8676, 86760G–86760G-12.
- Basavanhally, A. et al. (2011). „Incorporating domain knowledge for tubule detection in breast histopathology using O’Callaghan neighborhoods”. In: *SPIE Medical Imaging*. Vol. 7963, 796310–796310-15.
- Basavanhally, A. et al. (2012). „Multi-field-of-view strategy for image-based outcome prediction of multi-parametric estrogen receptor-positive breast cancer histopathology: comparison to oncotype DX”. In: *Journal of Pathology Informatics* 2.
- Basavanhally, A. et al. (2010). „Computerized image-based detection and grading of lymphocytic infiltration in HER2+ breast cancer histopathology”. In: *IEEE Transactions on Biomedical Engineering* 57, 642–653.
- Basavanhally, A. et al. (2013b). „Multi-field-of-view framework for distinguishing tumor grade in ER+ breast cancer from entire histopathology slides”. In: *IEEE Transactions on Biomedical Engineering* 60, 2089–2099.

- Beck, A. H. et al. (2011). „Systematic analysis of breast cancer morphology uncovers stromal features associated with survival”. In: *Science Translational Medicine* 3, 108ra113–108ra113.
- Belhomme, P. et al. (2011). „Towards a computer aided diagnosis system dedicated to virtual microscopy based on stereology sampling and diffusion maps”. In: *Diagnostic Pathology* 6.
- Beliën, J. A. M. et al. (1997). „Counting mitoses by image processing in feulgen stained breast cancer sections: the influence of resolution”. In: *Cytometry* 28, 135–140.
- Beucher, S. and F. Meyer (1993). „The morphological approach to segmentation: the watershed transformation”. In: *Mathematical Morphology in Image Processing*. Vol. 34, 433–481.
- Bilgin, C. et al. (2012). „Digitally adjusting chromogenic dye proportions in brightfield microscopy images”. In: *Journal of Microscopy* 245, 319–330.
- Bolton, K. L. et al. (2010). „Assessment of automated image analysis of breast cancer tissue microarrays for epidemiologic studies”. In: *Cancer Epidemiology Biomarkers and Prevention* 19, 992–999.
- Boucheron, L., B. Manjunath and N. Harvey (2010). „Use of imperfectly segmented nuclei in the classification of histopathology images of breast cancer”. In: *IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*, 666–669.
- Bray, F., P. McCarron and D. M. Parkin (2004). „The changing global patterns of female breast cancer incidence and mortality”. In: *Breast Cancer Research* 6, 229–239.
- Brüggmann, A. et al. (2012). „Digital image analysis of membrane connectivity is a robust measure of HER2 immunostains”. In: *Breast Cancer Research and Treatment* 132, 41–49.
- Bulte, J. P. et al. (2013). „One-day core needle biopsy in a breast clinic: 4 years experience”. In: *Breast Cancer Research and Treatment* 137, 609–616.
- Chan, T. and L. Vese (2001). „Active contours without edges”. In: *IEEE Transactions on Image Processing* 10, 266–277.
- Chekkoury, A. et al. (2012). „Automated malignancy detection in breast histopathological images”. In: *SPIE Medical Imaging*. Vol. 8315, 831515–831515-13.
- Chiusa, L., E. Margaria and A. Pich (2000). „Nuclear morphometry in male breast carcinoma: association with cell proliferative activity, oncogene expression, DNA content and prognosis”. In: *International Journal of Cancer* 89, 494–499.
- Ciresan, D. C., U. Meier and J. Schmidhuber (2012a). „Multi-column deep neural networks for image classification”. In: *Computer Vision and Pattern Recognition (CVPR)*, 3642–3649.
- Ciresan, D. C. et al. (2012b). „Deep neural networks segment neuronal membranes in electron microscopy images”. In: *Neural Information Processing Systems (NIPS)*, 2852–2860.
- Ciresan, D. C. et al. (2013). „Mitosis detection in breast cancer histology images with deep neural networks”. In: *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 411–418.
- Cong, G. and B. Parvin (2000). „Model-based segmentation of nuclei”. In: *Pattern Recognition* 33, 1383–1393.
- Cosatto, E. et al. (2008). „Grading nuclear pleomorphism on histological micrographs”. In: *International Conference on Pattern Recognition (ICPR)*, 1–4.
- Cui, Y. et al. (2007). „Nuclear morphometric features in benign breast tissue and risk of subsequent breast cancer”. In: *Breast Cancer Research and Treatment* 104, 103–107.
- Damianou, A. et al. (2012). „Manifold relevance determination”. In: *International Conference on Machine Learning (ICML)*, 145–152.
- Di Cataldo, S., E. Ficarra and E. Macii (2012). „Computer-aided techniques for chromogenic immunohistochemistry: status and directions”. In: *Computers in Biology and Medicine* 42, 1012–1025.
- Diest, P. J. van, J. C. Fleege and J. P. Baak (1992a). „Syntactic structure analysis in invasive breast cancer: analysis of reproducibility, biologic background, and prognostic value”. In: *Human Pathology* 23, 876–883.
- Diest, P. J. van, E. van der Wall and J. P. A. Baak (2004). „Prognostic value of proliferation in invasive breast cancer: a review”. In: *Journal of Clinical Pathology* 57, 675–681.

- Diest, P. van et al. (1992b). „Reproducibility of mitosis counting in 2,469 breast cancer specimens: results from the multicenter morphometric mammary carcinoma project”. In: *Human Pathology* 23, 603–607.
- Dowsett, M. et al. (2011). „Assessment of ki67 in breast cancer: recommendations from the international ki67 in breast cancer working group”. In: *Journal of the National Cancer Institute* 103, 1656–1664.
- Doyle, S. et al. (2008). „Automated grading of breast cancer histopathology using spectral clustering with textural and architectural image features”. In: *IEEE International Symposium on Biomedical Imaging (ISBI)*, 496–499.
- Elston, C. W. and I. O. Ellis (1991). „Pathological prognostic factors in breast cancer. i. the value of histological grade in breast cancer: experience from a large study with long-term follow-up”. In: *Histopathology* 19, 403–410.
- Fan, R.-E. et al. (2008). „LIBLINEAR: a library for large linear classification”. In: *Journal of Machine Learning Research* 9, 1871–1874.
- Fatakdawala, H. et al. (2010). „Expectation-maximization-driven geodesic active contour with overlap resolution (EMaGACOR): application to lymphocyte segmentation on breast cancer histopathology”. In: *IEEE Transactions on Biomedical Engineering* 57, 1676–1689.
- Ficarra, E. et al. (2011). „Automated segmentation of cells with IHC membrane staining”. In: *IEEE Transactions on Biomedical Engineering* 58, 1421–1429.
- Fitzgibbons, P. L. et al. (2000). „Prognostic factors in breast cancer. College of American Pathologists consensus statement 1999”. In: *Archives of Pathology and Laboratory Medicine* 124, 966–978.
- Fleege, J. C., P. J. van Diest and J. P. Baak (1990). „Computer assisted efficiency testing of different sampling methods for selective nuclear graphic tablet morphometry”. In: *Laboratory Investigation* 63, 270–275.
- Foran, D. J. et al. (2011). „ImageMiner: a software system for comparative analysis of tissue microarrays using content-based image retrieval, high-performance computing, and grid technology”. In: *Journal of the American Medical Informatics Association* 18, 403–415.
- Fuchs, T. J. and J. M. Buhmann (2011). „Computational pathology: challenges and promises for tissue analysis”. In: *Computerized Medical Imaging and Graphics* 35, 515–530.
- Garini, Y., I. T. Young and G. McNamara (2006). „Spectral imaging: principles and applications”. In: *Cytometry Part A* 69A, 735–747.
- Gavrielides, M. A. et al. (2011). „Observer variability in the interpretation of HER2neu immunohistochemical expression with unaided and computer-aided digital microscopy”. In: *Archives of Pathology and Laboratory Medicine* 135, 233–242.
- Gavrilovic, M. et al. (2013). „Blind color decomposition of histological images”. In: *IEEE Transactions on Medical Imaging* 32, 983–994.
- Ghaznavi, F. et al. (2013). „Digital imaging in pathology: whole-slide imaging and beyond”. In: *Annual Review of Pathology* 8, 331–359.
- Giusti, A. et al. (2013). „Fast image scanning with deep max-pooling convolutional neural networks”. In: *International Conference on Image Processing (ICIP)*, in press.
- Gudla, P. R. et al. (2008). „A high-throughput system for segmenting nuclei using multiscale techniques”. In: *Cytometry Part A* 73A, 451–466.
- Gurcan, M. N. et al. (2009). „Histopathological image analysis: a review”. In: *IEEE Reviews in Biomedical Engineering* 2, 147–171.
- Gutiérrez, R. et al. (2011). „A supervised visual model for finding regions of interest in basal cell carcinoma images”. In: *Diagnostic Pathology* 6.
- Hall, B. H. et al. (2008). „Computer-assisted assessment of the human epidermal growth factor receptor 2 immunohistochemical assay in imaged histologic sections using a membrane isolation algorithm and quantitative analysis of positive controls”. In: *BMC Medical Imaging* 8.

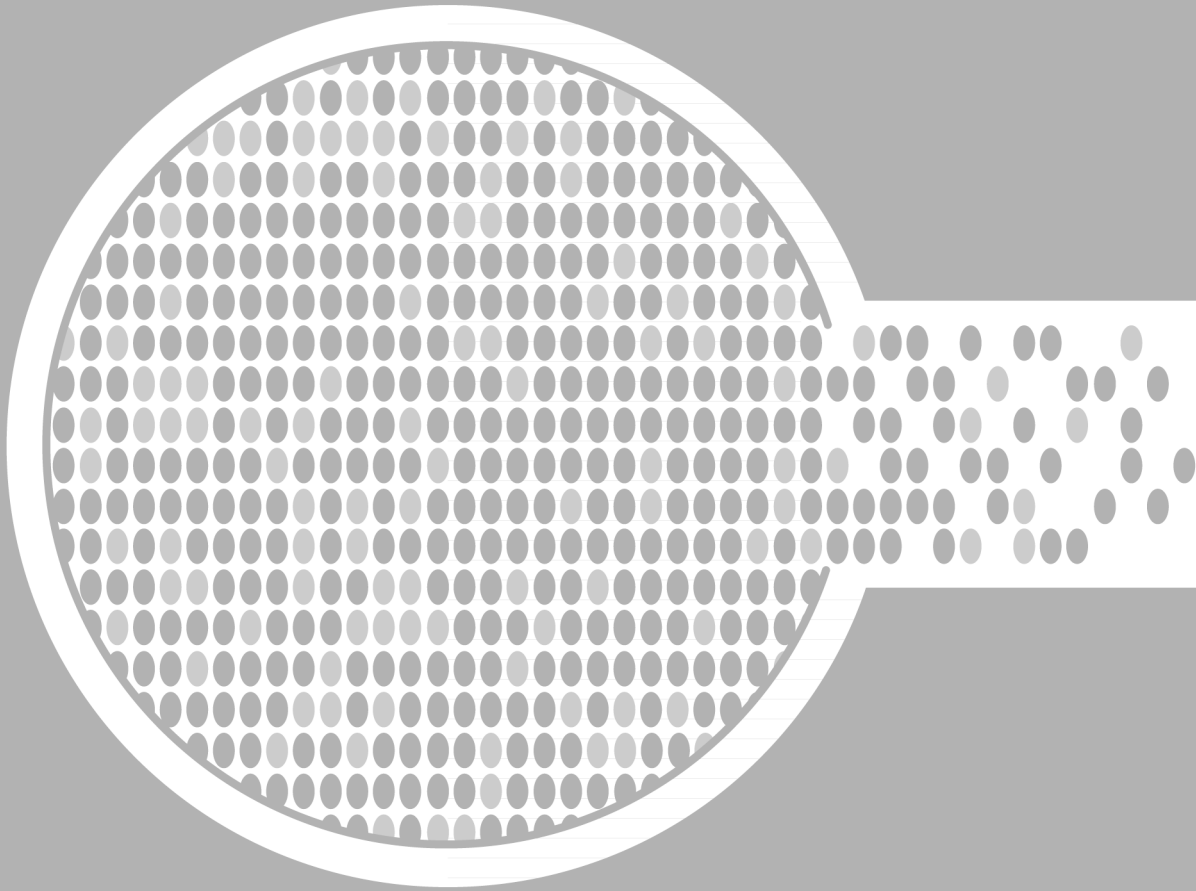
- Hammond, M. E. H. et al. (2010). „American Society of Clinical Oncology/College of American Pathologists guideline recommendations for immunohistochemical testing of estrogen and progesterone receptors in breast cancer”. In: *Journal of Oncology Practice* 6, 195–197.
- Hipp, J. D. et al. (2011). „Why a pathology image should not be considered as a radiology image”. In: *Journal of Pathology Informatics* 2.
- Huang, C.-H. and H.-K. Lee (2012). „Automated mitosis detection based on eXclusive independent component analysis”. In: *International Conference on Pattern Recognition (ICPR)*, 1856–1859.
- Huang, C.-H. et al. (2011). „Time-efficient sparse analysis of histopathological whole slide images”. In: *Computerized Medical Imaging and Graphics* 35, 579–591.
- Huang, P.-W. and Y.-H. Lai (2010). „Effective segmentation and classification for HCC biopsy images”. In: *Pattern Recognition* 43, 1550–1563.
- Huisman, A. et al. (2010). „Creation of a fully digital pathology slide archive by high-volume tissue slide scanning”. In: *Human Pathology* 41, 751–757.
- Irshad, H. (2013). „Automated mitosis detection in histopathology using morphological and multi-channel statistics features”. In: *Journal of Pathology Informatics* 4.
- Irshad, H. et al. (2013). „Automated mitosis detection using texture, SIFT features and HMAX biologically inspired approach”. In: *Journal of Pathology Informatics* 4.
- Al-Janabi, S., A. Huisman and P. J. Van Diest (2011). „Digital pathology: current status and future perspectives”. In: *Histopathology* 61, 1–9.
- Al-Janabi, S. et al. (2013). „Evaluation of mitotic activity index in breast cancer using whole slide digital images”. In: *PLoS ONE* 8, e82576.
- Jannink, I. et al. (1995). „At convenience and systematic random sampling: effects on the prognostic value of nuclear area assessments in breast cancer patients”. In: *Breast Cancer Research and Treatment* 36, 55–60.
- Jung, C. et al. (2010). „Unsupervised segmentation of overlapped nuclei using bayesian classification”. In: *IEEE Transactions on Biomedical Engineering* 57, 2825–2832.
- Kachouie, N. N. et al. (2010). „Constrained watershed method to infer morphology of mammalian cells in microscopic images”. In: *Cytometry Part A* 77A, 1148–1159.
- Kaman, E. J. et al. (1984). „Image processing for mitoses in sections of breast cancer: a feasibility study”. In: *Cytometry* 5, 244–249.
- Karaćali, B. and A. Tözeren (2007). „Automated detection of regions of interest for tissue microarray experiments: an image texture analysis”. In: *BMC Medical Imaging* 7.
- Karsnas, A., A. L. Dahl and R. Larsen (2012). „Learning histopathological patterns”. In: *Journal of Pathology Informatics* 2.
- Kate, T. K. ten et al. (1993). „Method for counting mitoses by image processing in feulgen stained breast cancer sections”. In: *Cytometry* 14, 241–250.
- Kayser, K. et al. (2008). „How to measure image quality in tissue-based diagnosis (diagnostic surgical pathology)”. In: *Diagnostic Pathology* 3.
- Kayser, K. et al. (2009). „Theory of sampling and its application in tissue based diagnosis”. In: *Diagnostic Pathology* 4.
- Khan, A. M. et al. (2014). „A nonlinear mapping approach to stain normalization in digital histopathology images using image-specific color deconvolution”. In: *IEEE Transactions on Biomedical Engineering* 61, 1729–1738.
- Khan, A. M. et al. (2013). „HyMaP: a hybrid magnitude-phase approach to unsupervised segmentation of tumor areas in breast cancer histology images”. In: *Journal of Pathology Informatics* 4.
- Al-Kofahi, Y. et al. (2010). „Improved automatic detection and segmentation of cell nuclei in histopathology images”. In: *IEEE Transactions on Biomedical Engineering* 57, 841–852.
- Kong, H., M. Gurcan and K. Belkacem-Boussaid (2011). „Partitioning histopathological images: an integrated framework for supervised color-texture segmentation and cell splitting”. In: *IEEE Transactions on Medical Imaging* 30, 1661–1677.

- Kononen, J. et al. (1998). „Tissue microarrays for high-throughput molecular profiling of tumor specimens”. In: *Nature medicine* 4, 844–847.
- Kornegoor, R. et al. (2011). „Molecular subtyping of male breast cancer by immunohistochemistry”. In: *Modern Pathology* 25, 398–404.
- Kornegoor, R. et al. (2012). „Fibrotic focus and hypoxia in male breast cancer”. In: *Modern Pathology* 25, 1397–1404.
- Kronqvist, P., T. Kuopio and Y. Collan (1998). „Morphometric grading of invasive ductal breast cancer. i. thresholds for nuclear grade.” In: *British Journal of Cancer* 78, 800–805.
- Kulikova, M. et al. (2012). „Nuclei extraction from histopathological images using a marked point process approach”. In: *SPIE Medical Imaging*. Vol. 8314, 831428–831428-8.
- Lahrman, B. et al. (2010). „Robust gridding of TMAs after whole-slide imaging using template matching”. In: *Cytometry Part A* 77A, 1169–1176.
- Larsen, A., J. Vestergaard and R. Larsen (2014). „HEp-2 cell classification using shape index histograms with donut-shaped spatial pooling”. In: *Medical Imaging, IEEE Transactions on* 33, 1573–1580.
- Laurinaviciene, A. et al. (2011). „Membrane connectivity estimated by digital image analysis of HER2 immunohistochemistry is concordant with visual scoring and fluorescence in situ hybridization results: algorithm evaluation on breast cancer tissue microarrays”. In: *Diagnostic Pathology* 6.
- Li, G. et al. (2008). „Segmentation of touching cell nuclei using gradient flow tracking”. In: *Journal of Microscopy* 231, 47–58.
- Linder, N. et al. (2012). „Identification of tumor epithelium and stroma in tissue microarrays using texture analysis”. In: *Diagnostic Pathology* 7.
- Lloyd, M. C. et al. (2010). „Using image analysis as a tool for assessment of prognostic and predictive biomarkers for breast cancer: how reliable is it?” In: *Journal of Pathology Informatics* 1.
- Loy, G. and A. Zelinsky (2003). „Fast radial symmetry for detecting points of interest”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 25, 959–973.
- Macenko, M. et al. (2009). „A method for normalizing histology slides for quantitative analysis”. In: *IEEE International Symposium on Biomedical Imaging (ISBI)*, 1107–1110.
- Madabhushi, A. et al. (2011). „Computer-aided prognosis: predicting patient and disease outcome via quantitative fusion of multi-scale, multi-modal data”. In: *Computerized Medical Imaging and Graphics* 35, 506–514.
- Malon, C. and E. Cosatto (2013). „Classification of mitotic figures with convolutional neural networks and seeded blob features”. In: *Journal of Pathology Informatics* 4.
- Malon, C. et al. (2012). „Mitotic figure recognition: agreement among pathologists and computerized detector”. In: *Analytical Cellular Pathology* 35, 97–100.
- Malpica, N. et al. (1997). „Applying watershed algorithms to the segmentation of clustered nuclei”. In: *Cytometry* 28, 289–297.
- Masmoudi, H. et al. (2009). „Automated quantitative assessment of HER2neu immunohistochemical expression in breast cancer”. In: *IEEE Transactions on Medical Imaging* 28, 916–925.
- McKenna, S. J. et al. (2013). „Immunohistochemical analysis of breast tissue microarray images using contextual classifiers”. In: *Journal of Pathology Informatics* 4.
- Meijer, G. A. et al. (1997). „Origins of ... image analysis in clinical pathology.” In: *Journal of Clinical Pathology* 50, 365–370.
- Meyer, J. S. et al. (2005). „Breast carcinoma malignancy grading by Bloom-Richardson system vs proliferation index: reproducibility of grade and advantages of proliferation index”. In: *Modern Pathology* 18, 1067–1078.
- Minot, D. M. et al. (2012). „Image analysis of HER2 immunohistochemical staining reproducibility and concordance with fluorescence in situ hybridization of a laboratory-validated scoring technique”. In: *American Journal of Clinical Pathology* 137, 270–276.

- Moelans, C. B. et al. (2009). „HER2neu amplification testing in breast cancer by multiplex ligation-dependent probe amplification in comparison with immunohistochemistry and in situ hybridization”. In: *Cellular Oncology* 31, 1–10.
- Mohammed, Z. M. A. et al. (2012). „Comparison of visual and automated assessment of HER2 status and their impact on outcome in primary operable invasive ductal breast cancer”. In: *Histopathology* 61, 675–684.
- Mommers, E. C. M. et al. (2001). „Prognostic value of morphometry in patients with normal breast tissue or usual ductal hyperplasia of the breast”. In: *International Journal of Cancer* 95, 282–285.
- Monaco, J. et al. (2012). „Image segmentation with implicit color standardization using spatially constrained expectation maximization: detection of nuclei”. In: *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 365–372.
- Mousses, S. et al. (2003). „RNAi microarray analysis in cultured mammalian cells”. In: *Genome Research* 13, 2341–2347.
- Naik, S. et al. (2008). „Automated gland and nuclei segmentation for grading of prostate and breast cancer histopathology”. In: *IEEE International Symposium on Biomedical Imaging (ISBI)*, 284–287.
- Nassar, A. et al. (2011). „Trainable immunohistochemical HER2neu image analysis: a multisite performance study using 260 breast tissue specimens”. In: *Archives of Pathology and Laboratory Medicine* 135, 896–902.
- Nguyen, K., A. Sarkar and A. K. Jain (2012). „Structure and context in prostatic gland segmentation and classification”. In: *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 115–123.
- Pantanowitz, L. et al. (2011). „Review of the current state of whole slide imaging in pathology”. In: *Journal of Pathology Informatics* 2.
- Peikari, M. et al. (2013). „A texture based approach to automated detection of diagnostically relevant regions in breast digital pathology.” In: *MICCAI Workshop on Breast Image Analysis*.
- Peng, Y. et al. (2011). „Segmentation of prostatic glands in histology images”. In: *IEEE International Symposium on Biomedical Imaging (ISBI)*, 2091–2094.
- Perez, E. A. et al. (2006). „HER2 testing by local, central, and reference laboratories in specimens from the North Central Cancer Treatment Group N9831 intergroup adjuvant trial”. In: *Journal of Clinical Oncology* 24, 3032–3038.
- Pienta, K. J. and D. S. Coffey (1991). „Correlation of nuclear morphometry with progression of breast cancer”. In: *Cancer* 68, 2012–2016.
- Qi, X. et al. (2012). „Robust segmentation of overlapping cells in histopathology specimens using parallel seed detection and repulsive level set”. In: *IEEE Transactions on Biomedical Engineering* 59, 754–765.
- Quinlan, J. R. (1996). „Bagging, boosting, and c4.5”. In: *Thirteenth National Conference on Artificial Intelligence*, 725–730.
- Rajpoot, N., H. Eldaly and A. Khan (2013). „A gamma-gaussian mixture model for detection of mitotic cells in breast cancer histopathology images”. In: *Journal of Pathology Informatics* 4.
- Rakha, E. A. et al. (2010). „Breast cancer prognostic classification in the molecular era: the role of histological grade”. In: *Breast Cancer Research* 12, 207.
- Rexhepaj, E. et al. (2008). „Novel image analysis approach for quantifying expression of nuclear proteins assessed by immunohistochemistry: application to measurement of oestrogen and progesterone receptor levels in breast cancer”. In: *Breast Cancer Research* 10, R89.
- Robbins, P. et al. (1995). „Histological grading of breast carcinomas: a study of interobserver agreement”. In: *Human Pathology* 26, 873–879.
- Rajo, M., G. Bueno and J. Slodkowska (2009). „Review of imaging solutions for integrated quantitative immunohistochemistry in the pathology daily practice”. In: *Folia Histochemica et Cytobiologica* 47, 349–354.

- Romo, D., E. Romero and F. Gonzalez (2011). „Learning regions of interest from low level maps in virtual microscopy”. In: *Diagnostic Pathology* 6.
- Roux, L. et al. (2013). „Mitosis detection in breast cancer histological images: an ICPR 2012 contest”. In: *Journal of Pathology Informatics* 4.
- Ruifrok, A. C. and D. A. Johnston (2001). „Quantification of histochemical staining by color deconvolution”. In: *Analytical and Quantitative Cytology and Histology* 23, 291–299.
- Schmitt, O. and M. Hasse (2008). „Radial symmetries based decomposition of cell clusters in binary and gray level images”. In: *Pattern Recognition* 41, 1905–1923.
- Seiffert, C. et al. (2010). „RUSBoost: a hybrid approach to alleviating class imbalance”. In: *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans* 40, 185–197.
- Sertel, O. et al. (2010). „Computer-aided detection of centroblasts for follicular lymphoma grading using adaptive likelihood-based cell segmentation”. In: *IEEE Transactions on Bio-Medical Engineering* 57, 2613–2616.
- Skacel, M. et al. (2002). „Tissue microarrays: a powerful tool for high-throughput analysis of clinical specimens: a review of the method with validation data”. In: *Applied Immunohistochemistry and Molecular Morphology* 10, 1–6.
- Skaland, I. et al. (2007). „Phosphohistone h3 expression has much stronger prognostic value than classical prognosticators in invasive lymph node-negative breast cancer patients less than 55 years of age”. In: *Modern Pathology* 20, 1307–1315.
- Sommer, C. et al. (2012). „Learning-based mitotic cell detection in histopathological images”. In: *International Conference on Pattern Recognition (ICPR)*, 2306–2309.
- Stathonikos, N. et al. (2013). „Going fully digital: perspective of a dutch academic pathology lab”. In: *Journal of Pathology Informatics* 4.
- Tambasco, M., M. Eliasziw and A. M. Magliocco (2010). „Morphologic complexity of epithelial architecture for predicting invasive breast cancer survival”. In: *Journal of Translational Medicine* 8, 140.
- Tambasco, M. and A. M. Magliocco (2008). „Relationship between tumor grade and computed architectural complexity in breast cancer specimens”. In: *Human Pathology* 39, 740–746.
- Tan, P. H. et al. (2001). „Correlation of nuclear morphometry with pathologic parameters in ductal carcinoma in situ of the breast”. In: *Modern Pathology* 14, 937–941.
- Taylor, C. R. and R. M. Levenson (2006). „Quantification of immunohistochemistry — issues concerning methods, utility and semiquantitative assessment II”. In: *Histopathology* 49, 411–424.
- Tek, F. B. (2013). „Mitosis detection using generic features and an ensemble of cascade ada-boosts”. In: *Journal of Pathology Informatics* 4.
- Titsias, M. K. and N. D. Lawrence (2010). „Bayesian gaussian process latent variable model”. In: *Journal of Machine Learning Research* 9, 844–851.
- Tuominen, V. J., T. T. Tolonen and J. Isola (2012). „ImmunoMembrane: a publicly available web application for digital image analysis of HER2 immunohistochemistry”. In: *Histopathology* 60, 758–767.
- Tuominen, V. J. et al. (2010). „ImmunoRatio: a publicly available web application for quantitative image analysis of estrogen receptor (ER), progesterone receptor (PR), and ki-67”. In: *Breast Cancer Research* 12, R56.
- Turashvili, G. et al. (2009). „Inter-observer reproducibility of HER2 immunohistochemical assessment and concordance with fluorescent in situ hybridization (FISH): pathologist assessment compared to quantitative image analysis”. In: *BMC Cancer* 9.
- Varma, M. and A. Zisserman (2005). „A statistical approach to texture classification from single images”. In: *International Journal of Computer Vision* 62, 61–81.
- Veta, M., P. J. van Diest and J. P. W. Pluim (2013a). „Detecting mitotic figures in breast cancer histopathology images”. In: *SPIE Medical Imaging*. Vol. 8676, 867607–867607-7.
- Veta, M. et al. (2012). „Prognostic value of automatically extracted nuclear morphometric features in whole slide images of male breast cancer”. In: *Modern Pathology* 25, 1559–1565.

- Veta, M. et al. (2013b). „Automatic nuclei segmentation in H&E stained breast cancer histopathology images”. In: *PLoS ONE* 8, e70221.
- Veta, M. et al. (2014a). „Assessment of algorithms for mitosis detection in breast cancer histopathology images”. Submitted for review.
- Veta, M. et al. (2014b). „Breast cancer histopathology image analysis: a review”. In: *IEEE Transactions on Biomedical Engineering* 61, 1400–1411.
- Veta, M. et al. (2011). „Marker-controlled watershed segmentation of nuclei in h amp;E stained breast cancer biopsy images”. In: *IEEE International Symposium on Biomedical Imaging (ISBI)*, 618–621.
- Vincent, L. (1993). „Morphological grayscale reconstruction in image analysis: applications and efficient algorithms”. In: *IEEE Transactions on Image Processing* 2, 176–201.
- Vink, J. et al. (2013). „Efficient nucleus detector in histopathology images”. In: *Journal of Microscopy* 249, 124–135.
- Wahlby, C. et al. (2004). „Combining intensity, edge and shape information for 2D and 3D segmentation of cell nuclei in tissue sections”. In: *Journal of Microscopy* 215, 67–76.
- Walker, R. A. (2006). „Quantification of immunohistochemistry — issues concerning methods, utility and semiquantitative assessment I”. In: *Histopathology* 49, 406–410.
- Walter, T. et al. (2007). „Automatic detection of microaneurysms in color fundus images.” In: *Medical Image Analysis* 11, 555–66.
- Walter, T. et al. (2010). „Automatic identification and clustering of chromosome phenotypes in a genome wide RNAi screen by time-lapse imaging.” In: *Journal of Structural Biology* 170, 1–9.
- Wang, C. et al. (2013). „Identifying survival associated morphological features of triple negative breast cancer using multiple datasets”. In: *Journal of the American Medical Informatics Association: JAMIA* 20, 680–687.
- Wang, C.-W. and A. Hunter (2010a). „A low variance error boosting algorithm”. In: *Applied Intelligence* 33, 357–369.
- (2010b). „Robust pose recognition of obscured human body”. In: *International Journal of Computer Vision* 90.
- Wang, C.-W. et al. (2011a). „Robust automated tumour segmentation on histological and immunohistochemical tissue images”. In: *PLoS ONE* 6, e15818.
- Wang, H. et al. (2014). „Cascaded ensemble of convolutional neural networks and handcrafted features for mitosis detection”. In: *SPIE Medical Imaging*. Vol. 9041, 90410B–90410B-10.
- Wang, Y. et al. (2011b). „A TMA de-arraying method for high throughput biomarker discovery in tissue research”. In: *PLoS ONE* 6, e26007.
- Wienert, S. et al. (2012). „Detection and segmentation of cell nuclei in virtual microscopy images: a minimum-model approach”. In: *Scientific Reports* 2.
- Wolff, A. C. et al. (2007). „American Society of Clinical Oncology/College of American Pathologists guideline recommendations for human epidermal growth factor receptor 2 testing in breast cancer”. In: *Archives of Pathology and Laboratory Medicine* 131, 18–43.
- Xu, J., J. P. Monaco and A. Madabhushi (2010a). „Markov random field driven region-based active contour model (MaRACel): application to medical image segmentation”. In: *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 197–204.
- Xu, J. et al. (2010b). „A weighted mean shift, normalized cuts initialized color gradient based geodesic active contour model: applications to histopathology image segmentation”. In: *SPIE Medical Imaging*. Vol. 7623, 76230Y–76230Y-12.
- Yang, X., H. Li and X. Zhou (2006). „Nuclei segmentation using marker-controlled watershed, tracking using mean-shift, and kalman filter in time-lapse microscopy”. In: *IEEE Transactions on Circuits and Systems I: Regular Papers* 53, 2405–2414.
- Yuan, Y. et al. (2012). „Quantitative image analysis of cellular heterogeneity in breast tumors complements genomic profiling”. In: *Science Translational Medicine* 4, 157ra143.



Nederlandse samenvatting

Pathologielaboratoria zijn vandaag de dag bezig over te gaan op een volledig digitale werkwijze. Naast het digitaal beheren van laboratoriumaanvragen en uitslagen, worden tegenwoordig ook de weefselpreparaten zelf gedigitaliseerd. Met behulp van weefselscanners worden preparaten opgeslagen en bekeken op een computerscherm. Die zou hierdoor de microscoop als het werkpaard van de patholoog kunnen vervangen. Dit proces is min of meer analoog aan de digitalisatie van de radiologische beeldvorming. Een van de grote voordelen van digitale weefselpreparaten in vergelijking met conventionele glaasjes is dat hierdoor een naadloze integratie van kwantitatieve, automatische beeldanalysetechnieken in de werkwijze van het pathologielab mogelijk is.

Een relatief hoog percentage van de weefselpreparaten die in pathologielaboratoria worden geanalyseerd behelzen borstkanker, daar dit de meest voorkomende vorm van kanker bij vrouwen is. Met behulp van deze preparaten bepalen pathologen de differentiatiegraad van de tumor en onderzoeken zij via immunohistochemie de aanwezigheid van hormoonreceptoren. Dit soort bepalingen zijn lastig uit te voeren en veelal onderhevig aan interbeoordelaarsvariabiliteit. De preparaten ondergaan allereerst een Haematoxylineosine-kleuring (H&E) en met het Bloom-Richardson systeem wordt vervolgens de differentiatiegraad bepaald. Hierbij wordt gekeken naar de celkern, de onderlinge rangschikking van de cellen en het aantal cellen dat mitose (celdeling) vertoont. De analyse met immunohistochemie wordt uitgevoerd om cellen die reageren met een specifiek antilichaam te lokaliseren en tellen.

Brede invoering van weefselscanners in pathologielaboratoria lijkt slechts een kwestie van tijd. De verspreiding van scanners wordt gedreven door de lagere kosten alsook door de vele voordelen die digitale beelden bieden aan de gebruiker. Deze voordelen omvatten makkelijke toegang tot de beelden, zelfs van een afstand, betere methoden voor annotatie van de beelden, de mogelijkheid gelijktijdig een aantal preparaten te bekijken en te vergelijken, en het gebruik van automatische beeldanalysetechnieken. Dergelijke beeldanalysetechnieken voor borstkankerhistopathologie zijn het focus van dit proefschrift

en dit hoofdstuk biedt een kort overzicht van de resultaten.

Hoofdstuk 2 bevat een literatuuroverzicht van beeldanalysetechnieken voor weefselpreparaten. De grootste voortgang is geboekt in de automatische analyse van immuunhistochemie, vooral omdat deze beelden wat eenvoudiger te analyseren zijn. Dit heeft geresulteerd in goedkeuring voor het gebruik van deze beeldanalysetechnieken in de kliniek. De automatische analyse van H&E preparaten verloopt wat moeizamer. De benodigde analyses hiervan zijn complexer en lastig om betrouwbaar uit te voeren. Inmiddels zijn er wel veel technieken ontwikkeld voor automatische segmentatie, diagnose en prognose op basis van deze beelden, maar de kwaliteit is nog niet klinisch acceptabel. De twee grootste knelpunten zijn: 1) het ontbreken van een goede gestandaardiseerde procedure voor het prepareren van het weefsel, waardoor grote verschillen tussen preparaten ontstaan die automatische analyse zwaar bemoeilijken en 2) onvoldoende beschikbaarheid van geannoteerde preparaten waarmee automatische technieken kunnen worden getraind en geëvalueerd.

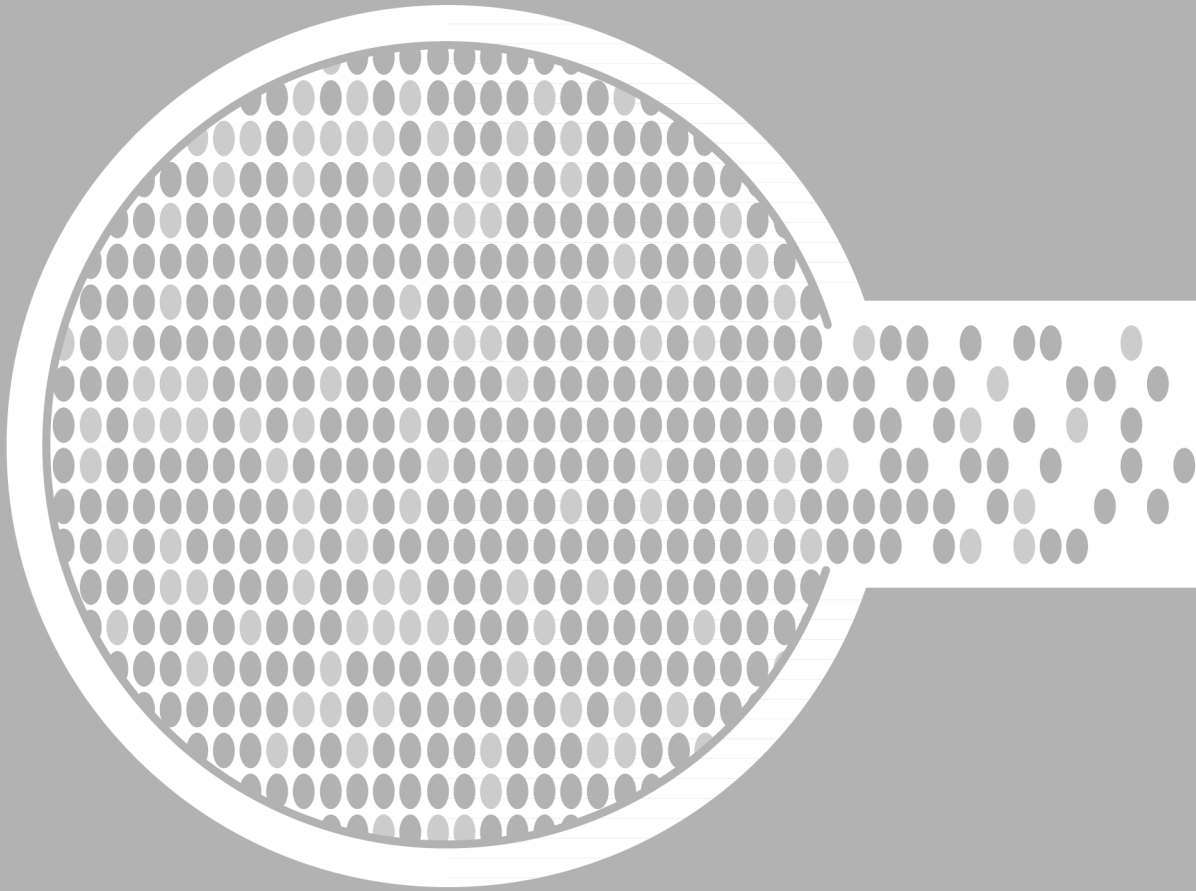
Een belangrijke eerste stap in automatische beeldanalyse is de detectie van celkernen in weefselpreparaten. In Hoofdstuk 3 wordt een automatische detectiemethode beschreven die werkt op H&E preparaten van borstkanker. De methode is gebaseerd op een marker-controlled watershed techniek, die werkt op een aantal schalen en met verschillende markertypes. De multischaal aanpak is geschikt om celkernen van verschillende groottes te detecteren. Dit is noodzakelijk, omdat de grootte van een celkern belangrijk is voor de prognose. De resultaten van de methode zijn geëvalueerd op 18 weefselpreparaten, waarin meer dan 2000 celkernen handmatig waren geannoteerd. De methode was zeer succesvol in het detecteren en segmenteren van de celkernen, evenals in het bepalen van de grootte van de cel. In Hoofdstuk 4 is deze techniek vervolgens gebruikt om verschillende vormkenmerken van cellen te onderzoeken in een populatie van 101 mannen met borstkanker. De grootte van de cel bleek een bepalende factor binnen een univariate overlevingsanalyse. Ook in een multivariate overlevingsanalyse werd dit bevestigd, samen met de grootte van de tumor en de formatie van tubulus.

Hoofdstukken 5 en 6 gaan beide over technieken voor het automatisch detecteren van mitose in H&E preparaten. Dit is een erg lastige en tijdrovende klus voor een patholoog en daarbij ook nog onderhevig aan een grote interbeoordelaarsvariabiliteit. Deze problemen ten spijt, de mitotische activiteit (uitgedrukt in aantal mitosen in een gebied van 2 vierkante millimeter) is echter wel een belangrijke prognostische factor. Automatische detectie van mitose kan de werkdruk van een patholoog verlichten en draagt bij aan een meer objectieve beoordeling.

Recentelijk hebben we een zogenaamde challenge georganiseerd met de titel "Assessment of Mitosis Detection Algorithms 2013" (AMIDA13). Het voor-

naamste doel van deze challenge was verschillende automatische technieken voor mitose detectie op een gemeenschappelijke dataset te evalueren en vergelijken. De dataset bestaat uit 12 preparaten om methoden te trainen en 11 preparaten waarop methoden getest worden, met in totaal meer dan 1000 mitosen aangegeven door verscheidenen beoordelaars. De resultaten van de challenge zijn opgenomen in Hoofdstuk 5. De beste automatische methode presteert vergelijkbaar met de interbeoordelaarsvariabiliteit. De door de methoden gemaakte fouten, de fout-positieve detecties, bleken bij een visuele inspectie veel op mitosen te lijken en ze zouden mogelijk ware mitosen kunnen zijn die gemist werden in deze moeilijke taak. In een herevaluatie bleek dat bijna 30% van de fout-positieven van de beste methode alsnog echte mitose was. Dit resultaat is een indicatie dat de beste automatische methoden van een niveau zijn dat ze in een semi-automatische werkwijze ingezet kunnen worden om de variabiliteit tussen pathologen te verminderen en ook om het proces te versnellen.

In Hoofdstuk 6 wordt een twee-staps methode beschreven voor de automatische lokalisering van mitose. De eerste stap bestaat uit een analyse op basis van de histologische kleuring: mitose is vaak donker gekleurd (hyperchromiciteit, absorptie van licht). Zo worden locaties waar in ieder geval geen mitose plaatsvindt eenvoudig geëlimineerd. De overige locaties worden in de tweede stap geclassificeerd als wel/geen mitose op basis van grootte, vorm, histologische kleuring en textuur. Dit gebeurt nadat de preparaten automatisch zijn bewerkt zodat ze een vergelijkbare, gestandaardiseerde histologische kleuring hebben. De behaalde resultaten zijn vergelijkbaar met die van de op een na beste methode van de AMIDA13 challenge.



Acknowledgements

There were many individuals who were instrumental in making my work a reality, providing guidance, advice, and friendship; this space is dedicated to you.

The completion of this thesis would not have been possible without the tremendous support from my promoters. Throughout my PhD studies they have been a model of excellence for me.

I owe a great debt of gratitude to my daily supervisor and promoter, prof. dr. Josien Plum.

Josien, it would be hard to imagine a more supportive supervisor. You provided me with valuable guidance while leaving me room to explore all my research ideas. The manuscripts I wrote only started to look like proper papers after integrating your comments and suggestions. You always had time to respond to my emails quickly and schedule an extra meeting if needed (and then you didn't even complain when I spent half the time discussing the peculiarities of Dutch life). Thank you very much for all these things.

I would like to thank my promoter, prof. dr. ir. Max Viergever, for giving me the opportunity to work at the Image Sciences Institute (ISI).

Max, you have supported and guided me from the very beginning of my PhD project. Your thoughtful advice helped me see things in a clearer way on many occasions. I consider you to be the person most responsible for the wonderful working atmosphere at the ISI, for which I am very grateful.

I am very grateful to my promoter prof. dr. Paul van Diest.

Paul, throughout my PhD project I was constantly inspired by your energy and enthusiasm. You are one of those rare persons that have both extensive medical and technical knowledge. Despite your busy schedule, you always had a time slot for me. I will always remember our mitosis annotation sessions with the soundtrack provided by Porcupine Tree.

Over the years I have had the pleasure of enjoying the company of many excellent co-workers.

André, thank you for introducing me to the field of Digital Pathology at the beginning of my PhD. Although we worked together only for a short period of

time your instructions were invaluable.

Nikolas, your support was essential for creating some of the datasets used in this research. For this and for all the useful discussions we had I am very grateful.

I am much indebted to all the pathologists from the UMCU Pathology Department that sacrificed part of their valuable time in order to provide annotations for my datasets. Robert, Stefan, Miangela, Anouk thank you. Petra, thank you for your help with arranging sample preparations.

I would like to thank all my co-workers at the ISI for helping me with my work and making my time there very enjoyable. Adriënnne, Alexander, Bas, Bruno, Chantal, Floor, Floris, Gerard, Gerrit, Harriet, Hendrik, Hugo, Ivan, Ivana, Jaap, Jacqueline, Jelmer, Jeroen, Job, Kajo, Kenneth, Koen, Lisette, Maartje, Marco, Marjan, Mark, Martijn, Mieke, Nynke, Paul B., Paul P., Pavel, Pieter, Pim (I prefer “Pym”), Renné, Robin (a.k.a. Thomas), Roel, Sascha, Sjoerd, Sjoerd G., Thessa, Wilbert, Yinghe, Yolanda, thank you.

Although my research topic was not image registration, I was always welcomed at the image registration group (AMOR) meetings. Floris, Harriet, Marijn, Paul P., Robin, Sascha, Yolanda, thank you for sitting through my off-topic presentations and giving me advice.

I would like to give a very special thanks to the secretaries of the ISI Jacqueline, Marjan and Renné for all their help. Gerard, thank you for all the help with setting up the website for the AMIDA challenge.

Willy, thank you for always quickly replying to my emails and arranging appointments with Paul (and for helping me locate that misplaced slide that was so crucial for one of my studies).

Harriet, you were the perfect officemate. Thank you for tolerating my weird taste in music.

It is always good to have a fellow electrical engineer nearby when you need one. Floris, I often busted into your office when I had an idea and needed someone to critically examine it. You always listened carefully and provided me with good advice and suggestions. Your patience has not gone unappreciated.

Settling in the Netherlands can be tricky. Hugo, you were a good instructor of all things Dutch, including such mysterious objects as the *flessenschraper*. I still owe you many beers for the entire help navigating through the Dutch bureaucracy.

Nynke, you are an expert in food which happens to be one of my hobby interests. I always enjoyed our lunchtime conversations on this topic.

Paul B., you were always there to cheer me up and support me with your friendship. Thank you for that and for introducing me to my new hobby of indoor RC helicopter flying.

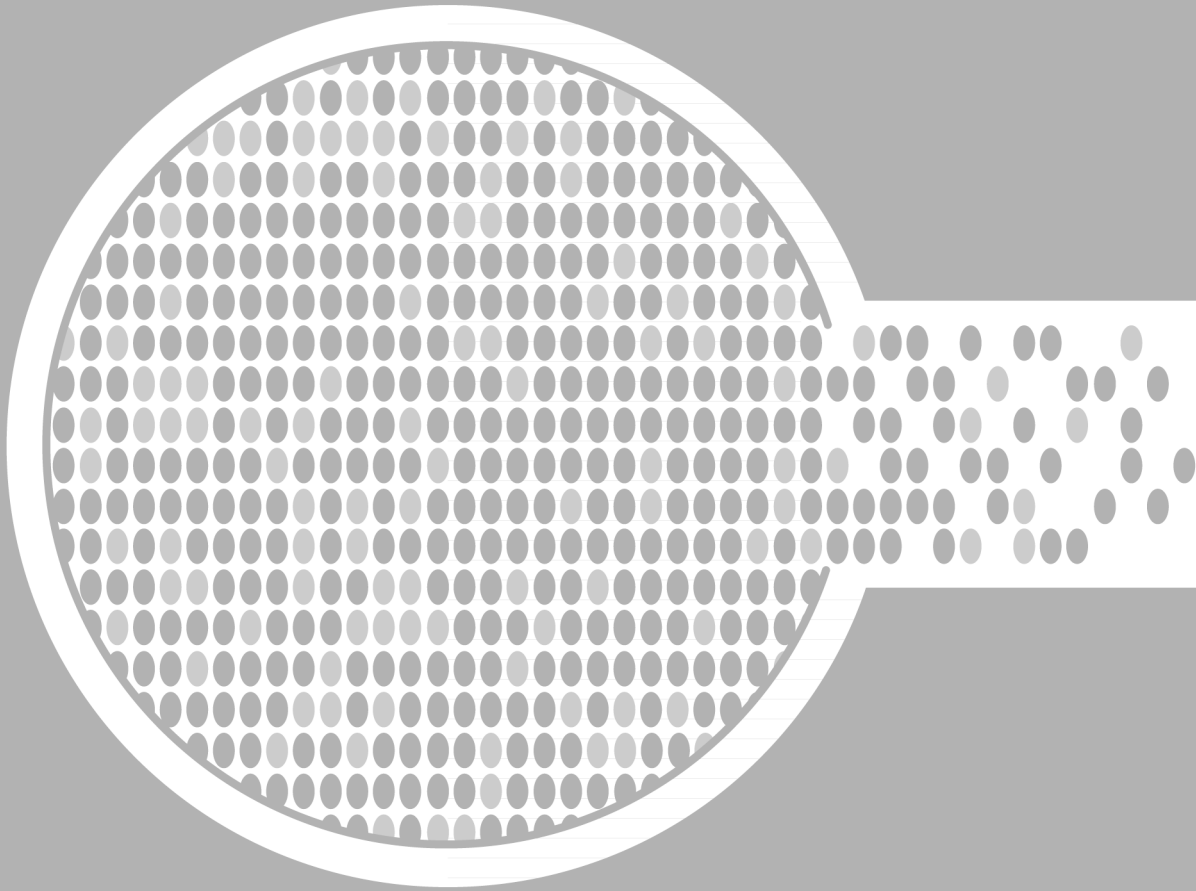
Hendrik and Gerrit, when I started my PhD project you were the “elders” in the PhD student corridor and thus I always sought your advice. Thank you for always finding the time to answer my questions.

I would like to thank Zoran Ivanovski, my supervisor during my work and Master’s studies at the University of SS Cyril and Methodius in Skopje, Macedonia, for encouraging me to pursue a career in science.

Too all my friends in Eindhoven, Utrecht, Skopje and Strumica — thank you for your support and your much-needed distractions from work.

I would like to thank my family for providing me constant support throughout my PhD studies. To my parents, Slobodan and Ljubinka, and my brother Ivan, thank you for always believing in me.

Finally, I would like to thank my best friend and partner Hristina for the constant love and encouragement. Without you this thesis would not have been possible.



List of publications

Papers in international journals

M. Veta, P.J. van Diest, S.M. Willems, H. Wang, Anant Madabhushi, A. Cruz-Roa, F. Gonzalez, A.B.L. Larsen, J.S. Vestergaard, A.B. Dahl, D.C. Cireşan, J. Schmidhuber, A. Giusti, L.M. Gambardella, F.B. Tek, T. Walter, C.W. Wang, S. Kondo, B.J. Matuszewski, F. Precioso, V. Snell, J. Kittler, T.E. de Campos, A.M. Khan, N.M. Rajpoot, E. Arkoumani, M.M. Lacle, M.A. Viergever, J.P.W. Pluim "Assessment of algorithms for mitosis detection in breast cancer histopathology images", 2014, Submitted for review

M. Veta, J.P.W. Pluim, P.J. van Diest, M.A. Viergever, "Breast Cancer Histopathology Image Analysis: A Review", IEEE Transactions on Biomedical Engineering, 2014, vol. 61, pp. 1400-1411

M. Veta, P.J. van Diest, R. Kornegoor, A. Huisman, M.A. Viergever, J.P.W. Pluim, "Automatic Nuclei Segmentation in H&E Stained Breast Cancer Histopathology Images", PLoS One, 2013, vol. 8, p. e70221.

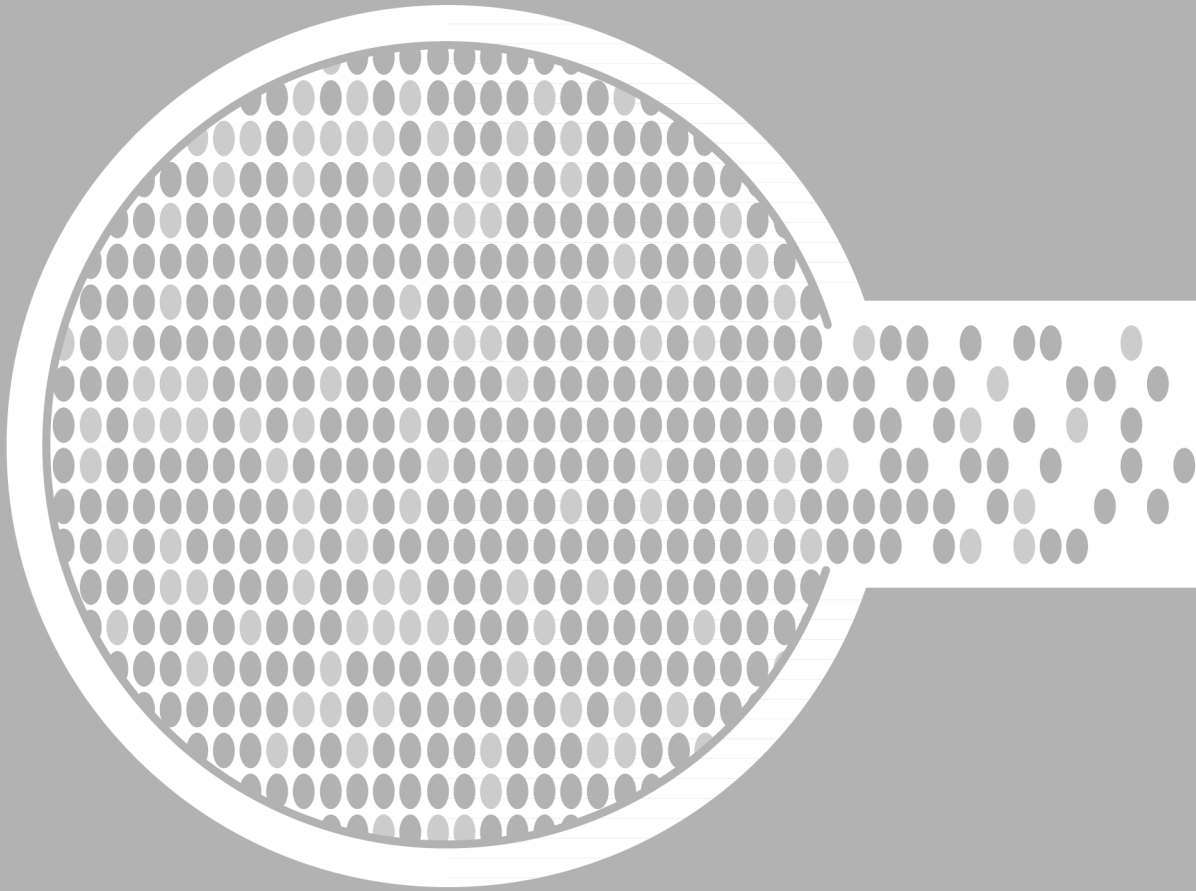
N. Stathonikos, **M. Veta**, A. Huisman, P.J. van Diest, "Going fully digital: Perspective of a Dutch academic pathology lab", Journal of Pathology Informatics, 2013, vol. 4

M. Veta, R. Kornegoor, A. Huisman, A.H.J. Verschuur-Maes, M.A. Viergever, J.P.W. Pluim, P.J. van Diest, "Prognostic value of automatically extracted nuclear morphometric features in whole slide images of male breast cancer", Modern Pathology, 2012, vol. 25, pp. 1559-1565.

Papers in conference proceedings

M. Veta, P.J. van Diest, J.P.W. Pluim, "Detecting mitotic figures in breast cancer histopathology images", in: SPIE Medical Imaging, 2013.

M. Veta, A. Huisman, M.A. Viergever, P.J. van Diest, J.P.W. Pluim, "Marker-controlled watershed segmentation of nuclei in H&E stained breast cancer biopsy images", in: IEEE International Symposium on Biomedical Imaging, 2011.



Curriculum vitae

Mitko Veta was born on August 20th, 1984 in Strumica, Macedonia. From 2003 to 2008 he studied Electrical Engineering at the Ss. Cyril and Methodius University in Skopje, Macedonia. In 2009 he received Master's degree in Digital Signal Processing from the same university with a thesis on digital video classification. Since 2010 he has been working as a PhD candidate at the University Medical Center in Utrecht, The Netherlands on the topic of automatic analysis of histopathology images. The results from this research are presented in this thesis.