

Aus dem Institut für Pathologie  
der Medizinischen Fakultät Charité – Universitätsmedizin Berlin

DISSERTATION

Quantitative Analyse von Zell- und Gewebeeigenschaften in  
histologischen Bilddaten

zur Erlangung des akademischen Grades  
Doctor rerum medicinalium (Dr. rer. medic.)

vorgelegt der Medizinischen Fakultät  
Charité – Universitätsmedizin Berlin

von

Stephan Wienert  
aus Cottbus

Datum der Promotion: 27.02.2015

# Inhaltsverzeichnis

Zusammenfassung .....	2
Abstract .....	3
Einführung.....	4
Zielstellungen .....	5
Methodik .....	6
Validierung der Minimum-Modell-Methode .....	6
Ergebnisse .....	7
Detektion und Segmentierung von Zellkernen in histologischen Bildern: Ein Minimum-Modell Ansatz.....	7
CognitionMaster: ein Bild-Objekt basiertes Bildanalyse Framework .....	9
Marginalzonen-Lymphome mit monozytoider Morphologie exprimieren T-bet und sind mit einer geringen Anzahl von extranodalen T-Zellen assoziiert.....	11
Mehrfarben-Abstammungsanalyse bei der Homöostase von Langerhans-Zellen.....	12
Abstammungsanalyse offenbart Ursprung und Funktion von follikulären Dendritischen Zellen des Lymphknotens .....	13
Diskussion .....	14
Literaturverzeichnis.....	16
Eidesstattliche Versicherung .....	17
Lebenslauf .....	19
Publikationsliste .....	20
Danksagung.....	22

# Zusammenfassung

## Einleitung

Aufgrund technischer Entwicklungen und einem wachsenden Bedarf an einer Standardisierung der qualitativen und quantitativen Auswertungen in der histopathologischen Forschung und Routinediagnostik hat das Forschungsfeld der automatisierten Mikroskop-Bild-Analyse in jüngster Zeit neue Impulse erhalten. Die automatische Detektion und Segmentierung der relevanten Strukturen, zumeist Zellkerne, ist die Grundlage für die computergestützte Messung von Zell- und Gewebeeigenschaften. Bis heute existiert hierfür allerdings keine universelle Standard-Lösung, die wichtigste Herausforderung ist hier die Heterogenität der verschiedenen Gewebe- und Tumortypen. Desweiteren ist die Pixel-basierte Modellierung von Bildanalyse-Datenstrukturen nach wie vor der weit verbreitete Standard, obwohl die Verwendung von Bild-Objekten als primäre Verarbeitungseinheit eher der menschlichen Wahrnehmung entspricht und besser für die Berechnung von Objektmerkmalen sowie für die Analyse von Nachbarschafts-Beziehungen geeignet ist.

## Ergebnisse

In dieser Arbeit wurde eine neue Methode zur Detektion und Segmentierung von Zellkernen entwickelt, die nur minimale a-priori Informationen verwendet und dadurch auch auf typischerweise heterogene histologische Bilddaten robust anwendbar ist. Weiterhin wurde die Open-Source Software-Plattform *CognitionMaster* entworfen und implementiert, die eine Bild-Objekt-basierte Modellierung der Datenstrukturen enthält und damit Limitierungen der konventionellen Pixel-basierten Systeme überwindet. Dieses Framework wurde für die computergestützte Bestimmung der Anzahl von CD3-positiven T-Lymphozyten in extranodalen Marginalzonen-Lymphomen, zur Untersuchung der Homöostase von Langerhans-Zellen der Epidermis sowie zur Analyse des Ursprungs und der Funktion von follikulären Dendritischen Zellen des Lymphknotens verwendet.

## Schlussfolgerungen

Es wurde gezeigt, dass die neu entwickelte Minimum-Modell-Methode eine robuste Detektion von Zellkernen auch bei morphologisch heterogenen Bilddaten ermöglicht. Mit der *CognitionMaster* Software-Plattform konnten für drei zum Teil sehr unterschiedliche Fragestellungen aus der Biologie und Pathologie erfolgreich Quantifizierungsverfahren implementiert werden, womit der proof-of-concept bereits für mehrere Problemklassen erbracht wurde.

# Abstract

## **Introduction**

Due to technical developments and an increased demand for a standardization of qualitative and quantitative analyses in histopathological research and routine diagnostics the field of automated analysis of microscopic images has recently received new impetus. The automatic detection and segmentation of the relevant structures, mostly cell nuclei, is pivotal for the computer-aided measurement of cell and tissue features. Until today, no universal standard solution exists for this task. Here, the heterogeneity of the different tissue and tumor types is the biggest challenge. Furthermore, pixel-based modeling of data structures is still the widespread standard in digital image analysis although the usage of image objects as primary processing unit relates well to human perception and is more suitable for the computation of object features and the analysis of neighbourhood relationships.

## **Results**

This thesis describes a novel approach for the detection and segmentation of cell nuclei. This method requires only minimal a-priori information and is therefore capable of dealing with typically morphologically heterogeneous histological image data. Additionally, the open source software platform *CognitionMaster* was designed and implemented. This framework features an image object-based modeling of data structures and therefore overcomes limitations of conventional pixel-based systems. The platform was used for the computer-aided quantification of CD3-positive T-lymphocytes in extranodal marginal zone lymphomas, for the analysis of the homeostasis of epidermal Langerhans cells and to study origin and dynamics of lymph node follicular Dendritic Cells.

## **Conclusions**

It was shown that the newly developed minimum-model method allows for a robust detection of cell nuclei in morphologically heterogeneous image data. Based on the *CognitionMaster* platform quantification systems were set up for three different tasks in biology and pathology whereby the proof-of-concept was already provided for multiple classes of problems.

## Einführung

Die automatische Mikroskop-Bild-Analyse hat in der letzten Zeit wieder eine erhöhte Aufmerksamkeit erhalten. Das liegt zum einen an der gestiegenen Verfügbarkeit von digitalen histologischen Bildern. Hier spielt die Entwicklung und Verbreitung von Mikroskop-Kameras und das Aufkommen der sogenannten digitalen Slide Scanner eine wesentliche Rolle. Zum anderen ist in den letzten Jahren der Bedarf an einer Standardisierung der qualitativen und quantitativen Auswertungen in der histopathologischen Forschung und Routinediagnostik gestiegen: Während die menschliche Wahrnehmung im Bereich der Mustererkennung sehr gut funktioniert, zeigen sich deutliche Schwächen bei der objektiven quantitativen Einschätzung von Merkmalen wie z.B. Form, Farbe, Größe oder Zellzahl. Diese Eigenschaften spielen bei der Beurteilung von histologischen Präparaten eine entscheidende Rolle. Viele Studien zeigen daher, dass eine rein visuelle, objektive und reproduzierbare Auswertung von Zell- und Gewebeeigenschaften schwierig ist [1–4]. In der Forschung, insbesondere der patientenorientierten klinischen Forschung und der Routine-Diagnostik kommt es aber darauf an, die relevanten Merkmale in verschiedenen Zentren zuverlässig und mit identischem Resultat zu bestimmen, um valide Grenzwerte zu etablieren. Im Vergleich zur rein visuellen Auswertung durch den Pathologen lassen sich Eigenschaften des histologischen Gewebes (z.B. Größe und Form von Zellen bzw. Zellkernen, Positivität einer immunhistochemischen Färbung) durch die Verwendung von computergestützten Methoden „exakt“ messen, die Voraussetzung hierfür ist allerdings die robuste Detektion der relevanten Strukturen, zumeist Zellkerne. Eine Vielzahl von Verfahren und Computersystemen sind in der Literatur beschrieben, z.B. [5–9], eine Standard-Lösung für diese Aufgabe existiert bis heute jedoch nicht. Ein Problem besteht darin, dass existierende Lösungen häufig durch eine Vielzahl von Parametern an das Bildmaterial angepasst bzw. „angelernt“ werden müssen oder Modellinformationen über eine „Idealzelle“ verwendet werden. Dies ist problematisch: zum einen wegen der hohen biologischen und technischen Variabilität, die eine Anpassung eines Verfahrens durch die Einstellung von Parametern mit steigender Größe (und damit Heterogenität) des Datensatzes schwieriger macht, und zum anderen, weil gerade die Veränderungen des Gewebes eine zentrale Rolle in der Diagnostik spielen. Hier könnte die Verwendung von Zell-Modellinformationen zu einem Bias bei der Zell-Erkennung führen.

Ein weiteres Problem ist das Fehlen einer Programmierplattform mit der sich die biologischen Strukturen im digitalen Bild adäquat beschreiben und verarbeiten lassen: Die Verarbeitung von Pixeln, der kleinsten Einheit des digitalen Bildes, ist bis jetzt der weitverbreitete Standard für Bildanalyse-Systeme. Dieses Konzept ist jedoch nur begrenzt geeignet, um z.B.

Objektmerkmale zu berechnen, Nachbarschafts-Beziehungen zwischen Bild-Objekten (z.B. Zellkern zu Zellplasma und Zellmembran) zu analysieren oder Nutzerinteraktionen auf der „Objekt-Ebene“ (z.B. Auswahl von bestimmten Zellen) umzusetzen [10]. Die übliche Verwendung von Markierungs-Bildern (Label-Bildern) und Werte-Listen zur Repräsentation von Segmentierungs- bzw. Klassifikations-Ergebnissen ist umständlich und entspricht zudem nicht den Standards der modernen objektorientierten Programmierung. Die Modellierung der Datenstrukturen zur Beschreibung von Bild-Objekten [10–13] (anstatt von Bild-Pixeln) entspricht eher der menschlichen Wahrnehmung und verspricht hier eine verbesserte Verständlichkeit und Wiederverwendbarkeit und damit insgesamt eine höhere Leistungsfähigkeit. Eine frei verfügbare Software-Plattform, die die Datenstrukturen für die digitale Bildanalyse Bild-Objekt-basiert nach Gesichtspunkten der modernen Software-Entwicklung (z.B. Datenkapselung und Wiederverwendbarkeit) modelliert und mit einer hohen Anzahl an Implementierungen von Bildanalyse-Algorithmen, einer anwenderfreundlichen Benutzeroberfläche sowie der Erweiterbarkeit durch leistungsstarke Schnittstellen verbindet, ist bisher nicht vorhanden.

### ***Zielstellungen***

Ziel dieser Arbeit war es eine Methode zur robusten Detektion und Segmentierung von Zellkernen zu entwickeln und zu validieren. Weiterhin sollte eine leistungsfähige Programmierplattform entworfen und implementiert werden, die eine Objekt-basierte Beschreibung und Verarbeitung der histologischen Strukturen erlaubt. Der proof-of-concept sollte mit der Implementierung von Quantifizierungsverfahren für verschiedene aktuelle, relevante Fragestellungen der Biologie und Pathologie erbracht werden.

## Methodik

Die in dieser Arbeit entwickelten Algorithmen und Datenstrukturen wurden implementiert in C# für Microsoft .NET. Als Entwicklungsumgebung wurde SharpDevelop [14] verwendet. Statistische Auswertungen wurden mit Microsoft EXCEL durchgeführt.

### **Validierung der Minimum-Modell-Methode**

Für die Validierung der Minimum-Modell-Methode (siehe *Ergebnisse*) wurde ein Goldstandard-Datensatz bestehend aus 7931 manuell markierten Zellen erstellt. Dazu wurden Großflächenschnitte und Tissue Micro Arrays (TMA) aus dem Institut für Pathologie, Charité, Universitätsmedizin Berlin genutzt. Die Patienten haben der Verwendung des histologischen Materials in der Forschung zugestimmt. Die TMAs wurden mit einem Micro-Arrayer (*Beecher Instruments*, Woodland, CA) erstellt. Alle Schnitte wurden 3µm dick geschnitten und Hämatoxylin&Eosin entsprechend dem Routine-Verfahren gefärbt. Die histologischen Schnitte wurden mit einem *Zeiss Mirax Slide Scanner (3DHitech, Ungarn)* mit einer Auflösung von 0,23µm x 0,23µm pro Pixel digitalisiert. Insgesamt wurden 7931 Zellkerne aus 36 zufällig ausgewählten Einzelbildern von je drei Pathologen markiert. Nur Zellen, bei denen Konsens bestand, wurden in den Datensatz übernommen. Die Bilddaten des Validierungs-Datensatzes wurden vollautomatisch mit der Minimum-Modell-Methode verarbeitet und die Zell-Positionsdaten verglichen. Richtig-positive (*tp*), falsch-positive (*fp*) sowie falsch-negative (*fn*) Ereignisse wurden gezählt und die folgenden statistischen Merkmale berechnet:

$$Precision=tp/(tp+fp)$$

$$Recall=tp/(tp+fn)$$

$$Conglomerate=(\text{Anzahl gefundener Objekte}-fp)/tp$$

## Ergebnisse

### ***Detektion und Segmentierung von Zellkernen in histologischen Bildern: Ein Minimum-Modell Ansatz***

*Wienert S, Heim D, Saeger K, Stenzinger A, Beil M, Hufnagl P, Dietel M, Denkert C, Klauschen F. Detection and segmentation of cell nuclei in virtual microscopy images: a minimum-model approach. Sci Rep 2012;2:503.*

Zur Detektion und Segmentierung von Zellkernen wurde die sog. Minimum-Modell-Methode entwickelt. Diese basiert auf der Verfolgung von Objekt-Konturen, deren Anfangspunkte durch zeilenweises Abtasten des digitalen Bildes auf Suchlinien gefunden werden, wie beschrieben in [15]. Im Gegensatz zum zuvor verwendeten Binärbild werden jetzt Bildzeilen des Graustufenbildes als eindimensionale Funktion  $I(x)$  interpretiert und die erforderlichen Parameter (Konturanfangspunkt, minimale und maximale Objektintensität) für die sich anschließende zweidimensionale Konturverfolgung vollautomatisch bestimmt (siehe Abbildung 2 und 3 in Publikation 1). Bei diesem Vorgehen (Primär-Segmentierung) können Objekt-Konturen mehrfach und in unterschiedlicher Qualität gefunden werden. Daher müssen die gefundenen Konturen bewertet und für die enthaltenen Bild-Objekte die jeweils am besten beschreibende Kontur bestimmt werden. Dafür werden der mittlere Gradient entlang der Kontur und der sogenannte *GradientFit* verwendet, d.h. die relative Anzahl der Konturpunkte die ein lokales Maximum im Gradienten-Bild sind. Da die jeweils lokal „prominenteste“ Kontur ausgewählt wird, können Objekte unabhängig von deren Form, Intensität und Gradient zur Objekt-Umgebung gefunden werden. Lediglich die Objektgröße muss eingeschränkt werden um die Erkennung von Super- bzw. Substrukturen zu vermeiden. Aus den bewerteten Konturen der Primär-Segmentierung werden anschließend eine nicht-überlappende Segmentierung des Originalbildes erstellt, nicht-kompakt liegende Objektpixel entfernt und Zellkonglomerate getrennt. Abschließend werden alle Objekte mittels eines Histogramm-basierten Schwellwertes [16] im Hämatoxylin-Signal [17] in „Zellkern“ und „Hintergrund“ klassifiziert.

Die Validierung der Minimum-Modell-Methode mit dem Goldstandard-Datensatz (siehe *Methodik*) ergab eine *Precision* von 0,908, einen *Recall* von 0,859 und einen *Conglomerate*-Wert von 0,953 bei einer mittleren Verarbeitungszeit von 725,4ms pro Bild. Weiterhin wurde ein Vergleich mit der Methode von Al-Kofahi et al. [9] durchgeführt. Dafür wurde eine vollautomatische Version (wie von den Autoren auf Nachfrage empfohlen) und eine Version



mit einem (von den Autoren nach Zusenden von Testdaten) optimiertem Parametersatz verwendet (Tabelle 1).

	<b>Minimum- Modell- Methode</b>	<b>Al-Kofahi et al. (vollautomatisch)</b>	<b>Al-Kofahi et al. (optimiert)</b>
<i>Precision</i>	<b>0,908</b>	<b>0,711</b>	<b>0,823</b>
<i>Recall</i>	<b>0,859</b>	<b>0,908</b>	<b>0,908</b>
<i>Conglomerate</i>	<b>0,953</b>	<b>0,964</b>	<b>0,966</b>
<b>Mittlere Verarbeitungszeit (ms)</b>	<b>725,4±182,3</b>	<b>678,2±49,5</b>	<b>1942,6±326,7</b>

**Tabelle 1: Vergleich der Minimum-Modell-Methode mit dem Verfahren von Al-Kofahi.**

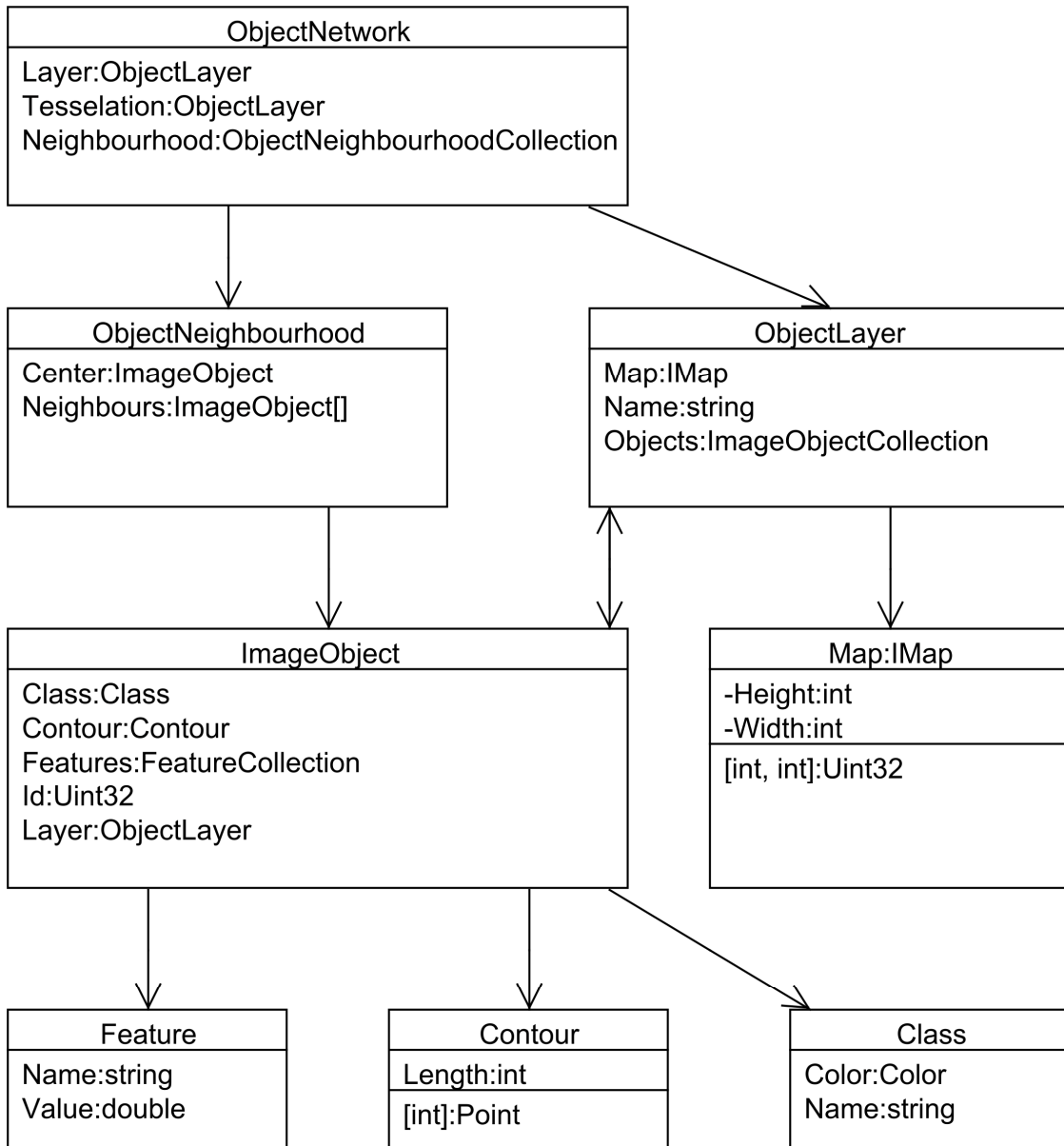
Quelle: Publikation 1.

Die vollautomatische Version von Al-Kofahis Methode ergab verglichen mit der Minimum-Modell-Methode eine niedrigere Verarbeitungszeit, aber eine um 0,197 niedrigere *Precision* ( $p=5,97 \cdot 10^{-10}$ ) und einen um 0,049 höheren *Recall* ( $p=2,03 \cdot 10^{-7}$ ). Mit dem optimierten Parametersatz blieb die *Precision* um 0,085 niedriger ( $p=5,60 \cdot 10^{-7}$ ), bei einem konstanten *Recall* (0,908,  $p=2,24 \cdot 10^{-7}$ ) und einer 2,6 fach höheren durchschnittlichen Verarbeitungszeit.

## ***CognitionMaster: ein Bild-Objekt basiertes Bildanalyse Framework***

*Wienert S, Heim D, Kotani M, Lindequist B, Stenzinger A, Ishii M, Hufnagl P, Beil M, Dietel M, Denkert C, Klauschen F. CognitionMaster: an object-based image analysis framework. Diagn Pathol 2013;8:34.*

Für die adäquate Beschreibung und Verarbeitung von histologischen Strukturen wurde die Open Source Bildanalyse-Plattform *CognitionMaster* entworfen und implementiert, die auf einer Bild-Objekt-basierten Modellierung der Datenstrukturen basiert (Abbildung 1): Die Klasse *ImageObject* repräsentiert eine kohärente Gruppe von Pixeln, die zusammen ein Segment bzw. Bild-Objekt darstellen. Eine Vielzahl von Eigenschaften können benutzt werden um ein Bild-Objekt zu beschreiben, wie z.B. *Contour* für die räumliche Position (der Objektgrenze) oder eine Liste von Schlüssel-Wert-Paaren (*Features*) für die Objekt-Eigenschaften (z.B. die mittlere Intensität). Die Klasse *ObjectLayer* wurde eingeführt, um alle Informationen zu dem Ergebnis eines Verarbeitungsschritts einer Bildanalyse-Verarbeitungskette strukturiert bereitzustellen. Dazu gehört eine Liste von *ImageObject* Instanzen sowie eine zweidimensionale Karte (*Map*) mit derselben Größe wie das korrespondierende Bild in der jeweils die ID des Bild-Objekts eingetragen ist, zu dem ein Pixel zugeordnet ist. Mittels diskreter Voronoi-Tessellation [18] können Nachbarschafts-Beziehungen zwischen Bild-Objekten berechnet werden. Diese werden dann in einem *ObjectNetwork* zusammengefasst und können verwendet werden, um z.B. die Anzahl der benachbarten oder berührenden Bild-Objekte einer bestimmten Klasse zu bestimmen. Die entstandene Software ist ein leistungsfähiges Werkzeug bestehend aus mehr als 250 Klassen (Bildanalyse-Algorithmen, Datenstrukturen, visuelle Komponenten), das als Basis für interaktive sowie vollautomatische Bildanalyse Anwendungen genutzt werden kann. Zum Funktionsumfang gehören u.a. Verarbeitungsketten, In-Program Scripting, Objekt-Modelle sowie interaktive 2D-Scatterplot Diagramme. Die Plug-In Schnittstelle erlaubt die Erweiterung der Software z.B. um zusätzliche Menüeinträge, Registerkarten oder unterstützte Bildformate. Plug-Ins können ebenfalls das Standard-Verhalten bei Maus- oder Tastatur-Ereignissen verändern.



**Abbildung 1:** Auszug aus dem Klassendiagramm mit den wichtigsten Datenstrukturen für die Bild-Objekt-basierte Datenverarbeitung: die Klasse *ImageObject* fasst all objektbezogenen Eigenschaften zusammen (z.B. Klassifikation (*Class*), Kontur (*Contour*) und Objekteigenschaften (*Feature*)). Alle Bild-Objekte (eines bestimmten Verarbeitungsschritts) ergeben eine Objektebene (*ObjectLayer*).  
Quelle: Publikation 2.

## ***Marginalzonen-Lymphome mit monozytoider Morphologie exprimieren T-bet und sind mit einer geringen Anzahl von extranodalen T-Zellen assoziiert***

*Lohneis P, Wienert S, Klauschen F, Ullrich A, Anagnostopoulos I, Jöhrens K. Marginal zone lymphomas with monocytoid morphology express T-bet and are associated with a low number of T cells in extranodal locations. Leuk Lymphoma 2013:1-16.*

Marginalzonen-Lymphome (MZL) umfassen eine heterogene Gruppe von indolenten Non-Hodgkin-Lymphomen: Nach der Klassifikation der World Health Organization (WHO) werden splenische, nodale und extranodale MZLs unterschieden. Das geringfügige Auftreten von infiltrierenden T-Lymphozyten ist ein charakteristisches Merkmal von monozytoiden B-Zell-Proliferaten in reaktiven Läsionen des Lymphknotens. Um zu untersuchen ob dieses Merkmal auch auf MZLs mit monozytoider B-Zell-Morphologie zutrifft, wurde die relative Anzahl CD3-positiver T-Lymphozyten bildanalytisch quantifiziert. Obwohl CD3 eine membranständige Färbung ist, war in vielen Präparaten eine Anfärbung des gesamten Zellkerns sowie eine starke unspezifische Färbung im Bildhintergrund zu beobachten. Dies führte in vielen Fällen dazu, dass in den entsprechenden Bildern Zellgrenzen auch visuell äußerst schwierig bzw. nicht sicher festzustellen waren, so dass die automatische Detektion und Segmentierung der Zellkerne nicht anwendbar war. Um die Auswertung dennoch computergestützt vornehmen zu können, wurde hier ein einfacherer heuristischer Ansatz gewählt: Ausschnittbilder der mit Hämatoxylin und DAB bzw. AP Red gefärbten Schnitte wurden bei 400x Vergrößerung erstellt und in 5x5 Pixel große Kacheln unterteilt. Kacheln mit einem positiven Hämatoxylin-Signal und negativem CD3-Signal wurden als CD3-negative „Zelle“ und Kacheln mit einem positiven Hämatoxylin-Signal sowie positivem CD3-Signal wurden als CD3-positive „Zelle“ definiert. Kacheln mit einem negativen Hämatoxylin-Signal wurden als Bildhintergrund verworfen. Das Hämatoxylin- sowie das CD3-Signal wurden mittels Color Deconvolution [17] aus dem RGB-Farbbild extrahiert. Der Schwellwert zur Bestimmung der Positivität wurde nach Otsu [16] bestimmt. Dabei galt eine Kachel als positiv, wenn mindestens 50% der enthaltenen Pixel einen Wert oberhalb des Schwellwerts aufwiesen. Der Quantifizierungs-Algorithmus wurde als Plugin für den *CognitionMaster* implementiert und interaktiv durch den auswertenden Pathologen verwendet. In dieser Studie wurde eine signifikant niedrigere Anzahl ( $p < 0,01$ ) von T-Lymphozyten in extranodalen MZLs mit monozytoider Morphologie gegenüber extranodalen MZLs mit anderer Morphologie nachgewiesen.

## **Mehrfarben-Abstammungsanalyse bei der Homöostase von Langerhans-Zellen**

Ghigo C, Mondor I, Jorquera A, Nowak J, Wienert S, Zahner SP, Clausen BE, Luche H, Malissen B, Klauschen F, Bajenoff M. Multicolor fate mapping of Langerhans cell homeostasis. *J Exp Med* 2013;184:2417.

Langerhans-Zellen (Langerhans Cell, LC) sind Zellen des Immunsystems in der Epidermis. LCs dienen der Antigenpräsentation und werden während der Embryogenese gebildet. Um eine konstante Anzahl von LCs in der Epidermis zu erhalten muss deren kontinuierliche Abwanderung in naheliegende Lymphknoten ausgeglichen werden. Da LCs nicht durch zirkulierende Vorläuferzellen erneuert werden, wird angenommen, dass LCs lokal in der Epidermis proliferieren [19–21], ein experimenteller Nachweis dieses Modells fehlt allerdings bisher. In dieser Studie wird ein Mehrfarben-Abstammungsanalyse-System (Fate-Mapping-System) basierend auf der *Cre-loxP* Technologie vorgestellt, das sich für die Analyse der Zellhomöostase eignet, da Zellen ihre einmal angenommene spezifische Expression von fluoreszierenden Reporter-Proteinen über die gesamte Lebensdauer erhalten und auch an ihre Tochterzellen weitergeben. Initial exprimieren alle Zellen das Fluoreszenzprotein dTomato. Durch ein Cyclization-Recombination-Event (*Cre*-Event) ändern die Zellen ihre Expression zufällig zu entweder YFP (~80%) oder CFP (~20%). Unmittelbar nach dem *Cre*-Event liegt also eine zufällige räumliche Verteilung von CFP<sup>+</sup> und YFP<sup>+</sup> Zellen vor. Würde jede Zelle eine in der Nachbarschaft entstandene Lücke durch Zellteilung wieder „füllen“, könnte erwartet werden, dass diese Verteilung über die Zeit konstant zufällig bleibt. Im Gegensatz dazu könnte erwartet werden, dass kompakte Konglomerate von CFP<sup>+</sup> und YFP<sup>+</sup> Zellen über die Zeit entstehen, wenn nur einige wenige Zellen lokal proliferieren. Für eine quantitative Auswertung wurde die Software *ClusterQuant* basierend auf dem *CognitionMaster* Framework entworfen und implementiert. CFP<sup>+</sup> und YFP<sup>+</sup> Zellen wurden in mikroskopischen Bildern manuell markiert. Die Zellpositionen wurden von *ClusterQuant* eingelesen und deren räumliche Nachbarschaftsbeziehungen mittels Voronoi-Tessellation [18] analysiert. Anschließend wurden zusammenhängende Cluster von benachbarten Zellen desselben Typs bestimmt. Als Maß für die Stärke der Clusterbildung wurde die Anzahl der Zellen pro Cluster berechnet und mit zufälligen Verteilungen aus je 10.000 Monte-Carlo-Simulationen verglichen. Zusammenfassend ergab die Untersuchung, dass die Clusterbildung der LCs mit  $p < 10^{-9}$  (Median) hoch signifikant nicht-zufällig ist. Die Experimente bestätigten damit das Modell, dass einige wenige LCs lokal in der Epidermis proliferieren, die eine Art Vorläufer-Zellen darstellen.

## ***Abstammungsanalyse offenbart Ursprung und Funktion von follikulären Dendritischen Zellen des Lymphknotens***

*Jarjour M, Jorquera A, Mondor I, Wienert S, Narang P, Coles MC, Klauschen F, Bajenoff M. Fate mapping reveals origin and dynamics of lymph node follicular Dendritic Cells. J Exp Med, Manuskript akzeptiert.*

Follikuläre Dendritische Zellen (Follicular Dendritic Cell, FDC) sind Zellen des Immunsystems, die in Primär- und Sekundärfollikeln von lymphatischem Gewebe vorkommen. Sie beeinflussen die Homöostase der B-Zellen sowie Immunreaktionen in sog. sekundären lymphatischen Organen. Über ihre Biologie ist bisher allerdings wenig bekannt. Das in der im vorangehenden Abschnitt beschriebenen Arbeit präsentierte Mehrfarben-Fate-Mapping-System wurde hier verwendet, um in verschiedenen Experimenten Einblicke in den Ursprung und die Funktion von FDCs zu erhalten. Dabei wurde die Clusterbildung der Zellen wiederum mit *ClusterQuant* basierend auf dem *CognitionMaster* Framework analysiert und mit jeweils 10.000 zufälligen Simulationen (Monte-Carlo-Simulation) verglichen. Die Untersuchungen haben ergeben, dass sich FDCs während einer Immunreaktion in Keimzentren anhäufen, jedoch weder die Rekrutierung von zirkulierenden Vorläufern noch die Proliferation von lokalen ausgereiften FDCs signifikant dazu beitragen. Weiterhin konnte nachgewiesen werden, dass FDCs aus der klonalen Expansion und Differenzierung von marginal-retikulären Zellen entstehen.

## Diskussion

Die Verwendung von Computersystemen für die qualitative und quantitative Analyse von Zell- und Gewebeeigenschaften in der histologischen Forschung und Diagnostik verspricht objektive und reproduzierbare Ergebnisse und damit eine Verbesserung gegenüber der rein visuellen Auswertung, z.B. durch den Pathologen. Dazu ist allerdings die automatische Detektion und Segmentierung der relevanten Strukturen im histologischen Bild, zumeist Zellkerne, erforderlich. Hierzu wurden seit mehreren Jahrzehnten viele Verfahren entwickelt und publiziert, z.B. [5–9]. Wegen der Komplexität und Vielfalt der Problemstellungen (z.B. Organ-spezifische Histoarchitektur, biologische Vielfalt, diverse Färbungen, technische Variabilität) ist der Anwendungsbereich dieser Verfahren in der Regel jedoch sehr eingeschränkt. In dieser Arbeit wurde eine Minimum-Modell-Methode entwickelt, die Zellkerne mit einer hohen Genauigkeit erkennt. Die Methode verwendet nur ein Minimum an a-priori Informationen und Parametern und ist daher in der Lage ohne eine Anpassung, Datensätze mit einer hohen morphologischen Vielfalt zu verarbeiten, und kann deswegen als Grundlage für die Forschung und Anwendung in der Pathologie dienen. Weiterhin wurde in dieser Arbeit das *CognitionMaster* Framework mit einer Bild-Objekt-basierten Modellierung der Datenstrukturen entworfen und implementiert. Dieses wurde verwendet, um Quantifizierungsverfahren für drei zum Teil sehr unterschiedliche Fragestellungen zu implementieren: Im Bereich der Pathologie, um mit der Quantifizierung von CD3-positiven T-Lymphozyten weiteren Aufschluss über Marginalzonen-Lymphome zu erhalten. Im Bereich der Biologie, um mit der Quantifizierung der Clusterbildung von Zellen im Rahmen der Abstammungsanalyse Erkenntnisse über die Homöostase von Langerhans-Zellen der Epidermis sowie den Ursprung und die Funktion von follikulären Dendritischen Zellen des Lymphknotens zu gewinnen. Die dargestellten Fragestellungen hätten zwar prinzipiell auch mit konventionellen Pixel-basierten Systemen untersucht werden können. Die theoretischen Vorteile des Objekt-basierten Konzepts haben sich allerdings (wenn auch schwer zu beziffern) in der praktischen Anwendung bestätigt. In der modernen Informatik geht man davon aus, dass mit einer Abbildung der „Realität“ in den Datenstrukturen (siehe Abb. 1), im Sinne der objektorientierten Programmierung, Software effizienter und verständlicher gestaltet werden kann. Dies zeigt sich auch in der Verbreitung vollständig objektorientierter Programmiersprachen, wie z.B. *Java* und *C#*. In konventionellen Pixel-basierten Systemen fehlt eben diese Abbildung der „Realität“ gänzlich. Die Leistungsfähigkeit der entwickelten Komponenten und Methoden zeigt sich außerdem in den sich anschließenden weiterführenden Projekten: die Minimum-Modell-Methode und das *CognitionMaster* Framework wurden

verwendet um den Proliferationsmarker Ki67 beim Mammakarzinom automatisch auszuwerten. Diese Scoring-Methode wurde an den Überlebensdaten von mehr als 1000 Patientinnen der *GeparTrio* Studie (NCT00544765) getestet (Manuskript eingereicht) und wird zusätzlich aktuell an der noch laufenden *Genevieve*-Studie (NCT01779479) prospektiv validiert. Weiterhin wurden die Minimum-Modell-Methode und die *CognitionMaster*-Plattform verwendet um CD15-positive neutrophile Granulozyten zur Diagnostik periprothetischer Infektionen zu quantifizieren (Manuskript eingereicht). Ein weiteres Einsatzfeld ist die Quantifizierung sogenannter Tumor-Infiltrating-Lymphocytes (TILs). Hier wurde die Anzahl von Lymphozyten in zuvor manuell markierten intra-tumoralen Stroma-Bereichen in den Bilddaten der *GeparSixto*-Studie (NCT01426880) bestimmt (Manuskript eingereicht). Zusammenfassend wurden in dieser Arbeit leistungsfähige Komponenten und Methoden entwickelt die auch in Zukunft für die quantitative Analyse von histologischen Bilddaten verwendet werden können.



## Literaturverzeichnis

1. Polley MC, Leung SCY, McShane LM et al. An international Ki67 reproducibility study. *J Natl Cancer Inst* 2013;105:1897-1906.
2. Vörös A, Csörgő E, Nyári T, Cserni G. An intra- and interobserver reproducibility analysis of the Ki-67 proliferation marker assessment on core biopsies of breast cancer patients and its potential clinical implications. *Pathobiology* 2013;80:111-118.
3. Varga Z, Diebold J, Dommann-Scherrer C et al. How reliable is Ki-67 immunohistochemistry in grade 2 breast carcinomas? A QA study of the Swiss Working Group of Breast- and Gynecopathologists. *PLoS ONE* 2012;7:e37379.
4. Umemura S, Osamura RY, Akiyama F et al. What causes discrepancies in HER2 testing for breast cancer? A Japanese ring study in conjunction with the global standard. *Am J Clin Pathol* 2008;130:883-891.
5. Bibbo M, Bartels PH, Dytch HE, Wied GL. Computed cell image information. *Monogr Clin Cytol* 1984;9:62-100.
6. Bengtsson E. The measuring of cell features. *Anal Quant Cytol Histol* 1987;9:212-217.
7. Naik S, Doyle S, Agner S et al. Automated gland and nuclei segmentation for grading of prostate and breast cancer histopathology. 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro. *ISBI 2008*;284-287.
8. Tuominen VJ, Ruotoistenmäki S, Viitanen A, Jumppanen M, Isola J. ImmunoRatio: a publicly available web application for quantitative image analysis of estrogen receptor (ER), progesterone receptor (PR), and Ki-67. *Breast Cancer Res* 2010;12:R56.
9. Al-Kofahi Y, Lassoued W, Lee W, Roysam B. Improved automatic detection and segmentation of cell nuclei in histopathology images. *IEEE Trans Biomed Eng* 2010;57:841-852.
10. Blaschke T, Strobl J. What's wrong with pixels? Some recent developments interfacing remote sensing and GIS. *GIS – Zeitschrift für Geoinformationssysteme*;2001:12-17.
11. Ghassemian H, Landgrebe D. Object-oriented feature extraction method for image data compaction. *IEEE Control Syst Mag* 1988;8:42-48.
12. Hay GJ, Castilla G, Wulder MA, Ruiz JR. An automated object-based approach for the multiscale image segmentation of forest scenes. *International Journal of Applied Earth Observation and Geoinformation* 2005;7:339-359.
13. Baatz M, Arini N, Schäpe A, Binnig G, Linssen B. Object-oriented image analysis for high content screening: Detailed quantification of cells and sub cellular structures with the Cellenger software. *Cytometry* 2006;69:652-658.
14. #Develop. <http://www.icsharpcode.net/OpenSource/SD/>
15. Hufnagl P, Voss K. Ein zeitoptimaler Konturfolgealgorithmus. "Digitale Bildverarbeitung", *Wiss Beitr d TU Dresden* 1983:18-26.
16. OTSU N. A Threshold Selection Method from Gray-Level Histograms. *IEEE Trans Syst Man Cybern* 1979;9:62-66.
17. Ruifrok AC, Johnston DA. Quantification of histochemical staining by color deconvolution. *Anal Quant Cytol Histol* 2001;23:291-299.
18. Velić M, May D, Moresi L. A Fast Robust Algorithm for Computing Discrete Voronoi Diagrams. *J Math Model Algor* 2009;8:343-355.
19. Chorro L, Sarde A, Li M et al. Langerhans cell (LC) proliferation mediates neonatal development, homeostasis, and inflammation-associated expansion of the epidermal LC network. *J Exp Med* 2009;206:3089-3100.
20. Merad M, Ginhoux F, Collin M. Origin, homeostasis and function of Langerhans cells and other langerin-expressing dendritic cells. *Nat Rev Immunol* 2008;8:935-947.
21. Merad M, Manz MG, Karsunky H et al. Langerhans cells renew in the skin throughout life under steady-state conditions. *Nat Immunol* 2002;3:1135-1141.

## **Eidesstattliche Versicherung**

Ich, Stephan Wienert, versichere an Eides statt durch meine eigenhändige Unterschrift, dass ich die vorgelegte Dissertation mit dem Thema: „Quantitative Analyse von Zell- und Gewebeeigenschaften in histologischen Bilddaten“ selbstständig und ohne nicht offengelegte Hilfe Dritter verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel genutzt habe.

Alle Stellen, die wörtlich oder dem Sinne nach auf Publikationen oder Vorträgen anderer Autoren beruhen, sind als solche in korrekter Zitierung (siehe „Uniform Requirements for Manuscripts (URM)“ des ICMJE -[www.icmje.org](http://www.icmje.org)) kenntlich gemacht. Die Abschnitte zu Methodik (insbesondere praktische Arbeiten, Laborbestimmungen, statistische Aufarbeitung) und Resultaten (insbesondere Abbildungen, Graphiken und Tabellen) entsprechen den URM (s.o) und werden von mir verantwortet.

Meine Anteile an den ausgewählten Publikationen entsprechen denen, die in der untenstehenden gemeinsamen Erklärung mit dem/der Betreuer/in, angegeben sind. Sämtliche Publikationen, die aus dieser Dissertation hervorgegangen sind und bei denen ich Autor bin, entsprechen den URM (s.o) und werden von mir verantwortet.

Die Bedeutung dieser eidesstattlichen Versicherung und die strafrechtlichen Folgen einer unwahren eidesstattlichen Versicherung (§156,161 des Strafgesetzbuches) sind mir bekannt und bewusst.

---

Datum, Unterschrift

### **Anteilerklärung an den erfolgten Publikationen**

Stephan Wienert hatte folgenden Anteil an den folgenden Publikationen:

Publikation 1: **Wienert S**, Heim D, Saeger K, Stenzinger A, Beil M, Hufnagl P, Dietel M, Denkert C, Klauschen F. Detection and segmentation of cell nuclei in virtual microscopy images: a minimum-model approach. Sci Rep 2012;2:503.

Eigener Anteil: Entwurf und Implementierung der Bildanalyse-Verfahren (90%), Entwurf und Auswertung der Vergleichsstudie (50%), Verfassen der Publikation (60%).

Publikation 2: **Wienert S**, Heim D, Kotani M, Lindequist B, Stenzinger A, Ishii M, Hufnagl P, Beil M, Dietel M, Denkert C, Klauschen F. CognitionMaster: an object-based image analysis framework. *Diagn Pathol* 2013;8:34.

Eigener Anteil: Entwurf und Implementierung der Software (90%), Verfassen der Publikation (70%).

Publikation 3: Lohneis P, **Wienert S**, Klauschen F, Ullrich A, Anagnostopoulos I, Jöhrens K. Marginal zone lymphomas with monocytoid morphology express t-bet and are associated with a low number of t cells in extranodal locations. *Leuk Lymphoma* 2013:1-16.

Eigener Anteil: Entwurf (50%) und Implementierung (90%) der Bildanalyse-Verfahren, Verfassen der Publikation (10%).

Publikation 4: Ghigo C, Mondor I, Jorquera A, Nowak J, **Wienert S**, Zahner SP, Clausen BE, Luche H, Malissen B, Klauschen F, Bajénoff M. Multicolor fate mapping of Langerhans cell homeostasis. *J Exp Med* 2013;184:2417.

Eigener Anteil: Entwurf (50%) und Implementierung (90%) der Bildanalyse-Verfahren, Durchführung der Bildanalyse (70%), Verfassen der Publikation (10%).

Publikation 5: Jarjour M, Jorquera A, Mondor I, **Wienert S**, Narang P, Coles MC, Klauschen F, Bajénoff M. Fate mapping reveals origin and dynamics of lymph node follicular Dendritic Cells. *J Exp Med*, Manuskript akzeptiert.

Eigener Anteil: Entwurf (50%) und Implementierung (90%) der Bildanalyse-Verfahren, Durchführung der Bildanalyse (70%), Verfassen der Publikation (10%).

Unterschrift, Datum und Stempel des betreuenden Hochschullehrers/der betreuenden Hochschullehrerin

---

Prof. Dr. Carsten Denkert

Unterschrift des Doktoranden/der Doktorandin

---

Stephan Wienert

## **Lebenslauf**

Mein Lebenslauf wird aus datenschutzrechtlichen Gründen in der elektronischen Version meiner Arbeit nicht veröffentlicht.

# Publikationsliste

## Originalarbeiten

- 1 **Wienert S**, Beil M, Saeger K, Hufnagl P, Schrader T. Integration and acceleration of virtual microscopy as the key to successful implementation into the routine diagnostic process. *Diagn Pathol* 2009;4:3.
- 2 **Wienert S**, Heim D, Saeger K, Stenzinger A, Beil M, Hufnagl P, Dietel M, Denkert C, Klauschen F. Detection and segmentation of cell nuclei in virtual microscopy images: a minimum-model approach. *Sci Rep* 2012;2:503.
- 3 **Wienert S**, Heim D, Kotani M, Lindequist B, Stenzinger A, Ishii M, Hufnagl P, Beil M, Dietel M, Denkert C, Klauschen F. CognitionMaster: an object-based image analysis framework. *Diagn Pathol* 2013;8:34.
- 4 Lohneis P, **Wienert S**, Klauschen F, Ullrich A, Anagnostopoulos I, Jöhrens K. Marginal zone lymphomas with monocytoid morphology express t-bet and are associated with a low number of t cells in extranodal locations. *Leuk Lymphoma* 2013:1-16.
- 5 Ghigo C, Mondor I, Jorquera A, Nowak J, **Wienert S**, Zahner SP, Clausen BE, Luche H, Malissen B, Klauschen F, Bajénoff M. Multicolor fate mapping of Langerhans cell homeostasis. *J Exp Med* 2013;184:2417.
- 6 Jarjour M, Jorquera A, Mondor I, **Wienert S**, Narang P, Coles MC, Klauschen F, Bajénoff M. Fate mapping reveals origin and dynamics of lymph node follicular Dendritic Cells. *J Exp Med*, Manuskript akzeptiert.
- 7 Denkert C, von Minckwitz G, Brase JC, Sinn BV, Gade S, Kronenwett R, Pfitzner BM, Salat C, Loi S, Schmitt WD, Schem C, Fisch K, Darb-Esfahani S, Mehta K, Sotiriou C, **Wienert S**, Klare P, André F, Klauschen F, Blohmer JU, Krappmann K, Schmidt M, Tesch H, Kümmel S, Sinn P, Jackisch C, Dietel M, Reimer T, Untch M, Loibl S. Tumor-infiltrating lymphocytes and response to neoadjuvant chemotherapy in breast cancer. Manuskript eingereicht.
- 8 Klauschen F, **Wienert S**, Schmitt WD, Loibl S, Gerber B, Blohmer J, Huober J, Rüdiger T, Erbstöber E, Mehta K, Lederer B, Denkert C, von Minckwitz G. Standardized Ki67 diagnostics using a novel automated scoring approach – clinical validation in the GeparTrio breast cancer study. Manuskript eingereicht.
- 9 Kölbel B, **Wienert S**, Dimitriadis J, Kendoff D, Gehrke T, Huber M, Frommelt L, Tiemann A, Saeger K, Krenn V. CD15-Fokus-Score und die Entwicklung einer morphometrischen Software (“CD15-Quantifier”) zur Diagnostik der periprotetischen Infektion. Manuskript eingereicht.

## Vorträge und Posterpräsentationen

- 1 Saeger K, **Wienert S**, Polus K, Emmel D, Hufnagl P. LIMS and workflow integration as impulse to establish Virtual Microscopy to daily clinical use. 2<sup>nd</sup> European Workshop on Tissue Imaging and Analysis. Jun 25-26, 2010; Heidelberg.
- 2 **Wienert S**, Klauschen F, Denkert C, Hufnagl P, Saeger K. Integrated image analysis and processing toolkit for whole slide imaging. 2<sup>nd</sup> European Workshop on Tissue Imaging and Analysis. Jun 25-26, 2010; Heidelberg.
- 3 **Wienert S**, Heim D, Denkert C, Saeger K. Automatic TMA de-arraying. 10th European Congress on Telepathology and 4th International Congress on Virtual Microscopy. Jul 1-3, 2010; Lithuania.
- 4 Klauschen F, **Wienert S**, Blohmer JU, et al. Automated computational Ki67 scoring in the GeparTrio breast cancer study cohort. Dec 4-8, 2012; San Antonio.

## **Danksagung**

In der Zeit in der diese Arbeit entstanden ist sind meine Kinder Jannick und Katharina um einiges größer geworden. Ihrem, wenn auch unbewusstem, Zurückstecken gilt mein größter Dank. Für die Betreuung und Überlassung des Themas danke ich Prof. Dr. Carsten Denkert und PD Dr. Frederick Klauschen. Die Zusammenarbeit war über die Jahre eine großartige Erfahrung. Desweiteren gilt mein Dank für die langjährige Unterstützung (in alphabetischer Reihenfolge) Prof. Dr. Michael Beil, M.Sc. Daniel Heim, Prof. Dr. Peter Hufnagl und Dipl.-Ing. Kai Saeger.

# Anhang





# Detection and Segmentation of Cell Nuclei in Virtual Microscopy Images: A Minimum-Model Approach

SUBJECT AREAS:  
BIOINFORMATICS  
SOFTWARE  
MEDICAL RESEARCH  
IMAGING

Stephan Wienert<sup>1,2</sup>, Daniel Heim<sup>2</sup>, Kai Saeger<sup>2</sup>, Albrecht Stenzinger<sup>3</sup>, Michael Beil<sup>4</sup>, Peter Hufnagl<sup>1</sup>, Manfred Dietel<sup>1</sup>, Carsten Denkert<sup>1</sup> & Frederick Klauschen<sup>1</sup>

Received  
11 April 2012

Accepted  
25 June 2012

Published  
11 July 2012

Correspondence and requests for materials should be addressed to F.K. (frederick.klauschen@charite.de)

<sup>1</sup>Institute of Pathology, Charité University Hospital Berlin, Charitéplatz 1, 10117 Berlin, Germany, <sup>2</sup>VMscope GmbH, Charitéplatz 1, 10117 Berlin, Germany, <sup>3</sup>Institute of Pathology, University Hospital Heidelberg, Im Neuenheimer Feld 220/221, 69120 Heidelberg, Germany, <sup>4</sup>Department of Medicine I, University of Ulm, Albert-Einstein-Allee 23, 89081 Ulm, Germany.

Automated image analysis of cells and tissues has been an active research field in medical informatics for decades but has recently attracted increased attention due to developments in computer and microscopy hardware and the awareness that scientific and diagnostic pathology require novel approaches to perform objective quantitative analyses of cellular and tissue specimens. Model-based approaches use a priori information on cell shape features to obtain the segmentation, which may introduce a bias favouring the detection of cell nuclei only with certain properties. In this study we present a novel contour-based “minimum-model” cell detection and segmentation approach that uses minimal a priori information and detects contours independent of their shape. This approach avoids a segmentation bias with respect to shape features and allows for an accurate segmentation (precision=0.908; recall=0.859; validation based on ~8000 manually-labeled cells) of a broad spectrum of normal and disease-related morphological features without the requirement of prior training.

The earliest attempts to use computers for the automated detection and analysis of cells in cyto-/histological specimens date back a few decades<sup>1,2</sup>. Since then the topic has established itself as an important area of research in biomedical informatics<sup>3–8</sup>. However, as the more recent literature shows the search for a robust, practically usable cell/nucleus segmentation module is still ongoing: Various different approaches have been proposed ranging from relatively simple thresholding techniques<sup>9,10</sup> to more sophisticated methods, such as, for instance, adaptive attention windows defined by the maximum cell size<sup>11</sup> or gradient flow tracking<sup>12</sup>. A “2D world decomposition” model combined with a priori texture and color information is used for the segmentation of specific cells (leukocytes) in microscopic images of bone marrow<sup>13</sup>. In<sup>14</sup> stepwise merging of rules based on mean-shift clustering and boundary removal rules with a gradient vector flow snake is used for the segmentation of blood cells. Another very recent study<sup>15</sup> uses a fuzzy C-means active contour algorithm that combines multi-spectral edge and region information through a vector multiphase level set framework. Further approaches including the separation of cell clusters can be found in<sup>16–21</sup>. This selection of publications exemplifies the methodological diversity employed to tackle this complex problem. However, despite these long-standing efforts in computer science it is only very recently that automated cell segmentation has also received increased attention on the application side in clinical pathology. The reasons for this are on one hand technology-based: Only in the last years computers, microscope cameras and whole slide scanners have become widely available and are able to generate, process and analyze high-quality digitized histological data sets. The other reasons are biomedical: While qualitative or at most semi-quantitative “visual” estimation of morphological features (for instance nuclear shape in cancer cells as a proxy of the grade of malignancy) or the expression of immunohistological markers are still considered the standard approach in diagnostic pathology, the requirements for cyto- and histological analyses are currently changing. More and more prognostic and predictive markers, i.e. markers that enable pathologists to assess the aggressiveness of a disease or its response to therapy, respectively, are being discovered with increasing requirements for standardized quantitative assessment<sup>22–25</sup>. Here, automated image analysis tools will not replace but assist pathologists to increase diagnostic precision and inter-observer reliability. Examples of potential applications of “computer-assisted diagnostics” include, for instance, the scoring of growth factor receptor expression levels in gastric and breast cancer to decide if a patient will benefit from antibody-based



inhibitory therapy<sup>26,27</sup>, the quantification of nuclear shape features in cancer cells (“grade”)<sup>28</sup> or the quantification of the tumor proliferation index as measured by Ki-67 immunohistochemistry<sup>29</sup>. Independent of the particular task to be performed all approaches share one algorithmic module pivotal for the accurate performance of any cellular segmentation and analysis approach: the robust detection (including localization and segmentation) of cell nuclei. To achieve this goal existing approaches often use a priori information about cellular/nuclear features<sup>13,14</sup>, which are used to build a generic cell model<sup>3,5,7,8</sup>. Such approaches perform well if cell populations are relatively homogeneous in terms of shape or size. However, especially in the pathologically most relevant cases, i. e. the segmentation of cancer cell nuclei in cytology or histology samples, this approach is problematic in two ways, because cancer cells have highly heterogeneous shape and size features and moreover are accompanied by stromal and immune cells of various shapes and sizes: On one hand this makes it difficult to define a generic contour model representative of a realistic distribution of cell nucleus features. Second, and this aspect is even more relevant, using a priori information and generating a model of contour features has a high risk of introducing a segmentation bias: Nuclei with extreme features would have a higher probability of not being covered by the model parameters and might thus be overlooked in comparison with more “regular” cell nuclei. Because cells with extreme shape deviations with respect to a “model cell” are very often the most malignant ones, such a bias might lead to significant diagnostic errors in all of the above mentioned applications and is therefore not acceptable. The only way around this issue would be to use very loose model parameters, which would, however, result in an unacceptably high number of false positives. In the study presented here, we describe a contour-based nucleus detection algorithm, which uses minimal a priori information and finds contours independent of their shape by combining a global contour search with local gradient information. This approach avoids a segmentation bias with respect to shape features and is capable of segmenting cell nuclei of various normal and cancerous cell types from breast, liver and bone marrow as well as kidney tissue and intestinal mucosa with a single set of parameters and no prior training.

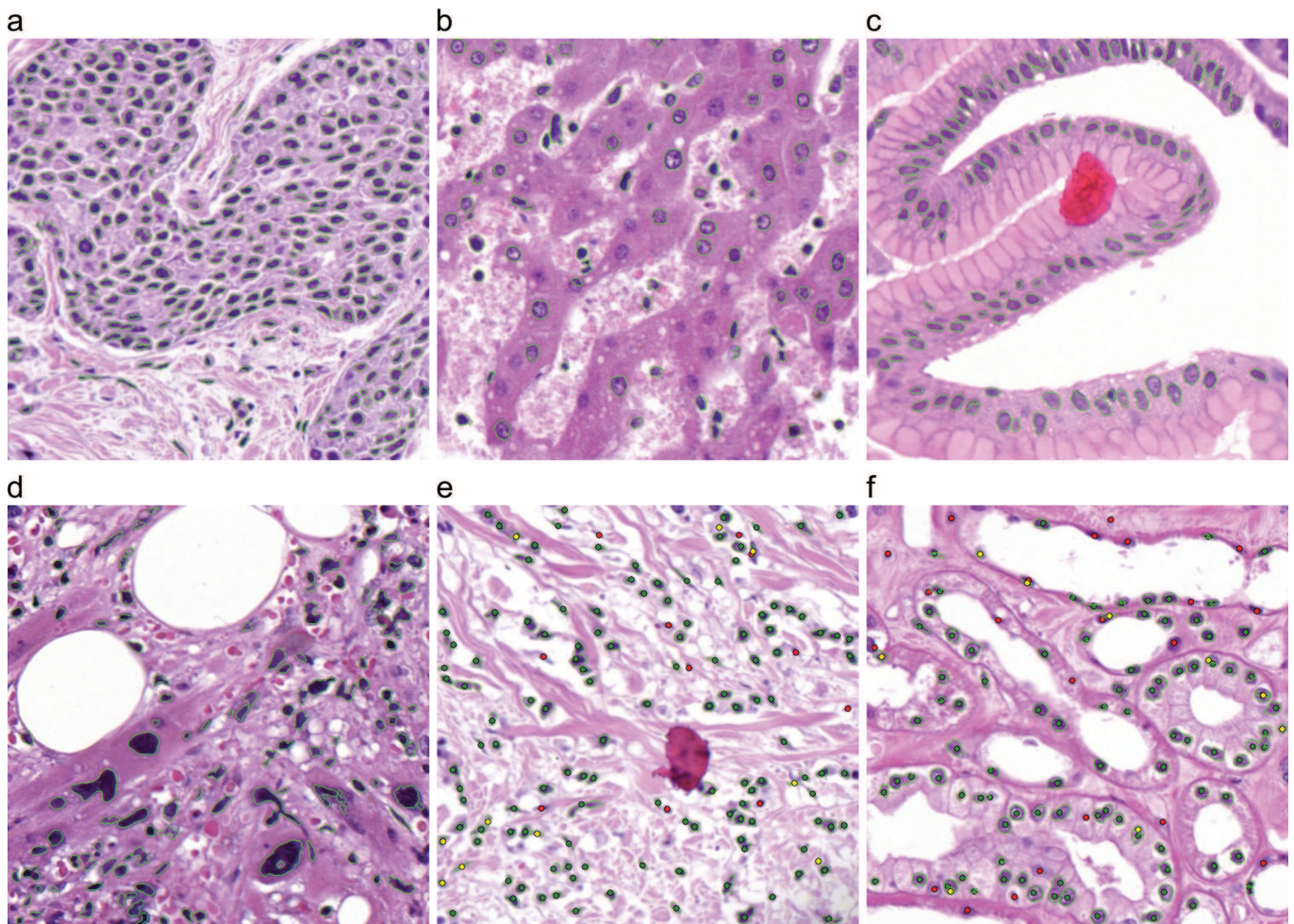
## Results

The cell nucleus segmentation approach we present in this study consist of six major modules: First, all possible closed contours present in the grayscale-transformed image are computed regardless of size, shape or whether they belong to hills or valleys in the intensity landscape. This step yields multiple, often overlapping contours, which are evaluated in a second module based on gradient features of the input image. We introduce a “contour value” as a measure to rank and select those contours that best represent the image objects. The contour value is a combination of the mean contour gradient that measures the relative “importance” of objects (in a group of overlapping objects) and the “gradient fit” that defines which of several alternative contours best represents a certain object. A non-overlapping segmentation is generated in the third step: The enclosed areas of the ranked contours are labeled in a two dimensional map. Subsequently, this segmentation is improved using a novel contour optimization method and an optional cluster separation step. Finally, cell nuclei are detected by assessing the (nucleus-specific) Hematoxylin within each contour area (for details see *Methods* section). To quantitatively evaluate the performance of the cell/nucleus detection we prepared a gold-standard data set using randomly-selected images containing 7931 cells manually labeled by a pathologist. The selected tissues comprise various breast cancer samples representing a broad morphological variety including normal tissue components, bone marrow with normal and pathologically altered cells, normal liver tissue as well as kidney tissue and intestinal mucosa acquired during routine diagnostics and processed with standard routine

laboratory protocols at our institute (see *Methods* for details). After the manual labeling, the original image data was segmented using the proposed method and the positional information available in the gold-standard data was compared to the automated segmentation result by automatically determining true positives, false negatives and false positives. This analysis yielded an overall precision of 0.908, a recall of 0.859 and a conglomerate score of 0.953. These data feature a broad range of cell/nucleus morphologies ranging from normal gastric mucosa (regular round/oval shaped nuclei) to bone marrow with primary myelofibrosis (highly irregular neoplastic megacaryocytes, Fig. 1d and Fig. 33 in Supplementary Notes S1 online). The analysis of these images shows similar segmentation accuracies irrespective of the different morphologies (Myelofibrosis: precision=0.88, recall=0.91; normal mucosa: precision=0.90, recall=0.89), which confirms the absence of a segmentation bias in our method. A comparison to Al-Kofahis method<sup>30,31</sup> was performed using first the default (fully automatic) mode (which was recommended by the authors upon request) and second, using an optimized parameter configuration file provided by the authors after submitting test data to them (Table 1). When using the automatic mode Al-Kofahis method required less processing time but yielded a by 0.197 lower precision ( $p=5.97e-10$ ) and by 0.049 higher recall ( $p=2.03e-7$ ). With optimized parameters the precision remained 0.085 less precise than our approach ( $p=5.60e-7$ ) with a constant recall (0.908,  $p=2.24e-7$ ) and requiring 2.6 times more mean processing time compared to our approach. The conglomerate score of Al-Kofahis method was only slightly better: 0.953 in our approach compared to 0.964 ( $p=3.75e-3$ ) in the automatic mode and 0.966 ( $p=1.16e-3$ ) with optimized parameters. All images (original and labeled) and the individual segmentation accuracies can be found side-by-side in Supplementary Note S1. Moreover, Supplementary Data S2 contains the original image data and the xml-coded positional information from the manual cell labeling.

## Discussion

Research in automated cell nucleus segmentation has recently regained attention because of an increased interest by the pathology community in standardized and quantitative evaluation of cyto- and histological specimens. This is on one side due to an increasing number of immunohistochemical diagnostic markers that need to be quantified to allow for a valid prognostic and/or predictive assessment. Moreover, the use of nuclear shape features for prognostic purposes in cancer diagnostics has recently received criticism due to a large inter-observer variability, especially in the context of novel genomic test (such as, for instance, OncotypeDX or MammaPrint) that compete with classical histopathology in the growing field of individualized predictive pathology. In this study, a contour-based segmentation method is described that is closely simulating the way the human eye detects cells, i. e. mainly based on the extent the objects are silhouetted against the background. Also, contours are detected independent of their shape. Consequently, very pleomorphic cell nuclei are found with the same accuracy as very “normal”, i. e. round cell nuclei and even contours with a contour-gradient of only one intensity value (difference between the objects contour and its surrounding) may be detected if these are the contours with the locally strongest contour-gradient. The method uses minimal a priori information about cell nucleus staining and size. Compared to the approach by Al-Kofahi et al.<sup>30,31</sup>, that uses 9 parameters, our method requires only 5 parameters that may be set within relatively loose bounds. Performing segmentation with few model assumptions on cell shape features can help avoid the introduction of a detection bias which would favour normal cells without nuclear atypia over pleomorphic malignant cells and therefore lead to significant diagnostic errors especially in the case of cancer tissues that display large morphological varieties. Our approach is capable of



**Figure 1** | H&E stained tissue samples: (a) breast cancer; (b) liver; (c) gastric mucosa; (d) bone marrow with primary myelofibrosis with highly pleomorphic megakaryocytes (without cell cluster separation); (e) connective tissue; (f) kidney tissue; (e, f) examples of the method validation (dots represent manually-assigned labels, green: labels classified as true positive, red: false negatives, yellow: false positives).

segmenting cell nuclei of various normal and cancerous cell types from breast, liver and bone marrow and other tissues with a single set of parameters and no prior training. Moreover, the described features render the method robust against variations in staining protocols within reasonable limits, because no absolute intensity information is used initially for contour detection. The method is also relatively robust against image blur (see Supplementary Note S3 online). Due to the general (minimum) model formulation our method is also capable of segmenting fluorescence microscopy data (e. g. confocal laser scanning microscopy images) or even electron microscopy images. It may therefore in the future also be useful for the registration of the different imaging modalities in the field of correlative light electron microscopy (CLEM)<sup>32,33</sup> (see Supplementary Note S4 online). Our method is therefore highly versatile and suitable to serve as a basis in form of a segmentation module for applications, for instance, designed to assess cancer malignancy (pathological grading) for which the variability of nucleus shapes is a central fea-

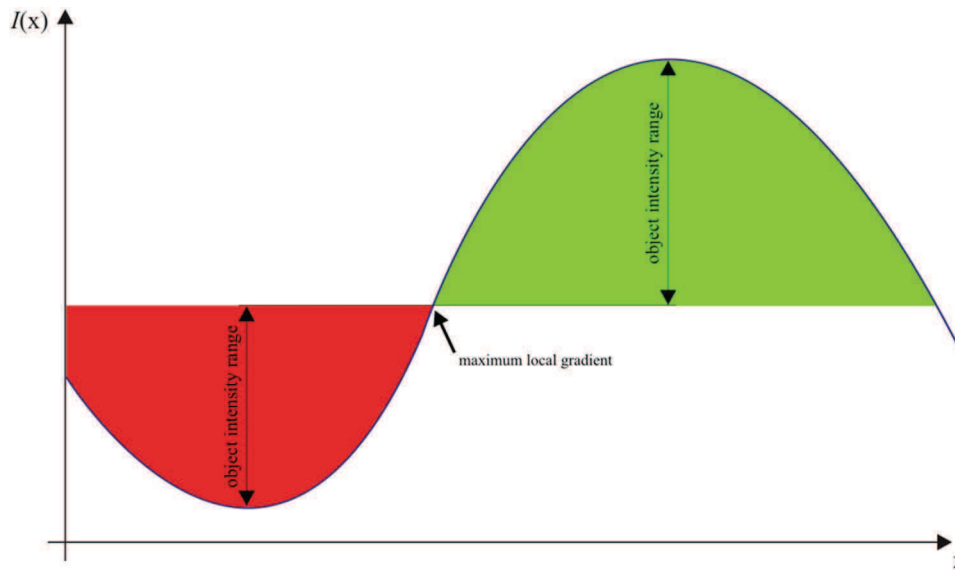
ture, or the immunohistochemical quantification of the proliferation index by immunostaining of the Ki-67 antigen, for which it is critical to detect all variants of cancer cells. In case of strong unspecific background Hematoxylin staining, false positive cell nuclei may occur, which, however, is rarely the case in the data we analysed. As our visual tests and quantitative validation with more than 7900 manually labeled cells show the proposed method performs well for a large variety of morphological cancer variants and normal tissue exhibiting a representative variation of nucleus shape features. To conclude, we propose a novel cell nucleus segmentation approach that minimizes the use of a priori knowledge and may therefore serve as the basis for segmentation modules in cancer nucleus feature quantification and classification tasks.

## Methods

The cell nucleus segmentation approach presented here consists of six internal processing steps:

**Table 1** | Comparison of the method presented in this paper and Al-Kofahis method

	Wienert	Al-Kofahi (automatic)	Al-Kofahi (optimized)
<b>Precision</b>	0.908	0.711	0.823
<b>Recall</b>	0.859	0.908	0.908
<b>Conglomerate</b>	0.953	0.964	0.966
<b>Mean processing time (ms)</b>	725.4±182.3	678.2±49.5	1942.6±326.7



**Figure 2** | One dimensional grayscale image function  $I(x)$  with one dark (red) and one bright (green) object. Initially, the minimum-model approach uses intensities to define objects (both hills and valleys in the intensity landscape can be objects).

- Detection of all possible closed contours
- Contour evaluation
- Generation of a non-overlapping segmentation
- Contour optimization
- Concave object separation
- Classification into cell nuclei and other objects

**Detection of all possible contours (a).** To find the best fitting contour for each nucleus within an image, initially, our method detects all possible closed contours irrespective of whether they actually belong to a nucleus or not. We call this the primary segmentation.

This contour detection is based on a conventional contour tracing approach for binary images<sup>34</sup>, which we extended for the use with grayscale images. The grayscale image is generated by computing the mean of the red, green and blue channel of the corresponding pixel in the original color image. While finding the contour start pixels (seeds) needed to initiate the contour tracing is straightforward for binary images, grayscale images require a different strategy: in our approach each image row is regarded as a one dimensional image function  $I(x)$  and the contour start pixels are defined as the positions at which the gradient between each pair of neighboring local minima/maxima or maxima/minima is maximal (Fig. 2). All pixels with grayscale values between the intensity at the contour start pixel and at the corresponding local minimum or the corresponding local maximum, respectively, are initially interpreted as object-pixels (Fig. 3). As a consequence each contour tracing is performed within a locally-adapted (“individual”) intensity range and objects may represent either “hills” or “valleys” in the intensity landscape. The detection of the contour start pixels and the determination of the corresponding intensity range is performed by scanning the image row-wise from left to right and storing all local minima and maxima with the corresponding maximum gradients in between (visual tests showed that the overall result is invariant with respect to the scan direction). Subsequently, the contour tracing follows the (potential) object contours clockwise using an 8-connected neighborhood (according to<sup>34</sup>). A contour is considered valid if and only if it reaches back to its start pixel. Otherwise, the contour tracing is terminated if a maximum contour length (225 pixels in our examples) is exceeded. Pseudocode for the contour start pixel and extremal value detection and an english description of the contour tracing algorithm can be found in the Supplementary Methods S5 and S6 online.

**Contour evaluation (b).** The contour detection approach (primary segmentation) results in multiple, often overlapping contours many of which do not represent actual objects at all. Thus, this primary segmentation is not composed of disjoint image segments and therefore heavily over-defined. To obtain a non-overlapping segmentation it has to be determined, which objects are more prominent than others within the same local area. An object is considered more important if it has a higher mean gradient along the objects contour (Equation 3), which largely corresponds to the (trivial) concept that an image region is more likely to correspond to an actual object if it is silhouetted sharply against the background (or other objects). Additionally, multiple contours from the primary segmentation may describe the same object and hence the decision has to be made, which of several alternative contours best represents a certain object: Contours are visually regarded as fitting best if they show a high concordance between contour pixels and the maximum local gradient. This “gradient fit” is defined as the relative number of contour pixels that represent a local maximum in a  $3 \times 3$  neighborhood in the gradient image (Equation

2) and is computed by using the Sobel operator  $S$  with its  $3 \times 3$  convolution kernels  $G_x$  and  $G_y$  (Equation 1) and the 2-D image function  $I$ . The mean gradient and the “gradient fit” are combined in the contour value (Equation 4). In equation (2) and (3) let  $C_i$  be the  $i$ th contour and  $p_{ij}$  its  $j$ th contour pixel.

$$|S| = \sqrt{(I * G_x)^2 + (I * G_y)^2} \quad G_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \quad G_y = \begin{bmatrix} -1 & 2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix} \quad (1)$$

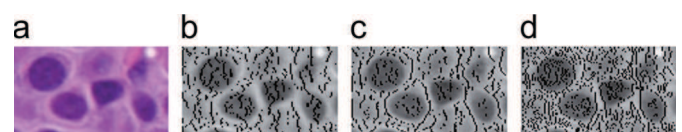
$$\text{GradientFit}_i = \frac{\sum_j p_{ij}^{\max}}{|C_i|}, \quad p_{ij}^{\max} := \begin{cases} 1 & \text{if } \max\{|S(p_{nm})|\} = |S(p_{ij})| \quad \forall n, m \quad \begin{matrix} x_i - 1 \leq n \leq x_i + 1 \\ y_i - 1 \leq m \leq y_i + 1 \end{matrix} \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

$$\text{MeanGradient}_i = \frac{\sum_j |S(p_{ij})|}{C_i} \quad (3)$$

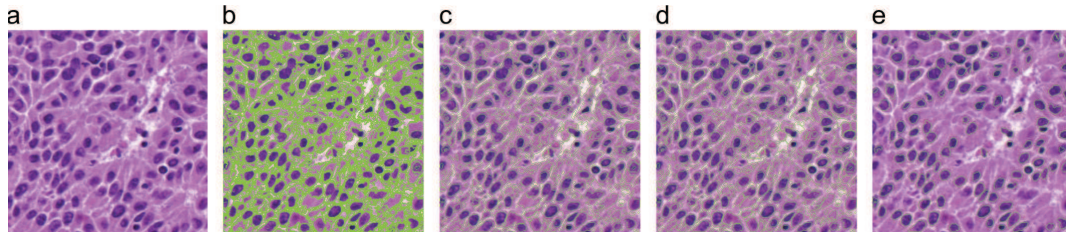
$$\text{ContourValue}_i = \text{MeanGradient}_i \bullet \text{GradientFit}_i \quad (4)$$

**Generation of a non-overlapping segmentation (c).** To generate a non-overlapping segmentation, initially, a two-dimensional map of the same size as the corresponding image is used to label the area within each contour with a unique identifier. To obtain the locally most prominent contours (as defined by the contour value defined in Eq. 4) the labeling process is performed in sorted order, beginning with the topmost valued contour and blocking overwriting of already assigned labels.

**Contour Optimization (d).** Due to the consecutive labeling process objects may have thin filaments on the object border (Fig. 4b). These filaments do not belong (are not compact) to the actual object. To remove these filaments a contour optimization processing step is applied (Fig. 4c): The distance transformation is computed based on the chamfering method introduced by<sup>35</sup> using the Manhattan metric. A distance value  $d$  is defined for testing the compactness of object pixels (3 in our examples). A Pixel  $p_i$  (with  $d_i$  the distance value of  $p_i$  and  $d_i < d$ ) is considered “compact” if it is connected to a pixel  $p_j$  (with  $d_j = d$ ) over  $d - d_i$  edges (Fig. 5). An efficient algorithm can be used to remove non-compact pixels by scanning the distance map row wise



**Figure 3** | (a) H&E stain of breast tissue, (b) local minima, (c) local maxima and (d) maximum local gradients of the horizontal image function  $I(x)$  marked with black pixels.

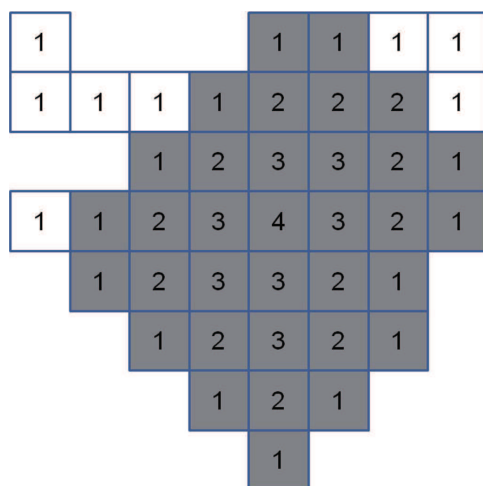


**Figure 4** | (a) 400×400 pixel H&E stained breast cancer tissue, (b) non-overlapping segmentation generated from full contour search (29815 contours in primary segmentation), (c) optimized contours, (d) results after concave object separation and (e) the final segmentation after removing all non-nuclei objects yielding 119 nuclei. The overall segmentation process took 0.39 seconds on a standard PC.

$d - 1$  times: A loop is set up to process pixels with a specific distance value  $d_i$  starting with  $d_i = d - 1$  down to 1. The whole distance map is scanned in each cycle. The distance value of pixels  $p_i$  with  $d_i = d_i$  is decremented by 1 if they do not have a paraxial neighbor with a distance value of  $d_i + 1$ . Finally, pixels  $p_i$  with  $d_i = 0$  are removed from the label map. Pseudocode for the non-compact pixel removal can be found as Supplementary Methods S7 online.

**Concave Object Separation (e).** To handle cases of cells forming clusters our approach is based on separating objects at concave borders by removing object pixels (labels) around a cutting line between two concavities (Fig. 4d). Two criteria are taken into account to select the optimal cutting line: The length of the cutting line between two concavities should be minimal compared to the depth of the corresponding concavities and the two corresponding contour concavities should be located opposite to each other. First, concave contour segments are detected in each object (Fig. 6): For that purpose, the convex hull is computed according to<sup>36</sup>. If at least one contour pixel  $C$  exists between two neighboring pixels of the convex hull  $A$  and  $B$  with distance from  $AB$  greater than 0 all contour pixels between  $A$  and  $B$  are considered concave points and form a concave contour segment. The angle  $\alpha$  (clockwise,  $-\pi \leq \alpha \leq \pi$ ) of the line segment  $AB$  to the x-axis  $+90^\circ$  ( $\pi/2$ ) is considered the angle of the concavity/concave contour segment. Next, all concave points of all concave contour segments of an object are combined to find the best fitting pair for a separation: This is the cutting line where the so called *SeparationScore* is minimal (Equation 7). The *SeparationScore* combines a minimization of the relative cutting line length  $r$  (Equation 5) with the “angle criterion” that improves if cutting lines are perpendicular to the convex hull (this helps avoid incorrect cut directions, Equation 6). In Equation (5)–(7) let  $C_1$  be the start point,  $C_2$  the end point and  $r$  the length of the potential cutting line  $C_1C_2$ ,  $depth_1$  is the distance of  $C_1$  from  $A_1B_1$ ,  $depth_2$  the distance of  $C_2$  from  $A_2B_2$ ,  $\alpha_1$  is the angle between the potential cutting line  $C_1C_2$  and the corresponding concave contour segment ( $A_1B_1$ ) and  $\alpha_2$  the angle between the potential cutting line  $C_2C_1$  and the corresponding concave contour segment ( $A_2B_2$ ).

$$LengthScore = \frac{r}{r + depth_1 + depth_2} \quad (5)$$



**Figure 5** | Image object (all pixels) with non-compact pixels on the object border (white pixels). Numbers represent the distance of the corresponding pixel to the nearest background (non-object) pixel with the Manhattan metric. A distance value  $d$  is given for the testing of compactness (3 in this example). Removing all pixels  $p_i$  (with  $d_i$  the distance value of  $p_i$  and  $d_i < d$ ) that are connected to a pixel  $p_j$  (with  $d_j = d$ ) over more than  $d - d_i$  edges results in a compact object (gray pixels).

$$AngleScore = \frac{\alpha_1 + \alpha_2}{2\pi} \quad (6)$$

$$SeparationScore = \frac{LengthScore + AngleScore}{2} \quad (7)$$

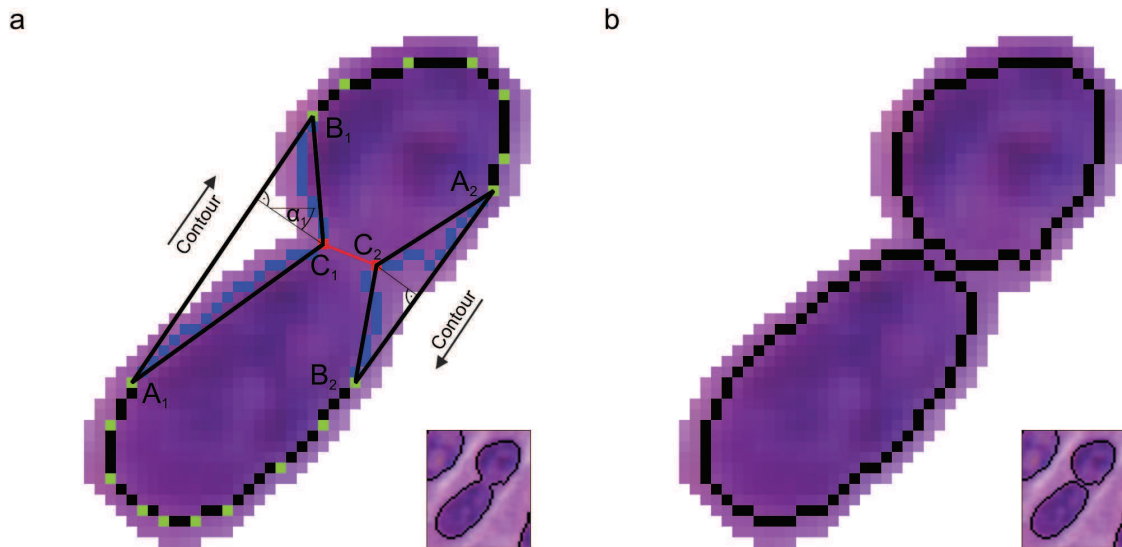
To avoid separations based on concavities with a low depth a minimum depth of concavity may be defined (4 pixels in our examples). Concave borders with a depth below this threshold are ignored in further processing. The separation process is repeated until no further cutting lines are detected. Pseudocode for the concave object separation can be found as Supplementary Methods S8 online.

**Classification into cell nuclei and other objects (f).** The classification into cell nuclei and other objects (background) is achieved by using nucleus-specific color information available from histochemical staining. Hematoxylin is a nuclear stain used in routine pathology characterized by a basophilic deep blue/lilac color and is used in combination with immunohistochemical dyes or cytoplasm-specific Eosin stain for conventional histology. To extract the Hematoxylin signal we use color deconvolution (according to<sup>37</sup>). Subsequently, the staining intensity is computed for each pixel and the optimal threshold is selected from the resulting distribution (according to<sup>38</sup>). All objects with a mean Hematoxylin staining intensity below this threshold are removed (Fig. 4e). Additionally tiny or artefactual objects with an object area below 50 pixels are removed.

**Computational optimizations.** For an effective computation the global contour search, contour evaluation, concave object separation and the color deconvolution filter are processed using parallel *for* loops: Each iteration loop is queued to the *Microsoft .NET ThreadPool* class. This component assigns, schedules and reuses threads for all queued tasks dependent on the system load and the number of available processors. If background contours (which are removed in the final segmentation step) are already removed in the contour evaluation step these contours do not need to be processed in the subsequent processing steps: Contours with a mean Hematoxylin intensity on contour pixels below the threshold described above are therefore rejected.

**Histological sections.** We used both whole slides and tissue microarrays derived from tissue samples available through routine diagnostics at the Institute of Pathology, Charité University Medicine Berlin. Patients had given prior consent to their tissue samples being used in scientific studies. The microarrays were constructed by selecting representative tumor areas and then punching out the respective region from formalin-fixed-paraffin-embedded (FFPE) tissue sample blocks and embedded into a new paraffin array block using a tissue microarrayer (Beecher Instruments, Woodland, CA). All sections used were cut 3- $\mu$ m thick and stained with H&E (Hematoxylin&Eosin) according to standard routine laboratory procedures.

**Digitization of histological specimens.** The histological slides were digitized using the *Zeiss Mirax Scan* slide scanner (produced by *3DHitech Ltd*, Hungary). The slide scanner was equipped with a *Zeiss Plan-Apochromat 20x* objective (numerical aperture=0.8) and an *AVT Marlin F-146C Firewire 1/2* CCD camera with 4.65  $\mu$ m x 4.65  $\mu$ m pixel size. Combined with the 20x objective and 1x C-mount adapter the resulting image resolution is 0.23  $\mu$ m x 0.23  $\mu$ m. All slides were scanned at 20x and all processing steps described here were performed at full resolution. No additional components/filters were used. The *Zeiss Mirax* slide scanner performs an automatic white balance, which corrects the white reference against an area on the glass slide without tissue. The tissue area to be scanned was selected manually. For the scan process we chose to have a focus correction every fourth image tile. The focus is found automatically by the *Mirax* slide scanning system. Stitching of image tiles was also performed automatically by the *Mirax* slide scanning system. No other pre-processing was performed. The resulting images were converted to the *Virtual Slide Format (VSF)* of *VMscope GmbH* with the actual image data encapsulated and in JPEG format with 85% JPEG quality. The image data is thus compressed, but has the typical JPEG quality and compression used in daily laboratory routine. The file sizes for the whole slide images ranged from 102 MB to 1.96 GB, with an average of 601 MB. The image dimensions ranged from 33280×29184 pixels to 70080×159000 pixels. The analyzed fields of view had a size of 600×600 pixels.



**Figure 6 | Cluster composed of two cell nuclei.** (a) Contour pixels (black), vertices of the convex hull (green), pixels of concave contour segments A<sub>1</sub>B<sub>1</sub> and A<sub>2</sub>B<sub>2</sub> (blue) and start and end point of potential cutting line C<sub>1</sub>C<sub>2</sub> (red). (b) Cell nuclei separation result.

**Hard- and Software.** The image segmentation algorithm was implemented in C# for Microsoft .NET using #Accessory.CognitionMaster<sup>®</sup>. Virtual Slide Access – SDK 4.0 of VMscope GmbH, Berlin, Germany, was used for the image I/O. All computations were performed on Windows 7 Enterprise x64 SP1 with Intel® Core™ i5-2520M Processor and 4 GB RAM, Barco MDRC-2124 24.1” color LCD-Monitor.

**Gold-standard data set preparation and method validation.** To assess the accuracy of the proposed method 7931 cells from 36 images were labeled by three pathologists (FK, AS, CD). All images were labeled by three pathologists and only cells for which consensus was achieved were included. All other cells were left unlabeled. The images (600×600 pixels, Hematoxylin-Eosin stained) were taken from previously digitized routine diagnostic cases. After manual labeling the images were segmented with the described method and true positive (tp), false negative (fn) and false positive (fp) events were automatically determined using the positional information (saved in xml-format) previously obtained in the manual “gold-standard” cell labeling. Statistical values were computed as follows:

$$\text{Precision} = \text{tp} / (\text{tp} + \text{fp})$$

$$\text{Recall} = \text{tp} / (\text{tp} + \text{fn})$$

$$\text{Conglomerate} = (\text{Number of detected cells} - \text{fp}) / \text{tp}$$

- Bibbo, M., Bartels, P. H., Dytch, H. E. & Wied, G. L. Computed cell image information. *Monogr Clin Cytol* **9**, 62–100 (1984).
- Bengtsson, E. The measuring of cell features. *Anal. Quant. Cytol. Histol* **9**, 212–217 (1987).
- Bamford, P. Unsupervised cell nucleus segmentation with active contours. *Signal Processing* **71**, 203–213 (1998).
- Bartels, P. H., Gahm, T. & Thompson, D. Automated microscopy in diagnostic histopathology: From image processing to automated reasoning. *Int. J. Imaging Syst. Technol* **8**, 214–223 (1997).
- Tianzi Jiang & Faguo Yang. An evolutionary tabu search for cell image segmentation. *IEEE Trans. Syst. Man, Cybern. B* **32**, 675–678 (2002).
- Latson, L., Sebek, B. & Powell, K. A. Automated cell nuclear segmentation in color images of hematoxylin and eosin-stained breast biopsy. *Anal. Quant. Cytol. Histol* **25**, 321–331 (2003).
- Naik, S. *et al.* Automated gland and nuclei segmentation for grading of prostate and breast cancer histopathology. *Biomedical Imaging: From Nano to Macro, 2008. ISBI 2008. 5th IEEE International Symposium on*, 284–287 (2008).
- Hai-Shan Wu, Barba, J. & Gil, J. A parametric fitting algorithm for segmentation of cell images. *IEEE Trans. Biomed. Eng.* **45**, 400–407 (1998).
- Korde, V. R., Bartels, H., Ranger-Moore, J. & Barton, J. Automatic segmentation of cell nuclei in bladder and skin tissue for karyometric analysis. *Anal Quant Cytol Histol* **31**, 83–89 (2009).
- Tuominen, V. J., Ruotoistenmaki, S., Viitanen, A., Jumppanen, M. & Isola, J. ImmunoRatio: a publicly available web application for quantitative image analysis of estrogen receptor (ER), progesterone receptor (PR), and Ki-67. *Breast Cancer Res* **12**, R56 (2010).
- Ko, B., Seo, M. & Nam, J.-Y. Microscopic cell nuclei segmentation based on adaptive attention window. *J Digit Imaging* **22**, 259–274 (2009).
- Li, G. *et al.* Segmentation of touching cell nuclei using gradient flow tracking. *J Microsc* **231**, 47–58 (2008).

- Reta, C., Gonzalez, J. A., Diaz, R. & Guichard, J. S. Leukocytes segmentation using Markov random fields. *Adv. Exp. Med. Biol* **696**, 345–353 (2011).
- Ko, B. C., Gim, J.-W. & Nam, J.-Y. Automatic white blood cell segmentation using stepwise merging rules and gradient vector flow snake. *Micron* **42**, 695–705 (2011).
- Bunyak, F., Hafiane, A. & Palaniappan, K. Histopathology tissue segmentation by combining fuzzy clustering with multiphase vector level sets. *Adv. Exp. Med. Biol* **696**, 413–424 (2011).
- Wittenberg, T., Grobe, M., Münzenmayer, C., Kuziela, H. & Spinnler, K. A semantic approach to segmentation of overlapping objects. *Methods Inf Med* **43**, 343–353 (2004).
- Li, S. Z. Recognizing multiple overlapping objects in image: an optimal formulation. *IEEE Trans Image Process* **9**, 273–277 (2000).
- Cristian Smochină. Doctoral Thesis. Technical University “GHEORGHE ASACHI”, 2011.
- Yang, L., Tuzel, O., Meer, P. & Foran, D. J. in *Lecture Notes in Computer Science*, edited by D. Hutchison *et al.* (Springer Berlin Heidelberg, Berlin, Heidelberg, 2008), pp. 833–841.
- Loke, R., Bayer, M., Mann, D. & Du Buf, J. in *Oceans '02 MTS/IEEE (IEEE2002)*, pp. 2457–2465.
- INDHUMATHI, C., CAI, Y., GUAN, Y. & OPAS, M. An automatic segmentation algorithm for 3D cell cluster splitting using volumetric confocal images. *Journal of Microscopy* **243**, 60–76 (2011).
- Freedman, L. Quantitative science methods for biomarker validation in chemoprevention trials. *Cancer Biomark* **3**, 135–140 (2007).
- Sullivan, C. A. W. & Chung, G. G. Biomarker validation: in situ analysis of protein expression using semiquantitative immunohistochemistry-based techniques. *Clin Colorectal Cancer* **7**, 172–177 (2008).
- Weigel, M. T. & Dowsett, M. Current and emerging biomarkers in breast cancer: prognosis and prediction. *Endocr. Relat. Cancer* **17**, R245–62 (2010).
- Yaziji, H. *et al.* Consensus recommendations on estrogen receptor testing in breast cancer by immunohistochemistry. *Appl. Immunohistochem. Mol. Morphol* **16**, 513–520 (2008).
- Rüschhoff, J. *et al.* HER2 diagnostics in gastric cancer-guideline validation and development of standardized immunohistochemical testing. *Virchows Arch* **457**, 299–307 (2010).
- Whenham, N., D'Hondt, V. & Piccart, M. J. HER2-positive breast cancer: from trastuzumab to innovative anti-HER2 strategies. *Clin. Breast Cancer* **8**, 38–49 (2008).
- Underwood, J. C. Nuclear morphology and grading in tumours. *Curr Top Pathol* **82**, 1–15 (1990).
- Gerdes, J. Ki-67 and other proliferation markers useful for immunohistological diagnostic and prognostic evaluations in human malignancies. *Semin. Cancer Biol* **1**, 199–206 (1990).
- Al-Kofahi, Y., Lassoued, W., Lee, W. & Roysam, B. Improved automatic detection and segmentation of cell nuclei in histopathology images. *IEEE Trans Biomed Eng* **57**, 841–852 (2010).
- Al-Kofahi, Y. *et al.* Cell-based quantification of molecular biomarkers in histopathology specimens. *Histopathology* **59**, 40–54 (2011).
- Vicidomini, G. *et al.* High Data Output and Automated 3D Correlative Light-Electron Microscopy Method. *Traffic* **9**, 1828–1838 (2008).
- Vicidomini, G. *et al.* A novel approach for correlative light electron microscopy analysis. *Microsc. Res. Tech* **73**, 215–224 (2010).



34. Hufnagl, P. & Voss, K. Ein zeitoptimaler Konturfolgealgorithmus (A time-optimal contour search algorithm). *Digitale Bildverarbeitung*, *Wiss. Beitr. d. TU Dresden*, 18–26 (1983).
35. BORGEFORS, G. Distance transformations in digital images. *Computer Vision, Graphics, and Image Processing* **34**, 344–371 (1986).
36. Eddy, W. F. A New Convex Hull Algorithm for Planar Sets. *ACM Trans. Math. Softw.* **3**, 398–403 (1977).
37. Ruifrok, A. C. & Johnston, D. A. Quantification of histochemical staining by color deconvolution. *Anal. Quant. Cytol. Histol* **23**, 291–299 (2001).
38. OTSU N. A Threshold Selection Method from Gray-Level Histograms. *IEEE Trans. Syst. Man, Cybern.* **9**, 62–66 (1979).
39. #Accessory.CognitionMaster. <http://sourceforge.net/projects/cognitionmaster/>.

## Acknowledgements

We thank Giuseppe Vicidomini and Carlo Tacchetti (MicroSCoBiO Research Center, University of Genoa, Italy) for providing the confocal fluorescence and electron microscopy data used in Supplementary Note S4. Funding was provided by the Charité Medical School Berlin and the Human Frontier Science Program (HFSP) Young Investigator Grant RGY0077/2011.

## Author contributions

FK, SW, DH and KS wrote the paper. SW and DH prepared the figures. FK, MD, CD, PH and MB planned and supervised the work. SW, DH and FK designed the algorithms. SW and DH implemented the algorithms and performed the data processing and analysis. FK, AS and CD annotated the validation data. FK and AS selected and provided histological samples. All authors reviewed the manuscript.

## Additional information

**Supplementary information** accompanies this paper at <http://www.nature.com/scientificreports>

**Competing financial interests:** The authors declare no competing financial interests.

**License:** This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-sa/3.0/>

**How to cite this article:** Wienert, S. *et al.* Detection and Segmentation of Cell Nuclei in Virtual Microscopy Images: A Minimum-Model Approach. *Sci. Rep.* **2**, 503; DOI:10.1038/srep00503 (2012).



SOFTWARE

Open Access

# CognitionMaster: an object-based image analysis framework

Stephan Wienert<sup>1,2</sup>, Daniel Heim<sup>2</sup>, Manato Kotani<sup>3,4</sup>, Björn Lindequist<sup>5</sup>, Albrecht Stenzinger<sup>6</sup>, Masaru Ishii<sup>3,4</sup>, Peter Hufnagel<sup>1,5</sup>, Michael Beil<sup>7</sup>, Manfred Dietel<sup>1</sup>, Carsten Denkert<sup>1</sup> and Frederick Klauschen<sup>1\*</sup>

## Abstract

**Background:** Automated image analysis methods are becoming more and more important to extract and quantify image features in microscopy-based biomedical studies and several commercial or open-source tools are available. However, most of the approaches rely on pixel-wise operations, a concept that has limitations when high-level object features and relationships between objects are studied and if user-interactivity on the object-level is desired.

**Results:** In this paper we present an open-source software that facilitates the analysis of content features and object relationships by using objects as basic processing unit instead of individual pixels. Our approach enables also users without programming knowledge to compose "analysis pipelines" that exploit the object-level approach. We demonstrate the design and use of example pipelines for the immunohistochemistry-based cell proliferation quantification in breast cancer and two-photon fluorescence microscopy data about bone-osteoclast interaction, which underline the advantages of the object-based concept.

**Conclusions:** We introduce an open source software system that offers object-based image analysis. The object-based concept allows for a straight-forward development of object-related interactive or fully automated image analysis solutions. The presented software may therefore serve as a basis for various applications in the field of digital image analysis.

**Keywords:** Software, Open source, Image analysis, Object-based image analysis

## Background

Image analysis has been an active research field since the development of scientific computing decades ago and since then many biomedical (and other) studies have been reported that rely on image analysis software to support their investigations, e.g. [1-6] just to name a few. Several open source imaging projects exist in that context with *ImageJ* [7] as one of the most popular. Some other projects, e.g. *CellProfiler* [8], *Fiji* [9] and *icy* [10] are based on *ImageJ* and provide powerful and flexible user interfaces including region-of-interest (ROI) selection, data analysis and plug-in interfaces. Other software frameworks like *OpenCV* [11] and *ITK* [12] are designed as libraries that provide algorithms for the integration into high-level applications. A common feature of the aforementioned solutions is that the basic

processing unit is the single image pixel, which is the smallest processable unit of a digital image containing intensity and color information (see [13] for the measurement of image information). By grouping pixels into distinct segments (referred to as objects) using the pixel intensity/color information or image texture biological meaningful objects and structures may be represented and used for an object-based analysis [14-16]. This can be advantageous when object-related properties (features) and relationships between objects (neighbourhood) have to be taken into account. As an example: After the initial segmentation of certain image objects (e. g. cells) a typical task in image analysis is to classify all objects according to certain features (e. g. dark and bright objects). With our approach the user may easily experiment interactively with the classification step by computing various features (e.g. size, shape localization and color) and then study correlations between different properties to define the optimal classification routine. This is all done on the object level and connected with

\* Correspondence: frederick.klauschen@charite.de

<sup>1</sup>Institute of Pathology, Charité University Hospital Berlin, Charitéplatz 1, Berlin 10117, Germany

Full list of author information is available at the end of the article



visualization of feature statistics. Another important question is the topological relationship between objects of the different classes. Evaluating topological problems could in principle be solved with the aforementioned pixel-based programs, too, but its implementation becomes less complicated if the basic processing unit is the image-object. Unfortunately, open source software tools that utilize object-based image analysis and cover a broad spectrum of functionality comparable to the flexible pixel-based tools mentioned above are not yet available. Here, we present an object-based open-source software platform that may be used for a broad spectrum of tasks in the field of (biomedical) image analysis: Ranging from algorithm development with an integrated C# compiler to (pipeline-style) one-click analysis provided through a powerful plug-in interface. An object layer structure was designed to handle image objects and their properties and therefore allow high-level formulations of image analysis algorithms. The tool provides flexible and interactive functionality with a variety of image analysis algorithms that may be combined in process chains, an object model editor and visualization/statistics functions. We demonstrate the utility of our approach with the design of an exemplary processing chain for quantifying Ki67 proliferation marker [17,18] in bright-field histological breast cancer. A second example shows the advantages of the object-based software design for the analysis of relationships between objects for the quantification of osteoclast–bone interactions in two-photon fluorescence microscopy data [19,20].

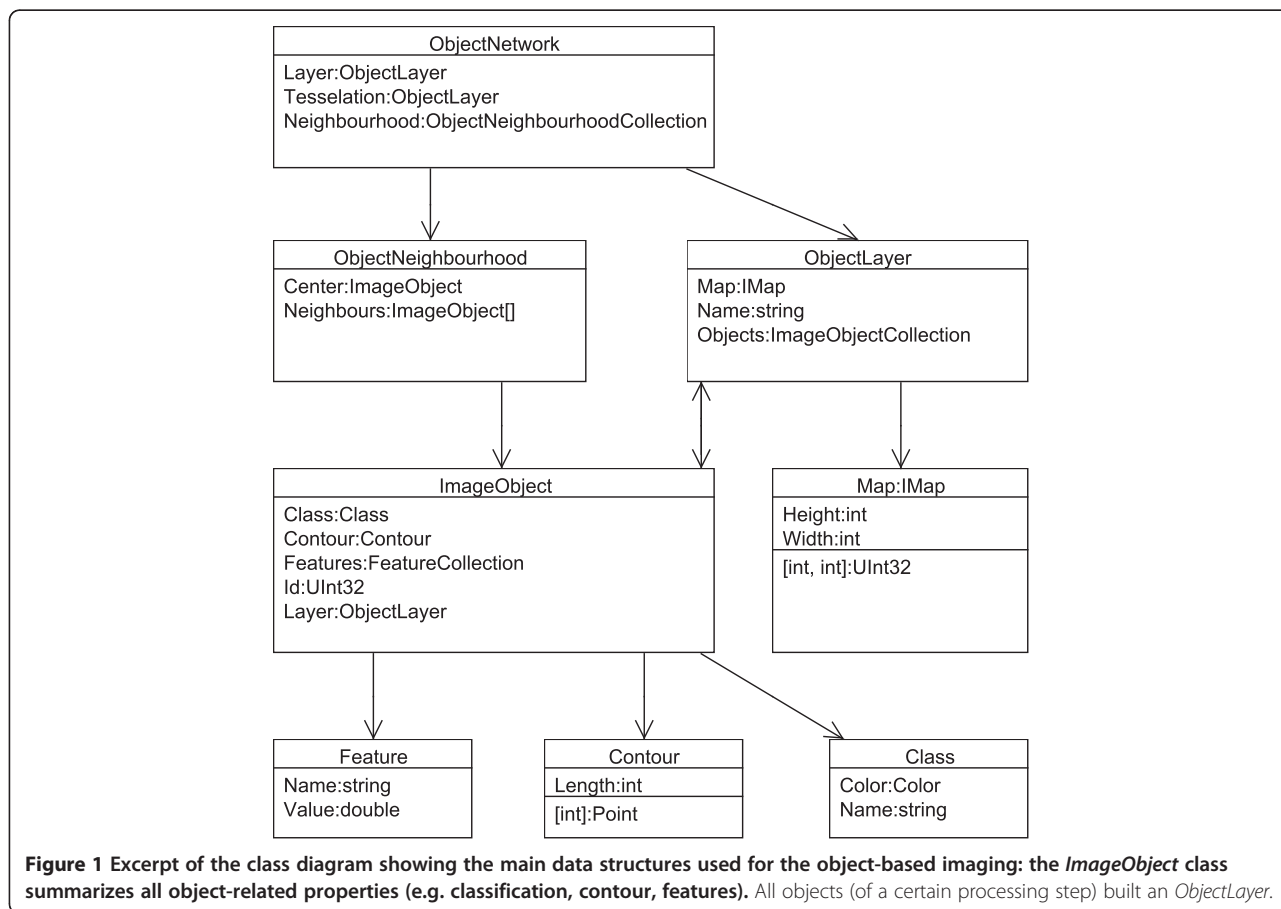
### Implementation

*CognitionMaster* was implemented in C# for Microsoft .NET. *SharpDevelop* [21] was used for the integrated editing of C# code. Icons from the *Tango!* [22] project were used for the graphical user interface. A documentation of the application programming interface (API) was generated using *doxygen* [23]. Figure 1 shows the main data structures, which were designed to organize objects/segments (instead of image pixels) and their properties and therefore allow for an object-based (with respect to handle image contents) modeling of software applications and algorithms. The class *ImageObject* represents a coherent group of neighbouring pixels (segment). Several properties can be used to describe an *ImageObject* as the contour for its spatial position or a collection of key value pairs (*Features*) for properties like mean intensity, object pixels etc. The *ObjectLayer* class was designed to summarize all information of one processing step: it contains a list of all *ImageObjects* and a two-dimensional map of the same size as the corresponding image which contains the identifier of the *ImageObject* a pixel belongs to. This map may be used for pixel-based access to objects or pixel-based operations (e.g. distance transformation).

In an analysis pipeline sequential processing modules obtain and pass on their input as *ObjectLayer* instances containing all objects and the respective features. Discrete Voronoi tessellation [24] on the object layers map is used to compute the topological relationships. These are represented in an *ObjectNetwork* class which contains the *ObjectNeighbourhood* for each object. The neighbourhood information can be used, e.g. to compute the length of the border to an object of a certain class or the number of neighbouring objects (or touching objects, respectively) of a certain class. The object-based design makes it easier to communicate between algorithms and handle user interactions: instead of passing a label image (where often the meaning of the labels is implicit) and several lists/dictionaries containing the related properties only an object layer or even a single object has to be passed. E.g. when the mouse is moved over a certain object the *MouseEnterObject* and *MouseLeaveObject* events are fired. The corresponding object is passed with the events data. Via this object reference all consumers of the event get access to all object properties and the corresponding object layer. The *MouseEnterObject* event, for example, is used to display the classification and features of this object (Figure 2A). Everything that happens is that in the event handler the name of the assigned class and the objects features are read from the corresponding *ImageObject* instance and printed to the popup. The object-based approach also allows for the (multi-) selection of objects in the image and scatter plot diagram. Selected objects are highlighted in both representations (Figure 2A and B). All algorithms can be used within the graphical user interface or in any other .NET application by referencing the *CognitionMaster* assemblies.

### Results

*CognitionMaster* is a software tool designed to serve as a basis for various interactive and/or fully automated image analysis tasks. The plug-in concept allows for the expansion of the software with new menu entries, tab pages and handling of special image formats, e.g. plug-ins for the support of DICOM<sup>a</sup> and whole slide images<sup>b</sup> (WSI) are available from the download page. Plug-ins even may extend the image rendering to present additional image (meta) information e.g. image sharpness maps [25] or visited regions [26] in virtual slides, an emerging field of medical imaging [27]. Additionally, plug-ins may change the default behavior of the software (e.g. reactions on mouse and key events) by overriding the corresponding event handler: the ROI manager plug-in enables users to assign labels to points or regions of interest by mouse-clicks. Besides the application *CognitionMaster* offers a C# component library with more than 250 classes (image analysis algorithms, data structures and visual components, etc.). Image analysis algorithms may be



selected from the *Process* menu (Figure 3), whereas results of prior executed algorithms may be used as input for consecutive processing steps. Processing steps may also be combined into processing chains (“analysis pipelines”). Objects are rendered as contours or filled, colored according to an assigned class (Figure 2 and 4) or with a default color if no class is assigned. A debug console may be used to output status messages when writing new algorithms (see in-program scripting section for details). For global data interchange (between plug-ins and/or in-program scripts) the *VarInterop* plug-in may be used: this plug-in manages and displays global variables. For data visualization a two-dimensional interactive scatter plot diagram is available. How objects can be classified best can be tested using the built-in object model editor: This plug-in enables the user to describe the ideal “model” object with features. This object model may be applied to a certain image and the user may check if his “object model” meets the real image.

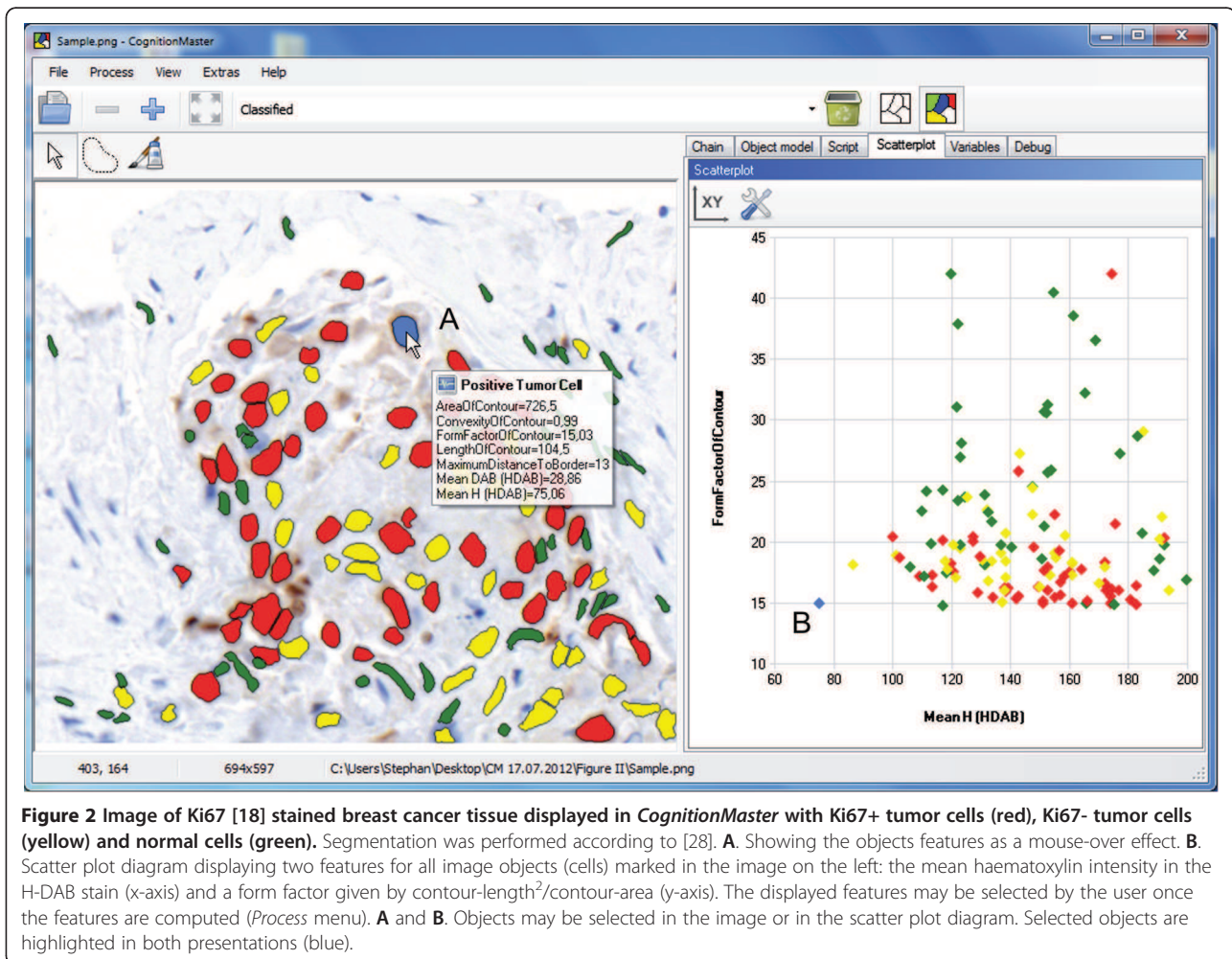
#### Ki67 proliferation assessment in breast cancer immunohistology samples

In the following section we demonstrate how different methods or hypothesis can be tested using a processing

chain for the quantification of proliferation marker assessment in breast cancer using Ki67 immunohistochemic staining [17,18]. Scoring proliferation through Ki67-immunohistochemistry is an important component in predicting response to chemotherapy in breast cancer patients. Therefore, positive and negative stained tumor cells have to be detected and counted by the pathologist in routine diagnostics. If one would like to set up a fully automated scoring system 4 major functions have to be implemented:

- a. Detection and segmentation of cell nuclei
- b. Classification into tumor and non-tumor cells
- c. Classification into Ki67-positive (Ki67+) and Ki67-negative (Ki67-) cells
- d. Counting of all Ki67+ and Ki67- tumor cells

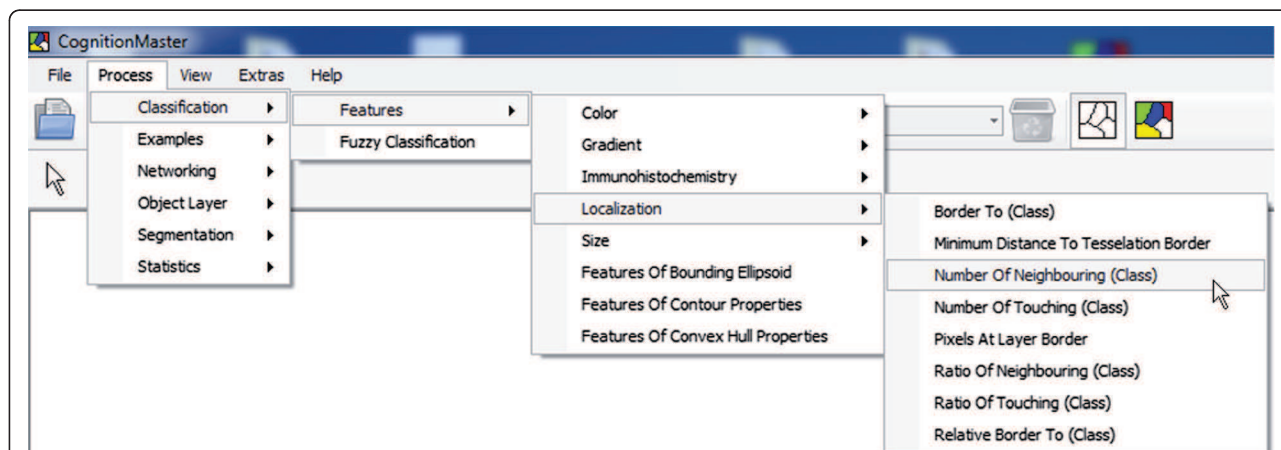
A method for the detection and segmentation of cell nuclei (a) was introduced recently in [28]. The modules described in this study (contour-based segmentation, non-compact pixel removal and concave object separation) are standard-components of the *CognitionMaster* library. For the second step of the scoring system (b) it may be interesting to test several hypothesis on how



tumor cells can be distinguished best from non-tumor cells: often non-tumor cells are smaller than tumor-cells or have a different aspect ratio (connective tissue cells are often spindle-shaped). For the validation of these hypotheses it is important to be able to easily get an impression of the feature values of a certain object or the distribution of a certain feature respectively. Therefore, users may select one or more features from the *Process* menu once the segmentation is executed. The selected features are computed for each object (cell nuclei). Then, features are displayed in a popup when the mouse cursor is moved over a certain object (Figure 2A). Additionally, the scatter plot diagram (Figure 2B) can be used to decide which feature and threshold is optimal. For the classification into Ki67+ and Ki67- cells features based on color deconvolution [29] may be used from the *Immunohistochemistry* group. Finally, the number of cells for each class are counted using *Num(Class)* algorithm available from the *Process* menu. The complete example including the sample image, the processing chain and a guidance file (readme.txt) can be found in Additional file 1 online.

### Cell interaction analysis

In the example described here, we demonstrate the advantages of our object-based approach for the analysis of relationships between objects. Therefore, we use image analysis to test the hypothesis that osteoclast attachment to bone surface as a proxy for bone decomposition is influenced by sphingosine-1-phosphate [19]. The main aspect of this approach is the analysis of interactions between bone tissue and osteoclasts. *CognitionMaster* contains a set of routines/algorithms that allow for the computation and analysis of object neighbourhoods e.g. the *ObjectNetwork* class (Figure 1) and the *NumberOfNeighbouring* and *NumberOfTouching* features. The first step of the analysis pipeline is the detection of the different types of tissue/cells (bone tissue, osteoclasts, monocytes). Here we re-implemented the detection algorithm including the threshold selection method and channel merging approach presented in the original publication [20] using in-program scripting (Figure 4, see in-program scripting section and the first processing step in the processing chain in Additional file 2 online).



**Figure 3** The *Process* menu of the main application window. This menu is generated on the applications startup: Therefore all available scripts and processing chains are loaded. Users may add own scripts to the respective directory resulting in additional menu entries.

After the segmentation the next step of the analysis is the computation of the interfacial-length between bone tissue and osteoclasts. This can easily be solved by using the *BorderTo* feature, which is a standard component of the software and computes the number of contour pixels with a neighbouring pixel of another object class. Finally, the user may export the results containing the object id, class name and features into a .csv file which he may then open with other statistics tools (e.g. Microsoft EXCEL) for further analysis. The complete example containing the sample image, the processing chain and a guidance file (readme.txt) can be found in Additional file 2 online.

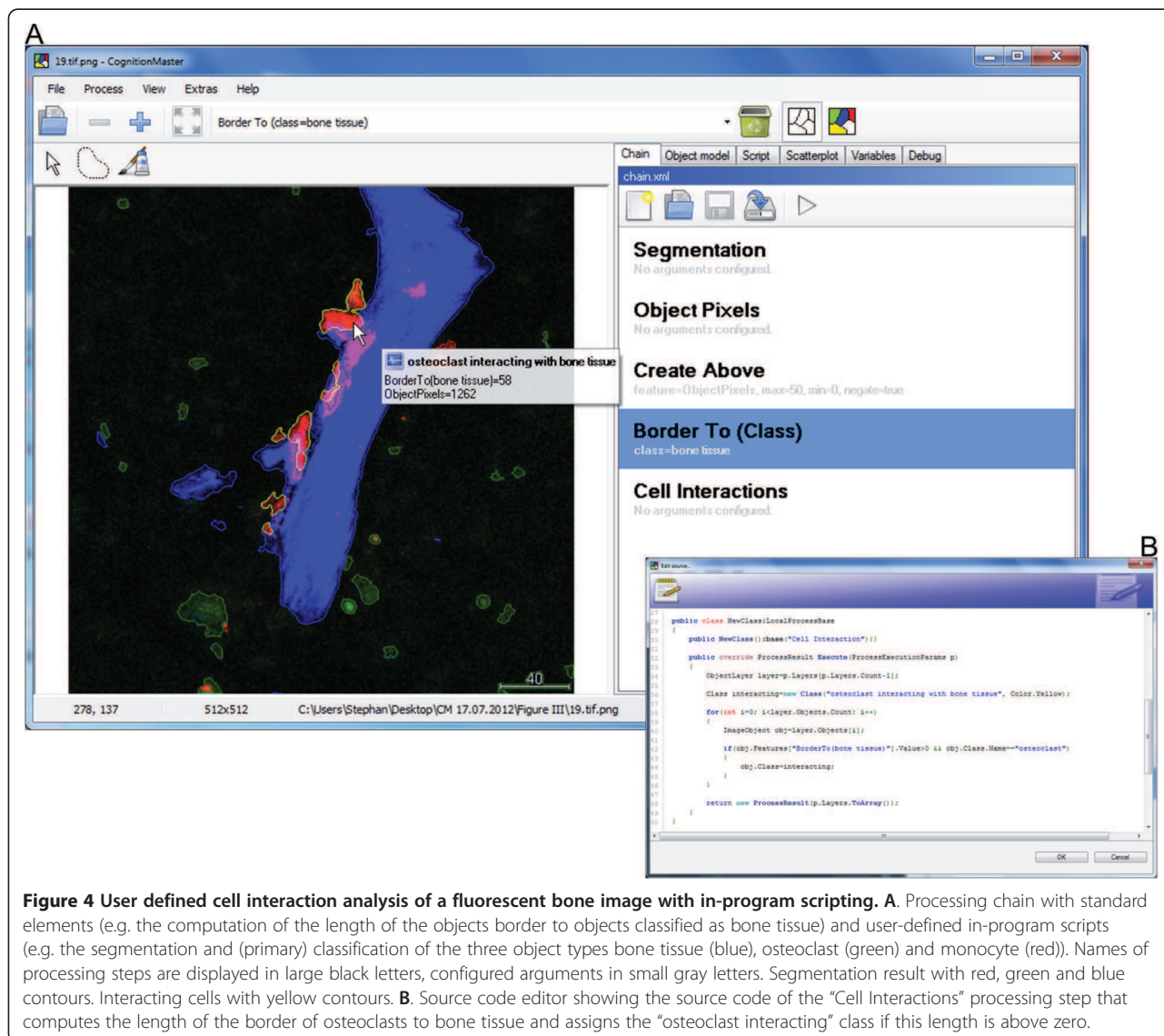
### In-program scripting

While many image analysis algorithms are already available, the development of new methods is an important field in digital image analysis. Furthermore many image analysis tasks require special pixel- or object-based operations that can hardly be part of common purpose image analysis software e.g. the specialized automatic threshold segmentation and channel merging technique used in the previous section. It can be very helpful to be able to write and execute user-specific source code without needing an extra software project or switching between different applications. A high-execution speed is important while testing the new idea and in the consecutive deployment to other users which should be, however, as simple as possible: it should be possible to use the new code directly among other (standard) components without the need for a re-implementation due to performance or deployment reasons. *CognitionMaster* comes with an integrated C# editor/compiler: this component enables the user to write, compile and execute standard C# code (so called scripts) within the main application (Figure 4B). For most algorithms contained in the *CognitionMaster* library

(e.g. watershed-segmentation and extraction of various features) a script comes with the installation and can be accessed via the *Process* menu (Figure 3), which is generated dynamically when the software is launched: Therefore all available scripts and processing chains are read from a certain file directory. This offers the opportunity to add own scripts to this menu. A major advantage of the in-program scripting technique is the high execution speed: the script is compiled using the standard .NET compiler and then executed with the same performance as a normal .NET executable (not an interpreted script). In contrast, the macro language of *ImageJ* is interpreted and therefore offers lower performance. Additionally, scripts can be combined in processing chains, whereby the source code of a script is copied into the processing chain and compiled when executing. Embedding the source code (instead of linking to it) offers the possibility to change the behavior (source code) of standard components in a certain processing chain. Furthermore, arguments can be denoted for each processing step, which avoids the necessity of using input dialogs at each module.

### Comparison with ImageJ

With *ImageJ* the built-in *ROI Manager* and *Particle Analyzer* can be used to analyze images on the object-level (see Additional file 3 online for details). Therefore a binary image is required that marks every pixel either as foreground or background. Unfortunately, differentiating two touching objects is not possible with this technique, which is a major drawback compared to our approach. Once the particles are found objects are shown as an overlay. Unfortunately, the user has to select preliminary how objects are rendered (e.g. object contours, ellipses or filled). It is not possible to easily switch



**Figure 4** User defined cell interaction analysis of a fluorescent bone image with in-program scripting. **A.** Processing chain with standard elements (e.g. the computation of the length of the objects border to objects classified as bone tissue) and user-defined in-program scripts (e.g. the segmentation and (primary) classification of the three object types bone tissue (blue), osteoclast (green) and monocyte (red)). Names of processing steps are displayed in large black letters, configured arguments in small gray letters. Segmentation result with red, green and blue contours. Interacting cells with yellow contours. **B.** Source code editor showing the source code of the “Cell Interactions” processing step that computes the length of the border of osteoclasts to bone tissue and assigns the “osteoclast interacting” class if this length is above zero.

“online” between different rendering modes. In contrast, *CognitionMaster* allows “online” switching of the rendering mode (contours, filled, none) and moreover, supports visualization of classification results by rendering each object with the color of the assigned class. *ImageJ* only allows for the selection of certain objects by using the object list provided by the *ROI Manager* plug-in. Free-hand selection or selecting individual objects by mouse-clicks are not supported. Mouse-over feature statistics for a certain object are also not supported by *ImageJ* as well as the interactive scatter plot diagram, which are both available in *CognitionMaster* (see Figure 2). Finally, *CognitionMaster* allows for overriding or extending user-interactions by using the corresponding events (e.g. *MouseEnterObject*, *MouseLeaveObject* or *SelectedObjects Changed*), which is not possible with *ImageJ* because of the lack of a dedicated object model.

## Discussion

Conventional image analysis applications were designed with the single image pixel as basic processing unit. This concept is advantageous when image-to-image transformations (filters) are the main purpose of the software application, but has limitations when image objects and their properties and relationships between objects are considered. In this case the object-based approach allows for the use of high-level features, such as the interactivity with image objects like mouse-over effects (Figure 2A) and data-analysis on the object level through a GUI-based pipeline assembly such as, for instance, the evaluation of spatial relationships between cells. Object-based approaches, however, require *a priori* object segmentation/detection, for which the *CognitionMaster* offers a variety of established approaches and the option to implement alternative methods using in-program scripting.

Moreover, using object-based data structures to process object-based image contents facilitates the design of a software architecture compliant to state-of-the-art object-oriented-programming (OOP). This favors clarity of design and especially the reusability of available analysis modules not only in processing pipelines or in-program scripts within the *CognitionMaster*, but also as components in other .NET projects. In contrast to that, several label maps (or one label map and several lists) are required to store “pixel” to “segment”, “segment” to “features” and “segment” to “class” mappings in pixel-based approaches, which can hardly be reused (e.g. by inheritance).

## Conclusions

We introduce the open source software system *CognitionMaster* that features object-based image analysis and implements a large variety of image analysis algorithms that may be extended using a plug-in interface or in-program scripting. Therefore, *CognitionMaster* may serve as a basis for a broad range of applications in the field of digital image analysis.

## Availability and requirements

*CognitionMaster* is released under the terms of the GNU General Public License (GPL). The installation packages, source code files and documentation for the main application and available plug-ins can be found on <http://sourceforge.net/projects/cognitionmaster/>. *CognitionMaster* can be installed on all *Microsoft Windows* PCs (*Windows XP* or newer, 32-bit or 64-bit). *Microsoft .NET 2.0* is required. A documentation of the application programming interface (API) is provided with the installation packages.

## Endnotes

<sup>a</sup>Using ClearCanvas, free

<sup>b</sup>Using Virtual Slide Access – SDK 4.0 from VMscope GmbH (Germany), commercial

## Additional files

**Additional file 1: Example 1.**

**Additional file 2: Example 2.**

**Additional file 3: Comparison with ImageJ.**

## Competing interests

The authors declare no conflict of interests.

## Authors' contributions

SW, FK and DH wrote the paper. SW implemented the software. BL wrote the software manual. FK, AS, CD, PH, MD and MB supervised the software design and architecture and algorithm implementation. DH, AS and FK performed software tests. FK and CD selected and provided histological samples. MI and MK provided the two-photon microscopy data and advised on the image analysis. All authors reviewed the manuscript. All authors read and approved the final manuscript.

## Acknowledgements

This work was partly supported by the Grant from the International Human Frontier Science Program (RGY0077/2011) to FK and MI.

## Author details

<sup>1</sup>Institute of Pathology, Charité University Hospital Berlin, Charitéplatz 1, Berlin 10117, Germany. <sup>2</sup>VMscope GmbH, Charitéplatz 1, Berlin 10117, Germany. <sup>3</sup>Laboratory of Cellular Dynamics, WPI-IFReC, Osaka University, 3-1 Yamada-oka, Suita 5650871 Osaka, Japan. <sup>4</sup>CREST, Japan Science and Technology Agency (JST), 5 Sanbancho, Chiyoda-ku 1020075 Tokyo, Japan. <sup>5</sup>University of Applied Sciences Berlin, Wilhelminenhofstraße 75A, Berlin 12459, Germany. <sup>6</sup>Institute of Pathology, University Hospital Heidelberg, Im Neuenheimer Feld 220/221, Heidelberg 69120, Germany. <sup>7</sup>Department of Medicine I, University of Ulm, Albert-Einstein-Allee 23, Ulm 89081, Germany.

Received: 11 January 2013 Accepted: 17 February 2013

Published: 27 February 2013

## References

- Graham J, Taylor CJ: Automated chromosome analysis using the magiscan image analyser. *Anal Quant Cytol* 1980, **2**:237–242.
- Salazar I, Cifuentes JM, Sánchez Quinteiro P, García Caballero T: Structural, morphometric, and immunohistological study of the accessory olfactory bulb in the dog. *Anat Rec* 1994, **240**:277–285.
- Portet S, Vassy J, Beil M, Millot G, Hebbache A, Rigaut JP, Schoevaert D: Quantitative analysis of cytokeratin network topology in the MCF7 cell line. *Cytometry* 1999, **35**:203–213.
- Brinkmann M, Lütkemeyer D, Gudermann F, Lehmann J: New technologies for automated cell counting based on optical image analysis; The Cellscreen. *Cytotechnology* 2002, **38**:119–127.
- Gutiérrez R, Gómez F, Roa-Peña L, Romero E: A supervised visual model for finding regions of interest in basal cell carcinoma images. *Diagn Pathol* 2011, **6**:26.
- Grimm M, Krimmel M, Alexander D, Munz A, Kluba S, Keutel C, Hoffmann J, Poligkeit J, Reinert S, Hoefert S: Prognostic value of histamine H1 receptor expression in oral squamous cell carcinoma. *Clin Oral Investig* 2012, Epub ahead of print.
- Schneider CA, Rasband WS, Eliceiri KW: NIH Image to ImageJ: 25 years of image analysis. *Nat Meth* 2012, **9**:671–675.
- Jones TR, Kang I, Wheeler DB, Lindquist RA, Papallo A, Sabatini DM, Golland P, Carpenter AE: Cell profiler analyst: data exploration and analysis software for complex image-based screens. *BMC Bioinforma* 2008, **9**:482.
- Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B, Tinevez J, White DJ, Hartenstein V, Eliceiri K, Tomancak P, Cardona A: Fiji: an open-source platform for biological-image analysis. *Nat Meth* 2012, **9**:676–682.
- de Chaumont F, Dallongeville S, Chenouard N, Hervé N, Pop S, Provoost T, Meas-Yedid V, Pankajakshan P, Lecomte T, Le Montagner Y, Lagache T, Dufour A, Olivo-Marin J: Icy: an open bioimage informatics platform for extended reproducible research. *Nat Meth* 2012, **9**:690–696.
- OpenCV. <http://opencv.org>.
- Yoo TS, Ackerman MJ, Lorensen WE, Schroeder W, Chalana V, Aylward S, Metaxas D, Whitaker R: Engineering and algorithm design for an image processing Api: a technical report on ITK—the insight toolkit. *Stud Health Technol Inform* 2002, **85**:586–592.
- Kayser K, Görtler J, Borkenfeld S, Kayser G: How to measure diagnosis-associated information in virtual slides. *Diagn Pathol* 2011, **6**(Suppl 1):S9.
- Blaschke T, Strobl J: What's wrong with pixels? Some recent developments interfacing remote sensing and GIS. *GIS-Zeitschrift für Geoinformationssysteme* 2001, **06**:12–17.
- Hay GJ, Castilla G, Wulder MA, Ruiz JR: An automated object-based approach for the multiscale image segmentation of forest scenes. *Int J Appl Earth Obs Geoinf* 2005, **7**:339–359.
- Baatz M, Arini N, Schäpe A, Binnig G, Linssen B: Object-oriented image analysis for high content screening: detailed quantification of cells and sub cellular structures with the cellenger software. *Cytometry A* 2006, **69**:652–658.
- Bouzubar N, Walker KJ, Griffiths K, Ellis IO, Elston CW, Robertson JF, Blamey RW, Nicholson RI: Ki67 immunostaining in primary breast cancer: pathological and clinical associations. *Br J Cancer* 1989, **59**:943–947.

18. Gerdes J: **Ki-67 and other proliferation markers useful for immunohistological diagnostic and prognostic evaluations in human malignancies.** *Semin Cancer Biol* 1990, **1**:199–206.
19. Ishii M, Egen JG, Klauschen F, Meier-Schellersheim M, Saeki Y, Vacher J, Proia RL, Germain RN: **Sphingosine-1-phosphate mobilizes osteoclast precursors and regulates bone homeostasis.** *Nature* 2010, **465**:966.
20. Klauschen F, Ishii M, Qi H, Bajénoff M, Egen JG, Germain RN, Meier-Schellersheim M: **Quantifying cellular interaction dynamics in 3D fluorescence microscopy data.** *Nat Protoc* 2009, **4**:1305–1311.
21. #Develop. <http://www.icsharpcode.net/>.
22. Tango. <http://tango.freedesktop.org/>.
23. Doxygen. <http://doxygen.org>.
24. Velić M, May D, Moresi L: **A fast robust algorithm for computing discrete voronoi diagrams.** *J Math Model Algor* 2009, **8**:343–355.
25. Zerbe N, Hufnagl P, Schlüns K: **Distributed computing in image analysis using open source frameworks and application to image sharpness assessment of histological whole slide images.** *Diagn Pathol* 2011, **6**:S16.
26. Romo D, Romero E, González F: **Learning regions of interest from low level maps in virtual microscopy.** *Diagn Pathol* 2011, **6**:S22.
27. Kayser K, Görtler J, Borkenfeld S, Kayser G: **Interactive and automated application of virtual microscopy.** *Diagn Pathol* 2011, **6**:S10.
28. Wienert S, Heim D, Saeger K, Stenzinger A, Beil M, Hufnagl P, Dietel M, Denkert C, Klauschen F: **Detection and segmentation of cell nuclei in virtual microscopy images: a minimum-model approach.** *Sci Rep* 2012, **2**:503.
29. Ruifrok AC, Johnston DA: **Quantification of histochemical staining by color deconvolution.** *Anal Quant Cytol Histol* 2001, **23**:291–299.

doi:10.1186/1746-1596-8-34

**Cite this article as:** Wienert et al.: CognitionMaster: an object-based image analysis framework. *Diagnostic Pathology* 2013 **8**:34.

**Submit your next manuscript to BioMed Central and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)



**Marginal zone lymphomas with monocytoid morphology express T-bet and are associated with a low number of T cells in extranodal locations.**

*Lohneis P, Wienert S, Klauschen F, Ullrich A, Anagnostopoulos I, Jöhrens K. Leuk Lymphoma 2013;1-16.*

<http://dx.doi.org/10.3109/10428194.2013.797082>



## ***Multicolor fate mapping of Langerhans cell homeostasis.***

*Ghigo C, Mondor I, Jorquera A, Nowak J, Wienert S, Zahner SP, Clausen BE, Luche H, Malissen B, Klauschen F, Bajenoff M. J Exp Med 2013;184:2417.*

<http://dx.doi.org/10.1084/jem.20130403>

***Fate mapping reveals origin and dynamics of lymph node  
follicular Dendritic Cells.***

*Jarjour M, Jorquera A, Mondor I, Wienert S, Narang P, Coles MC, Klauschen F, Bajenoff M.  
J Exp Med, Manuskript akzeptiert.*

<http://dx.doi.org/10.1084/jem.20132409>