

Formalisierung und Quantifizierung verbaler Beschreibungen von Zellanordnungen für die computergestützte zytologische Krebsdiagnose

Stefanie Mehl¹, Timna Schneider¹, Dietrich Meyer-Ebrecht¹
und Alfred Böcking²

¹Lehrstuhl für Bildverarbeitung, RWTH Aachen, 52056 Aachen

²Institut für Cytopathologie, Heinrich-Heine-Universität Düsseldorf,

Moorenstr. 5, 40225 Düsseldorf

Email: {sm,ts}@lfm.rwth-aachen.de

Zusammenfassung. In der Zytopathologie dienen mit minimalinvasiven Methoden gewonnene Zellen als Material für die Diagnostik. Bislang werden für computergestützte Untersuchungen in der Regel einzelne Zellen herangezogen. Wichtige Informationen für die Diagnose liefert aber auch die Analyse typischer Zellanordnungen. Es wird ein Ansatz zur Beschreibung und Unterscheidung einiger dieser Anordnungen vorgestellt. Untersucht werden verschiedene geometrische Merkmale sowie Merkmale aus der Delaunay-Triangulation. Die Validierung erfolgt anhand von Scatterplots und einem Gütekriterium auf der Basis von Scattermatrizen.

1 Einleitung

Die Zytopathologie bietet eine Möglichkeit zur frühen Krebsdiagnose. Dabei werden Zellen aus Abstrichen, Körperflüssigkeiten oder Punktaten auf Objektträger aufgebracht, eingefärbt und unter dem Lichtmikroskop begutachtet. Durch verschiedene Färbungen können verschiedene Eigenschaften der Zellen besonders hervorgehoben werden. Im Rahmen des Projektes *Multimodale Zellanalyse (MMZA)* [1] wird ein System entwickelt, welches es dem Mediziner ermöglicht, verschiedene Untersuchungen in unterschiedlichen Färbungen an denselben Zellen durchzuführen. Der Vorteil des Verfahrens besteht sowohl in einer Verbesserung der Diagnosegenauigkeit als auch im verminderten Bedarf an Zellmaterial für die Diagnose.

Zur Zeit erfolgt die Suche nach auffälligen Zellen bzw. Regionen auf dem Objektträger manuell durch den Mediziner. Dies ist sehr zeitaufwändig und begrenzt den Einsatz in der Routinediagnostik. Zur weiteren Unterstützung des Mediziners ist es wünschenswert, diese Suche automatisch durchzuführen, sowie dem Mediziner relevante Anordnungen zur weiteren Begutachtung präsentieren zu können. Dazu müssen die von Mediziner vorliegenden Beschreibungen der Zellanordnungen formalisiert werden.

Insbesondere bei der Untersuchung von Ergüssen der serösen Körperhöhlen ist nicht nur die Betrachtung der Zellen für sich, sondern auch die Art ihrer An-

ordnung diagnoserelevant. Handelt es sich dabei um fest verbundene Zellanordnungen (Verbände), bleiben diese auch bei den verschiedenen Präparationsarten erhalten.

2 Stand der Forschung

Es existiert eine Vielfalt möglicher Zellanordnungen in Ergüssen. In der Literatur werden diese mit bildhaft beschreibenden Namen bezeichnet (z.B. als „Rosette“, „Zeile“, „Ball“), welche abhängig von der subjektiven Vorstellung des Mediziners und deshalb nicht immer übereinstimmend gewählt sind [2, 3].

Bislang gibt es keine umfassenden Versuche, diese Beschreibungen mathematisch zu fassen und sie für eine automatische Erkennung sowie für eine automatische Auswahl zu untersuchender Regionen in der morphologischen Färbung zu nutzen. Es ist lediglich ein Ansatz bekannt, in Schilddrüsenpräparaten die follikulär angeordneten Zellgruppen zu finden [4].

Andererseits sind in der technischen Literatur [5] eine Reihe von Merkmalen zu finden, die zur Analyse von Anordnungen (z.B. Galaxien, Sternbilder) oder Beschreibung bestimmter Formen (z.B. Blätter, einzelne Zellen [6]) verwendet werden. Dazu gehören beispielsweise einfache Formbeschreibungen, graphenbasierte Merkmale und verschiedene statistische Merkmale. Eine Schwierigkeit der meisten Merkmale liegt in ihrer Anwendung auf Objekte mit hoher Gestaltvariabilität, wie sie bei allen biologischen Objekten und damit auch bei Zellanordnungen gegeben ist.

3 Merkmale

Als Grundlage für eine formale und quantitative Beschreibung von Anordnungen wurden zunächst Beschreibungen für einige relevante Anordnungstypen (Abschnitt 5) zusammengestellt.

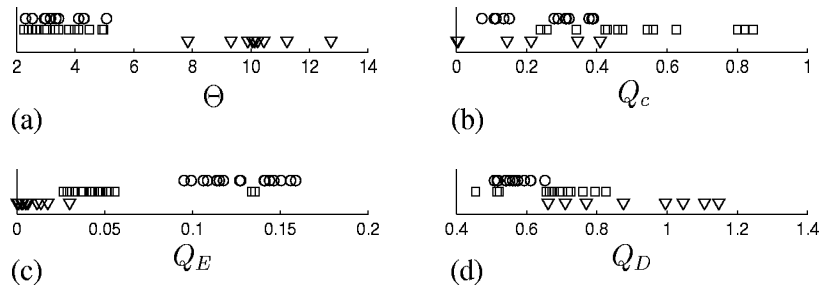
Plasma und Zellkerne der Anordnungen liegen für die Analyse in Form von diskreten Konturen vor.

Für die Untersuchung der Zellanordnungen werden bei den Zellkernen die Schwerpunkte als Berechnungsgrundlage herangezogen, für das Zellplasma sowohl die aus der Kontur gewonnene Maske als auch die Kontur selbst.

Für die Zellkernanordnungen werden Abstandsmerkmale in Bezug auf den gemeinsamen Schwerpunkt, die Ausgleichsgerade durch die Anordnung sowie Best-Fit-Ellipsen aus den Schwerpunkten der Kerne berechnet. Weitere topologische Untersuchungen basierend auf der Delaunay-Triangulation umfassen eine Analyse der Dreiecksflächen, der Kantenlängen und der Anzahl der nächsten Nachbarn, repräsentiert durch die zugehörigen Kanten.

Für das Zellplasma werden vorrangig rein geometrische Merkmale wie der Formfaktor oder der Quotient aus dem halben maximalen Abstand zwischen zwei Pixeln auf der Aussenkontur und dem Radius des maximalen einbeschriebenen Kreises bestimmt. Weitere Merkmale bauen auf der Kombination von Kernschwerpunkten und Zytoplasma auf.

Abb. 1. Beispiele für eindimensionale Scatterplots: (a) Θ , (b) Q_c , (c) Q_E , (d) Q_D . Legende: Kreis=Mosaik, Quadrat=Rosette, Dreieck=Linie



4 Validierung

Die Trennfähigkeit der Merkmale wird zunächst an Scatterplots (Abbildung 1) untersucht. Für vielversprechende Merkmale wird als Gütekriterium für die Trennfähigkeit der Quotient aus der Varianz zwischen den Klassen und der Summe der Varianzen innerhalb der Klassen berechnet [7]. Dieser Wert ist klein für eine schlechte Trennbarkeit und groß, wenn sich die Klassen deutlich unterscheiden. Die besten Merkmale sind in der Tabelle 1 aufgeführt.

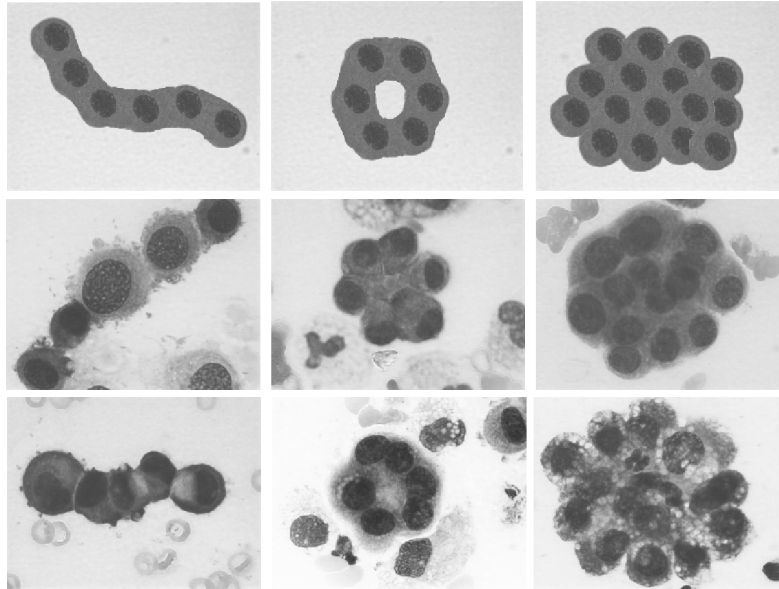
5 Ergebnisse

Von den möglichen Anordnungen können bislang Rosetten, Mosaiken und Zeilen untersucht werden, wobei es sich bei Rosetten um eher ringförmige, bei Mosaiken um flächige, aber zusammenhängende Anordnungen von Zellen und bei Zeilen um linienförmig verbundene Zellen handelt (Abbildung 2).

Basierend auf den formalen Beschreibungen wurde ein Satz von Testbildern erzeugt. Dieser enthält aus realen Zellen künstlich zusammengesetzte Anordnungen. Dieser Testdatensatz besteht aus insgesamt 50 Objekten, jede Klasse ist ungefähr gleich stark vertreten. Eine gezielte Variation der „Idealform“ dient der differenzierten Merkmalsanalyse in den Scatterplots. Weiterhin steht eine Auswahl aus ca. 80 realen, handklassifizierten Aufnahmen von aus Ergüssen gewonnenen Dünnschichtpräparaten zur Verfügung.

Die Analyse der Scatterplots (Abbildung 1) liefert schon für relativ unkomplizierte Merkmale eine gute Trennfähigkeit. Insgesamt wurden 33 Merkmale für die weiteren Untersuchungen ausgewählt. Davon erwiesen sich folgende Merkmale als besonders vielversprechend: Das Verhältnis des Umkreisradius zum Radius des maximalen einbeschriebenen Kreises Θ ; der Quotient aus maximalem zu minimalem Abstand der Kerne von ihrem gemeinsamen Schwerpunkt Q_c ; das Verhältnis der mittleren Kantenlänge der Delaunay-Dreiecke zum mittleren Kernumfang Q_D ; der Quotient aus der Standardabweichung des Kernabstandes zur Plasmaaußenkante und dem mittleren Kernumfang Q_E .

Abb. 2. Zellanordnungen, v. l. n. r.: Linie, Rosette, Mosaik. Obere Reihe: Beispiele für Testanordnungen. Mittlere und untere Reihe: Beispiele für reale Aufnahmen.



Θ ist das bislang beste Merkmal für die Beschreibung von Linien. Zur Trennung von Rosetten und Mosaiken zeigen sich Q_c , Q_D und Q_E geeignet.

In Tabelle 1 ist zu erkennen, dass die Werte des Gütekriteriums stark schwanken. Den mit Abstand höchsten Wert erreicht das Merkmal Θ , was leicht am Scatterplot in Abbildung 1(a) nachzuvollziehen ist.

Die Untersuchung der realen Aufnahmen dient der Berücksichtigung natürlicher Gestaltvariabilität. Auf diese Weise wird auch die Abgrenzung der analysierten Klassen gegen andere Anordnungen untersucht. Bei den natürlichen Bildern ist in den Scatterplots eine Abgrenzung der Klassen weniger deutlich zu erkennen. Dies spiegelt sich auch in den Werten des Gütekriteriums wieder. Das liegt daran, dass die unspezifischen Anordnungen teilweise in einzelnen Merkmalen Rosetten oder Mosaiken ähnlich sind und eine Trennung erst durch Kombination mehrerer Merkmale möglich werden kann.

Tabelle 1. Einige Merkmale und ihre Scatterwerte

Merkmal	Scatterwert	
	test	real
Q_c	1,3173	0,3026
Q_D	1,5858	1,0716
Q_E	5,1581	0,6506
Θ	13,6568	1,8873

6 Diskussion

Die Ergebnisse zeigen, dass bereits recht einfache und intuitive Merkmale eine hohe Relevanz bezüglich entscheidender Formeigenschaften besitzen, und somit eine Vertrauensbasis für die automatische Selektion relevanter Zellanordnungen schaffen können. Der Zeitaufwand für die Berechnung dieser Merkmale ist vergleichsweise gering. Bislang fehlen jedoch ausreichend viele sicher klassifizierte natürliche Trainingsmuster, um die tatsächliche Trennfähigkeit der Merkmale auch mittels statistischer Methoden oder geeigneter Klassifikatoren zu untersuchen. Auch für die Analyse weiterer Anordnungen und ihrer mathematischen Beschreibung ist eine größere Trainingsmenge erforderlich. Die untersuchten Merkmale bilden eine Ausgangsbasis für die Suche nach weiteren Merkmalen.

Weiterhin ist eine gute Segmentierung wesentlich für die Anwendbarkeit der Merkmale. Es gibt allerdings dreidimensionale Anordnungen mit sehr vielen Kernen, die sich nicht ausreichend segmentieren lassen, beispielsweise so genannte „Bälle“. Diese können nicht für die gängigen Analyseverfahren wie DNA-Zytometrie oder AgNOR-Analyse betrachtet werden, ihr Auftreten an sich ist jedoch bereits pathologisch. Eine mögliche Beschreibung kann zum Beispiel auf Formmerkmalen der Plasmakontur (bei „Bällen“ Ellipsenförmigkeit) und der aufgrund der großen Anzahl enthaltener Kerne geringen Lichtdurchlässigkeit basieren.

Danksagung. Das Projekt wird vom Viktor-und-Mirka-Pollak-Fonds unterstützt.

Literaturverzeichnis

1. Böcking A, Stockhausen J, Meyer-Ebrecht D. Towards a single cell cancer diagnosis. Multimodal and Monocellular Measurements of Markers and Morphology (5M). *Cell Oncol* 2004;26:73–79.
2. Feichter G, Dalquen P. *Zytopathologie*. Springer; 2000.
3. Freudenberg N. *Zytopathologie*. Schattauer; 1988.
4. Jütting U, Gais P, Rodenacker K, et al. Analyse von Objekttagglomeraten in Bildern. In: Kazmierczak H, editor. *Mustererkennung 1983*. vol. 35 of VDE-Fachberichte. DAGM, VDE; 1983. p. 137–142.
5. da Fontoura Costa L, Marcondes Cesar Jr R. *Shape Analysis and Classification: Theory and Praxis*. Image Processing Series. CRC Press LLC; 2001.
6. Rodenacker K, Bengtsson E. A feature set for cytometry on digitized microscopic images. *Anal Cell Pathol* 2003;25:1–36.
7. Fukunaga K. *Statistical Pattern Recognition*. 2nd ed. Academic Press; 1990.