

Graphen als Modelle für Gewebestrukturen

Felix Riehm

Hochschule für Technik und Wirtschaft

Berlin, Deutschland

Felix.Riehm@student.htw-berlin.de

Abstract—Die computerunterstützte Analyse von Gewebe ist in der Pathologie zu einem festen Bestandteil geworden. Man spricht von der digitalen Pathologie, die Zeitaufwände verringern und zu einer quantitativen Aussage über Gewebe verhelfen soll. In dieser Arbeit richten wir unser Augenmerk auf Graphen als Modelle für Gewebestrukturen. Es wird insbesondere das Voronoi-Diagramm betrachtet, welches die Basis für viele graphenbasierte Methoden zur Gewebeanalyse ist. Wir werden es formal beschreiben und eine Pseudocode Umsetzung zeigen. Eine effiziente algorithmische Implementation wird anhand des Sweep-Line Algorithmus von Fortune im näheren Detail deskriptiv erklärt und eine Aussage zur Performance gegeben. Des Weiteren widmen wir uns kurz zu den Merkmalen von Graphen, der Knotenidentifizierung und der aktuellen Entwicklung von Graphen als Modelle für Gewebestrukturen.

Index Terms—digital histopathology, voronoi diagram, fortunes sweep line algorithm, graph-based methods, delaunay triangulation

I. EINLEITUNG

Die Analyse von Gewebestrukturen in der digitalen Pathologie ist divers. Dazu gehören Texturanalyse, Spektralanalyse, Analysen basierend auf Farbwerten und graphenbasierte Modelle. Graphenmodelle eignen sich besonders gut, weil sie die topologischen Eigenschaften von Gewebe effizient darstellen können [2]. Wir wollen diese daher in den nächsten Abschnitten genauer betrachten. Für die Anwendung eines solchen Graphenmodells, müssen zuerst Knoten und Verbindungen aufgebaut werden. Im ersten Abschnitt "Identifizierung von Knoten" werden wir beschreiben, wie Knoten gefunden werden können und welche relevanten Merkmale in einem Gewebe hierfür vorhanden sein können. Sind die Knoten bekannt und mit Informationen versehen worden, beispielsweise mit der Koordinate eines Knotens, müssen Verbindungen zwischen den Knoten gefunden werden, um eine Graphenstruktur aufbauen zu können. Hierzu gibt es mehrere Möglichkeiten die wir im zweiten Abschnitt "Verfahren zur Bestimmung von Verbindungen" näher kennenlernen werden. Wir werden einen besonderen Fokus auf das Voronoi-Diagramm und der dazugehörigen Delaunay-Triangulation legen und diese als Beispiel näher erklären. Im Anschluss werden wir im Abschnitt "Ermittlung von Merkmalen" Methoden erläutern, wie Eigenschaften aus Graphen extrahiert werden können, um eine Diagnose zu unterstützen oder festzulegen. Diese Methoden werden wir in fünf verschiedene Kategorien unterteilen und einzeln erläutern. Schlussendlich wird im letzten Abschnitt ein kurzer Überblick der aktuellen Entwicklung von Graphenmodelle zur Analyse von Gewebestrukturen der letzten fünf Jahre gegeben.

II. IDENTIFIZIERUNG VON KNOTEN

Traditionell kennzeichnen Pathologen die Knoten auf einem Gewebebild mit einer Software per Hand. Allerdings ist dies eine zeitaufwändige Arbeit, die mit modernen Bildverarbeitungs-Algorithmen automatisiert werden kann. Es stellt eine schwierige Aufgabe für Algorithmen dar, denn die Strukturen von Gewebe sind komplex und besitzen Störfaktoren (Elemente die ähnlich aussehen wie die Zielobjekte, aber nicht gesucht werden). Zuerst muss die Frage geklärt werden, welche Objekte in einem Gewebe relevant sind. Dies ist für jedes Gewebe und dessen Ursprungsorgan unterschiedlich und variiert nach der Fragestellung, was man beschreiben will. Am verbreitetsten ist die Zuordnung von Knoten zu Zellkernen (Fig. 1.a). Zellkerne können weiterhin in andere Zellarten unterschieden werden, z.B. in Tumorzellen oder Immunzellen. Andere Zielobjekte können Drüsenstrukturen sein (Fig. 1.b), Cluster von Zellen (Fig. 1.c) oder Geweberegionen (Fig. 1.d). Die Knoten besitzen Informationen, wie der Koordinate auf

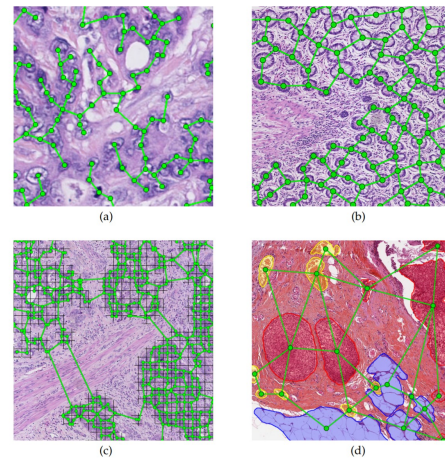


Fig. 1. Verschiedene Typen der Knotenidentifizierung. Zellkerne (a), Drüsenstrukturen (b), Cluster von Zellen (c) und Geweberegionen (d) [1].

der 2D-Ebene, und je nach Knotenrepräsentation weitere, wie Zelltyp bei Zellkernen, Zuordnung, Größe oder Form bei Drüsenstrukturen [1].

III. VERFAHREN ZUR BESTIMMUNG VON VERBINDUNGEN

Wenn die Knoten eines Gewebes definiert sind, muss eine Methode gefunden werden, mit der man eine Struktur aufbauen kann. Hierfür wurden verschiedene Graphen erforscht,

wie zum Beispiel der O'Callaghan Graph, der minimal spannenden Baum, der Delaunay Graph oder der Gabriel Graph. Je nach Anwendungsbereich sind diese unterschiedlich geeignet (siehe Fig. 2) [1]. Als vertiefendes Beispiel wollen wir uns das Voronoi-Diagramm anschauen.

Organ	Graph	Vertices	Description	Stain	Year
Methods based on Syntactic Structure Analysis					
lung	O'Callaghan	epithelial cells	classifications of carcinoma (discriminant analysis)	H&E	1985
lung	O'Callaghan	epithelial cells	classifications of carcinoma (discriminant analysis)	H&E	1987
cervix	Voronoi - MST - O'Callaghan	nuclei	characterization of normal and dysplastic tissue	H&E	1988
lung	Johnson-Mehl - Voronoi	nuclei	discriminant analysis for carcinoma grading	Feulgen	1989
lung	MST	nuclei	characterization of lung carcinomas	Feulgen	1990

Fig. 2. Ausschnitt von Gewebeanalyseverfahren und ihre Eigenschaften aus [1].

A. Voronoi-Diagramm

Voronoi-Diagramme eignen sich besonders gut um Gewebe einzuteilen, weil der Aufbau der komplementären Tessellation die Zellen klar visualisiert und es ein metrisches System bezüglich der Nachbarschaftsstruktur einführt [2]. Es entsteht dabei ein typisches wiedererkennbares Voronoi-Muster, das aus vielen verbundenen Zellen besteht (siehe Fig. 3). Das Voronoi-Diagramm ist allerdings noch kein Graph. Dazu kann eine Delaunay-Triangulation durchgeführt werden, die eine duale Struktur zum Voronoi-Diagramm ist. Dazu später mehr. Aus dem Voronoi-Diagramm können räumliche Informationen herausgelesen werden, wie zum Beispiel der Fläche oder die Orientierung der Regionen. Um ein Voronoi-Diagramm

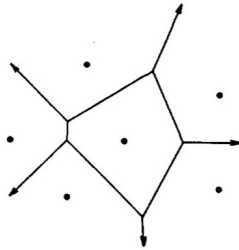


Fig. 3. Beispiel eines Voronoi-Diagramms. Die Punkte nennt man Sites und die Eckpunkte der Trennlinien Vertices [6].

aufzubauen wird eine Punktmenge $S \subset \mathbb{R}^2$ (Sites) mit mindestens drei Punkten betrachtet. Außerdem gilt $q, p \in S$. Die Halbebenen werden mit

$$D(p, q) = \{x \in \mathbb{R}^2 \mid d(p, x) < d(q, x)\} \quad (1)$$

bestimmt, wobei p und q Punkte aus der Menge S sind und die Distanzfunktion $d(p, x)$ die Euklidische Distanz

$$d(p, x) = \sqrt{(p_1 - x_1)^2 + (p_2 - x_2)^2} \quad (2)$$

ist. Die Grenze zwischen den Halbebenen, genannt Bisektor, wird mit

$$B(p, q) = \{x \in \mathbb{R}^2 \mid d(p, x) = d(q, x)\} \quad (3)$$

gebildet. Eine Voronoi-Region wird durch

$$VR(p, S) = \bigcap_{q \in S, q \neq p} D(p, q) \quad (4)$$

definiert. Hier muss die Teilmenge genommen werden, da es zu Schnittmengen kommt, die nicht Teil der Region sind. Schlussendlich lassen sich die Regionen durch

$$V(S) = \bigcup_{p, q \in S, q \neq p} \overline{VR(p, S)} \cap \overline{VR(q, S)} \quad (5)$$

zu einem Voronoi-Diagramm vereinen [3]. Man muss beachten das bei dieser Definition stets alle Punkte betrachtet werden, was hinsichtlich der Laufzeit einer algorithmischen Implementierung nicht optimal ist.

Würde man diese formale Beschreibung in einen Algorithmus umsetzen, würde dies den Pseudocode aus dem Algorithmus 1 ergeben. Dies ist eine naive Variante und hat eine

Algorithm 1 Pseudocode für die Erstellung eines Voronoi-Diagramms [4]

```

for each point in plane do
  voronoi[point]  $\leftarrow \emptyset$ ;
  for each site in list_sites do
    tempdist  $\leftarrow$  distance(point, site);
    if (voronoi[point] =  $\emptyset$  or tempdist < distance(point,
    voronoi[point])) then
      voronoi[point]  $\leftarrow$  site;
    end if
  end for
end for

```

Laufzeit von $O(p \times s)$, wobei p die Punkte und s die Sites auf der Ebene sind. Da alle Pixel oder Punkte auf der Ebene durchlaufen werden kann dies zu einer hohen Laufzeit führen, weshalb man in der Regel den Algorithmus von Fortune verwendet [4].

Optimierungsbedarf besteht darin das nicht jeder Pixel geprüft werden muss und das nicht jeder Bisektor eine Kante im finalen Voronoi-Diagramm ergibt. Der Algorithmus macht sich die Idee der Eigenschaft zu Nutze, dass ein Punkt q auf der Kante zweier Vertices liegt, wenn der Rand des größten leeren Kreises ausgehend von q , die Sites, die von dieser Kante (Bisektor) getrennt sind, berührt. Dies führt aus (3) herbei. Außerdem ist jeder leere Kreis ausgehend von einem Vertex genau dann ein Vertex, wenn der Rand mindestens drei Sites berührt [5]. Diese Eigenschaft werden wir bei der Delaunay-Triangulation wiederfinden.

B. Fortune's Sweep-Line Algorithmus

Der Sweep-Line Algorithmus von Fortune definiert eine Sweep-Line und eine Beach-Line. Die Sweep-Line scannt eine Ebene mit Sites von oben nach unten und baut hinter sich das

Voronoi-Diagramm auf, während die Sites, die vor der Sweep-Line liegen, vorerst nicht berücksichtigt werden. Die Beach-Line ist eine Menge von Parabeln, die die Sweep-Line von den Sites trennt. Die Trennung erfolgt so, dass alle Punkte q in der Parabelfläche, die von Site p_i ausgeht, näher zu p_i als zu der Sweep-Line sind (siehe Fig. 4). Die Parabeln bewegen sich mit der Sweep-Line also mit und werden durch die fortlaufende Sweep-Line breiter. Entlang der Schnittpunkte der Parabeln bilden sich die Kanten des Voronoi-Diagramms. Verschmelzen zwei Schnittpunkte, entsteht ein Vertex [5].

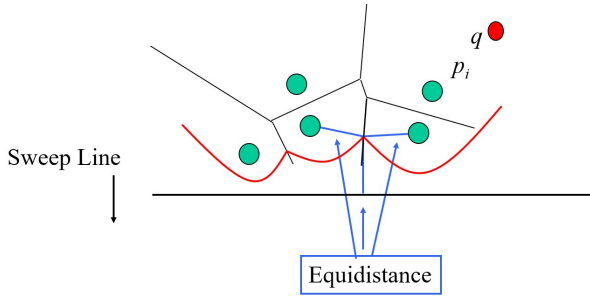


Fig. 4. Ein möglicher Zustand des Sweep-Line Algorithmus. Rot ist die Menge der Parabeln und grün sind die Vertices [5].

Möchte man diesen Algorithmus implementieren, braucht man dafür die folgend beschriebene Datenstrukturen. Für das Voronoi-Diagramm braucht man eine Doppel-Verkettete Liste mit Halbkanten (eine Halbkante ist die Aufteilung einer Kante in zwei entgegengesetzte Richtungen, notwendig für den entstehenden Binärbaum), eine Repräsentation der Vertices und die Repräsentation der Voronoi-Regionen. Es muss weiterhin der Zustand der Beach-Line gespeichert werden, d.h. die Schnittpunkte der Parabeln und die vorhandenen Parabeln selbst (als balancierten Binärbaum). Außerdem muss für die Sweep-Line eine Prioritätsliste mit Events gespeichert werden, die absteigend der y-Koordinate der Events, sortiert ist. Events können entstehen, wenn ein neuer Site von der Sweep-Line entdeckt wurde oder wenn die Sweep-Line einen leeren Kreis, der drei oder mehr Sites mit seiner Außenlinie durchläuft, am Boden dieses Kreises berührt. Die Site-Events werden beim Start des Algorithmus initialisiert, während die Events für leere Kreise beiläufig berechnet werden. Bei $O(n)$ Site-

Algorithm 2 Pseudocode für den Sweep-Line Algorithmus [6]

```

event queue  $Q \leftarrow$  all site events
binary search tree  $T \leftarrow \emptyset$ 
double linked list  $D \leftarrow \emptyset$ 
while  $Q \neq \emptyset$  do
    next event  $E \leftarrow \text{removeFirst}(Q)$ 
     $\text{handleEvent}(E, T, D)$ 
end while

```

und Kreis-Events und einem Event-Handler der mit einem balancierten Binärbaum implementiert wurde, kommen wir auf eine Laufzeit von $O(n \log n)$, denn die Operationen im Event-Handler (Bearbeitungsoperationen im Binärbaum und in der

Doppel-Verketteten Liste und das Hinzufügen neuer Events zur Eventliste) sind konstant und die Suche im Binärbaum ist $O(\log n)$ [5]. Siehe hierzu auch den Pseudocode aus Algorithmus 2. Eine visuelle Veranschaulichung und eine Implementierung mit JavaScript ist bei [7] zu finden.

C. Delaunay-Triangulation

Mit der Delaunay-Triangulation kann ein planarer Graph für ein Voronoi-Diagramm erstellt werden. Der Graph gehört zu der Gruppe der Nachbarschaftsgraphen. Das bedeutet, dass zwei Knoten nur miteinander verbunden werden, wenn eine geometrische Bedingung von zwei Knoten erfüllt ist [1]. Im Falle des Delaunay-Graphs ist die Bedingung folgend beschrieben: Drei Knoten werden miteinander verbunden, wenn diese auf dem Rand eines Kreises liegen und dieser Kreis keine anderen Knoten in seiner Fläche miteinschließt. Man nennt dies einen leeren Kreis (siehe Fig. 5). Der Ursprung dieses Kreises ist ein Vertex des Voronoi-Diagramms. Somit sind die Delaunay-Triangulation und das Voronoi-Diagramm duale Strukturen. Den Knoten des Delaunay-Graphen kann man nun Umfang oder Radius des leeren Kreises als geometrischen Wert zuordnen.

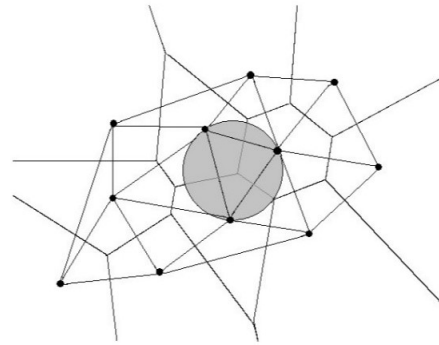


Fig. 5. Leerer Kreis bei der Delaunay-Triangulation [2].

IV. ERMITTLUNG VON MERKMALEN

In der Bildanalyse wird für die Ermittlung von Merkmalen, Werte, wie zum Beispiel Farbe, Textur oder Morphologie, verwendet. Anders ist dies bei Graphen, dort sucht man nach Merkmalen in der Graphenstruktur. Man kann diese Verfahren gliedern in syntaktische Strukturanalyse, Netzwerkanalyse, spektrale Analyse, persistente Homologie oder mathematische Morphologie (siehe Fig. 6). Wobei die ersten drei die verbreitetsten und am meist gebräuchlichsten sind. Syntaktische Strukturanalyse basiert auf geometrischen Eigenschaften des Graphen, dazu gehören Fläche, Länge und Winkel. Für die Auswertung werden in der Regel statistische Messungen gemacht, die zum Beispiel auf Standard-Abweichungen oder auf Verhältnissen basieren. Bei der Netzwerkanalyse wird der Grad, die Pfadlänge oder ein Cluster von Knoten betrachtet. Die Spektralanalyse konzentriert sich auf die Eigenwerte und Eigenvektoren von Matrizen, die sich aus den Graphen ergeben können, wie zum Beispiel die Verbundenheitsmatrix oder die

Inzidenzmatrix. Die persistente Homologie und mathematische Morphologie sind in Bezug auf die Merkmalsextraktion aus Graphen relativ unerforscht und beschäftigen sich mit der Verwendung von Betti Zahlen (Teilgebiet der Topologie) bzw. morphologischen Operationen [1].

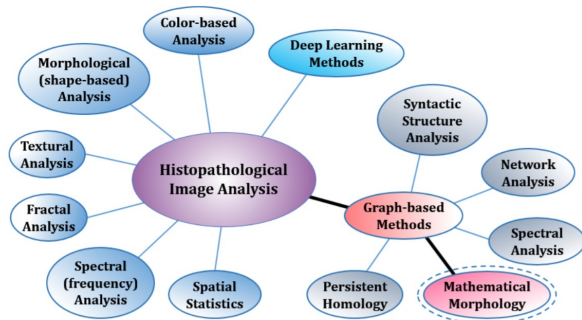


Fig. 6. Überblick auf die verschiedenen Methoden der Gewebeanalyse [1].

V. AKTUELLE ENTWICKLUNGEN

Neue Arbeiten stützen sich vermehrt auf "Graph Convolution" Netzwerke [8] [9]. In [8] aus dem Jahr 2020 richten sich die Autoren auf GCNNs (Graph Convolution Neural Network). GCNNs nehmen einen Graphen als Input und berechnen daraus einen neuen. Sie entwerfen entweder räumliche Informationen, zum Beispiel aus Nachbarschaftsbeziehungen, oder entnehmen eine Spektralanalyse mit Filtern, die aus der Signalverarbeitung von Graphen stammen. Aus dem resultierenden Graphen wird ein Merkmalsvektor generiert, der mittels maschinellen Lernens das Gewebe klassifiziert. Das vorgestellte Framework dient dazu zwei Subtypen von Lungenkrebs zu klassifizieren [8].

Eine andere Arbeit aus 2019 [10] versucht eine Pipeline zu entwickeln, die aus Gewebebilddern histopathologische Informationen von Kollagenstrukturen heraus extrahiert. Dazu wird mittels eines "Convolutional Neural Network" ein relationaler Graph aufgebaut und benutzt, um topologische und relationale Informationen der Kollagen zu gewinnen [10].

Eine weitere Arbeit aus 2019 [11] beschäftigt sich mit der Klassifizierung von Darmdrüsen mittels eines Nachbarschaftsgraphen. Mehrere Graphen von Drüsenstrukturen wurden erstellt und mit der "graph edit distance", kurz GED, mit einer Referenzmenge verglichen und anschließend, als normale Drüse oder abnormale Drüse, mit der Nächste-Nachbarn-Klassifikation klassifiziert (siehe Fig. 7) [11]. Die GED bietet ein Vergleich von Graphen indem ein Graph in einen anderen transformiert wird. Dazu werden Bearbeitungsoperationen wie Einfügen, Löschen oder Substituieren verwendet [12].

VI. FAZIT

Graphenmodelle in der digitalen Pathologie finden weitreichende Anwendungen und sind aufgrund der räumlichen Informationen eines Gewebes gut dafür geeignet. So ist es mit Graphenmodellen möglich, Krebsarten in verschiedenen Organen einstufen bzw. charakterisieren. Am verbreitetsten ist

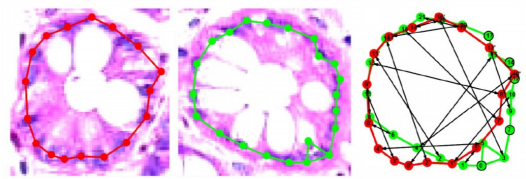


Fig. 7. Vergleich zweier Graphen mittels GED Transformationen. Die schwarzen Pfeile sind Substitutionen und die schwarzen Kreise Löschen- bzw. Einfüge-Operationen [11].

die Anwendung des Voronoi-Diagramms. Dieses bildet die Grundlage für viele andere Modelle, wie dem Delaunay-Graph oder dem Gabriel-Graph. Voronoi-Muster lassen sich vermehrt in der Natur entdecken, sowie in der Biologie des Menschen. Finden aber auch Anwendungen in der Computergrafik oder der Stadtplanung. Die Erstellung eines Voronoi-Diagramms mit dem Sweep-Line Algorithmus von Fortune ist mit $O(n \log n)$ logarithmisch, und daher für die Anwendung aus Sicht der Performance geeignet. Die unterschiedlichen Typen der Merkmalsextraktion von Graphen bieten je nach Gewebeart oder Problemstellung viel Spielraum für umfangreiche Analysen. Aktuelle Entwicklungen zeigen das Graphen auch in anderen Modellen eingesetzt werden können, indem diese einfach mit anderen Modellen verbunden oder verschachtelt werden, um dadurch ein neues komplexeres Framework zur Analyse von Gewebestrukturen zu kreieren.

REFERENCES

- [1] Ben Cheikh, Bassem. "Morphologie mathématique sur les graphes pour la caractérisation de l'organisation spatiale des structures histologiques dans les images haut-contenu: application au microenvironnement tumoral dans le cancer du sein", Diss., Paris, 2017.
- [2] Sharma, Harshita, et al. "A review of graph-based methods for image analysis in digital histopathology." Diagnostic pathology, 2015
- [3] Aurenhammer, Franz, and Rolf Klein. "Handbook of computational geometry - Voronoi diagrams.", 2000, pp. 201–290.
- [4] codingame.com, "Voronoi diagrams", <https://www.codingame.com/playgrounds/243/voronoi-diagrams/what-are-voronoi-diagrams>, 03.05.2020
- [5] Toni Sellarès, "Geometria Computacional - Voronoi diagram", http://ima.udg.es/~sellares/ComGeo/Vor2D_1.ppt, Universitat de Girona, 2010, 03.05.2020
- [6] Steven Fortune, "A Sweepline Algorithm for Voronoi Diagrams", Algorithmica, Springer-Verlag New York Inc., 1987, pp. 153–174
- [7] Raymond Hill, "Javascript implementation of Steven J. Fortune's algorithm to compute Voronoi diagrams", <http://www.raymondhill.net/voronoi/rhill-voronoi.html>, 03.05.2020
- [8] Adnan, Mohammed, Shivam Kalra, and Hamid R. Tizhoosh. "Representation Learning of Histopathology Images using Graph Neural Networks." arXiv preprint arXiv:2004.07399, 2020.
- [9] Ye, Haili, et al. "Improving Histopathological Image Segmentation and Classification using Graph Convolution Network." Proceedings of the 2019 8th International Conference on Computing and Pattern Recognition, 2019, pp. 192–198
- [10] Jung, Hyun, et al. "Integration of Deep Learning and Graph Theory for Analyzing Histopathology Whole-slide Images", IEEE Applied Imagery Pattern Recognition Workshop (AIPR), 2018
- [11] Studer, Linda, et al. "Intestinal gland classification from colorectal cancer tissue images using graph-based methods", Proceedings of the 4th Joint Annual Meeting of the Swiss and Austrian Societies of Pathology, 7-9 November 2019, Luzern, Switzerland; Der Pathologe, 2019
- [12] Riesen, Kaspar. "Structural pattern recognition with graph edit distance.", Advances in computer vision and pattern recognition, Springer International Publishing, 2015