Eine hierarchische Wasserscheidentransformation für die Spotdetektion in 2D Gelelektrophorese-Bildern

S. Wegner, A. Sahlström, K.-P. Pleißner, H. Oswald, E. Fleck

Deutsches Herzzentrum Berlin und
Virchow-Klinikum der Humboldt-Universität Berlin,
Augustenburger Platz 1, D-13353 Berlin, Germany

Zusammenfassung Eine hierarchische Wasserscheidentransformation (WST) zur Spotdetektion in 2D Gelelektrophorese-Bildern wird beschrieben. Die WST wird auf dem Gradientenbild durchgeführt und resultiert folglich in Regionen mit unterschiedlicher Krümmung. Da sich Spots durch eine konvexe Krümmung auszeichnen, können Spotregionen sofort im Ergebnis der WST lokalisiert werden. Weil jedoch Spots durch mehrere Regionen charakterisiert sein können, müssen diese Teilspotregionen noch zusammengefaßt werden.

1 EINLEITUNG

Ein wichtige Voraussetzung für molekularbiologische Untersuchungen auf Proteinebene ist das Detektieren von Proteinspots in 2D Gelelektrophorese-Bildern. Da in diesen Gelbildern oftmals mehr als 1000 Spots gefunden werden müssen, sollte für ein effizientes und reproduzierbares Ergebnis eine automatische Bildanalyse durchgeführt werden. In der Literatur wird hierfür eine markerbasierte Wasserscheidentransformation (WST) beschrieben [3]. In diesem Verfahren wird aufgrund von a priori Wissen zuerst eine grobe Segmentierung durchgeführt, um sogenannte Marker zu bestimmen, auf denen dann die Anwendung der WST beruht. Im Gegensatz dazu verwenden wir eine hierarchische WST. Diese WST wird ohne vorherige Bestimmung von Markern angewandt und die entstandenen Regionen werden aufgrund bildbasierter Kriterien analysiert. Die WST wird direkt auf dem Gradientenbild durchgeführt, so daß Regionen mit konkaver und konvexer Krümmung voneinander getrennt werden. Da sich Proteinspots durch eine konvexe Krümmung auszeichnen, können nun Spotregionen sofort im Ergebnis der WST lokalisiert werden.

Zuerst beschreiben wird das Prinzip der WST. Die dabei entstehenden Spotregionen werden durch ein Preprocessing und eine Merkmalsanalyse lokalisiert. Durch diese beiden Schritte werden alle Spotregionen gefunden. Da jedoch Spots auch durch mehrere Regionen charakterisiert sein können, müssen diese Teilspotregionen noch zusammengefaßt werden.

2 SEGMENTIERUNG MIT DER WST

Um die WST als Segmentierungsverfahren anwenden zu können, wird das Gradientenbild als Grauwertgebirge interpretiert, indem jeder Gradientenwert als entsprechende Erhebung repräsentiert wird. Dabei wird angenommen, daß Regionen im Bild durch einen geringen Kontrast und folglich durch niedrige Gradienten, die Täler, charakterisiert werden; während hohe Gradienten, die Berggrate bzw. Wasserscheiden, auf starke, kontrastreiche Konturen im Grauwertbild zurückzuführen sind.

Für die Bestimmung von Wasserscheiden im Gradientenbild wird die topographische Oberfläche mit Wasser, das aus verschiedenen Quellen kommt, geflutet [1]. Beginnend mit den kleinsten regionalen Minima, füllen sich die Täler des Gradientengebirges mit Wasser und bilden Regionen. Wasserscheidendämme werden an solchen Stellen errichtet, an denen Regionen zusammenfließen würden. Am Ende dieses Flutungsprozesses begrenzen diese Wasserscheiden bzw. Konturen die einzelnen, jeweils durch ein Minimum entstandenen Regionen (Abb. 1).

Die WST ergibt ein Mosaikbild mit gut lokalisierten, geschlossenen Grenzen, das jedoch stark übersegmentiert ist, d.h. neben den relevanten Regionengrenzen werden auch Grenzen gefunden, die hauptsächlich auf Bildrauschen zurückzuführen sind. In diesem Ergebnis müssen nun die Regionen gefunden werden, die einem Spot entsprechen.

3 LOKALISIEREN VON SPOTREGIONEN

Im Mosaikbild wird aufgrund der starken Übersegmentierung nicht jeder Spot als eine einzelne Region erkannt. In vielen Fällen setzt sich ein Spot noch aus mehreren Teilregionen zusammen. Es müssen folglich zwei unterschiedliche Arten von Regionen lokalisiert werden (Abb. 2):

1. Die Region entspricht bereits dem vollständigen Spot.
2. Die Region beschreibt nur einen Teil des Spots.

Diese Spot- bzw. Teilspotregionen werden nun durch zwei Schritte bestimmt: einem Preprocessing und einer Merkmalsanalyse bzw. Krümmungsuntersuchung.

Im ersten Preprocessing-Schritt wird die Anzahl der möglichen Spotregionen durch einfache Schwellwertkriterien reduziert, indem a priori Wissen verwendet wird, daß es sich bei Spots um dunkle Flecken auf hellem Hintergrund handelt.

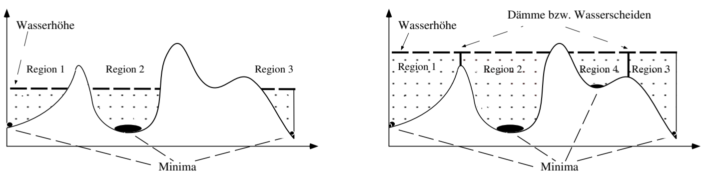

Abb. 1. Prinzip des Flutungsalgorithmus (1D).

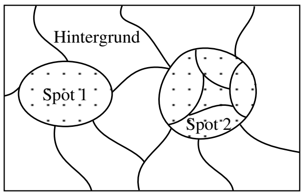

Abb. 2. Mosaikbild mit den zu lokalisierenden Spottypen. Spot 1 wird bereits durch eine Region separiert, Spot 2 besteht noch aus mehreren Regionen.

Jede Region wird im Mosaikbild zusätzlich durch ein Merkmal repräsentiert. Im folgenden wird als Merkmal der Mittelwert der Grauwerte der Region im Gelbild verwendet. Die verwendeten Schwellwerte werden in Abhängigkeit von den Grauwerten im aktuellen Gelbild bestimmt.

Das erste Kriterium basiert auf der Annahme, daß ein Spot durch signifikant höhere Grauwerte im Gelbild dargestellt wird. Jede mögliche Spot- bzw. Teilspotregion grenzt an eine Hintergrundregionen mit deutlich niedrigerem Grauwert. Folglich kann es sich nur dann um eine Spot- bzw. Teilspotregion handeln, wenn eine Nachbarregion mit wesentlich niedrigeren Grauwerten vorhanden ist. Für alle weitere Betrachtungen sei:

- N_r die Anzahl der Regionen im Mosaikbild
- r_i, i = 1, ..., N_r, die jeweilige Region
- $Neighbor_{r_i}$ die Nachbarregionen einer Region r_i
- μ_{r_i} der Mittelwert der Grauwerte einer Region r_i

Kriterium 1: Der Abstand D(r_i) einer Region r_i, i = 1, ...,N_r, zu seinen Nachbarregionen wird gegeben durch :

$$D(r_i) = inf\left\{d_{ij} = \mu_{r_i} - \mu_{r_j} : \forall j \in Neighbor_{r_i}\right\} \qquad \forall i = 1, ..., N_r \tag{1}$$

Eine mögliche Spot- bzw. Teilspotregion ist eine Region r_i, wenn

$$D(r_i) > T_1, \tag{2}$$

wobei der Schwellwert T_1 aufgrund von 20% aller Grauwerte im Gelbild bestimmt wird.

Die oben gemachte Annahme ist jedoch nicht für alle Teilspotregionen erfüllt. Hierbei handelt es sich oftmals um Teilspots, die sich in der Mitte eines Spots befinden (siehe die mittlere Region von Spot 2 in Abb. 2). Der mittlere Grauwert solcher Regionen ist jedoch immer sehr hoch. Hieraus ergibt sich das zweite Kriterium.

Kriterium 2: Eine Region r_i ist ein Kandidat für eine Spot- bzw. Teilspotregion, wenn

$$\mu_{r_i} > T_2, \tag{3}$$

wobei T_2 sich durch 80% aller auftretenden Grauwerte im Gelbild ergibt.

Obwohl durch die obigen Kriterien die Anzahl der möglichen Spotregionen erheblich reduziert wird, sind diese nicht spezifisch genug, um nur Spotkandidaten zu finden. Im zweiten Schritt wird nun als ein signifikantes Spotmerkmal die Krümmung einer Region bestimmt.

Wird ein Gelbild als topographische Oberfläche betrachtet, zeichnen sich Spots durch eine konvexe Krümmung aus. Diese Eigenschaft gilt ebenfalls für Teilspotregionen. Da die WST auf der ersten Ableitung ausgeführt wird, werden bereits Regionen mit konvexer und konkaver Krümmung separiert.

Krümmungskriterium: Die Krümmung einer Region r_i, i=1,..., N_r, wird beschrieben durch:

$$C_{r_i} = \sum_{j=1}^{N_{r_i}} f''(p_{ij}), \tag{4}$$

wobei p_{ij}, j = 1, ...,N_{r_i} die Pixel einer Region r_i sind und f'' die zweite Ableitung des Bildes ist. Eine mögliche Spotregion bzw. Teilspotregion ist eine Region r_i, wenn

$$C_{r_i} > 0. \tag{5}$$

Es verbleiben nur solche Regionen, die einem Spot oder einem Teilspot entsprechen. Nur diese Teilspotregionen müssen noch zusammengefaßt werden.

4 ZUSAMMENFASSEN VON TEILSPOTREGIONEN

Das Zusammenfassen von Teilspotregionen basiert auf zwei Eigenschaften von Spots:

1) Ein Spot sollte ungefähr eine elliptische Form haben.
2) Grenzen zwei Spots aneinander, ist die lokale Krümmung entlang ihrer Grenze konkav.

Da jede dieser beiden Eigenschaften nicht für alle Regionen hinreichend genug bestimmt werden kann, wird hier die Kombination beider Kriterien betrachtet.

Um die Ähnlichkeit einer Region zu einer Ellipse beurteilen zu können, wird eine Ellipse anhand des umschreibenden Rechtecks der Region konstruiert. Dann wird mit dem χ^2-Test beurteilt, inwieweit die Region der konstruierten Ellipse ähnelt. Eine Region sollte mit einer Nachbarregion zusammengefaßt werden, wenn die Ähnlichkeit der einzelnen Regionen zu einer Ellipse kleiner ist als die Ähnlichkeit der zusammengefaßten Region zu einer Ellipse.

Für die Bestimmung der lokalen Krümmung zwischen zwei Regionen wird entlang der Wasserscheidenlinie zwischen diesen Regionen die zweite Ableitung bestimmt. Die Richtung ergibt sich hierbei aufgrund einer Linie, die zwischen den Schwerpunkten der Regionen bestimmt wird. Teilspotregionen eines Spots zeichnen sich durch eine lokale positive Krümmung um die Wasserscheidenlinie aus. Ist die lokale Krümmung entlang der Wasserscheidenlinie hingegen konkav, handelt es sich um zwei getrennte Spots, die folglich nicht zusammengefaßt werden sollten.

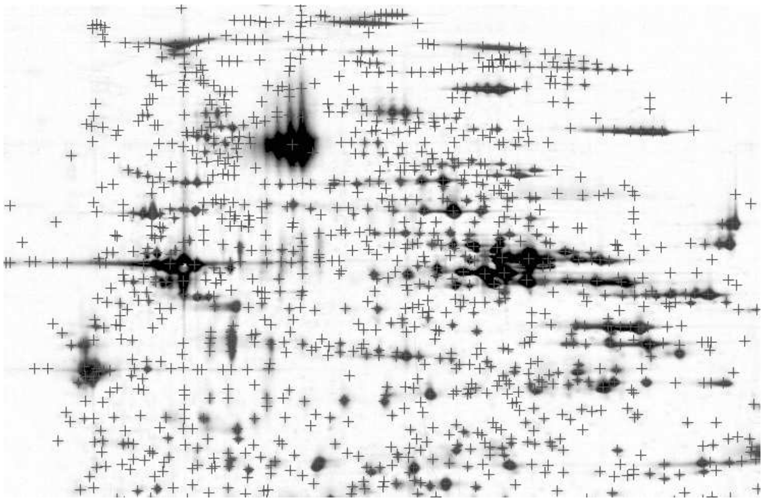

Abb. 3. Ergebnis der Spotdetektion.

5 ERGEBNISSE UND DISKUSSION

Die Spotdetektion mit der WST wurde anhand von mehreren Gelbildern für Herzproteine getestet. Nach der Anwendung der WST ist das Ergebnis zunächst sehr stark übersegmentiert, so daß sehr viele Spots durch mehrere Teilregionen beschrieben werden. Diese müssen entsprechend zusammengefaßt werden. Ein Ergebnis einer Spotdetektion zeigt Abb. 3. In den getesteten Gelbildern wird der überwiegende Teil der Spots gefunden. Zu Problemen kann es jedoch bei gesättigten Spots oder aufgrund von Bildstörungen durch Streifen kommen. Der Vergleich mit der Spotdetektion eines Gelanalysesystems (PDQUEST [2]) zeigte, daß mit dem hier vorgestellten Verfahren wesentlich mehr Spots gefunden werden konnten. Besonders deutlich waren die Unterschiede in Gelbildern mit geringem Grauwertumfang zu sehen, bei denen das PDQUEST-System nur sehr wenige Spots separieren konnte. Ein weiterer Vorteil des beschriebenen Verfahrens ist, daß auch überlappende Spot, die vielfach auftreten, gefunden und voneinander getrennt werden können.

References

1. L. Vincent, P. Soille : Watersheds in Digital Spaces: An Efficient Algorithm Based on Immersion Simulation. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.13, No.6, (1991) 583–598
2. P.J. Collins, S.H. Blose: Todays densitometer- Multi-application densitometry for biotechnology. Biotech Forum Europe, International Journal Biotechnology, Vol.9 (1992) 202-205
3. F. Meyer, S. Beucher: Morphological Segmentation. Journal of Visual Communication and Image Representation, Vol. 1, No. 1, (1990) 21-46