

## Abstract zur Dissertation

### *Die pixelbasierte Clusterung von Luftaufnahmen im Rahmen von Erosionsuntersuchungen*

Dem Fachbereich Statistik der Universität Dortmund  
vorgelegt von Matthias Zerbst aus Hagen

Dortmund 2001

Die Arbeit beschäftigt sich mit Clusterverfahren zur Bearbeitung von Luftaufnahmen aus der Erosionsforschung und -überwachung. Dabei werden die Bildpunkte (Pixel) der Aufnahmen aufgrund ihres RGB-Farbwertes geclustert. Es entstehen Bilder, die als Ausgangsbasis für thematische Karten dienen. Zusätzlich werden umfangreiche grafische Darstellungen und Tabellen erstellt, die eine Beurteilung der Erosionssituation aufgrund der Clusterung der Luftaufnahme erleichtern.

Für eine Clusterung wird eine Menge von Werten in möglichst homogene Gruppen eingeteilt. Zur Beurteilung der Homogenität wird die Intra-Class-Varianz der Clusterung verwendet. Je kleiner deren Wert, desto besser ist die Clusterung. Es wird gezeigt, dass in der Erosionsforschung zusätzliche Bedingungen wichtig sind, die von diesem Kriterium nicht berücksichtigt werden.

Zur Berechnung von Clusterungen wird der neu entwickelte Maximum-Linkage-Algorithmus verwendet. Anhand dieses Verfahrens werden einzelne Farbwerte aus der Menge aller Farbwerte des Bildes ausgewählt. Diese ausgewählten Werte dienen als Zentroide der späteren Klassen. Aufgrund der Zentroide werden die übrigen Werte klassifiziert und eine Clusterung erstellt. Die Zentroide werden dabei unter Berücksichtigung der empirischen Häufigkeitsverteilung der Farbwerte dergestalt ausgewählt, dass der Abstand zwischen den einzelnen Werten möglichst groß ist. Das Maximum-Linkage-Verfahren zeichnet sich dadurch aus, dass die entstehende Clusterung eine besonders große Trennschärfe zwischen den einzelnen Klassen aufweist. Damit wird zum einen die Interpretation erleichtert. Zum anderen wird es einfacher möglich die Klassen in der realen Umgebung zu finden, wenn z. B. zusätzlich zu dem Bildmaterial terrestrische Meßdaten erhoben werden sollen. Des weiteren werden durch das Maximum-Linkage-Verfahren auch Klassen mit seltenen Elementen identifiziert, die nur durch wenige Pixel repräsentiert sind. Solche Klassen werden von Standard-Verfahren, wie z. B. dem k-Mittelwert-Algorithmus oft nicht gefunden. Die Elemente dieser Klassen werden dabei oft in andere, größere Gruppen mit vielen Elementen klassifiziert. Aufgrund ihrer geringen Gewichtung gehen diese seltenen Werten in den großen Klassen unter und stehen für die Interpretation nicht mehr zur Verfügung. Gerade in der Erosionsforschung sind diese seltenen Elemente wichtig, weil sie oft eine große Auswirkung auf die Erosionsentwicklung haben. So

beeinflussen einige wenige Sträucher in einer ansonsten wüstenartigen Landschaft die Erosion deutlich. Werden solche Objekte nicht in eigenen, deutlich von anderen getrennten Klassen identifiziert, fehlen wesentliche Schlüsselemente für die Interpretation.

Als Vergleichsverfahren zum Maximum Linkage werden verschiedene neuronale Netze verwandt. Ein neuronales Netz besteht aus Elementen, sogenannten Neuronen, zwischen denen gerichtete Verbindungen bestehen. Die Netzstruktur wird über Gewichte an den Verbindungen ausgeprägt. Diese Gewichte werden in einer Lernphase herausgebildet, in der die einzelnen Pixel als Lerneingabe für das Netz geschaltet werden. Die Netzstruktur wird durch Modifikation (Adaption) der Gewichte an die entsprechende Lerneingabe angepaßt. Im Laufe der Lernphase bildet sich ein an die Ausgangsdaten angepaßtes Netz heraus. Die Gewichte dieses Netzes werden dann als Zentroide für eine Klassifikation der übrigen Werte verwendet.

Bei den verwandten Netztypen handelt es sich zunächst selbstorganisierende Karten. Diese Netze sind auf Kohonen zurückzuführen. Diese Netze können allerdings nicht die hohe Trennschärfe zwischen den Klassen wie der Maximum-Linkage-Algorithmus garantieren. Auch das Auffinden seltener Elemente in eigenen Klassen gelingt in vielen Fällen nicht. Zum dritten ist die Berechnungszeit einer Clusterung bei einem neuronalen Netz etwa 50 mal höher als beim Maximum Linkage. Aus diesem Grund sind diese neuronalen Netze durch Entwicklung spezieller Lernraten und Nachbarschaftsfunktionen verbessert worden. Die Berechnungsdauer wurde dadurch deutlich gesenkt. Das Problem der geringen Trennschärfe zwischen den Klassen und der Nichtidentifizierung seltener Elemente konnten dadurch jedoch nicht vollständig gelöst werden. Durch eine Vereinfachung des Netztyps kann hier jedoch eine Verbesserung erzielt werden. Eine relativ brauchbare Alternative zum Maximum-Linkage-Algorithmus ist damit entwickelt worden. Die guten Eigenschaften des Netzes für die Analyse von Luftaufnahmen aus der Erosionsforschung wird durch die Entwicklung einer von dem üblichen Prinzip abweichenden Nachbarschaftsfunktion erreicht.

Um mit verschiedenen Algorithmen erstellte Clusterungen vergleichen und beurteilen zu können, sind Qualitätsbeurteilungen für Clusterungen nötig. Dazu wird neben dem bekannten Kriterium der minimierten Intra-Class-Varianz ein zusätzliches Kriterium eingeführt. Dieses Kriterium berechnet sich über den Mittelwert der minimalen Abstände aller Zentroide zum jeweils nächsten Nachbarn. Die Kombination dieser beiden Kriterien erlaubt eine dem Anwendungsgebiet angemessene Bestimmung der optimalen Clusterung.

Zu einer Clusterung gehört nicht nur die Berechnung der Klassen sondern auch eine entsprechende Ergebnisdarstellung, die eine einfache Interpretation ermöglicht. Daher werden im Rahmen dieser Arbeit Methoden zur Ergebnisdarstellung und –aufbereitung präsentiert.

Ein wichtiger Bestandteil der Arbeit besteht in der Umsetzung der vorgestellten Algorithmen und Darstellungsmöglichkeiten in ein Rechnerprogramm. Das C++ basierte Programm Namens PicAna wird bereits von Anwendern in der Erosionsforschung eingesetzt. Zum einen verwendet das Remote Sensing Laboratory des Institut des Regions Arides, Medenine/Tunesien das Programm bei der Auswertung von Bildmaterial zur Überwachung und Erforschung von Erosionsproblemen im Süden Tunesiens. Zum anderen bestehen aufgrund dieser Arbeit gemeinsame Projekte mit dem Fachbereich für Landwirtschaft, Internationale

Agrarentwicklung und Ökologische Umweltsicherung der Universität Gesamthochschule Kassel.

Das Remote Sensing Laboratory versucht die Auswertung von Luftaufnahmen und Satellitenbildern anhand der neuen Algorithmen zu verfeinern. Die zur Erosionsbeobachtung verwendeten Luftbilder wurden bisher anhand von Beobachtungen am Boden von Hand in schematische Karten verwandelt. Mit Hilfe der neuen Verfahren werden wesentliche Teile der Bildung schematischer Karten automatisiert, präzisiert und vereinfacht. Zwei Studien in Tunesien werden zeigen, ob sich die im Rechnerprogramm PicAna umgesetzten Algorithmen gegenüber einer Bildinterpretation von Hand bewähren. Es wird untersucht, ob die Verfahren erfolgreich sind oder nur teilweise die verschiedenen Bodenarten identifizieren können. Ferner wird sich zeigen, ob sich die Überwachung der Bodendegradierung mit Hilfe der neuen Verfahren verbessern und vereinfachen läßt. Verschiedene Testgebiete, mit unterschiedlichen Arten von Bodenerosion im Süden Tunesiens, werden mit den neuen Algorithmen untersucht, auch um die Verfahren weiter zu verbessern und Probleme der Praxis bei Weiterentwicklungen zu berücksichtigen.

Der Fachbereich für Landwirtschaft und Internationale Agrarentwicklung und Ökologische Umweltsicherung der Universität Gesamthochschule Kassel verwendet die Verfahren, um die Auswertung bisher nur schwer zu interpretierenden Datenmaterials aus dem Niger/Afrika zu vereinfachen und zu optimieren. Dabei wird sich zeigen, ob sich die neuen Verfahren bewähren und sich die bisherigen Erkenntnisse bestätigen lassen. Des weiteren bieten sich Anwendungsmöglichkeiten der Bildanalyseverfahren bei einer Studie in einem Trockengebiet im Südwesten Chinas. Eine neu beginnende Studie im Oman/Arabien wird so angelegt, dass Daten speziell unter Berücksichtigung der später anzuwendenden Clusterverfahren erhoben werden.

Diese Dissertation legt den Grundstein zu einer Reihe weitere Forschungen. So werden die oben genannten Projekte zur weiteren Verbesserung der entwickelten Verfahren dienen. Des weiteren sind weitreichendere Weiterentwicklungen im Rahmen hierauf aufbauender Projekte geplant.