

Ansätze zur Übertragung von Theorien der kognitiven Wahrnehmung auf die rechnerische Interpretation von Fernerkundungsszenen

JOCHEN SCHIEWE, Vechta

Keywords: remote sensing, classification, cognitive perception, feature extraction, multi-sensor systems, scene interpretation

Zusammenfassung: Das übergeordnete Ziel dieses Beitrages besteht in der Weiterentwicklung von Methoden zur rechnerischen Interpretation von Fernerkundungsszenen. Hierbei wird der Ansatz verfolgt, einen Brückenschlag zu den Theorien der kognitiven Wahrnehmungsforschung herzustellen. Anhand ausgewählter Auswerteschritte der Interpretation (Merkmalsverarbeitung und Klassifizierung) werden Theorien der Wahrnehmungspsychologie sowie entsprechende Auswertemethoden der Szeneninterpretation betrachtet. Hierbei wird der Schwerpunkt auf die Auswertung hoch auflösender und multi-sensoraler Daten gelegt. Aus dieser Gegenüberstellung werden offene Forschungsfragen bzw. grundsätzliche Entwicklungsrichtungen für Methoden der Szeneninterpretation abgeleitet.

Summary: *Concepts for the transfer of theories of cognitive perception to automatical interpretation of remotely sensed scenes.* The general goal of this contribution is the improvement of methods for the automatical interpretation of remotely sensed scenes. We follow the idea of bridging these methods with theories of cognitive perception. Concentrating on selected interpretation steps (feature processing, classification) theories from cognitive perception and corresponding implemented automatical interpretation methods will be compared. The emphasis will be laid upon processing high resolution and multi-sensoral data. From that comparison research tasks as well as general development directions will be derived.

1 Einleitung

Digitale Fernerkundungsdaten stellen aufgrund ihrer potenziellen Aktualität, ihres geringen Generalisierungsgrades sowie ihrer großen Flächenleistung eine wichtige Quelle zum Aufbau bzw. der Fortführung von Datenbeständen Geographischer Informations-Systeme (GIS) dar. Eine wirtschaftliche und effiziente Auswertung dieser Daten bedingt einen möglichst vollständigen digitalen und automatischen Datenfluss. In diesem Zusammenhang muss allerdings festgehalten werden, dass sich zuverlässige rechnerische Methoden – insbesondere für die Interpretation räumlich hoch auflösender sowie multisensoraler digitaler Daten – noch in der Entwicklungsphase befinden,

bzw. dass das Potenzial, das Fernerkundungssensoren gegenüber der menschlichen Signal- bzw. Reizerfassung aufweisen (z.B. größere spektrale Bandbreite, spektrale Trennbarkeit, Aufzeichnung von exakten Höhenwerten), in der nachfolgenden Interpretation nicht vollständig ausgenutzt wird.

Von diesen Defiziten ausgehend soll im Folgenden die generelle Zielsetzung verfolgt werden, einen stärkeren Brückenschlag zwischen der rechnerischen Interpretation von Fernerkundungsdaten einerseits und den Theorien der kognitiven Wahrnehmungsforschung andererseits herzustellen. Dieser gedankliche Ansatz begründet sich darin, dass deutliche Analogien zwischen diesen beiden Bereichen evident sind (siehe auch Abschnitt 3), und dass aktuelle Theorien der

Wahrnehmungspsychologie in der jüngeren Vergangenheit kaum Beachtung in der Fernerkundungs-Literatur gefunden haben – zu den wenigen Ausnahmen gehören z.B. FÖRSTNER (2002) oder DONNER (2002). Zur Vertiefung werden in Abschnitt 4 einige Theorien der Wahrnehmungsforschung für die Teilbereiche der Merkmalsextraktion und -verarbeitung sowie der Klassifizierung herausgegriffen. Aus der Gegenüberstellung zu existierenden Methoden der rechnerischen Szeneninterpretation, die bereits einige Aspekte der Wahrnehmungstheorien aufgreifen, werden offene Forschungsfragen sowie künftige Entwicklungsrichtungen skizziert. Der Schwerpunkt bei den hier vorgestellten rechnerischen Methoden liegt auf eigenen Arbeiten, die eine Interpretation der Bild- und Laserscanning-Daten des Multi-Sensor-Systems TopoSys II zum Ziel haben (siehe auch Abschnitt 2).

2 Auswertebispiel

Im Folgenden soll der Transfer einiger Ansätze aus der Wahrnehmungsforschung in die rechnerische Szeneninterpretation schwerpunktmäßig an Beispielen eigener Arbeiten demonstriert werden. Der hierfür verwendete **Datensatz** besteht aus Bild- und Höhendaten, die mit dem Flugzeuggestützten Multi-Sensorsystem TopoSys II (jetzt: FALCON; TopoSys, 2003) simultan aufgezeichnet wurden und einen Ausschnitt der Stadt Ravensburg abbilden (siehe Abb. 1).

Die multispektralen Bilddaten werden mit einem elektro-optischen Zeilenscanner aufgezeichnet, der bei einer Flughöhe von 1000 m eine Bodenelementgröße von ca. 0.5 m sowie eine Streifenbreite von ca. 350 m ergibt. Bei einer radiometrischen Auflösung von 8 bit decken die vier Kanäle den sichtbaren sowie den nahen Infrarot-Bereich des elektromagnetischen Spektrums ab. Die Höhendaten werden mit einem Laserscanning-System erzeugt, das nach dem Faserbündel-Prinzip operiert. Die Bodenpixelgröße des regelmäßigen Gitters nach der Prozessierung beträgt 0.5 m, die Höhengenaugigkeit wird mit 0.15 m angegeben. Es liegen first und last pulse Lasermessungen, jedoch kein normalisiertes Digitales Oberflächen-Modell vor.

Das Ziel der **Szeneninterpretation** ist die Gliederung in die Objektklassen „Gebäude“, „Straßen“ (inkl. gepflasterter Wege und Flächen), „Grünland“ sowie „Büsche/ Bäume“. Im Folgenden wird der generelle Ablauf der Auswertung im Überblick skizziert (siehe auch Abb. 2), um eine Einordnung der beispielhaften Umsetzungen – die teilweise mit vorhandener, teilweise mit eigener Software erfolgt sind – in die punktuellen, theoretischen Betrachtungen (Abschnitt 4) zu ermöglichen.

Diese Klassen, deren Merkmale und Relationen sowie die zugehörigen Auswertemethoden werden in einem *Konzeptmodell* in Form eines semantischen Netzes beschrieben. Aufgrund des hohen Detaillier-



Abb. 1: Ausschnitt des multi-sensoralen Datensatzes: Multispektrale Bilddaten (links; Abdeckung ca. 400 m x 300 m) und Perspektivansicht nach Draping des Bildes auf das Digitale Oberflächen-Modell (rechts).

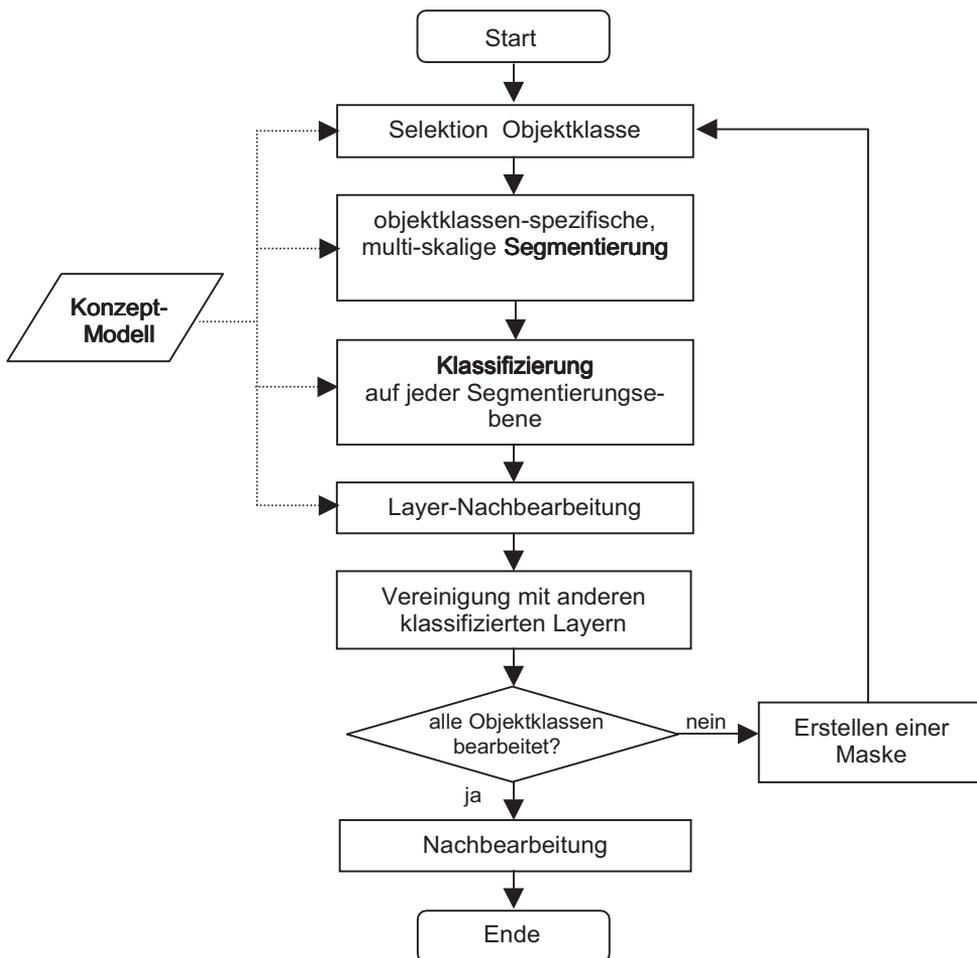


Abb. 2: Ablaufdiagramm der vorgestellten hybriden und multi-skaligen Szeneninterpretation.

rungsgrades der Eingangsdaten ist eine *Segmentierung* notwendig, die mit Hilfe des Programmpaketes eCognition (Definiens-Imaging, 2003) mit individuellen Parametern für jede Objektklasse (Gebäude: Höhen-Krümmung, Rest: Normalisierter Differenzen-Vegetations Index, NDVI) sowie in mehreren Generalisierungsgraden erfolgt, sodass eine durch eine hierarchische Struktur miteinander verbundene Segmentpyramide aufgebaut werden kann. Die *Klassifizierung* geschieht in jeder einzelnen Ebene dieser Pyramide durch einen Vergleich der tatsächlichen mit den im Konzeptmodell definierten

Segmentmerkmalen für die jeweils betrachtete Objektklasse. Eine Nachbearbeitung (z.B. die Eliminierung von Inselbereichen) schließt die Bearbeitung jeder Ebene ab. Nach Zusammenführung aller klassifizierten Layer (siehe auch Ergebnis in Abb. 3) muss für die weitere Verwendbarkeit noch eine zusätzliche, teilweise interaktive Nachbearbeitung (z.B. zur Linienglättung) stattfinden. Eine Bewertung der erzielten Ergebnisse ist nicht Gegenstand dieses Beitrages (siehe hierzu SCHIEWE 2003), da der Schwerpunkt hier auf die Auswertestrategie an sich gelegt werden soll.

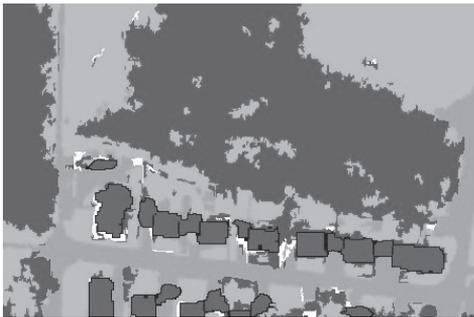


Abb. 3: Ergebnis der Szeneninterpretation (ohne interaktive Nachbearbeitungsschritte; vgl. Abb. 1).

3 Wahrnehmungsforschung und Szeneninterpretation

Ziel dieses Abschnittes ist es, ausgehend von der Definition der kognitiven Wahrnehmung die Analogie zwischen dieser sowie der rechnerischen Interpretation von Fernerkundungsszenen darzulegen und somit den in diesem Beitrag verfolgten Brückenschlag zu motivieren.

Die Wahrnehmungsforschung hat bisher noch keine einheitliche und vollständige Theorie entwickeln können; auf Grundlage der unterschiedlichen Ansätze gibt es daher auch eine Reihe von Definitionen zur Wahrnehmung. Nach GUSKI (1989) und ZIMBARDO (1988) wird unter *Wahrnehmung* nicht nur die Aufnahme, sondern auch die stufenweise Verarbeitung von Informationen über Objekte und Ereignisse der Umwelt verstanden. Die Aufnahme (Wahrnehmung im engeren Sinn) wird durch eine Reizung der Sinnesorgane durch physikalische Objekte und Ereignisse hervorgerufen. Im Kontext der visuellen Wahrnehmung werden diese Reize zu zweidimensionalen und geordneten Reiz-Repräsentationen (Perzepten) organisiert und münden über mehrere Zwischenstufen in die Erkennung des Reizgegenstandes (Wahrnehmung im weiteren Sinn).

Beim Ansatz *der kognitiven Wahrnehmung* wird zusätzlich davon ausgegangen, dass im Gedächtnis gespeicherte und abgerufene Informationen zum Erkennen des Reizgegenstandes genutzt werden. Es erfolgt ein Wech-

selspiel von Reiz- und Wissensdaten, bei dem je nach Verarbeitungsrichtung von *bottom-up-* bzw. *top-down-Prozessen* gesprochen wird. Im Gegensatz zu dieser hybriden Betrachtungsweise formuliert der in der Fernerkundungs-Literatur oft angeführte, algorithmische Ansatz von MARR (1982) die Objektwahrnehmung lediglich als bottom-up-Prozess. Ferner haben die aktuellen Erkenntnisse aus der Neuroanatomie und -physiologie hohe Anteile von *Rückkopplungsmechanismen* zwischen den einzelnen Verarbeitungsstufen aufgedeckt (GOLDSTEIN 2002), sodass die streng sequenzielle und algorithmische Herangehensweise von MARR auch hier nicht mehr als adäquat angesehen werden kann.

Schon aus der Definition der kognitiven Wahrnehmung wird die Analogie zum Auswertebau der Szeneninterpretation in der Fernerkundung bzw. allgemeiner im Bereich der *Computer Vision (Machine Vision)* evident. Auch hier besteht die Aufgabe in der Rekonstruktion und Interpretation 3-dimensionaler Szenen aus 2-dimensionalen Abbildungen von Signalen (entsprechen den Reizen), wobei Techniken der Bildverarbeitung, Mustererkennung und Künstlichen Intelligenz kombiniert werden (HARALICK & SHAPIRO 1992) und hierfür datengetriebene (bottom-up) sowie modellgetriebene (top-down) Strategien bekannt sind. Zur Veranschaulichung dieser Analogien stellt Abb. 4 den generellen Auswerteprozess der kognitiven Wahrnehmung dem der Interpretation von Fernerkundungsdaten gegenüber.

4 Transfer von Ansätzen

Ziel dieses Abschnittes soll es sein, ausgewählte Ansätze der Wahrnehmungsforschung vorzustellen, hiervon einzelne Transfers in den Bereich der Szeneninterpretation zu präsentieren und aus erkannten Defiziten offene Forschungsfragen bzw. potenzielle Entwicklungstendenzen abzuleiten. Während in diesem Kontext in der Vergangenheit ein Schwerpunkt auf dem Aspekt der *Wissensrepräsentation* durch *neuronale Netze*, die die physiologischen Funktionen des menschlichen Gehirns kopieren sollen, ge-

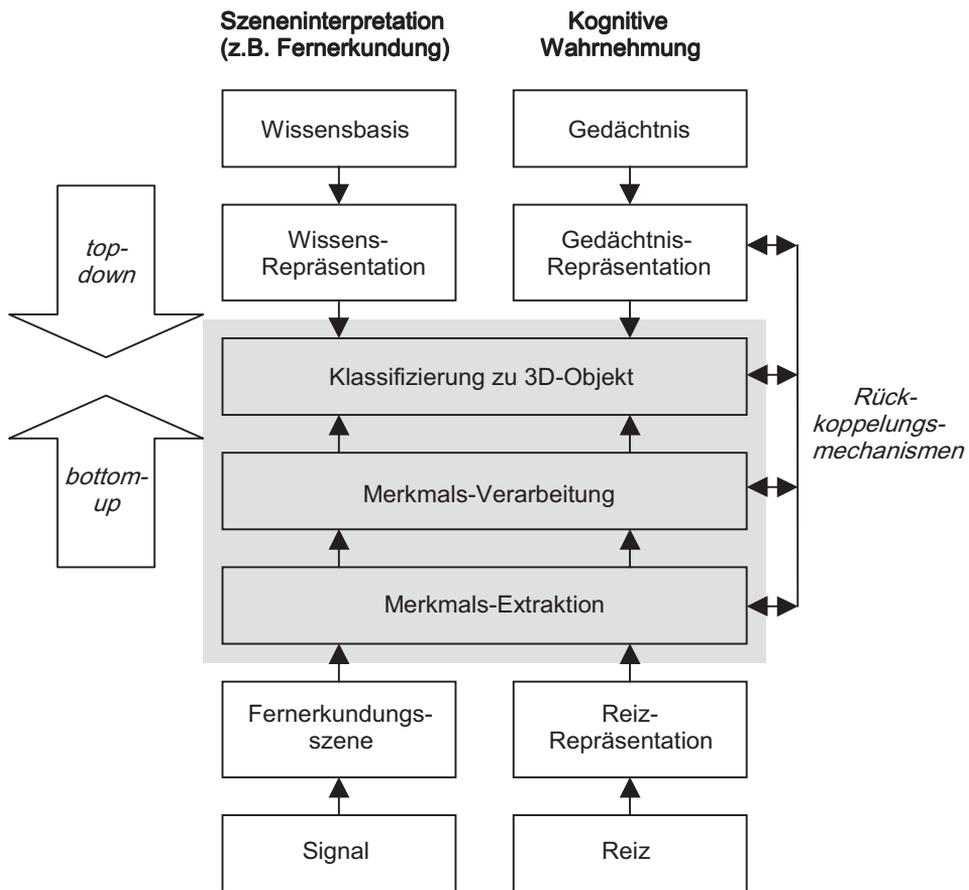


Abb. 4: Analogie der generellen Abläufe zur Interpretation von Fernerkundungsdaten sowie der kognitiven Wahrnehmung.

legt worden sind, sollen im Folgenden die in Abb. 4 grau unterlegten Prozesse der Extraktion und Verarbeitung von Merkmalen (Abschnitt 4.1) sowie der Klassifizierung (Abschnitt 4.2) im Mittelpunkt stehen.

4.1 Extraktion und Verarbeitung von Merkmalen

Eine Szeneninterpretation erfolgt i.d.R. nicht anhand der aufgezeichneten Reize oder Signalwerte selbst, sondern über hieraus abgeleitete *Merkmale (features)*. Die Extraktion und Weiterverarbeitung dieser Attribute stehen im Mittelpunkt dieses Teilabschnittes.

4.1.1 Ansätze in der Wahrnehmungsforschung

Setzt man einen stufenweisen Ablauf der menschlichen Objektwahrnehmung voraus (vgl. z.B. GOLDSTEIN 2002), so werden nach MARR oder TREISMAN in einem ersten Schritt – der *präattentiven Verarbeitung* – aus den erhaltenen Reizen in sehr kurzer Zeit *Elementarmerkmale* abgeleitet, die eine Trennung in Texturbereiche ermöglichen können. Zu diesen Basismerkmalen, die bei JULESZ auch als *Textone* bezeichnet werden, gehören

- Orientierungen (z.B. Linieneigungen),
- Linienbögen und -endpunkte,

- kleine Punkte,
- geschlossene Flächen unterschiedlicher Größe und Form (Kleckse, Kreise, etc.),
- spektrale Eigenschaften (Farbe, Helligkeit, Kontrast) oder
- Bewegungen.

Die Regeln bzw. Heuristiken, nach denen insbesondere die räumliche Zusammenfassung zu (Textur-)Bereichen erfolgt, werden z.B. in der *Gestalttheorie* nach WERTHEIMER u.a. beschrieben. Diese Theorie konstatiert zum Einen, dass sich die genannten elementaren Merkmale gegenseitig beeinflussen (Grundregel: „Das Ganze ist mehr als die Summe seiner Teile“). Zum Anderen benennt sie die Faktoren, die zu den Bereichs-Gruppierungen führen. Hierzu gehören insbesondere der *Faktor der Ähnlichkeit* (z.B. von Farbton, Orientierung oder Größe), aber auch der der Nähe (bzw. Nachbarschaft), der Einfachheit (Erzeugung möglichst einfacher Strukturen), der Linienfortsetzung (z.B. aus Punkten gedachte, sanft geschwungene Linien), des gemeinsamen Schicksals (z.B. der gleichen Bewegungsrichtung) oder der Vertrautheit.

Alternativ zu den Textonen sieht die Theorie von BIEDERMAN (1987) elementare Teilkörper als Grundlage des Wahrnehmungsprozesses an, die auch als geometrische Ionen (*Geone*) bezeichnet werden und als Komponenten dreidimensionaler Objekte aufzufassen sind.

Die in diesem ersten Wahrnehmungsschritt erhaltenen Basismerkmale, die noch unabhängig und nebeneinander in unterschiedlichen neuronalen Kanälen bestehen, müssen nun in einer zweiten – weniger gut erforschten – Stufe (der so genannten *gerichtete Aufmerksamkeit*) wieder miteinander verknüpft werden. Nach TREISMANS Merkmalsintegrationstheorie besteht eine zentrale Aufgabe hierbei in der Verbindung zwischen den attributiven und den Orts-Informationen, wobei verstärkt vorhandenes Wissen eingesetzt wird (top-down-Prozess).

Unabhängig von der Art der Merkmale muss einschränkend festgehalten werden, dass von diesen aufgrund der begrenzten Verarbeitungskapazität und -zeit der

menschlichen Wahrnehmung tatsächlich nur ein kleiner Teil weiterverarbeitet wird. So stellt z.B. ALLEN (1982) fest, dass i.d.R. nur zwei Informationsquellen in den Prozess der gerichteten Aufmerksamkeit einfließen (z.B. die nahezu rechtwinklige Fläche sowie der Helligkeitsunterschied zur Umgebung bei der Detektion von Häusern aus Luftbildern). Auch diese *Selektion* basiert auf heuristischen Entscheidungen sowie vorhandenen Gedächtnisinformationen.

Ein weiteres Phänomen im Rahmen der Merkmals-Verarbeitung stellt die menschliche Fähigkeit zur *Abstraktion* dar, d.h., dass auch ohne die Vollständigkeit und Korrektheit aller Merkmale ein Objekt richtig erkannt werden kann. Dies wird z.B. im Fall der dreidimensionalen Objekt-Rekonstruktion auf Basis von Geonen deutlich: So kann z.B. ein Gebäude in einem Luftbild erkannt werden, obwohl aufgrund von Verdeckungen, Abschattungen oder Aufnahme Fehlern nicht jedes Gebäudebestandteil identifizierbar ist.

4.1.2 Rechnerische Methoden

Im Kontext der rechnerischen Szeneninterpretation erfolgt die Extraktion und Verarbeitung von Merkmalen durch die Anwendung spezieller Funktionen auf die originären Daten bzw. auf bereits bestimmte Merkmale. Im Folgenden sollen sowohl die generellen Strategien zur Merkmalsverarbeitung, als auch Art und Umfang der extrahierten Merkmale betrachtet werden.

Hinsichtlich der generellen **Strategien** zur Merkmalsverarbeitung können punkt- bzw. pixel-, linien- oder Regionen-basierte Methoden unterschieden werden. So sind *linienhafte Auswertungen* in der Lage, einen Teil der beschriebenen *Textone* zu extrahieren: Insbesondere Kontrastunterschiede in den Bilddaten werden zur Definition von Kanten genutzt, die weiterhin – nach dem Gestaltfaktor der Linienfortsetzung – auf Basis von Konturverfolgungs-Algorithmen zu Linien zusammengesetzt werden können. Insbesondere durch die Verbesserung der räumlichen Auflösungen der aktuellen Fernerkundungssensoren steigt die Bedeutung *Re-*

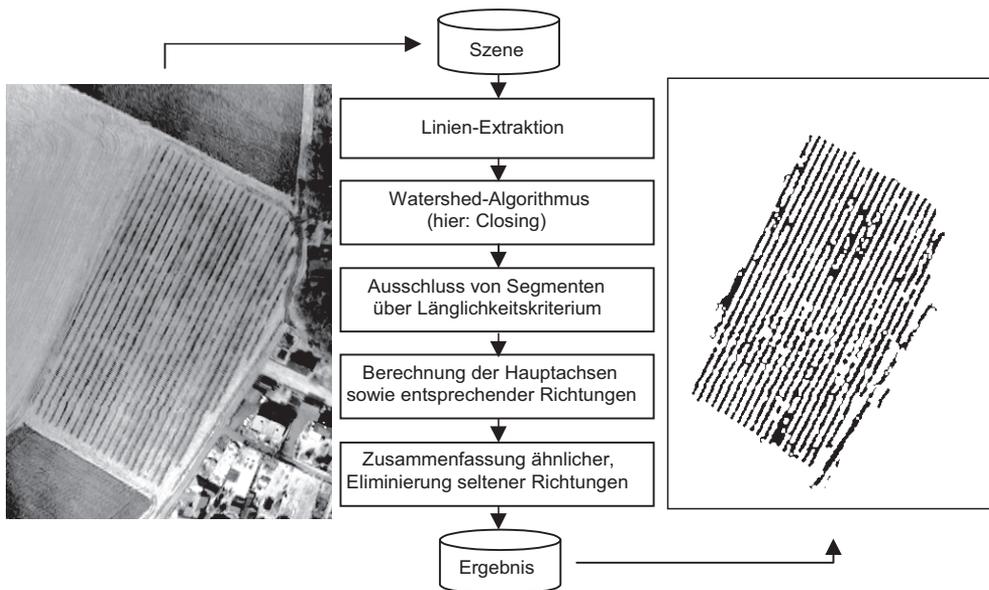


Abb. 5: Einfacher Algorithmus (Mitte) zur Extraktion von flächenhaften Mustern (rechts) aus gegebenen Bilddaten (naher Infrarot-Kanal, Ausschnitt ca. 160 m × 420 m; links).

gionen-basierter Auswertungen signifikant an, da nahezu alle Objekte als flächenhafte Elemente modelliert und Signal- oder Merkmals-Heterogenitäten innerhalb einzelner Entitäten beachtet werden müssen. Bei diesen *Segmentierungen* werden i.d.R. die Gestaltfaktoren der Ähnlichkeit und Nähe (bzw. Nachbarschaft) für spektrale Ausprägungen sowie Formeigenschaften (z.B. Kompaktheit oder Länglichkeit) herangezogen, um homogene Regionen zu bilden.

Auch die in Abschnitt 2 skizzierte Auswertung ist ein Beispiel für eine Regionenbasierte Strategie. Hier ist es durch die Variation der Skalierungsparameter bzw. die Bildung von Segmentpyramiden auch möglich, Merkmale in verschiedenen Generalisierungsstufen abzuleiten. Zur Reduktion des Verarbeitungsaufwandes findet ferner auch eine objektklassen-spezifische *Selektion* der Merkmale statt, indem im Konzeptmodell nach sinn-logischen Gesichtspunkten definiert wird, welche Merkmale signifikant für die Detektion jeder einzelnen Objektklasse sind.

Hinsichtlich der **Art und des Umfangs der zu verarbeitenden Merkmale** stellt sich bei

unserer Beispiel-Auswertung heraus, dass aufgrund der Variabilität der vorhandenen Objekte die üblichen Formeigenschaften des Grundrisses (Kompaktheit, Länglichkeit) einen geringen Aussagewert besitzen. Neben den spektralen Eigenschaften (hier insbesondere dem NDVI) werden daher verstärkt aus den Laserscanning-Daten abgeleitete Höhen-Informationen in den Auswerteprozess einbezogen: So weisen sehr hohe Gradienten bzw. Krümmungen bei der gegebenen räumlichen Auflösung sehr sicher auf Gebäudewände hin. Dies ist ein erstes Beispiel für die Erweiterung des Merkmalsraumes, die durch die simultane, multi-sensorale Datenerfassung ermöglicht wird.

Ein weiteres Beispiel betrifft die Berücksichtigung der *Orientierung* von schmalen Elementen bzw. von linearen Mustern. So wird aus Abb. 5 (links) deutlich, dass für die vorhandene Reihenfeldaufbaufläche das konventionelle Erkennungsmerkmal des mittleren NDVI aufgrund der internen, schmalen vegetationsfreien Flächen ungeeignet ist. Für eine Detektion und Beschreibung dieser Regionen reichen eine Texturfilterung oder eine Kantenextraktion allein

nicht aus, weil so noch keine Gruppierung der zwar schmalen, aufgrund der hohen geometrischen Auflösung aber flächenhaften Elemente erfolgt. Auch hier ist also eine Segmentierung notwendig, die im vorgestellten, sehr einfachen Algorithmus (vgl. Abb. 5, Mitte) aber nicht auf ein region growing-Verfahren aufbaut, da ein solches aufgrund der Schmalheit homogene Flächen nur schwer identifizieren kann. Stattdessen wird eine Methode verwendet, die nach der Detektion von Linien (statt diffuser Kanten) einen einfachen Watershed-Algorithmus durchführt und über die Parameter Länglichkeit und Hauptrichtung die parallelen Muster des Reihenfeldanbaus extrahiert (Abb. 5, rechts).

In den Kontext der BIEDERMAN'schen *Geonen*-Theorie als Grundlage einer alternativen Merkmalsart sind z.B. die Arbeiten zur Gebäude-Extraktion aus Luftbildern im Rahmen des DFG-Bündelprojektes „Semantische Modellierung“ einzuordnen (siehe zur Übersicht FISCHER et al. 1998). KOLBE (2000) hat hierbei nicht nur eine bestmögliche Zuordnung der extrahierten Bildmerkmale (d.h. der Ecken und Kanten, die zu Gebäude-Komponenten verbunden werden) zu einer Reihe von Gebäudehypothesen (bzw. unterschiedlichen Gebäudetypen) durchgeführt, sondern auch explizit die Unbeobachtbarkeit und Unschärfen der Merkmale und Relationen berücksichtigt und somit die Fähigkeit zur *Abstraktion* abgebildet.

4.1.3 Forschungsfragen

Bezüglich der **Strategie** zur Merkmalsextraktion und -verarbeitung ist festzustellen, dass konventionelle Auswerteverfahren der Szeneninterpretation entweder punkt-, linien- oder Regionen-basiert vorgehen. Eine *hybride Strategie* im Hinblick auf die Merkmalsträger, so wie sie der menschliche Auswerter in Abhängigkeit von der betrachteten Objektklasse und dem aktuellen Generalisierungsgrad durchführt, findet dagegen nur sehr selten statt – ein Beispiel zur simultanen Extraktion polymorpher Merkmale gibt FUCHS (1998). Ferner wurden auch

Kombinationen von Segmentierungsverfahren basierend auf Kantendetektionen sowie Regionenwachstum entwickelt (siehe z.B. PAVLIDIS & LIOW 1990, GAMBA et al. 1997, BHALERO & WILSON 2001), allerdings wurde deren Eignung kaum, i.d.R. aber gar nicht mit Fernerkundungsdaten getestet, die bekanntermaßen einen komplexen Anwendungsfall darstellen (siehe hierzu auch FUCHS 1998).

Aufgrund der Forderung nach Anwendung hybrider Verfahren sowie des erweiterten Merkmalsumfangs (siehe unten) erhöht sich die Komplexität der Auswertung. In diesem Zusammenhang ist daher auch eine stärkere Auseinandersetzung mit *heuristischen* statt ausschließlich mit algorithmischen Auswerteanätzen anzuregen (siehe auch Abschnitt 4.2.3). Im Kontext der Merkmalsextraktion bedeutet dies, dass Wissensrepräsentationen wie *heuristische Entscheidungsbäume*, *heuristische Entscheidungstabellen* oder *Diagnose-Scores* verwendet werden können, um die Komplexität des sogenannten Problemraumes (bestehend aus Merkmalen, Objektklassen und ihren Relationen) zu reduzieren. Die genannten Darstellungsformen können eine Auswahl sowie eine Reihen- oder Rangfolge der Merkmale im Hinblick auf die abzuleitenden Objektklassen definieren und besitzen die Eigenschaft, dass Merkmale auch unsicher oder sogar unbekannt sein dürfen.

Hinsichtlich der **Art und des Umfangs der Merkmale** besteht die statistisch begründete Annahme, dass mit der Anzahl der zur Verfügung stehenden Informationen auch die Wahrscheinlichkeit eines richtigen Klassifizierungsergebnisses ansteigt. Gerade hier ist ein großes Entwicklungspotenzial erkennbar, da konventionelle Auswerteverfahren i.d.R. nur einen kleinen Ausschnitt der Texturen (i.d.R. Farbe und Textur) aus dem tatsächlich vorhandenen, mit Multi-Sensor-Systemen sogar noch erweiterten Merkmalsumfang verwenden. Unter Berücksichtigung der Kosten einer erhöhten Komplexität sollte daher der Formalisierung weiterer Merkmale sowie der Untersuchung ihres Mehrwertes in Zukunft eine bedeutende Rolle zukommen.

Ein Beispiel hierfür wurde in Abschnitt 4.1.2 mit der Beschreibung der *Orientierung* von schmalen Elementen bzw. von linearen Mustern gegeben. Das hier verwendete, sehr einfache Verfahren benötigt sicherlich noch einige Verfeinerungen. Zu den Merkmalen, die ebenfalls eine intensivere Betrachtung verdienen, gehören auch die *Schattenbereiche*, die durch die Integration der vorhandenen, hoch genauen Höhendaten auch zuverlässig als solche prädiert (siehe z.B. HARTL & CHENG 1995; für SPOT-Daten) und somit als nicht eigenständige Regionen interpretiert werden können. Ferner sind verstärkt *semantische Nachbarschaften* zu formalisieren, aus deren Definition einige Klassenzuweisungen angenommen oder verworfen werden können. Einen Versuch in diese Richtung unternahm ZHAN et al. (2002), die aus extrahierten Segmenten strukturelle Beziehungen ableiten und mit gegebenem Vorwissen städtische Grob-Landnutzungsklassen definieren.

4.2 Klassifizierung

Die *Klassifizierung* im Kontext der Szeneninterpretation verfolgt die Aufgabe, den Daten oder Merkmalsträgern (d.h. Punkten, Linien oder Regionen) die Instanz aus einer vordefinierten Menge drei-dimensionaler Objekte zuzuweisen. Ausgangspunkt dieses Prozesses ist das Aufstellen einer Hypothese zur Zugehörigkeit des Merkmalsträgers zu einer bestimmten Objektklasse, wobei entweder zwischen zwei Alternativen (z.B.: Objekt ist ein Haus oder nicht) oder mehreren Alternativen (z.B.: Objekt ist entweder Haus, See, Wald oder Straße) entschieden werden muss. Die Entscheidungsstrategien zur Verifizierung oder Falsifizierung dieser Hypothesen stehen im Mittelpunkt dieses Teilabschnittes.

4.2.1 Ansätze in der Wahrnehmungsforschung

Der Prozess der Entscheidungsfindung im Kontext einer Hypothesen-Bewertung lässt sich in mehrere **Phasen** unterteilen: Nach der Erfassung von Ausprägungen diverser

Merkmale (siehe Abschnitt 4.1.1) erfolgt deren Bewertung und Gewichtung, um schließlich zu einer Berechnung eines gesamten Zugehörigkeitsmaßes zu gelangen. In Abhängigkeit von den gegebenen Voraussetzungen sowie den angewandten Methoden werden in der Psychologie zwei Richtungen zur Untersuchung dieses Prozesses unterschieden – die *präskriptive* sowie die *deskriptive Entscheidungstheorie*.

Die **präskriptive (oder normative) Entscheidungstheorie** setzt eine *rationale* Entscheidungsfindung voraus. Dies bedeutet, dass alle notwendigen Informationen, ein eindeutig formuliertes (Objekt-)Modell sowie zuverlässige Lösungsalgorithmen vorhanden sein müssen. Dabei wird durchaus berücksichtigt, dass es Faktoren gibt, die sich einer vollständigen Beherrschung durch die Entscheidungsträger entziehen (sogenannte *Umwelteinflüsse*), sodass es je nach Kenntnisstand Entscheidungen „unter Sicherheit“ oder „unter Risiko“ gibt. Für die Bearbeitung der Aufgaben in den oben beschriebenen Phasen existiert eine Reihe von Vorgehensweisen (siehe z.B. POSCHMANN 1999). Hierzu gehören z.B. das Scoring-Verfahren (auch: Nutzwertanalyse), die multi-attributive Werttheorie (MAVT) für Entscheidungen „unter Sicherheit“ sowie die multi-attributive Nutzentheorie (MAUT) für Entscheidungen „unter Risiko“. Die entsprechenden axiomatischen Vorgehensweisen des präskriptiven Ansatzes sind offensichtlich gut durch algorithmische Ansätze der Informationsverarbeitung abzubilden.

Die **deskriptive (oder empirische) Entscheidungstheorie** geht dagegen davon aus, dass aufgrund der begrenzten kognitiven und zeitlichen Kapazitäten der menschlichen Wahrnehmung vollständig rationale Entscheidungen nicht möglich sind und daher das *tatsächliche* Verhalten analysiert werden sollte (daher auch: Ansatz des „natürlichen Treffens von Entscheidungen“). In einigen Disziplinen findet bereits eine intensive Auseinandersetzung mit diesem Ansatz statt – so z.B. in den Wirtschaftswissenschaften, wo die Nobelpreisträger im Jahr 2002, KAHNEMANN und SMITH, das irrationale bzw.

kalkulierte Interesse einzelner Individuen im Gegensatz zu den normativen Wirtschaftstheorien untersuchen.

Auch im Kontext der Szeneninterpretation werden die rationalen Defizite, die die deskriptive Theorie begründen, in allen genannten Phasen der Entscheidungsfindung deutlich. So wurde schon in Abschnitt 4.1.1 darauf hingewiesen, dass der Mensch tatsächlich nur eine Teilmenge der vorhandenen Merkmalsausprägungen erfasst und weiter verarbeitet. Ferner weisen z.B. WIKKENS & HOLLANDS (2000) darauf hin, dass *Negativ-Merkmale* (z.B.: „dieses Objekt ist kein Haus“) bei der menschlichen Entscheidungsfindung kaum eine Rolle spielen, und dass menschliche Bewertungen oft nicht mit objektiven Wahrscheinlichkeiten übereinstimmen (z.B. die Wahrscheinlichkeit seltener Ereignisse überschätzt wird). Auch der Aufwand für eine Gewichtung nach Kriterien sowie eine nachvollziehbare Bestimmung von gesamten Zugehörigkeitsmaßen ist selten feststellbar. Stattdessen genügen oft hinreichend gute Ausprägungen zur Entscheidungsfindung (*satisfying principle*), oder es werden Wahlalternativen aufgrund nicht zufriedenstellender Ausprägungen in einem Eliminationsverfahren frühzeitig ausgeschlossen (*Eliminationsschlüssel*).

Dass trotz der begrenzten kognitiven und zeitlichen Kapazitäten der menschlichen Wahrnehmung i.d.R. dennoch zufriedenstellende Klassifizierungsergebnisse erzielt werden, lässt sich nicht zuletzt darauf zurückführen, dass der Mensch ein offenes Informationssystem praktiziert, in dem Entscheidungen nicht in einer sequenziellen Abfolge sondern in Schleifen getroffen werden, d.h., dass diverse **Rückkoppelungsmechanismen** mit den Gedächtnisrepräsentationen vollzogen und Entscheidungen somit optimiert werden können (siehe auch Abschnitt 3 bzw. Abb. 4).

4.2.2 Rechnerische Methoden

Von den in Abschnitt 4.2.1 beschriebenen **Phasen** der Entscheidungsfindung wurde die *Erfassung* und *Verarbeitung* von Merkmalen im Rahmen einer rechnerischen Szenenin-

terpretation bereits in Abschnitt 4.1.2 diskutiert. Für die *Bewertung* der Merkmale existiert eine Reihe von wohl definierten Methoden, die z.B. auf der Wahrscheinlichkeits- oder der Fuzzy Logic-Theorie beruhen. Hierauf aufbauend sind auch *Gewichtungen* der einzelnen Merkmale (z.B. über bedingte Wahrscheinlichkeiten mit dem Bayes'schen Klassifikator oder über Schwellwerte) sowie die Zusammenfassung zu einem gesamten *Zugehörigkeitsmaß* möglich. Diesen Methoden liegt grundsätzlich die Idee des vollständig rationalen, d.h. **präskriptiven Ansatzes** zugrunde. Entscheidungen „unter Risiko“, wobei Umwelteinflüsse hier Limitierungen oder Aufzeichnungsfehler der Sensoren darstellen, können prinzipiell auch mit (bedingten) Wahrscheinlichkeiten oder Regeln der Fuzzy Logic modelliert werden.

Aber auch im Zusammenhang mit der rechnerischen Szeneninterpretation ist eine Auseinandersetzung mit dem **deskriptiven Ansatz** der Entscheidungstheorie angebracht, da analog zur menschlichen Wahrnehmung ebenfalls häufig Bedingungen der Rationalität (siehe Abschnitt 4.2.1) verletzt werden. Neben der Tatsache, dass nicht alle zur Entscheidungsfindung notwendigen Merkmalsausprägungen tatsächlich vorhanden sind, sind auch die Voraussetzungen zur Anwendung der Lösungsalgorithmen für die folgenden Phasen nicht immer eindeutig gegeben. So ist z.B. die Bewertung auf Basis von Wahrscheinlichkeiten oft schwierig bzw. streng genommen gar nicht erlaubt, weil z.B. keine normalverteilten Zufallsgrößen vorliegen. Auch die Bestimmung von Schwellwerten erfolgt i.d.R. subjektiv und ist kaum auf andere Anwendungsfälle übertragbar.

Ein pragmatischer Ansatz zur Reduktion des Suchraumes sowie zur Vermeidung von interaktiv festzulegenden Schwellwerten im Sinne eines deskriptiven bzw. empirischen Vorgehens ist die Nutzung des Vorwissens aus Datenbeständen vorhandener Geographischer Informations-Systeme (GIS). EHLERS et al. (1989) bzw. EHLERS (2000) stellen ein entsprechendes regelbasiertes Konzept vor, das einerseits das Vorwissen der GIS-

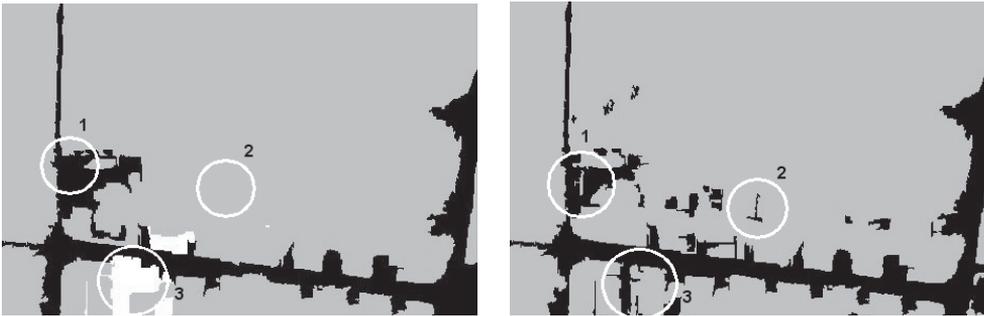


Abb. 6: Mehrwert einer Straßen-Klassifizierung durch Verwendung *aller* Segment-Ebenen (rechts) im Gegensatz zu *einer* Ebene (links; Straßen in schwarz; basierend auf den Daten aus Abb. 1): Bessere Identifizierung von Trennflächen (1) und schmalen Wegen (2) sowie geringere Anzahl nicht klassifizierter Flächen (in weiß; 3).

Daten, andererseits die Aktualität der Fernerkundungsdaten für die Objektextraktion nutzt. BALTSAVIAS & HAHN (2000) geben einen umfassenden Überblick über entsprechende Implementierungsbeispiele. Es sollte aber nicht übersehen werden, dass bei dieser Vorgehensweise Komplexitäts- und Übertragbarkeitsprobleme einer operationellen Verwendung im Wege stehen. So stellt KONECNY (1995) fest, dass der GIS-Einsatz bei der Automation bestehender Aufgabe nur marginale Vorteile erbringt und signifikante Kosteneinsparungen nur bei längerfristiger Anwendung erzielbar sind. Einen alternativen Ansatz zur Vermeidung der Formalisierung von Schwellwerten beschreibt HELLWICH (1999), der die Schätzung von Modellparametern im Rahmen einer Bildinterpretation durch ein Simulated Annealing durchführt.

Rückkoppelungsmechanismen zwischen den einzelnen Klassifizierungsphasen sowie vorhandenen Gedächtnis-Repräsentationen wurden bisher kaum in rechnerischen Verfahren implementiert. Ein Realisierungsbeispiel hierfür demonstriert die in Abschnitt 2 beschriebene Auswertung, die im Gegensatz zu konventionellen Strategien eine Klassifizierung auf Basis *mehrerer* Segmentierungsebenen durchführt, indem für *jede* Segmentierungsebene ein Vergleich der Szenenmerkmale mit den im Konzeptmodell definierten Objektmerkmalen stattfindet, bevor die nächste Ebene bearbeitet wird. Hiermit wird die Tatsache berücksich-

tigt, dass sich die Objekte auch innerhalb einer Klasse bezüglich ihrer Eigenschaften (z.B. Größe, Helligkeiten) unterscheiden und somit auch am besten in unterschiedlichen Ebenen abgebildet werden können (objekt-spezifische statt objektklassen-spezifische Vorgehensweise).

Das Problem dieser multi-skaligen Vorgehensweise besteht offensichtlich darin, dass kleinere Segmente, die z.B. als Einzelhäuser oder -bäume klassifiziert worden sind, in einem größeren Segmentierungs-Level anhand der dortigen Merkmale zu anwendungsspezifisch unerwünschten, größeren Regionen (z.B. einem Gebäudeblock oder Waldgebiet) aggregiert werden könnten. Um dies zu verhindern, muss in einer feineren Klassifizierungs-Ebene der Zwischenraum zwischen Objekten bereits einer Klasse zugeordnet oder festgelegt werden, dass an diesem Ort bestimmte Objekte niemals auftreten dürfen (*Negativ-Klassifizierung*). Auf einem größeren Level erfolgt eine Klassifizierung für ein Segment nur dann, wenn *alle* Subsegmente der aktuell betrachteten Klasse angehören (Aggregation) oder bisher nicht (positiv oder negativ) klassifiziert werden konnten (Neuzuweisung). Abb.6 demonstriert den Mehrwert dieser Vorgehensweise unter Berücksichtigung *aller* Ebenen: So werden beim konventionellen Verfahren mit nur einer Ebene durch vorherige Aggregationen solche kleinen Segmente nicht berücksichtigt, die entweder den positiven Merkmalsanforderungen entsprechen wür-

den (im Beispiel: schmale Wege) oder die eine negative Klassifizierung bedeuten würden (im Beispiel: Trennflächen zwischen Wegen oder Straßen).

4.2.3 Forschungsfragen

Die Ausführungen in den vorangegangenen Abschnitten haben das Dilemma der rechnerischen Szeneninterpretation verdeutlicht: Im Kontext des präskriptiven Ansatzes der Entscheidungstheorie sind geeignete Verfahren vorhanden, die im Gegensatz zum menschlichen Vorgehen eine objektive Bewertung, Gewichtung und Bestimmung des Zugehörigkeitsmaßes erlauben. Prinzipiell besteht ferner auch die Möglichkeit, den Merkmalsumfang mit modernen Sensoren weiter zu steigern, um somit das Entscheidungsrisiko zu minimieren bzw. die Ergebnisqualität zu steigern (siehe auch Abschnitt 4.1.2).

Auf der anderen Seite ist jedoch der Aufwand für eine Formalisierung der objekt-spezifischen Modellparameter oder auch für eine Integration von GIS-Daten i.d.R. nicht vertretbar, weil aufgrund der Variabilität der Merkmale eine Übertragbarkeit selten gewährleistet ist (siehe auch BALTSAVIAS & HAHN 2000). Demgegenüber bedeutet die Verfolgung des deskriptiven Ansatzes zwar eine Einsparung von Kapazitäten, bietet aber auch nur sub-optimale Klassifizierungsergebnisse.

Die Folgerung aus diesem Dilemma – insbesondere der Tatsache, dass praktikable Lösungen für das Formalisierungs-Problem nicht in Sicht sind – besteht unserer Meinung darin, basierend auf heuristischen Wissensrepräsentationen (siehe Abschnitt 4.1.3) verstärkt *heuristische Auswerte- bzw. Suchverfahren* anzuwenden (z.B. Greedy-, A*-, WA*-Suche oder Bergsteigen-Verfahren). Die erzielten Zwischenergebnisse können dann mit einer objektiven Bewertung ausgewählter Merkmale sowie mit *Rückkoppelungsmechanismen* optimiert werden.

5 Zusammenfassung

Die Motivation dieses Beitrages bestand im unbefriedigenden Status der Methoden zur rechnerischen Interpretation von Fernerkundungsszenen. Hierzu wurde der Ansatz verfolgt, einen stärkeren Brückenschlag zu den Theorien der kognitiven Wahrnehmungsforschung herzustellen. Diese Idee begründet sich einerseits in den offensichtlichen Analogien zwischen der kognitiven Wahrnehmung und der rechnerischen Szeneninterpretation sowie andererseits in der Tatsache, dass dieser Ansatz in der jüngeren Vergangenheit in der Fernerkundungs-Literatur vernachlässigt wurde.

Anhand ausgewählter Auswerteschritte der Interpretation – der Extraktion und Verarbeitung von Merkmalen sowie der Klassifizierung – wurden einigen Theorien der Wahrnehmungspsychologie entsprechende Auswertemethoden der Szeneninterpretation gegenübergestellt. Der Schwerpunkt wurde hierbei auf die Beschreibung eigener Arbeiten zur Auswertung von simultan aufgezeichneten Bild- und Laserscanning-Daten des Multi-Sensorsystems TopoSys II gelegt. Aus dieser Gegenüberstellung lassen sich zwei grundsätzliche Entwicklungsrichtungen für Methoden der rechnerischen Szeneninterpretation ableiten:

- Eine Richtung verfolgt das Ziel, zur Verbesserung der Interpretations-Zuverlässigkeit den *Merkmalsumfang* zu steigern (hier wurden entsprechende Beispiele zu den Merkmalen Orientierung und Höhen-Gradient aufgezeigt) sowie Objekt-klassen-spezifische *hybride Auswertestrategien* anzuwenden (hier wurde ein Beispiel zur linien-basierten Segmentierung für schmale Feldmuster vorgestellt). Bei dieser Entwicklungsrichtung stehen dem Vorteil einer angepassten Merkmalsverarbeitung mit einer streng objektiven Bewertung die wachsende Komplexität und der – aufgrund der fehlenden Übertragbarkeit – selten vertretbare Formalisierungs-Aufwand als Nachteile gegenüber.
- Die andere – in der Szeneninterpretation bisher selten umgesetzte – Entwicklungsrichtung versucht die tatsächliche, *heuris-*

tische bzw. deskriptive Herangehensweise des Menschen abzubilden. Diese beinhaltet auch die Möglichkeit von *Rückkopplungsmechanismen* zwischen den einzelnen Klassifizierungsphasen und vorhandenen Wissens-Repräsentationen, deren Mehrwert am Beispiel einer Auswertung in mehreren Generalisierungsebenen (statt in nur einem Level) demonstriert wurde. Grundsätzlich stehen sich bei dieser Vorgehensweise der Nachteil von nur sub-optimalen Interpretationsergebnissen und der Vorteil eines angemessenen zeitlichen und damit auch wirtschaftlichen Aufwandes gegenüber.

Um die jeweiligen Vorteile miteinander zu kombinieren, ist eine Integration dieser Entwicklungsrichtungen anzustreben: Ausgehend von der Tatsache, dass eine praktikable Lösung für das Formalisierungs-Problem nicht in Sicht ist, kann ein Kompromiss darin bestehen, verstärkt heuristische Ansätze zu entwickeln und diese mit einer objektiven Bewertung ausgewählter Merkmale sowie mit Rückkopplungsmechanismen zu optimieren. Diese Integrations-Idee und insbesondere der Verzicht auf vollständig algorithmische Ansätze bedeutet zweifelsohne einen gewissen Paradigmenwechsel, der aufgrund der Defizite der rechnerischen Interpretation von Fernerkundungsszenen aber auch notwendig erscheint und den es zu diskutieren gilt.

Danksagung

Dank gilt der Firma TopoSys für die Bereitstellung des multi-sensoralen Datensatzes.

Literatur

- ALLEN, R. B., 1982: Cognitive factors in human interaction with computers. – In: BADRE, A. & SHNEIDERMAN, B. (Hrsg.): *Directions in Human/Computer Interaction*. – Ablex Publishing Corporation, Norwood (USA).
- BALTSAVIAS, E. & HAHN, M., 2000: Integrating spatial information and image analysis – one plus one makes ten. – *International Archives of Photogrammetry and Remote Sensing*, **33** (B2): 63–74.
- BHALERO, A. & WILSON, R., 2001: Unsupervised image segmentation combining region and boundary estimation. – *Image and Vision Computing*, **19**: 353–368.
- BIEDERMAN, I., 1987: Recognition-by-components: A theory of human image understanding. – *Psychological Review*, **94**: 115–147.
- Definiens-Imaging, 2003: www.definiens-imaging.com, letzter Zugriff: 10.02.2003.
- DONNER, R., 2002: Über den Beitrag der Geoinformatik zum Verständnis des Systems Erde. – *Photogrammetrie-Fernerkundung-Geoinformation*, **2002** (6): 443–450.
- EHLERS, M., EDWARDS, G. & BEDARD, Y., 1989: Integration of Remote Sensing with Geographic Information Systems: A Necessary Evolution. – *Photogrammetric Engineering and Remote Sensing*, **55** (11): 1619–1627.
- EHLERS, M., 2000: Integrated Geographic Information Systems: From data integration to integrated analysis. – *International Archives of Photogrammetry and Remote Sensing*, **33** (B2): 172–178.
- FISCHER, A., KOLBE, T. H., LANG, F., CREMERS, A. B., FÖRSTNER, W., PLÜMER, L. & STEINHAGE, V., 1998: Extracting Buildings from Aerial Images Using Hierarchical Aggregation in 2D and 3D. – *Computer Vision and Image Understanding*, **72** (2): 185–203.
- FÖRSTNER, W., 2002: Kognitive Luftbildinterpretation zur Kartierung auf Anfrage. – Vortrag anlässlich der 22. Jahrestagung der DGPF, 24.09.2002.
- FUCHS, C., 1998: Extraktion polymorpher Bildstrukturen und ihre topologische und geometrische Gruppierung. – DGK-C 502, München.
- GAMBA, P., LODOLA, R. & MECOCCHI, A., 1997: Scene interpretation by fusion of segment and region information. – *Image and Vision Computing*, **15**: 499–509.
- GOLDSTEIN, E. B., 2002: *Wahrnehmungspsychologie*. – 2. dt. Auflage, Spektrum Akademischer Verlag, Heidelberg, Berlin.
- GUSKI, R., 1989: *Wahrnehmung. Eine Einführung in die Psychologie der menschlichen Informationsaufnahme*. – Kohlhammer, Stuttgart u.a.
- HARALICK, R. M. & SHAPIRO, L. G., 1992: *Computer and Robot Vision*. – Addison-Wesley, Reading.
- HARTL, P. & CHENG, F., 1995: Delimiting the building heights in a city from the shadow on a panchromatic SPOT image: Part 2: Test of a complete city. – *International Journal of Remote Sensing*, **16** (15): 2829–2842.
- HELLWICH, O., 1999: Model Parameter Estimation for Digital Image Analysis Using Simula-

- ted Annealing. – Photogrammetrie Fernerkundung Geoinformation, **1999** (6): 371–380.
- KOLBE, T.H., 2000: Identifikation und Rekonstruktion von Gebäuden in Luftbildern mittels unscharfer Constraints. – Shaker-Verlag, Aachen.
- KONECNY, G., 1995: Geographische Informationssysteme aus der Sicht der Photogrammetrie. – In: BUZIEK, G. (Hrsg.): GIS in Forschung und Praxis. – 30–42, Wittwer-Verlag.
- MARR, D., 1982: Vision. – W.H. Freeman, San Francisco.
- PAVLIDIS, T. & LIOUW, Y.T., 1990: Integration Region-Growing and Edge Detection. – IEEE Transactions on Pattern Analysis and Machine Intelligence, **12** (3): 225–233.
- POSCHMANN, G., 1999: Grundbegriffe der präskriptiven Entscheidungstheorie. – Forschungsbericht 99-1 am Zentrum Mensch-Maschine-Systeme (ZMMS) der Technischen Universität Berlin.
- SCHIEWE, J., 2003: Auswertung hoch auflösender und multi-sensoraler Fernerkundungsdaten – Entwicklung von Methoden und Transfer in die Lehre. – Habilitationsschrift, Hochschule Vechta, Materialien Umweltwissenschaften Vechta (erscheint ca. Mai 2003).
- TopoSys, 2003: www.toposys.com, letzter Zugriff: 10.02.2003.
- WICKENS, C.D. & HOLLANDS, J.G., 2000: Engineering Psychology and Human Performance. – Prentice Hall, New Jersey.
- ZHAN, Q., MOLENAAR, M. & TEMPFLI, K., 2002: Hierarchical image object-based structural analysis towards urban land use classification using high-resolution imagery and airborne LIDAR data. – Proceedings of 3rd International Symposium Remote Sensing of Urban Areas: 251–258.
- ZIMBARDO, P.G., 1988: Lehrbuch der Psychologie. – Springer-Verlag, Berlin, Heidelberg, New York.

Anschrift des Verfassers:

PD Dr.-Ing. habil. JOCHEN SCHIEWE
Hochschule Vechta, Forschungszentrum Geoinformatik und Fernerkundung (FZG)
PF 1553, D-49364 Vechta
Tel.: 04441-15558, Fax: 04441-15583
e-mail: jschiewe@fzg.uni-vechta.de

Manuskript eingereicht: Februar 2003
Angenommen: März 2003