

Kombination von terrestrischen Aufnahmen und Fernerkundungsdaten mit Hilfe der k-Nächste-Nachbarn-Methode zur Klassifizierung und Kartierung von Wäldern

WOLFGANG STÜMER, Dresden & MICHAEL KÖHL, Hamburg

Keywords: remote sensing, terrestrial surveying, forest planning and management, k-nearest neighbours method (kNN)

Zusammenfassung: Im vorliegenden Beitrag wird die k-Nächste-Nachbarn-Methode (kNN-Methode), die terrestrische Aufnahmen mit Fernerkundungsdaten kombiniert, beschrieben. An Hand der zwei Merkmale *Grundfläche* (metrische Daten) und *Totholz* (kategoriale Daten) wurden umfangreiche Berechnungen durchgeführt, wobei verschiedenste Einstellungen der Distanzfunktion, der Wichtungsfunktion und der Anzahl k-Nächsten Nachbarn für die kNN-Methode berücksichtigt wurden.

Als Fernerkundungsdatenquellen kamen Landsat- und Hyperspektraldaten zum Einsatz, die sich sowohl von ihrer spektralen wie auch ihrer räumlichen Auflösung unterscheiden. Mit zwei Landsat-Szenen eines Gebietes vom September 1999 und 2000 wurde der multitemporale Ansatz berücksichtigt.

Die terrestrische Datengrundlage setzt sich aus Felddatenaufnahmen mit verschiedenen Aufnahmedesigns zusammen, wobei ein wichtiges Kriterium die gleichmäßige Verteilung von Merkmalswerten (z. B. Grundflächenwerten) über den Merkmalsraum darstellt.

Die Verifizierung der Ergebnisse wurde mithilfe des prozentualen Root Mean Square Error durchgeführt.

Die erzielten Genauigkeiten für das Merkmal *Grundfläche* liegen zwischen 35 % und 67 % (Landsat) bzw. zwischen 65 % und 67 % (HyMapTM). Für das Merkmal *Totholz* liegen die Übereinstimmungen zwischen den kNN-Schätzern und den Referenzwerten zwischen 60,0 % und 73,3 % (Landsat) und zwischen 60,0 % und 63,3 % (HyMapTM).

Mit den erreichten Genauigkeiten bietet sich die kNN-Methode für die Klassifizierung von Beständen bzw. für die Integrierung in Klassifizierungsverfahren an.

Summary: *Terrestrial Survey and use of Remote Sensing Data in combination with the k-nearest neighbours (kNN) method for classification and mapping of woods.* Mapping forest variables and associated characteristics is fundamental for forest planning and management. The paper describes the k-nearest neighbours (kNN) method for improving estimates and for producing maps of the attributes basal area (metric data) and deadwood (categorical data). Several alternatives of the kNN-method were tested, including: distance metric, weighting function and number of neighbours. As sources of remote sensing Landsat TM satellite images and hyper spectral data (HyMapTM) were used, which differ both from their spectral as well as their spatial resolutions. Two Landsat scenes from the same area acquired September 1999 and 2000 allowed for multitemporal approaches.

The field data for the kNN-method comprise tree field measurements campaigns which were collected from the test site Tharandter Wald (Germany). The three field data campaigns are characterized by three different sampling designs. The relative root mean square errors (RMSE) and the Bootstrap method were utilized in order to find optimal parameters. The estimated accuracy for the attribute basal area is between 35 % and 67 % (Landsat) and 65 % and 67 % (HyMapTM). For the attribute deadwood the accuracy is between 60 % and 73 % (Landsat) and 60 % and 63 % (HyMapTM).

Recommendations for operational applications the kNN method for mapping and regional estimation are provided.

Einleitung

Für die Durchführung von Waldinventuren ist der steigende Bedarf an Informationen über Wälder für die nationale und internationale Forst- und Umweltpolitik, die Holzmarktpolitik und zur Erfüllung der Verpflichtungen, die im Rahmen internationaler Verträge, Programme und Initiativen (z. B. Klimaschutz, Pan-Europäischen Prozess zum Schutz der Wälder – Helsinki-Prozess, Zertifizierung, FFH-Richtlinien, Biotopkartierung etc.) zu beobachten. Andererseits werden die finanziellen Ressourcen zur Durchführung von Waldinventuren immer knapper. Die Ergänzung der herkömmlichen Waldinventuren durch Fernerkundungsdaten kann dafür eine Lösung bieten, wie beispielsweise die Landesforstinventur (LFI) in der Schweiz zeigt, wo durch den Einsatz von Fernerkundungsdaten die Kosten gesenkt wurden (WSL, 1999).

An diesem Punkt setzt die hier vorgestellte Untersuchung zur kNN-Methode an (MUONEN & TOKKOLA 1990, TOMPPU 1991, 1993, 1997a, 1997b, FAZAKAS & NILSSON 1996, NILSSON 1997, MALTAMO & KANGAS 1998). Es soll eine Verfahrensalternative weiterentwickelt und verifiziert werden, die über die spektrale Information von Scannerdaten eine räumliche Interpolation von punktförmig erhobenen, terrestrischen Daten ermöglicht und somit eine kartenmäßige Darstellung der räumlichen Struktur durch die Kombination verschiedenster, terrestrisch erhobener Merkmale und aus Satellitendaten abgeleiteter Informationen gestattet. Die Kombination von terrestrischen Daten und Scannerdaten zur Erstellung detaillierter Bestandskarten erfolgt über die kNN-Methode.

Übersicht über die kNN-Methode

Die k-Nächste-Nachbarn-Methode (kNN-Methode) ist ein Verfahren des Maschinellen Lernens, wobei als Lernen jede Änderung der Reaktionsweise eines Systems in Richtung eines optimalen Verhaltens verstanden wird (NIEMANN 1983). Dieses optimale Verhalten wird durch eine Zielfunktion

beschrieben, die durch eine gegebene endliche Stichprobe möglichst gut approximiert werden soll (MITCHELL 1997). Daher kann das Lernen als Suche nach der Hypothese, die die Zielfunktion am besten beschreibt, aufgefasst werden (HESSENMÖLLER 2002). Wird die kNN-Methode zur Klassifizierung verwendet, beschreibt OELLIEN (2002) sie wie folgt: „Die k-Nächste-Nachbarn-Methode ist ein Klassifizierungsverfahren bei dem für einen gegebenen Einzelpunkt der Datenwert gesucht wird, der die größte Ähnlichkeit zu einem bereits klassifizierten Datenpunkt aufweist. Dieser so genannte nächste Nachbar wird dabei mithilfe von Distanz- und Kombinationsfunktionen (in der Regel *Euklidische* Distanzen) ermittelt. Das unbekannte Datenobjekt wird letztendlich derjenigen Klasse zugeordnet, die die höchste Anzahl an ähnlichen bzw. nächsten Nachbarn enthält“.

Bei der Auswertung von Satellitendaten werden für die kNN-Methode Referenzpixel ausgewählt, die mit terrestrischen Erhebungen übereinstimmen. Für ein Pixel ohne entsprechende terrestrische Informationen werden aus der Menge der Referenzpixel die k-nächsten Nachbarn im spektralen Merkmalsraum mithilfe der euklidischen Distanz bestimmt. Für die ausgewählten k Referenzpixel werden die zugehörigen terrestrisch erhobenen Merkmalswerte mit der Distanz im spektralen Merkmalsraum gewichtet und dem Pixel zugeordnet. Dies wird für jedes Pixel wiederholt, bis jedem Pixel der Szene ein Schätzwert des Merkmals zugeordnet ist. Die Klassifizierung von Bilddaten in der Fernerkundung ist nur ein Beispiel für die Anwendung der kNN-Methode. Eine Vielzahl von Objekten in verschiedenen Fachgebieten können mithilfe der kNN-Methode klassifiziert werden. Beispielsweise wird in der Medizin die kNN-Klassifizierung zur Auswertung von MRI-Daten (Magnetic resonance imaging data) und biochemische Substanzen eingesetzt (WARFIELD 1996, QI 2002). Auf dem Fachgebiet der Chemie werden chemische Verbindungen mithilfe der kNN-Methode analysiert (DOWNS & BARNARD 2001) und in der Informatik Datenbankabfragen, wie beispielsweise die Suche

nach Bildern, Videos, Sprache und Dokumenten, realisiert (vgl. DE VRIES et al. 2002, TUNCEL & ROSE 2002).

Die Vorteile der kNN-Modelle gegenüber funktionellen, auf empirischen Parametern beruhenden Verfahren liegt in der Tatsache, das alle zur Verfügung stehenden Beobachtungen in der Schätzung des fehlenden Merkmals direkt berücksichtigt werden. Nachteilig zeigt sich allerdings der enorm hohe Rechenaufwand und die mangelnde Extrapolationsfähigkeit des Ansatzes (HESSENMÖLLER & ELSENHANS 2002).

In der Forstwissenschaft wird die kNN-Methode zur Klassifizierung von Beständen eingesetzt. Im Zusammenhang mit Forstinventuren wurde sie zur Kombination von Satellitendaten und terrestrischen Erhebungen erstmals von KILKKI & PÄIVINEN (1986) beschrieben. TOMPPU (1991, 1993, 1997a, 1997b) entwickelte die kNN-Methode weiter und führte sie in die finnische nationale Forstinventur ein. Dort wird die kNN-Methode zur Herleitung von Resultaten für Verwaltungseinheiten (Forestry Board Districts) verwendet, indem Landsat-TM-Daten mit den terrestrisch erhobenen Daten der finnischen nationalen Forstinventur kombiniert werden. Des Weiteren wurde die Methode im Rahmen der Landeswaldinventur Nordrhein-Westfalen 1997 erprobt (TOMPPU & PEKKARINEN 1997). Die kNN-Methode wurde ursprünglich zur Herleitung von Karten für metrische Daten entwickelt. Da kategoriale Daten keine Mittelwertbildung erlauben, wurde die kNN-Methode im Rahmen des ProSmart-Projektes an der TU Dresden für die Anwendung auf kategoriale Daten erweitert (KÖHL et al. 2000).

kNN-Methode für metrische Daten

Ausgehend von Pixel, deren geografische Position mit den Stichprobenpunkten bei den Feldaufnahmen übereinstimmen, werden für alle Pixel – auch solche, die nicht mit terrestrischen Proben koinzidieren – Variablenwerte (z.B. Holzvorrat, Stammzahl oder Totholzanteil) berechnet. Die Variablenwerte eines Pixels ergeben sich als Mittel-

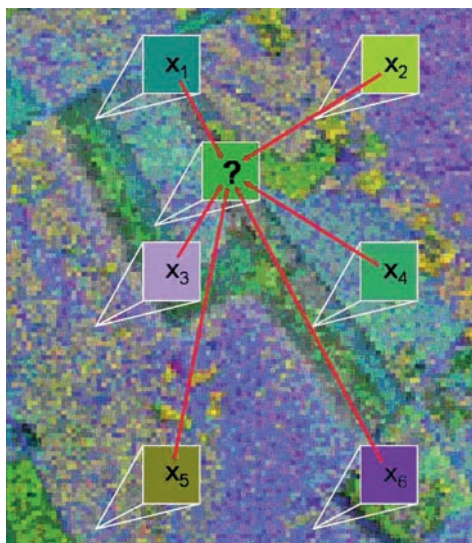


Abb. 1: kNN-Methode. Über bekannte Merkmale von Pixels (x_1 bis x_6) wird über einen Schätzer das unbekannte Pixel (?) klassifiziert.

werte von k benachbarten Proben, die über ein Distanzmaß, das die spektrale Ähnlichkeit beschreibt, gewichtet werden (Abb. 1).

Jedes Pixel enthält pro Kanal einen digitalen Wert als spektrale Information. Die spektrale Differenz zwischen zwei Pixel kann mit Hilfe einer einfachen Metrik bestimmt werden. Ein gebräuchliches Distanzmaß ist die Euklidische Distanz, $d_{(i)p}$, die im spektralen Raum von Pixel p , das klassifiziert werden soll, zu jedem Pixel i , für das eine terrestrische Beobachtung vorliegt, berechnet wird. Falls x_1 und x_2 die Merkmalsvektoren zweier Pixel sind, deren Ähnlichkeit überprüft werden soll, ist die Euklidische Distanz $d(x_1, x_2)$ zwischen ihnen

$$d_{(x_1, x_2)} = \left[\sum_{j=1}^N (x_{1j} - x_{2j})^2 \right]^{1/2} \quad (1)$$

wobei N die Anzahl der spektralen Komponenten (z. B. verwendete Kanäle) bezeichnet. Eine Generalisierung der Euklidischen Distanz ist die Minkowski- r -Distanz (BORTZ 1993). Wird statt des Exponenten 2 (bzw. $1/2$) der Exponent r (bzw. $1/r$) eingesetzt, erhält man eine Verallgemeinerung der

mit der Gleichung (1) beschriebenen Euklidischen Distanz:

$$d_{(x_1, x_2)}^n = \left[\sum_{j=1}^N (x_{1j} - x_{2j})^r \right]^{1/r} \quad (2)$$

Aus $r = 1$ resultiert die City-Block-Metrik (BORTZ 1993), bei der sich die Distanz zweier Punkte aus der Summe der Merkmalsdifferenzen ergibt.

Durch das Variieren des Metrikkoeffizienten werden Merkmalsdifferenzen unterschiedlich gewichtet. Mit $r = 1$ werden alle Merkmalsdifferenzen unabhängig von ihrer Größe gleich gewichtet. Für $r = 2$ erhalten größere Differenzen ein stärkeres Gewicht als kleinere Differenzen. Wird $r = \infty$ gesetzt, wird die größte Merkmalsdifferenz mit 1 gewichtet und alle übrigen erhalten ein Gewicht von 0.

Objekte zeigen einen unterschiedlichen Einfluss auf die Intensität, räumliche Variabilität und den Spektralbereich der Reflexion. Die Spektralbereiche mit einer großen Variabilität in der Rückstrahlung sind meist besser zur Differenzierung von Klassen eines Merkmals geeignet. Die Variabilität der spektralen Information ist in den einzelnen Kanälen unterschiedlich stark ausgeprägt. Um Kanäle mit einer großen Variabilität in der Rückstrahlung und den damit verbundenen Einfluss auf die Unterscheidung von Merkmalsklassen eines Attributes stärker zu wichten, wird in die Distanzformel der Parameter a_j für die Wichtung der Kanäle eingefügt (FRANCO-LOPEZ et al. 2001).

$$d_{(x_1, x_2)} = \left[\sum_{j=1}^N a_j^r (x_{1j} - x_{2j})^r \right]^{1/r} \quad (3)$$

Wählt man den Parameter a_j für $j = 1, \dots, N$ gleich 1, so haben alle Kanäle bei der Berechnung der Distanz das gleiche Gewicht. Es kann aber auch jedem Kanal ein angepasstes Gewicht a_j zuteil werden.

Die $k = 1$ bis $k = 50$ nächsten spektralen Nachbarn, d. h. Pixel mit korrespondierenden terrestrischen Beobachtungen, werden für die weiteren Betrachtungen verwendet. Dabei kommen die Pixel zur Auswahl, die die Bedingung

$$d_{(1),p} \leq d_{(2),p} \leq \dots \leq d_{(k),p} \leq \dots \leq d_{(n),p} \quad (4)$$

im spektralen Raum erfüllen, wobei $d_{(k),p}$ die Distanz der k nächsten Nachbarn und n die Anzahl der verfügbaren Pixel mit korrespondierenden terrestrischen Daten ist. Alle Pixel mit Distanzen im spektralen Raum größer $d_{(k),p}$ vom betrachteten Pixel p werden ignoriert. Bei $k = 1$ wird nur das Pixel mit der geringsten spektralen Differenz für die weitere Berechnung berücksichtigt. Je höher k ist, desto mehr Pixel mit zugeordneten terrestrischen Informationen beeinflussen den zu berechnenden Merkmalswert des Pixels ohne terrestrische Informationen. Die Verwendung von k Stichproben erlaubt es, die beispielsweise durch Signalstörung verursachte, zufällige Streuung einzuschränken.

Die Distanzwerte spiegeln nur die Unterschiede zwischen den spektralen Informationen zweier Pixel wieder. Um die Werte eines bei terrestrischen Aufnahmen erhobenen Merkmals, welche den k nächsten Pixel zugeordnet sind, in die weitere Berechnung einfließen zu lassen, müssen diese entsprechend ihrer spektralen Distanz gewichtet werden. Deshalb wird für jedes ausgewählte Pixel ein Gewicht $w_{(i),p}$ kalkuliert.

$$w_{(i),p} = \frac{1}{d_{(i),p}^2} \cdot \frac{1}{\sum_{i=1}^k d_{(i),p}^2} \quad (5)$$

Je ähnlicher die spektrale Information ist, desto höher ist das Gewicht und damit der Einfluss auf den zu berechnenden Merkmalswert.

MALTAMO & KANGAS (1998) modifizieren (5) zur Bestimmung des Pixelgewichtes

$$w'_{(i),p} = \frac{\left(\frac{1}{1 + d_{(i),p}} \right)^t}{\sum_{i=1}^k \left(\frac{1}{1 + d_{(i),p}} \right)^t} \quad (6)$$

wobei k die Anzahl der nächsten Nachbarn festlegt und t die Wichtung der Distanz be-

einflusst. Je größer t gewählt wird, umso größer ist das Gewicht der Pixel mit geringer spektraler Distanz. Die Summe aller Gewichte $w'_{(i),p}$ beträgt stets 1.

Für ein Pixel wird mit Hilfe der auf terrestrischen Probeeinheiten erfassten Merkmalswerte und den zugehörigen, aus spektralen Daten abgeleiteten Gewichten der k nächsten Pixel ein Wert für das Merkmal bestimmt.

\hat{m}_p ist ein Schätzer des Merkmals M für Pixel p

$$\hat{m}_p = \sum_{i=1}^k w_{(i),p} m_{(i),p} \quad (7)$$

wobei $m_{(j),p}$ die terrestrisch erfassten Werte der $i = 1, \dots, k$ Pixel sind, die im spektralen Raum am nächsten zu Pixel p liegen. Der Prozess wird für jedes Pixel wiederholt und resultiert in Abhängigkeit von der Größe der Inventurregion und der räumlichen Auflösung des Sensors in intensiven Berechnungen.

Durch die Wahl der Parameter k , r , t und a_j ist eine Einflussnahme auf den Schätzer möglich. In der Untersuchung wurden über 650 Berechnungen mit verschiedensten Kombinationsmöglichkeiten der Parameter k , r und t durchgeführt. Die Parameter a_1, a_2, \dots, a_j erhielten den Wert 1, da sonst die direkte Vergleichbarkeit der unterschiedlichen Sensoren unter praxisnahen Bedingungen nicht mehr möglich wäre. Um die optimalen Einstellungen für die Variablen zu erhalten, wurden die Fehler der kNN-Ergebnisse mit dem prozentualen Root Mean Square Error (%RMSE) (Formel 8) bestimmt und die Berechnungsalternative ermittelt, die zum kleinsten %RMSE führte.

$$\text{RMSE}\% = \frac{\sqrt{(\text{Varianz von } \hat{\mu}) + (\text{bias})^2}}{\hat{\mu}} * 100\% \quad (8)$$

kNN-Methode für kategoriale Daten

Die oben gezeigten Schätzprozeduren sind nur für intervall- und absolut skalierte Daten definiert und können unter der Voraussetzung der Permutationsinvarianz nicht auf

rang- oder nominal skalierte Daten übertragen werden. Als Ansatz zur Lösung dieses Problems wird auf eine gewichtete Mittelwertbildung verzichtet und stattdessen die dem betrachteten Pixel zuzuordnende Merkmalsklasse über Wahrscheinlichkeiten hergeleitet.

Zunächst wird wie im metrischen Fall der Abstand der k nächsten Nachbarn im spektralen Raum zu Pixel p bestimmt und die Gewichte $w_{(i),p}$ hergeleitet (Formel 1 bis 6). Jedem der k Referenzpixel ist ein Merkmal j mit s Merkmalsklassen zugeordnet. Für jede der s Merkmalsklassen von j wird die Summe der Gewichte $w_{(i),p}, w_{p,j}$ berechnet.

$$w_{(i),p,j} = \begin{cases} w_{(i),p}, & \text{falls Pixel } p \text{ Attribut } j \text{ zugeordnet ist} \\ 0, & \text{in allen anderen Fällen} \end{cases} \quad (9)$$

$$w_{p,j} = \sum_{i=1}^k w_{(i),p,j} \quad (10)$$

mit $p = 1, \dots, N$; $i = 1, \dots, k$

Aus der Menge der s Gewichte $w_{p,j}$ wird das Maximum bestimmt und das entsprechende Merkmal j mit der entsprechenden Merkmalsklasse s wird Pixel p zugeordnet.

Untersuchungsgebiet und Datengrundlage

Das Untersuchungsgebiet Tharandter Wald liegt ca. 15 km südwestlich von Dresden und ist durch seine große zusammenhängende Waldfläche, das moderate Relief und der strukturreichen Waldformation prädestiniert für die forstliche Fernerkundung.

Als terrestrische Datengrundlage standen drei terrestrische Datenquellen zur Verfügung, die sich durch ihr Aufnahmedesign unterscheiden. Die Aufnahmen werden im folgendem mit Aufnahme 1 bis 3 bezeichnet. Die Aufnahmen 1 und 2 verwenden ein einphasiges Stichprobenverfahren, in dem konzentrische Probeflächen systematisch in einem Raster angeordnet sind. Die Probepunkte der Aufnahme 1 sind zusätzlich in Clustern zusammengefasst. Dem gegenüber steht ein zweiphasiges Stichprobenverfah-

ren, welches im Rahmen der Aufnahme 3 erhoben wurde. Die verschiedenen Aufnahmedesigns ermöglichen eine Optimierung der terrestrischen Datengrundlage für die kNN-Methode und die damit verbundene Senkung der Aufnahmekosten.

Als Fernerkundungsdatenquellen wurden neben einer Landsat-5- (Juli 1989) und zwei Landsat-7-Szenen (September 1999 und 2000) noch jeweils eine HyMapTM- und HRSC-A-Szene verwendet.

Für die Landsat-Szenen wurde am Lehrstuhl für Biometrie und Forstliche Informatik, TU Dresden eine Atmosphärenkorrektur und eine Georeferenzierung durchgeführt (HEHMKE 2001). Für alle drei Landsat-Datensätze wurde eine Hauptkomponentenanalyse (Principle Component Analysis, PCA) realisiert und die vier neu generierten Kanäle dem Ausgangsbild angefügt. Außerdem wurde für alle drei Landsat-Szenen der Normalized Difference Vegetation Index (NDVI) bestimmt und als Kanal dem Ausgangsbild hinzugefügt. Der NDVI nutzt charakteristische Reflexionsunterschiede zwischen sichtbarem Licht und nahem Infrarot (HILDEBRANDT 1996).

$$NDVI = \frac{IR - R}{IR + R} \quad (11)$$

Die Hyperspektraldaten wurden im August 1998 mit dem Hyperspektralscanner HyMapTM der australischen Firma Integrated Spectronics aufgenommen. Der Scanner zeichnet Daten in einem Wellenlängenbereich von 0,4 bis 2,4 μm in 128 Spektralbändern auf, die geometrische Auflösung betrug bei der gewählten Flughöhe 6 m. Zusammen mit der High Resolution Stereo Camera (HRSC-A) des Instituts für Planetenerkundung, DLR Berlin (LEHMANN et al. 1999) wurde der Hyperspektralscanner in einer Cessna 208 Caravan montiert. Prozessierung, Kalibrierung und Atmosphärenkorrektur der HyMapTM Daten wurden von der DLR, Institut für Optoelektronik durchgeführt (MÜLLER et al. 1999). Die Geokodierung und eine Minimum Noise Fraction Transformation (MNF) der HyMapTM-Daten wurde von der Gesellschaft für Ange-

wandte Fernerkundung (GAF) in München durchgeführt (KÖHL & LAUTNER 2001). Bei der MNF-Transformation wurden die ersten 20 Hauptkomponenten ausgewählt und eine Hyperspektralszene mit 20 MNF-Kanälen erzeugt.

Ergebnisse

Fehleranalyse – Grundfläche

Für die Fehlerwertberechnung wurde eine Vielzahl von Szenarien berücksichtigt. Dazu zählen, unterschiedliche Kanalkombinationen und die Wahl der Parameter k , r und t , wobei k als Schlüsselvariable die Anzahl der verwendeten nächsten Nachbarn definiert. Der Parameter r steht als Exponent bei der Euklidischen Distanz und ermöglicht die Wichtung der spektralen Differenzen der einzelnen Kanäle. Mit dem Parameter t ist wiederum eine Einflussnahme auf die Wichtung der Distanz möglich. Der Einsatz von Stichproben von den drei unterschiedlichen Aufnahmen erlaubt die Untersuchung des Einflusses verschiedener Stichprobenverfahren auf die kNN-Ergebnisse.

Als erstes wurde die Anzahl der zu verwendenden k -nächsten Nachbarn untersucht. Für die Berechnung des %RMSE wurden verschiedene k -Werte in Verbindung mit unterschiedlichen Kanalkombinationen der Landsat-Szenen berücksichtigt. Unter Verwendung der nur vom Sensor bereitgestellten Kanäle ist der kleinste %RMSE von 38,56% bei der Landsat-Szene von 1999 zu beobachten. Werden diese Kanäle durch die PCA- und NDVI-Kanäle ergänzt, wird bei der multitemporalen Szene von 1999/2000 ein Minimum von 37,07% erreicht (Tab. 1). In Abb. 2 werden die Ergebnisse der kompletten k -nächsten Nachbarn-Wertereihen für die Landsat-Szenen unter Verwendung aller Kanäle inkl. PCA- und NDVI-Kanäle dargestellt. Die Zusammenführung zweier Landsat-Szenen zu einer multitemporalen Szene und die Verwendung dieser bei den kNN-Auswertungen führen zu einem Genauigkeitsgewinn, der bei Verwendung aller Kanäle sein Maximum erreicht.

Bei den Hyperspektraldaten konnte kein temporaler Ansatz untersucht werden, da

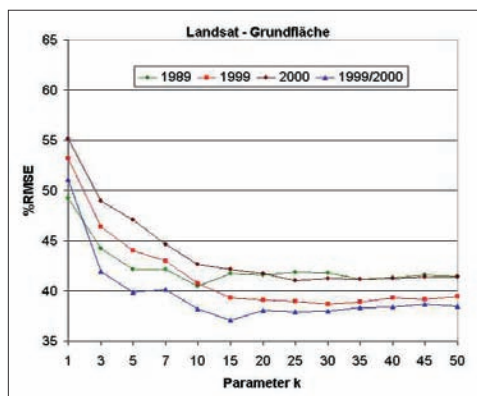


Abb. 2: %RMSE der Grundfläche von vier Landsat-Szenen (alle Kanäle inkl. PCA und NDVI) für verschiedene k -Werte (alle Stichproben; $r = 2$, $t = 2$).

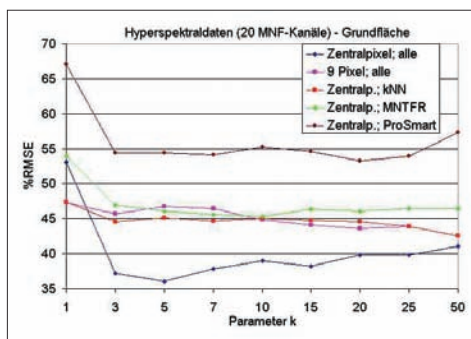


Abb. 3: %RMSE der Grundfläche der transformierten Hyperspektralszene (20 MNF-Kanäle) für die Variante „Zentralpixel“ und „9 Pixel“ (MNTFR = Aufnahme 1; kNN = Aufnahme 2; ProSmart = Aufnahme 3; alle = alle Stichproben der Aufnahmen 1 bis 3).

nur eine Szene zur Verfügung stand. Der Szene mit 128 Kanälen wurde die transformierte Szene mit 20 MNF-Kanälen gegenübergestellt und die Fehlerwerte für die *Grundfläche* verglichen. Der kleinste %RMSE für die 128-kanalige-Szene beträgt 41,79% bei $k = 75$ und für die 20-kanalige-Szene 36,05% bei $k = 5$. Die transformierte

Szene bringt bei den kNN-Rechnungen einen Genauigkeitsgewinn (Tab. 1).

Als nächstes wurde der Unterschied zwischen den zwei Varianten Zentralpixel und Mittelwertbildung der 9 Pixel über der Stichprobe untersucht. Die Ergebnisse zeigen, dass die Zentralpixelvariante zu den kleineren %RMSE-Werten führt (Abb. 3).

Tab. 1: Kleinste beobachtete %RMSE der Grundflächenschätzung für Landsat und HyMap™ von verschiedenen kNN-Varianten (bei den Varianten wurden alle Probenpunkte der drei Aufnahmen verwendet).

Szene	Landsat 1999	Landsat 1999/2000	Landsat 1999/2000	HyMap	HyMap	HyMap
Kanäle	1–5; 7	alle; PCA; NDVI	alle; PCA; NDVI	128	20 MNF	20 MNF
Parameter k	25	15	9	75	5	4
Parameter r	2	2	10	2	2	2
Parameter t	2	2	0,6	2	2	0,1
Grundfläche [m ² /ha]	37,59	37,59	37,59	34,02	34,02	34,02
Grundfläche (kNN) [m ² /ha]	38,03	36,58	35,71	35,39	34,57	35,02
Bias [m ² /ha]	– 5,96	– 4,51	– 3,63	– 3,32	– 2,50	– 2,95
Standardfehler	13,40	12,79	11,49	14,41	12,21	12,00
MSE [m ² /ha] ²	215,15	183,95	145,33	218,78	155,30	152,81
RMSE [m ² /ha]	14,67	13,56	12,06	14,79	12,46	12,36
%RMSE	38,56	37,07	33,76	41,79	36,05	35,29

Die terrestrisch erhobenen Stichproben haben durch ihre Anzahl und Verteilung einen entscheidenden Einfluss auf die kNN-Ergebnisse. Bei dem Attribut *Grundfläche* kamen die drei terrestrischen Aufnahmen 1 bis 3 und eine vierte Variante mit allen Stichproben zum Einsatz. Für die Hyperspektraldaten sind die %RMSE-Werte der vier Varianten in Abb. 3 gegenübergestellt und zeigen bei Verwendung aller Stichproben die kleinsten Fehlerwerte. Für die kNN-Methode wird eine Verteilung der Stichprobenpunkte über die gesamte Merkmalsklasse benötigt, was näherungsweise durch die Verwendung aller Stichprobenpunkte (354 Punkte bei Landsat bzw. 263 bei HyMapTM) erreicht wird. Die unterschiedlichen Aufnahmedesigns haben wegen ihrem anteilmäßig geringen Stichprobenumfang nur einen untergeordneten Einfluss auf die Ergebnisse. Beim Vergleich der Varianten treten allerdings erhebliche Unterschiede zwischen den Aufnahmen auf. Das schlechte Abschneiden der Aufnahme 1 ist auf das Fehlen von Stichproben in Jungbeständen und der damit verbundenen Überschätzung der Grundfläche zurückzuführen. Die Verteilung der Proben erstreckt sich nicht über den gesamten Merkmalsraum, sondern deckt vor allem Bestände mit einer mittleren und hohen Grundfläche ab. Auch die beiden anderen Aufnahmen beschränken sich auf einen Teil des Merkmalraums, wobei die aufgenommenen Proben unterschiedliche Bereiche abdecken. Durch die Zusammenführung der Stichproben wird ein breiter Bereich abgedeckt und damit ein entsprechender Genauigkeitserfolg erzielt.

Wird bei der Euklidischen Distanz der konstante Exponent durch den Exponent r ersetzt, erhält man einen Metrikoeffizienten, der die Wichtung von spektralen Differenzen der einzelnen Kanäle ermöglicht. Für die Berechnungen wurden verschiedene r -Werte untersucht. Der Einfluss des Parameters r auf die Wichtung einzelner Kanäle und die damit verbundene Gesamtdistanz bewegt sich in sehr engem Rahmen. Die maximale Differenz des minimalen und maximalen %RMSE für die verwendeten Szenen beträgt 4,06 %. Trotz des geringen Einflusses

wird bei $r = 10$ eine Minimierung des %RMSE auf 36,63 % (Landsat 1999/2000; $k = 10$; $t = 2$) bzw. bei $r = 4$ auf 36,01 % (HyMapTM; $k = 5$; $r = 2$) erreicht.

Für weitere Berechnungen wurden die Ergebnisse der kNN-Berechnungen mit den k -, r - und t -Wertreihen zusammengeführt, um eine optimale Kombination dieser drei Parameterwerte zu finden (Tab. 1). Unter Verwendung aller terrestrischer Proben wurde für die multitemporale Landsat-Szene von 1999/2000 bei $k = 9$, $r = 10$ und $t = 0,6$ ein %RMSE von 33,76 % gefunden. Bei den Hyperspektraldaten wurde bei Verwendung der 20 MNF-Kanäle und der Auswahl von vier nächsten Nachbarn bei $t = 0,1$ ein Minimum des %RMSE mit 35,29 % gefunden (Tab. 1). Eine weitere Minimierung des Fehlers ist mit den gegebenen Einstellungen und Daten nicht möglich.

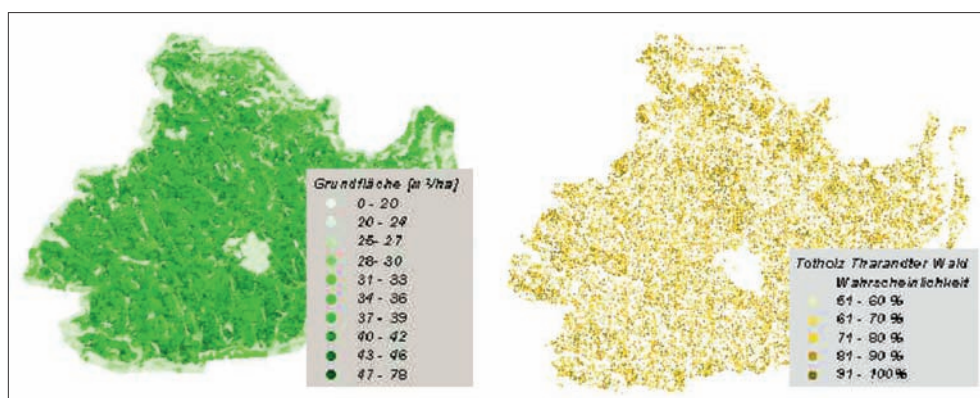
Fehleranalyse – Totholz

Da das Attribut *Totholz* nicht metrisch skaliert ist, wird statt des %RMSE ein alternativer Ansatz zur Fehleranalyse verwendet. Es wurden 30 terrestrisch erhobene Werte, die als Referenzpunkte zufällig ausgewählt wurden und nicht in die kNN-Berechnungen einfließen, mit den über den kNN-Ansatz berechneten Werten verglichen und über die Anzahl der Übereinstimmungen eine Genauigkeitsaussage getroffen. Die Aussage über das Auftreten oder des Nichtauftretens von *Totholz* wird über die Berechnung der Eintrittswahrscheinlichkeit mit Hilfe der kNN-Methode getroffen. Die Wahrscheinlichkeitsschwelle für die Zuordnung zur Merkmalsklasse *Totholz* wurde bei allen Ergebnissen bei 50 % festgelegt.

Die größte Anzahl Übereinstimmungen wurde bei Verwendung der Landsat-Szene vom September 2000 mit Integrierung aller Kanäle (inkl. PCA und NDVI) und aller Stichproben mit 22 Übereinstimmungen (73,3 %) erreicht (Tab. 2). Beim Vergleich der Ergebnisse bezogen auf die verwendeten Kanäle erhält keine Variante den Vorzug, da sich bei den Landsat-Szenen von 1989, 1999 und 1999/2000 die Ergebnisse bei Auswahl aller Kanäle verschlechtert und bei der Szene von 2000 verbessert. Die Parameter r und

Tab. 2: Totholz-Fehlerwerte unter Verwendung aller Stichproben ($p^2 = 50\%$)

Szene/Shift-Algorithmus (HyMap™)	Landsat 1989	Landsat 99/00	Landsat 2000	HyMap™ Zentralp.	HyMap™ 9 Pixel
Kanäle	1–5,7	1–5, 7, 14–18, 20	alle	20 MNF	20 MNF
Parameter k	50	5	2	15	1
Parameter r	2	5	2	2	2
Parameter l	2	2	2	2	2
Übereinstimmungen (max. 30 möglich)	21	21	22	18	19
Übereinstimmungen in Prozent	70,0	70,0	73,3	60,0	63,3

**Abb. 4:** kNN-Karten für die Merkmale *Grundfläche* und *Totholz* vom Tharandter Wald (Ausgangsdaten Landsat-7-Szene von 1999/2000 alle Kanäle + PCA + NDVI).

t bringen einen Genauigkeitsgewinn, der bei der Szene von 1999/2000, die den multitemporalen Ansatz vertritt, nachgewiesen werden konnte. Allerdings ist zu beachten, dass sich die Abweichungen der Anzahl der Übereinstimmungen zwischen den verschiedenen Fehlerrechnungsvarianten zwischen 1 (3,3%) und max. 3 (10%) in einem kleinen Rahmen bewegen (Tab. 2).

kNN-Karten

Die mit Hilfe der kNN-Methode errechneten Karten zeigen für jedes einzelne Pixel den über die Grauwerte der Kanäle ermittelten Wert für die Attribute *Grundfläche* und *Totholz*. Zur besseren Darstellung wurden Klassen gebildet und diese aussagefähigen

Farben zugeteilt. Bei *Totholz* sind die beiden Merkmalsklassen *vorhanden* oder *nicht vorhanden* möglich, weshalb die aus der kNN-Berechnung geschätzte Eintrittswahrscheinlichkeit für das *Totholz* zur Klassenbildung herangezogen wurde.

kNN-Grundflächen- und -Totholzkarten vom Tharandter Wald entstanden aus der Landsat-Szene 1999/2000 sind in Abb. 4 und aus den Hyperspektraldaten in Abb. 5 zu finden.

Diskussion

Ziel dieser Untersuchung war es, eine umfangreiche Analyse der kNN-Methode in Hinblick auf die Eignung und Genauigkeits-

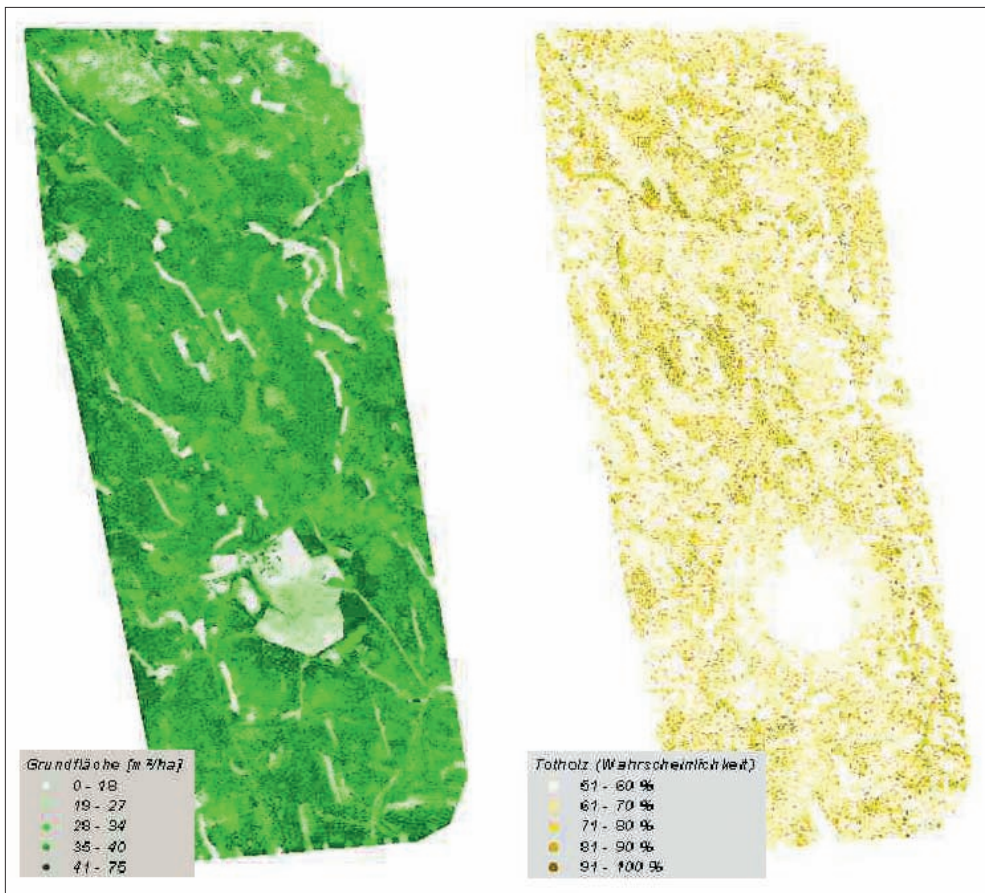


Abb. 5: kNN-Grundflächen- und Totholz-Karte für den Teil des Tharandter Waldes, der von der Hyperspektralszene (20 MNF-Kanäle; Zentralpixel) abgedeckt wird.

aussage zur Herstellung von Bestandeskarten vorzunehmen. Als terrestrische Datengrundlage kamen drei Datenquellen zum Einsatz, die sich durch ihr Aufnahmedesign unterscheiden. Insgesamt wurden 542 Proben im Tharandter Wald, der einer Fläche von ca. 6000 ha aufweist, aufgenommen, was einer Auswahlintensität von ca. 0,1 Proben pro Hektar entspricht. Diese Proben-dichte kann als ausreichend für aussagekräftige kNN-Ergebnisse betrachtet werden (vgl. FRANCO-LOPEZ et al. 2001, NILSSON 1997). Bei einer separaten Betrachtung der Aufnahme 1 stehen 260 Proben für die kNN-Berechnungen zur Verfügung. Dies entspricht einem Durchschnitt von 0,04 Pro-

ben pro Hektar. Auch diese Anzahl an Proben kann als ausreichend für die kNN-Methode betrachtet werden. Voraussetzung dafür wäre allerdings die Verteilung der Proben über den gesamten Merkmalsraum, wie beispielsweise die Gleichverteilung der Proben auf jede Grundflächenklasse. Dies ist aber nicht gegeben, da nur sehr wenige Stichproben in Jungbeständen und eine überproportionale Anzahl in Altbeständen liegen. Der hohe Grundflächenmittelwert von 44,9 m²/ha der Stichproben der Aufnahme 1 bestätigt diese Verteilung. Proben mit einem Grundflächenwert kleiner 15 m²/ha sind nur begrenzt vorhanden. Diese Konstellation führt bei den kNN-Berech-

nungen zur Überschätzung der Grundflächenwerte, da für die k -nächsten Nachbarn Proben mit großen Grundflächenwerten herangezogen werden. Dies lässt den Schluss zu, dass für die Erstellung von k NN-Karten eines Gebietes ohne Über- bzw. Unterschätzung des Merkmals die repräsentative Stichprobenerhebung über den gesamten Merkmalsraum mit der an den k -nächsten Nachbarn angepassten Anzahl eine entscheidende Rolle spielt. Dieses Problem könnte mit einer sehr hohen Anzahl von Stichproben, beispielsweise einer Probe pro Hektar, begegnet werden. Zu beachten ist aber, dass eine hohe Anzahl nicht gleichzusetzen ist mit einer gleichmäßigen Verteilung über den gesamten Merkmalsraum. Mit Hilfe eines zweiphasigen Stichprobenverfahrens, wie bei der Aufnahme 3 bewerkstelligt, kann eine Optimierung hinsichtlich der Gleichverteilung der terrestrischen Datengrundlage für die k NN-Methode und die damit verbundene Senkung der Aufnahmekosten erfolgen.

Bei einer Überschätzung der Grundfläche bei den k NN-Berechnungen liegen die k NN-Schätzer über den realen Werten. Dies wird bei der Fehleranalyse durch den Vergleich der Schätzer mit den Referenzproben geprüft. Mit Hilfe des Bias, der die Treffgenauigkeit wiedergibt, steht ein Maß für die systematische Abweichung zwischen den Schätzern und den Referenzwerten zur Verfügung. Bei den Fehlerberechnungen der Aufnahme 1 liegt der Bias zwischen -7 und $-12 \text{ m}^2/\text{ha}$ und auch bei Verwendung aller Proben der drei Aufnahmen erreicht er noch Werte zwischen -2 und $-5 \text{ m}^2/\text{ha}$. Somit steht mit dem Bias ein Maß zur Kontrolle der Verteilung der Proben über den Merkmalsraum zur Verfügung. Bei einer optimalen Verteilung sollte der Bias bei Null liegen.

Die in der Untersuchung erzielten Genauigkeiten für die Grundfläche liegen zwischen 35 % und 67 % bei Verwendung der Landsat-Daten und zwischen 65 % und 67 % bei Hyperspektraldaten. Vergleicht man die mit den beiden gewählten Auflösungen erreichten Ergebnissen dieser Untersuchung mit Ergebnissen aus anderen Untersuchungen, ist eine differenzierte Betrachtung

ungsweise nötig. Die Einschätzung der Eignung von Landsat-TM-Daten zur Erkundung des Waldes und der dabei erzielbaren Genauigkeiten schwankt stark. Beispielsweise gibt SCHARDT (1998) in einer Abhandlung über die Eignung von Landsat-TM-Daten für die Erfassung forstlicher Parameter in Abhängigkeit vom Untersuchungsgebiet für bestimmte Klassenschemata der Artenzusammensetzung und der natürlichen Altersklasse eine Genauigkeit von ca. 85 % an. Dagegen findet sich in der umfassenden Rückschau von HOLMGREN & THURESSON (1998) eine weitaus pessimistischere Einschätzung der Nutzbarkeit multispektraler Satellitendaten für die forstliche Planung. HOLMGREN & THURESSON stellen unter anderem fest, dass die erreichbaren Genauigkeiten für verschiedene Merkmale im Allgemeinen zwischen 65 % und 85 % liegen. Die Genauigkeiten im Rahmen dieser Untersuchung liegen in diesem Bereich. Es ist davon auszugehen, dass mit einer Optimierung der terrestrischen Aufnahmen und einer optimalen radiometrischen und geometrischen Korrektur der Fernerkundungsdaten eine höhere Genauigkeit möglich ist.

Für das Merkmal *Totholz* liegen bei der Fehlerrechnung die Übereinstimmungen zwischen den k NN-Schätzern und den Referenzwerten zwischen 60,0 % und 73,3 % (Landsat) bzw. zwischen 60,0 % und 63,3 % (HyMapTM). Auf dieser Genauigkeitsbasis stellen die mit der k NN-Methode erstellten Totholzkarten eine gute Arbeitsgrundlage dar.

Bei der Anzahl der k -nächsten Nachbarn wurde für das Merkmal *Grundfläche* bei Verwendung der Landsat-Daten und Auswahl des kleinsten prozentualen RMSE (37,07 %; $r = 2$; $t = 2$) ein k -Wert von 15 ermittelt. Dies steht im Einklang mit anderen Untersuchungen, die 10 bis 15 nächste Nachbarn als Optimalwert herausfanden (FRANCO-LOPEZ et al. 2001, NILSSON 1997, TOKOLA et al. 1996, TOMPPU 1996). Bei Verwendung der HyMapTM-Daten und einem kleinsten erreichten %RMSE von 36,05 % ($r = 2$; $t = 2$) sind 5-nächste Nachbarn ausreichend. Die durch die spektrale wie räumliche höhere Auflösung vorhandene differen-

ziertere und genauere Informationsdichte scheint es zu ermöglichen, in der spektralen Information fast identische Nachbarn zu finden und somit mit wenigen Nachbarn auszukommen. Bei Betrachtung des Merkmals *Totholz* liegen sowohl für die Landsat-Daten wie für die Hyperspektraldaten die meisten Übereinstimmungen zwischen den kNN-Schätzern und den Referenzproben bei k -Werten zwischen 1 und 7.

Dies lässt den Schluss zu, dass durch eine höhere Übereinstimmung der spektralen Reflexion einzelner Nachbarn bei den kNN-Berechnungen mit wenigen k -nächsten Nachbarn genauere Ergebnisse erzielbar sind.

Bei der Distanzfunktion, die bei der kNN-Methode die spektrale Differenz zwischen zwei Pixel bestimmt, kann durch den Parameter r eine unterschiedliche Wichtung einzelner Kanäle erfolgen. Die erzielten Ergebnisse und die damit verbundene Gesamtdistanz bewegen sich in sehr engem Rahmen. Die maximale Differenz des minimalen und maximalen %RMSE für Landsat- wie Hyperspektraldaten beträgt 4,06 % (Grundfläche). Trotz des geringen Einflusses wird bei $r = 10$ eine Minimierung des Fehlers auf 36,63 % (Landsat; $k = 10$; $t = 2$) bzw. bei $r = 4$ auf 36,01 % (HyMapTM; $k = 5$; $t = 2$) erreicht. Mit einer Anhebung des r -Wertes fließen hohe spektrale Differenzen einzelner Kanäle mit einer größeren Wichtung in der Gesamtdistanz ein und dies führt zu einer Verschlechterung der Gesamtdistanz. Umgekehrt bedeutet dies, dass Pixel mit relativ homogenen spektralen Differenzen der einzelnen Kanäle bevorzugt werden. Dies führt bei der Gesamtbetrachtung der kNN-Methode zu genaueren Ergebnissen.

Mit der Einführung des Parameters t ist eine Einflussnahme auf die Gewichtung der als k -nächsten Nachbarn ausgewählten Pixel und ihrem zugeordneten Merkmalswert über die Distanz möglich. Je größer t gewählt wird, umso größer ist das Gewicht der Pixel mit geringer spektraler Distanz. Für die Landsat-Daten wurde ein Optimum des t -Wertes von 0,6 gefunden. Die den als k -nächsten Nachbarn ermittelten Pixeln zugeordneten Merkmalswerte fließen mit nur

geringem Unterschied, abhängig von der spektralen Distanz, in den Merkmalswertschätzer ein. Eine Heraushebung einzelner Pixel der k -nächsten Nachbarn wegen ihrer geringen spektralen Distanz führt zu einer Verschlechterung der kNN-Ergebnisse.

Die in dieser Untersuchung gewonnenen Erkenntnisse können als Richtwerte für den Einsatz der kNN-Methode gelten. Trotzdem sollte bei einem Einsatz der kNN-Methode in anderen Untersuchungsgebieten und bei Ermittlung anderer Merkmalswerte Probe-rechnungen durchgeführt werden, um für die jeweils gegebenen Bedingungen die Optimalwerte zu erreichen.

Dank

Diese Studie wurde im Rahmen des Projektes „Kombination von terrestrischen Aufnahmen und Fernerkundungsdaten mit Hilfe der kNN-Methode zur Kartierung von Waldökosystemen“ durchgeführt, das vom Bundesministerium für Bildung und Forschung und vom Deutschen Zentrum für Luft- und Raumfahrt initiiert und finanziell gefördert (FKZ 50EE0037) wurde. Wir danken Dr. LÜTTENBERG, DLR Bonn, einen Teil der Ergebnisse der Studie in diesem Rahmen darstellen zu dürfen. Für die Bereitstellung der Hyperspektraldaten im Rahmen des Vorhaben ProSmart danken wir der Infoterra GmbH und der Astrium GmbH, Friedrichshafen.

Literatur

- BORTZ, J., 1993: Statistik für Sozialwissenschaftler. – Berlin, Springer.
- DOWNES, G.M. & BARNARD, J.M., 2002: Clustering of very large datasets. – Sheffield, UK BCI Barnard Chemical Information Ltd.
- WSL (Hrsg.) 1999: Das Landesforstinventar. – In: Schweizerisches Landesforstinventar, Ergebnisse der Zweitaufnahme 1993–1995, Haupt Verlag, Bern.
- FAZAKAS, Z. & NILSSON, M., 1996: Volume and forest cover estimation over southern Sweden using – a VHRR data calibrated with TM data. – Remote Sensing Environment **17**: 1701–1709.

- FRANCO-LOPEZ, H., EK, A. R. & BAUER, M. E., 2001: Estimation and mapping of forest stand density, volume, and cover type using the k-nearest neighbors method. – *Remote Sensing of Environment* **77**: 251–274.
- HEHMKE, M., 2001: Untersuchung von Landschaftsveränderungen an der ehemaligen innerdeutschen Grenze mit Hilfe von Landsat TM-Daten. – Diplomarbeit am Lehrstuhl für Biometrie und Forstliche Informatik TU Dresden, 62 S.
- HESSENMÖLLER, D., 2002: Modelle zur Wachstums- und Durchforstungssimulation im Göttinger Kalkbuchenwald. – Dissertation, Forstwissenschaftlicher Fachbereich der Georg-August-Universität Göttingen.
- HESSENMÖLLER, D. & ELSENHANS, A.S., 2002: Zur Schätzung des Zuwachses bei Rotbuche *Fagus sylvatica* L. – Ein Vergleich parametrischer Verfahren mit der k-nearest neighbour Methode. – *Allg. Forst- u. J.-Ztg.*, 173. Jg., 11-12.
- HILDEBRANDT, G., 1996: Fernerkundung und Luftbildmessung. – Wichmann Verlag, Heidelberg.
- HOLMGREN, P. & THURESSON, T., 1998: Satellite remote sensing for forestry planning – a review. – *Scandinavian Journal Forest Research* **13**: 90–110.
- KILKKI, P. & PÄIVINEN, R., 1986: Weibull function in the estimation of the basal-area dbh-distribution. – *Silva Fennica* **20**: 149–156.
- KÖHL, M. & LAUTNER M., 2001: Erfassung von Waldökosystemen durch Hyperspektraldaten. – *Photogrammetrie Fernerkundung Geoinformation* **2001**(2): 107–117.
- KÖHL, M., STÜMER, W., SCHEUBER, M., ZIESE, H., HÄUSSLER, T. & LAUTNER M., 2000: Combining hyperspectral data and ground surveys for the assessment of non-wood goods and services of forests. – In: BUCHROITHNER, M.F.: A Decade of Trans-European Remote Sensing Cooperation. – Proceedings of the 20th EARSeI Symposium, Balkema Publishers, Tokyo.
- LEHMANN, F., PISCHEL, R. & WEWEL, F., 1999: Teilvorhaben HRSC-A. – In: ProSmart Endbericht, Daimler Chrysler Aerospace, Dornier Satellitensysteme, Friedrichshafen: 5/1 – 5/33.
- MALTAMO, M. & KANGAS, A., 1998: Methods based on k-nearest neighbor regression in the prediction of basal area diameter distribution. – *Canadian Journal Forest Research* **28**: 1107–1115.
- MITCHELL, T., 1997: Machine learning. – McGraw Hill.
- MUONEN, E. & TOKKOLA, T., 1990: An application of remote sensing for communal forest inventory. – In: The usability of remote sensing for forest inventory and planning. – Proceedings from the SNS/IUFRO workshop in Umea, 26–28 February, 1990, 35–42.
- MÜLLER, A., RICHTER, R. & HAUSOLD, A., 1999: Teilvorhaben HyMap. – In: ProSmart Endbericht, Daimler Chrysler Aerospace, Dornier Satellitensysteme, Friedrichshafen, 3/1 – 3/32.
- NIEMANN, H., 1983: Klassifikation von Mustern. – Springer-Verlag, Berlin.
- NILSSON, M. 1997: Estimation of forest variables using satellite image data and airborne lidar. – Doctoral thesis, Swedish University of Agricultural Sciences, Umea.
- QI, H., 2002: Feature Selection and kNN Fusion in Molecular Classification of Multiple Tumor Types. – Electrical and Computer Engineering Department, University of Tennessee.
- SCHARDT, M., 1998: Erfassung forstlicher Parameter mittels Landsat TM-Daten. – *AFZ – Der Wald* **24**: 1461–1463.
- TOMPPA, E., 1991: Satellite image-based national forest inventory of Finland. – *International Archives of Photogrammetry and Remote Sensing* **28** (7-1): 419–424.
- TOMPPA, E., 1993: Multi-source national forest inventory of Finland. – In: Proceedings of Ilvesalo symposium on National Forest Inventories. – The Finnish Forest Research Institute, Research papers 444: 52–59.
- TOMPPA, E., 1997a: Application of remote sensing in Finnish national forest inventory. – In: Application of remote sensing in european forest monitoring. – International workshop proceedings, Vienna, Austria: European Commission, CL-NA-17685-EN-C, 14–16 October 1996, 375–388.
- TOMPPA, E., 1997b: Recent status and further development of the Finnish multi-source forest inventory. – In: Managing the resources of the world's forests, Lectures given at the 1997 Marcus Wallenberg Prize Symposium, The Marcus Wallenberg Foundation Symposia Proceedings, Falun, Sweden, 53–69.
- TOMPPA, E. & PEKKARINEN, A., 1997: Methodenerprobung der Finnischen Nationalen Multiquellen-Waldinventur in Nordrhein-Westfalen. – In: Landeswaldinventur – Konzeption, Verfahrenstest, Ergebnisse, Schriftenreihe der Landesforstverwaltung Nordrhein-Westfalen Heft 5, 52–67.
- TUNCEL, E. & ROSE, K., 2002: Towards Optimal Data Clustering for Approximate Similarity Searching. – In: IEEE International Conference on Multimedia and Expo, Lausanne, Switzerland, August 2002.

VRIES, A. P. DE, MAMOULIS, N., NES, N. & KERSTEN, M., 2002: Efficient k-NN Search on vertically decomposed data. – ACM SIGMOD 2002, Madison, Wisconsin, USA.

WARFIELD, S., 1996: Fast k-NN Classification for Multichannel Image Data. – Pattern Recognition Letters, **17**: 713–721.

Anschriften der Autoren

Dr. WOLFGANG STÜMER

Technische Universität Dresden

Institut für Waldwachstum und Forstliche Informatik

Pienner Str. 18, D-01737 Tharandt

Tel.: +49 351 463 31620

Fax: +49 351 463 31632

e-mail: wolfgang.stuemer@forst.tu-dresden.de

Prof. Dr. MICHAEL KÖHL

Universität Hamburg

Fachbereich Biologie, Zentrum Holzwirtschaft,

Arbeitsbereich Weltforstwirtschaft

Leuschnerstr. 91, D-21031 Hamburg

Tel.: +49 40 7396 2100

Fax: +49 40 42891 2665

e-mail: m.koehl@holz.uni-hamburg.de

Manuskript eingereicht: Oktober 2004

Angenommen: November 2004