Aspekte der Segmentierung von Bildmaterial aus multispektralen Merkmalsräumen am Beispiel von Satellitendaten

T. Koch, T. Kubertschak, Dr.-Ing. K.-H. Franke

Zentrum für Bild- und Signalverarbeitung e.V. Ilmenau
Gustav-Kirchhoff-Straße 5
98693 Ilmenau
eMail:
torsten.koch@zbs-ilmenau.de,
tim.kubertschak@zbs-ilmenau.de,
karl-heinz.franke@zbs-ilmenau.de
URL: http://www.zbs-ilmenau.de

Zusammenfassung. Zur Erkennung und Bewertung von Eigenschaften der Landoberfläche werden seit geraumer Zeit satellitengestützte Sensoren eingesetzt, die simultan in mehreren spektralen Bereichen, Bildmoden und mitunter auch in unterschiedlicher Auflösung die Landoberfläche bildhaft erfassen.
Hauptzweck für die Datenerfassung sind Kartierungen diverser Eigenschaften (z.B. Biomasse, Feldfruchtarten,...) bzw. die Erkennung von Änderungen derselben. Ein hierfür wichtiger Verarbeitungsschritt ist eine Klassifikation des Bildmaterials. Diese kann pixelweise aber auch segmentbasiert erfolgen. Eine segmentbasierte Klassifikation bietet dabei den Vorteil, dass durch die Verarbeitung mittlerer Segmenteigenschaften modentypisch, stochastische Störeinflüsse, wie Rauschen oder Speckle, reduziert und damit die Klassifikationsgüte erhöht werden kann.
Im Rahmen des Verbundprojektes *ENVILAND-2*[1] entsteht dafür eine Verarbeitungspipeline, welche eine automatisierte Prozessierung von hochaufgelösten Satellitendaten (einschließlich Georeferenzierung, Registrierung multimodaler Daten, Segmentierung, automatische Ableitung von Trainingsgebieten, Klassifikation) realisieren soll.
In diesem Beitrag sollen Aspekte der hierfür eingesetzten mehrkanaligen Segmentierung näher beleuchtet und Teilergebnisse vorgestellt werden. Dies betrifft sowohl die synergetisch multimodale Verarbeitung (SAR- und optisches Bildmaterial) als auch die Verarbeitung in unterschiedlichen Auflösungen.

1 Einleitung

Segmentierungsverfahren sind in der Bildverarbeitung weit verbreitet. In den meisten Fällen werden Bilder auf Basis weniger qualitativ gleichartiger (z.B. drei RGB-) Bildmoden in homogene Regionen zerlegt.

[1] gefördert mit Mitteln des Bundesministeriums für Wirtschaft und Technologie (**BMWi**) durch das Deutsche Zentrum für Luft- und Raumfahrt (**DLR**) unter der Fördernummer 50 EE 0845;
weitere Partner: Jena-Optronik AG, FSU-Jena und ZFL-Bonn

In diesem Beitrag sollen jedoch Aspekte im Zusammenhang mit einer Segmentierung auf (prinzipiell) beliebig vielen Kanälen aufgezeigt werden, die sich hinsichtlich ihrer Modalität und Auflösung unterscheiden können. Das realisierte methodische Vorgehen ist hierbei auch auf den speziellen Anwendungsfall der Farbbildverarbeitung übertragbar.
Eine derartige Segmentierung wird im Projekt Enviland-2 dazu verwendet, mehrkanaliges Satellitenbildmaterial zu segmentieren, um darauf aufbauend eine Klassifikation zu betreiben. Dieses Vorgehen ermöglicht einerseits die synergetische Nutzung von Bildmaterial unterschiedlicher Modalität. Andererseits wird durch die integrativen Eigenschaften einer segmentbasierten Klassifikation im Folgeschritt der Einfluss diverser Störgrößen reduziert.
Diese Segmentierung ist Bestandteil einer automatisch ablaufenden Prozesskette (Abbildung 1), die neben der Klassifikation auch Vorverarbeitungsstufen und eine strukturbasierte Registrierung enthält.

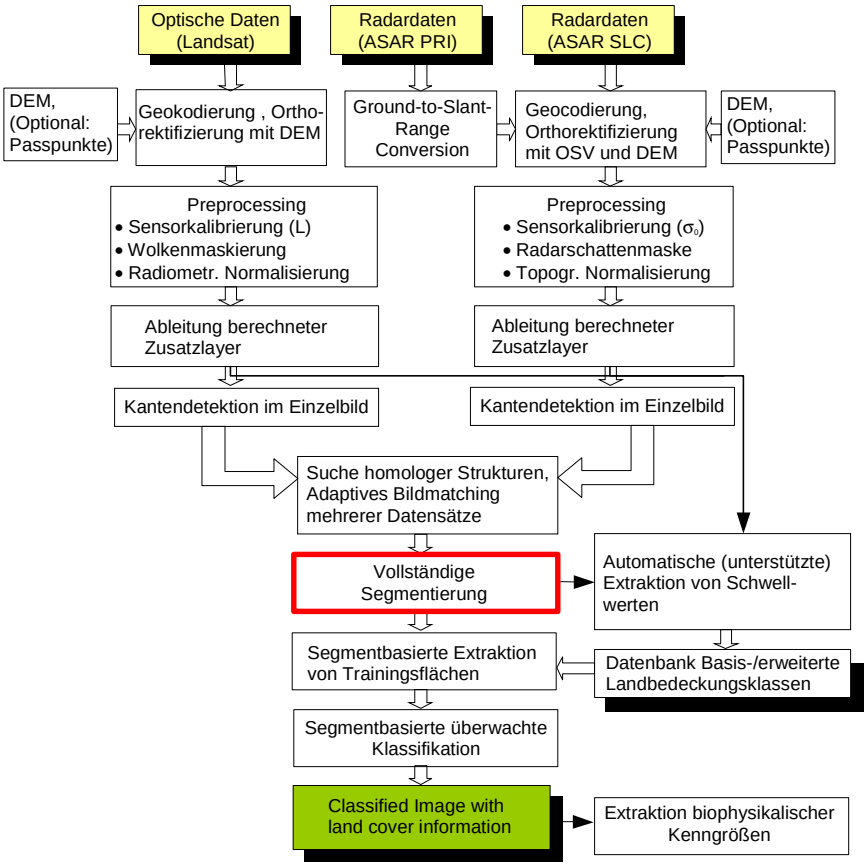


Abbildung 1 Prozesskette des Enviland-2-Projektes

Das mehrkanalige Bildmaterial kann sowohl von verschiedenen Satelliten bzw. Sensoren stammen, als auch zu verschiedenen Zeitpunkten aufgenommen worden sein.
Die Segmentierung erhält atmosphärisch korrigierte, entzerrte, registrierte und geocodierte Eingangsbilder. Insbesondere sind sensortypische Bildveränderungen bzw. -verzerrungen also (weitestgehend) eliminiert bzw. ausmaskiert. Bei SAR-Aufnahmen betrifft dies im Wesentlichen foreshortening-, layover- und Schatten-Effekte an Hangneigungen, während bei optischen Aufnahmen perspektivische Verzerrungen sowie atmosphärische Einflüsse (z.B. Luftfeuchtigkeit, Wolken) eine Rolle spielen.
Durch Berücksichtigung mehrerer Kanäle wird außerdem die Auswertung von Zeitserien ermöglicht. Dies wird beispielsweise für die Erkennung von Feldfruchtarten anhand mehrerer Beobachtungen während der Wachstumsphase angewandt.
Grundsätzlich liefert die Segmentierung eine "Kartierung" des Eingangsbildstapels (in Form von Labelbildern), die jeweils zusammenhängende Bildbereiche mit insgesamt geringer Varianz über alle Kanäle des Bildstapels darstellen.

2 Grundsätzliche Arbeitsweise der Segmentierung

Voraussetzung für die Segmentierung ist registriertes Bildmaterial (im GeoTIFF-Format). Das bedeutet, dass jedes Bild eine Transformationsvorschrift zwischen Pixel- und geografischen Koordinaten besitzt. Diese Transformationsvorschrift wird in der vorhergehenden Verarbeitungsstufe "adaptives Bildmatching" ermittelt [6]. Da diese strukturbasiert arbeitet, ist sie auch in der Lage, Bildmaterial unterschiedlicher Modalität miteinander zu registrieren.
Die Segmentierung selbst arbeitet in mehreren Stufen. Zunächst werden alle Eingangsbilder anhand ihrer Transformationsvorschriften und unter Zuhilfenahme eines geeigneten Resamplingverfahrens in ein gemeinsames Pixelkoordinatensystem transformiert. Dies betrifft entweder die gesamte Szene oder einen definierbaren (geographischen) Ausschnitt. Erst dadurch wird ein Vektorfeld (n-dimensionaler Bildstapel[i], bei Farbe: n=3) auf gemeinsamer Rasterbasis[ii] erzeugt. Denn Bilder, die zu unterschiedlichen Zeitpunkten oder von verschiedenen Satelliten aufgenommen wurden, haben prinzipiell eine unterschiedliche geografische Zuordnung/ Rasterbasis.)
Um noch vorhandene, sensorspezifische Störungen (Speckle, Rauschen) zu reduzieren, können daraufhin verschiedene Filterverfahren zur Preprozessierung angewandt werden. Die Auswahl des geeigneten Filterverfahrens sowie dessen Parametrierung erfolgt dabei in Abhängigkeit von der Modalität, dem Aufnahmezeitpunkt und dem Sensor-Band.
Auf Basis dieser gefilterten Eingangsbilder lassen sich noch weitere Merkmalsbilder ableiten, wodurch sich der Merkmalsraum entsprechend vergrößert. Zudem können die Eingangskanäle unterschiedlich gewichtet werden.
Im eigentlichen Segmentierungsprozess werden dann - auf der Grundlage der gesamten (mehrkanaligen) Auswertung des Eingangsbildstapels - Positionen für Saatpunkte (in möglichst homogenen Bereichen, nicht auf Kanten) für das Growing-Verfahren ermittelt und in einem iterativen Verfahren das Regionen-Wachstum durchgeführt.

$$u^2 = \frac{\sum_{\{c \in C\}} \min\left((p_c(x,y) - \mu_c)^2, d_{F_c}^2\right) \cdot \omega_c^2}{\sum_{\{c \in C\}} \omega_c^2}$$

Gleichung 1 Abstand u (C=Menge aller Kanäle; F_c=Menge aller Merkmale des Kanals c; $p_z(x,y)$=Pixelwert an Position (x,y) in Kanal/Merkmal z; μ_c=Mittelwert der bisher gewachsenen Region in Kanal c; ω_c=Gewicht von Kanal c; d_{Fc}=Abstand bezüglich der Merkmale F_c)

Alle lateral zusammenhängenden Punkte, welche im Merkmalsraum innerhalb eines definierten Abstandes u (Gleichung 1) liegen, werden dabei zu einer gemeinsamen Region zusammengefasst.
Benachbarte Regionen, deren integrale Merkmale den Abstandskriterien für die Bildung eines gemeinsamen Segmentes genügen, werden fusioniert. Jede dieser so gefundenen Regionen wird in einem Labelbild mit einem eindeutigen Wert (Segmentnummer) gekennzeichnet.
Aufgrund des erheblichen Datenumfangs bei Satellitenbildern werden für die Segmentierung Strategien eingesetzt, welche prinzipiell den Umgang mit Bildmaterial in beliebiger Größe und Anzahl ermöglichen.

3 Aspekte der mehrkanaligen Segmentierung

Die Motivation einer mehrkanaligen Segmentierung im gegebenen Kontext resultiert aus der nachfolgenden Klassifikation. Da diese mit möglichst vielen signifikanten und dekorrelierten Merkmalen bzw. Kanälen betrieben wird, sollte auch die Segmentierung auf Basis dieser Merkmale erfolgen, um möglichst homogene Bereiche zu erhalten.
Zur Veranschaulichung soll im Folgenden vom physikalischen Hintergrund ausgehend, der Vorteil einer mehrkanaligen Herangehensweise herausgearbeitet werden.
Der Signalwert F_n des Sensors für Band n ergibt sich gemäß Gleichung 2 aus der auf den Sensor eintreffenden Strahlung φ und der Sensorempfindlichkeit f (Abbildung 2). Wobei die spektrale Verteilung (Signatur) der Strahlung φ von der der Remission β des beobachteten Objektes, der atmosphärischen Transmission τ und der Bestrahlung S (durch die Sonne) abhängig ist.
Letztere wird im Folgenden als konstant angenommen. Die atmosphärische Transmission τ wurde bereits in einer vorhergehenden Stufe der Verarbeitungspipeline (Abbildung 1) eliminiert.

$$F_n = \int_{\lambda=0}^{\infty} f_n(\lambda) \cdot \varphi(\lambda) d\lambda \quad \text{mit } \varphi(\lambda) = \beta(\lambda) \cdot \tau(\lambda) \cdot S(\lambda)$$

Gleichung 2 Signalwert F_n des Sensors; mit f_n spektrale Empfindlichkeit des Sensors; φ spektrale Verteilung der auf den Sensor auftreffenden Strahlung [**2**]

Bei Berücksichtigung *nur eines Spektral-Bandes* kann es vorkommen, dass verschiedene Beobachtungsobjekte O_1 und O_2 mit jeweils unterschiedlicher spektraler

Signatur $\varphi_1(\lambda)$ und $\varphi_2(\lambda)$ einen gleichen Signalwert hervorrufen (Beobachtermetamerie). Damit sind diese Objekte - hinsichtlich einer Klassenzugehörigkeit - nicht unterscheidbar.
Um dieses Problem zu entschärfen, werden Beobachtungen meist in *mehreren Spektral-Bändern*[iii] vorgenommen. Unter solchen Bedingungen führen unterschiedliche spektrale Signaturen $\varphi_1(\lambda)$ und $\varphi_2(\lambda)$ meist zu unterschiedlichen Signalwerten $\underline{F}_1$ und $\underline{F}_2$. Die Objekte O_1 und O_2 sind damit unterscheidbar und können eindeutig verschiedenen Segmenten (und somit später auch verschiedenen Klassen) zugeordnet werden!

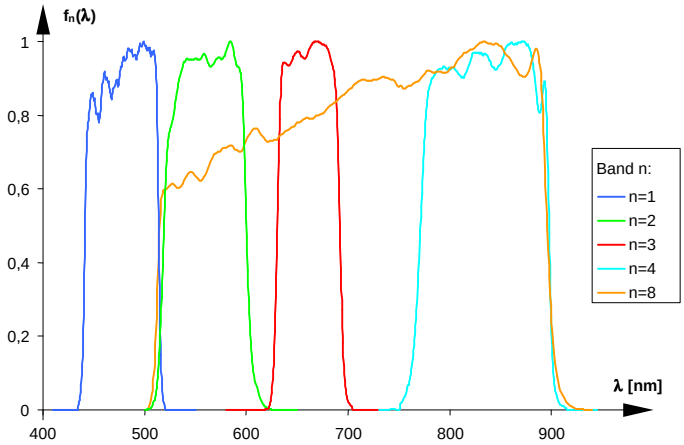
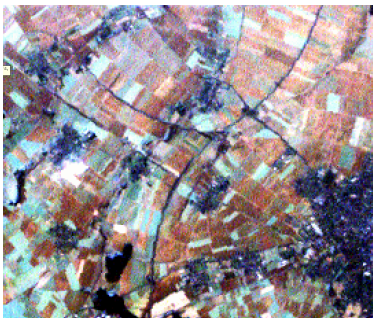
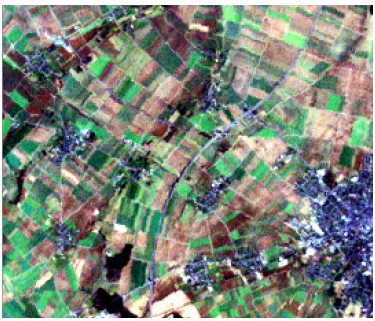


Abbildung 2 spektrale Sensorempfindlichkeiten ausgewählter Spektral-Bänder des ETM+-Sensors von Landsat 7 (Quelle: [1])

Es kann aber auch vorkommen, dass verschiedene Objekte O_1 und O_2 - in den betreffenden Spektralbereichen - sehr ähnliche spektrale Signaturen $\varphi_1(\lambda)$ und $\varphi_2(\lambda)$ hervorrufen. Daraus würden ebenfalls ähnliche Signalwerte $\underline{F}_1$ und $\underline{F}_2$ resultieren, was eine Unterscheidbarkeit von O_1 und O_2 erschwert.
Dieses Problem kann meist durch Berücksichtigung weiterer spektraler Bänder gelöst werden. Sind dennoch keine signifikanten Unterschiede erkennbar, müssen andere Strategien, wie die nachfolgend genannten, angewandt werden.
Eine bekannte Möglichkeit ist die Berücksichtigung der zeitlichen Änderung des spektralen Remissionsverhaltens $\beta(\lambda,t)$ der Beobachtungsobjekte. Bei diesen sog. multitemporalen Verarbeitungsprinzip werden zu anderen Zeitpunkten gewonnene Aufnahmen als zusätzliche Kanäle berücksichtigt. Die Veränderung der Remission zwischen den Aufnahmezeitpunkten t_1 und t_2 führt dann (meist) zu unterschiedlichen spektralen Signaturen $\varphi_1(\lambda,t_1)$, $\varphi_2(\lambda,t_2)$ und somit zu verschiedenen Signalwerten $\underline{F}_1$ und $\underline{F}_2$, wodurch eine Unterscheidung von O_1 und O_2 möglich wird.
Zeitliche Änderungen des Remissionsverhaltens sind meist bei biologischen Objekten gegeben. In Abbildung 3 ist besonders deutlich ist zwischen c und d der Unterschied durch blühende Rapsfelder zu erkennen.


(a)

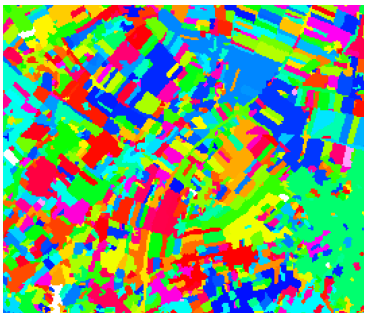

(b)

(c)

(d)

Abbildung 3 RGB-Komposit einer LS7-Szene in der Nähe von Euskirchen (vom 17.1.2001 (a), 6.3.2001 (b), 9.5.2001 (c), 25.5.2001 (d))

In Abbildung 4 sind Ergebnisse einer multitemporalen und einer monotemporalen Segmentierung gegenübergestellt. Beide Berechnungen erfolgten auf Landsat-7-Bildmaterial (pro Zeitpunkt jeweils Bänder 1-5 und 7). Die multitemporale Prozessierung erfolgte auf Basis von 9 verschiedenen Zeitpunkten[iv], während die monotemporale Ausführung nur auf Basis der Aufnahme vom 25.5. erfolgte. In Abbildung 3 sind RGB-Kompositionen (jeweils nur Bänder 3,2,1) einiger dieser Zeitpunkte dargestellt.


(a)

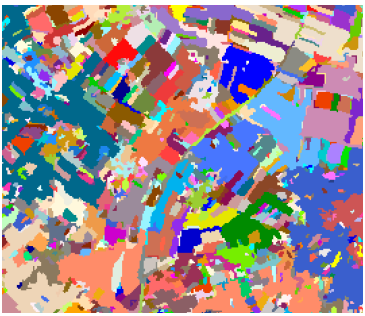

(b)

Abbildung 4 Segmentierungsergebnisse (LS-7, Bänder 1-5,7): (a) multitemporal, auf Basis von 9 Aufnahmezeitpunkten; (b) monotemporal, 25.05.2001

Die Ergebnisse zeigen bei multitemporaler Segmentierung eine deutlich höhere Detaillierung als bei monotemporaler Segmentierung.
Eine andere Möglichkeit zur Verbesserung der Diskriminanzfähigkeit einer segmentbasierten Klassifikation, ist die zusätzliche (multimodale) Verwendung von andersartigen Sensoren bzw. Aufnahmeprinzipien wie z.B. SAR[v] (siehe [4]).

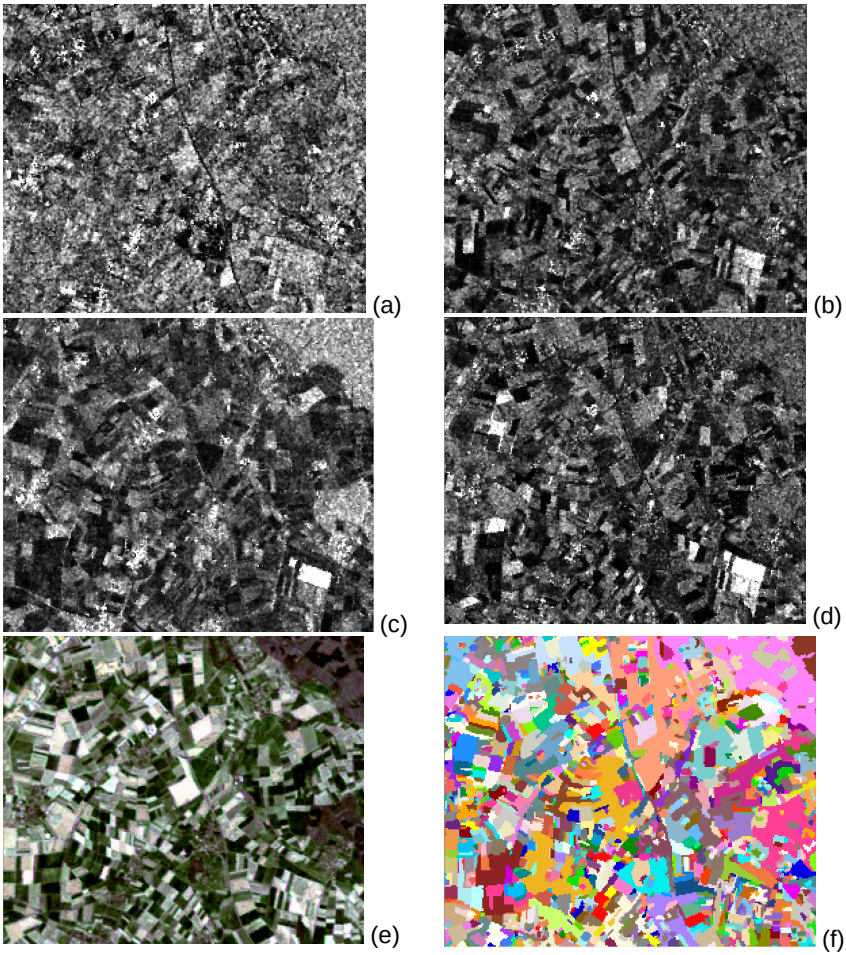


Abbildung 5 SAR- und optische Aufnahmen aus der Region zwischen Bonn und Euskirchen,2005; Envisat-ASAR vom: 04.08.(a), 30.06. (b), 26.05. (c), 21.04 (d); LS5-TM[vi] vom 03.04. (e); Segmentierungsergebnis (f)

Mit Hilfe von SAR-Beobachtungen können Eigenschaften abgeleitet werden, die durch optische Sensoren nicht erfassbar sind. Beispiele hierfür sind dielektrische Parameter (hauptsächlich beeinflusst durch den Feuchtigkeitsgehalt) und geome-

trische Parameter oder Rückstreuquerschnitt (beeinflusst durch Wuchshöhe, Blattform) [3].
Die Signifikanz, mit der diese Parameter aus dem Bildinhalt abgeleitet werden können, hängt in der Regel von Beobachtungswinkel, Polarisation (Abstrahlung und Empfang) und der Wellenlänge ab [5].
Durch die Unabhängigkeit von der natürlichen Beleuchtung durch die Sonne und der Wolkenbedeckung hat SAR das Potential, zeitliche Beobachtungslücken zu schließen, wodurch ein verlässlicherer Beobachtungszyklus realisiert werden kann.
In Abbildung 5 ist ein Szenario einer multimodalen Segmentierung (ausschnittsweise) aufgeführt. Es wurden eine Landsat-5-Szene (Bänder 1-5 und 7) sowie vier ASAR-Aufnahmen (VV-Polarisation) zusammen prozessiert. Die Größe der prozessierten Szene beträgt 800x1000 Pixel und wurde in 1:35 min verarbeitet[vii].
Eine nachfolgende segmentbasierte Klassifikation[viii] auf Grundlage dieses Segmentierungsergebnisses erbrachte eine Genauigkeit von ca. 89%.
Dabei kam exemplarisch ein SVM-Klassifikator zum Einsatz.

4 Fazit

Die bisher im Rahmen des Projektes Enviland-2 entwickelte Software zeigt, dass es möglich ist, durch Ausnutzung der Synergien von optischen und SAR-Bildmaterial, mehrkanaliges Bildmaterial derartig zu segmentieren, dass eine nachfolgende segmentbasierte Klassifikation qualitative Vorteile gegenüber der Verarbeitung einzelsensorischen Bildmaterials erzielt [4].
Aufgrund der angewandten bildinhaltsgetriebenen Suchstrategie für Saatpunkte (im Vorfeld des Region-Growings), konnte ein determinierter Verarbeitungsprozess realisiert werden. Hierdurch werden reproduzierbare Ergebnisse ermöglicht.
Die positiven Ergebnisse resultieren vor allem aus der verbesserten Klassentrennbarkeit (insbesondere durch die multitemporale und multimodale Verarbeitung) sowie der Reduktion des negativen Einflusses von Rausch- und Speckleeffekten aufgrund der integralen segmentspezifischen Merkmale.
Sowohl die Anzahl der Eingangskanäle als auch die Bildgrößen sind im Rahmen des vorgestellten Segmentierungsverfahrens prinzipiell nicht begrenzt.
Daraus resultiert die universelle Übertragbarkeit des methodischen Vorgehens auf andere Applikationen. Im Fall von herkömmlichen RGB-Bildmaterial, bedeutet dies lediglich eine Beschränkung auf die typischen Farbvalenzen von RGB-Aufnahmesystemen.

5 Quellen

[1] NASA-Webseite: http://landsathandbook.gsfc.nasa.gov/handbook/handbook_htmls/chapter8/chapter8.html

[2] K.-H. Franke: Bildverarbeitung I (WS 2004/2005); ISSN 1432-3346

[3] V. Heinzel (2006): “Synergetic use of optical and ERS-2 data for crop yield retrieval”, 2nd Workshop of the EARSeL Special Interest Group on Land Use and Land Cover 2006, Bonn.

[4] B. Waske, V. Heinzel, M. Braun, G. Menz: Classification of Segmented SAR and Multispectral Satellite Imagery using Support Vector Machines; 52. Internationales Wissenschaftliches Kolloquium, Proceedings, Vol. 2; Ilmenau, 2007; ISBN 978-3-939473-17-6

[5] V. Heinzel: Retrieval of biophysical parameters (winter wheat) using multi-sensoral and multi-temporal imagery

[6] Koch, T.; Franke, K.-H.: Aspekte der strukturbasierten Fusion multimodaler Satellitendaten und der Segmentierung fusionierter Bilder; Computer science meets automation: 52. IWK; proceedings; 10 - 13 September 2007, Technische Universität Ilmenau, Faculty of Computer Science and Automation; S. 269; ISBN 978-3-939473-17-6

---

[i] jedes Bild hat die gleiche Pixel-Höhe und -Breite

[ii] gleiche Pixelpositionen in unterschiedlichen Bildern (dieses Stapels) beziehen sich auf die gleiche geographische Position

[iii] also mit Sensoren unterschiedlicher spektraler Empfindlichkeit *f*

[iv] Aufnahmen vom: 05.07., 22.08., 10.11., 17.10., 06.03., 09.05., 25.05., 26.06., 29.08. (jeweils 2005)

[v] SAR: Synthetic Arperture Radar

[vi] dargestellt sind Bänder 1-3 als RGB-Tripel

[vii] Standard-PC (3 GHz, min. 1GB RAM)

[viii] Klassifikation in vier Grundklassen: (Wasser, Wald, Landwirtschaft, urbane Region)