52nd Internationales Wissenschaftliches Kolloquium
Technische Universität Ilmenau
10 – 13 September 2007

T. Koch / K.-H. Franke

Aspekte der strukturbasierten Fusion multimodaler Satellitendaten und der Segmentierung fusionierter Bilder

Einführung

Im Rahmen verschiedener Missionen zur Fernerkundung der Erdoberfläche entsteht multimodales und multitemporales Bildmaterial sowohl aus dem optischen Spektralbereich als auch Radar-Daten unterschiedlicher Polarität und Auflösung.
Diese Daten können zur Lösung vielfältiger Problemstellungen, wie z.B. die automatische Kartierung von Feldfruchtarten, die Ermittlung biophysikalischer Kenngrößen und die Erfassung der Änderungsdynamik, eingesetzt werden (s. [5],[6]).
Im Rahmen des Verbundprojektes ENVILAND[1] entstand ein Software-Prototyp, der diese Eingangsdaten durch eine Kette mehrerer Verarbeitungsstufen zu einem klassifizierten Labelbild verarbeitet. Wesentliche Teilaspekte[2] sind hierbei die Fusion des multimodalen Bildmaterials und dessen anschließende Segmentierung.
Dieser Beitrag soll einige Ergebnisse darstellen, die im Rahmen des Verbundprojektes ENVILAND bei dem *Zentrum für Bild- und Signalverarbeitung e.V. (ZBS)* entstanden sind.

Problemstellung

Das von den Herstellern (ESA,NASA,...) gelieferte Datenmaterial kann zwar (meist manuell) geocodiert bezogen werden, jedoch ist diese Geocodierung in der Regel nicht so genau, dass damit eine pixelgenaue Registrierung durchgeführt werden kann. Dieses Problem verschärft sich insbesondere bei Verwendung von Bildmaterial verschiedener Hersteller oder unterschiedlicher Modalität (SAR, VIS, IR, ...), da eine jeweils verschiedene Grundlage[3] zur Geocodierung eingesetzt wird.
Die Fusion ist in der vorgesehenen Verarbeitungspipeline nach einer Geocodierung

---

[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 0406;
weitere Partner: *Jena-Optronik AG*, *FSU-Jena*, *ZFL-Bonn* und *Desotron GmbH*
[2] die Verarbeitungsstufen *Kantenextraktion*, *Fusion* und *Segmentierung* wurden durch das ZBS erstellt
[3] bezüglich Vorverarbeitungs-Algorithmen, digitales Geländemodell, manuell gesetzte Passpunkte, ...

und anschließender Kantenextraktion eingegliedert. Sie soll das geocodierte (und orthorektifizierte[4]) Bildmaterial pixelgenau registrieren (Bestimmung von Translation, Skalierung und Rotation eines affinen Modellansatzes) und dadurch einen Bildstapel für die anschließende Segmentierung liefern.

Aufgabe dieser Segmentierung ist die Erzeugung einer Karte homogener Regionen, die dazu dient, einer anschließenden segmentbasierten Klassifikation Regionenmerkmale (inkl. abgeleiteter Texturmerkmale) zu liefern. Mit Hilfe dieser Klassifikation werden die gesuchten Klassenzugehörigkeiten (z.Z. vier Grundklassen) ermittelt.

Hauptproblem der Registrierung von Bildmaterial unterschiedlicher Modalität ist die Tatsache, dass die physikalische Ursache der Bilder und die verwendeten Sensoren völlig verschieden sind. Daraus resultiert eine geringe Korrelation zwischen Bildern unterschiedlicher Modalität. Aus diesem Grund sind herkömmliche grauwertbasierte Korrelationsverfahren ungeeignet und fehleranfällig (Abbildung 4, links).

Weiterhin sind unterschiedliche Auflösungen sowie Rauscheinflüsse (insb. bei SAR) zu bewältigen. Typische zu prozessierende Bildgrößen sind 12000x12000 Pixel! Daraus ergeben sich spezielle Erfordernisse sowohl an die Komplexität der eingesetzten Algorithmen als auch an das Speicherhandling.

Prinzip-bedingte (und durch die Geocodierung nur teilweise kompensierte) lokale Verzeichnungen sollen möglichst nur geringen störenden Einfluss zeigen. Diese lokalen Verzeichnungen resultieren aus abbildungstypischen Eigenschaften der eingesetzten Sensorik (z.B. bei Radar: foreshortening und perspektivische Verzerrungen bei optischen Abbildungssystemen) die im Zusammenhang mit unterschiedlichen Geländehöhen entstehen.

Weitere spezielle sensortypische Störungen sind Wolken (VIS) und Radar-Schatten (SAR). Diese sollen während der Geocodierung erkannt und ausmaskiert werden.

Eine spezielle Herausforderung sind urbane Regionen. Diese sind gekennzeichnet durch eine hohe Variabilität, was zu Problemen in der Fusion und Segmentierung[5] führt.

Lösungsansatz zur Bildfusion

Um das oben erwähnte Problem der geringen Korrelation zu lösen, müssen Strukturinformationen verwendet werden, welche in allen Modalitäten extrahierbar

[4] Herstellung eines homogenen Bezugs zwischen Pixelposition und geografischer Koordinate

[5] macht Einsatz geeigneter Texturfilter (wie z.B. NGLD [5]) oder spezieller Nachbearbeitung erforderlich

sind und in dem jeweils anderen Bildkanal mit hoher Wahrscheinlichkeit wiedergefunden werden können. Als eine solche gemeinsame Strukturinformation können die Übergangsbereiche (Kanten) an Segmentgrenzen benutzt werden. Um diese zu extrahieren, wird eine Filterung mit einem Deriche-Gradienten-Filter [4] durchgeführt. Eine *Non-Maximum*-Unterdrückung und die anschließende *Scan-Along*-Konturverfolgung (s. [1]) führen zur Kanteninformation.

In [3] wird ein Verfahren beschrieben, das eine kantenbasierte Korrelation durchführt. Es korreliert einzelne Kantenpunkte eines Bildkanals mit einem Invers-Distanz-Bild des jeweils anderen Kanals. Durch Variation der Transformationsparameter und anschließender Korrelation wird das beste Transformationsmodell ermittelt. Der Autor selbst schildert aber die „anziehende“ Wirkung stark strukturierter Regionen, die durch sehr hohe Korrelationswerte zur ungerechtfertigten Bevorzugung eben dieser Regionen führt. Die Ursache hierfür ist die glättende Eigenschaft der Invers-Distanz-Funktion, die über die Regionen verschmierte Korrelationswerte verursacht. Außerdem ist das Verfahren nicht determiniert, da nur zufällig gewählte Einzelpixel für die Korrelation Verwendung finden.

Um diese Probleme zu umgehen, wurde eine Alternative entwickelt, die auf dem in [3] beschriebenen Verfahren aufbaut.

Ausgangspunkt ist die Überlegung, dass Fehlkorrelationen hauptsächlich durch einander kreuzende statt durch deckungsgleich liegende Kanten gekennzeichnet sind. Deshalb wird im Gegensatz zu [3] nicht der Kantenort, sondern eine *kombinierte Betrags- und Richtungskorrelation* der jeweiligen Gradienten an den Kantenorten benutzt, die den Korrelationsbeitrag der Kanten minimiert, welche mit unterschiedlicher Orientierung überlagert würden.

Weiterhin werden jetzt alle Konturpixel in beiden Bildern berücksichtigt.

Das Neue an dem hier entwickelten Verfahren ist die mathematische Realisierung der Richtungs- bzw. Winkelkorrelation im Fourierbereich. Unter Zuhilfenahme einer schnellen Fouriertransformation (z.B. Radix-2, FFTW, o.ä.) wird dadurch eine Komplexität des Algorithmus von $O(N \cdot \log N)$ erreicht (s. [2]). (Bisherige Verfahren berechnen eine Winkelkorrelation im Ortsbereich und erzielen dadurch eine Komplexität von $O(N^2)$!) Erst dadurch wird es möglich, die geforderten Bildgrößen in angemessener Zeit zu verarbeiten.

Konkret wird der Umstand ausgenutzt, dass eine Kreuz-Korrelation (KKF) als Faltung eines Bildes mit einem rotierten Kern aufgefasst werden kann. Diese stellt im

Fourierbereich eine Multiplikation des fouriertransformierten Bildes mit dem konjugiert-komplexen, fouriertransformierten Kern dar (Gleichung 1; [1], ff. ME5, ME15).

$$C(dx,dy) = \sum_{\forall x}\sum_{\forall y} A(x,y)B(x+dx,y+dy) = A(x,y) * B(-x,-y)$$

$$C(dx,dy) = A(x,y) * B(-x,-y) \;\; \circ\!\!-\!\!\bullet \; F\{A(x,y)\} \cdot F^*\{B(x,y)\}$$

Gleichung 1 Äquivalenz der KKF im Fourierbereich; mit $F\{\bullet\}$ als Fouriertransformation; A,B als Quellbilder; C als Korrelationsbild

$$Q(x,y) = \frac{C(x,y) - lm_C(x,y)}{\min\limits_{\forall x,y} C(x,y)}$$

Gleichung 2 Gütemaß Q (lm_C ist der Mittelwert von C in einer lokalen Umgebung um (x,y))

Die Translationsparameter lassen sich durch Suche nach dem Optimum (hinsichtlich eines Gütemaßes (Gleichung 2)) im Korrelationsbild bestimmen. Die Skalierungs- und Rotationsparameter werden durch wiederholte Ausführung der Korrelation, bei sukzessiver Variation derselben Parameter und gezielter Auswahl der besten Korrelationsvariante ermittelt.

Wenn bereits eine initiale (aber ungenaue) Pre-Registrierung vorhanden ist, ist es meist nicht nötig, das komplette Korrelationsbild zu berechnen. Vielmehr ist ein kleiner Ausschnitt um die erwartete Position des Maximums von Interesse. Dieser kann bestimmt werden, indem beide Eingangsbilder in Kacheln zerlegt, jeweils zugehörige Kacheln der beiden Kanäle korreliert und die Korrelationsergebnisse anschließend aufakkumuliert werden. Hierbei ist aber zu beachten, dass eine einfache (nicht-überlappende) Kachelung eine unvollständige Korrelation in dem Sinne darstellt, dass Informationen in Randbereichen von Kacheln nicht mit denen der jeweiligen Nachbarkacheln in Beziehung gesetzt werden. Die Folge ist eine vom Zentrum des Korrelationsbildes zum Rand hin abnehmende Signifikanz der Korrelationswerte. Um dieses Problem zu lösen, wurde eine überlappende Kachelung eingesetzt, die in einer definierten Umgebung (um die Mitte des Korrelationsbildes) korrekte Werte ermittelt. Hierbei bestimmt die Breite der Überlappung die Größe des Signifikanzbereiches um das Zentrum des Korrelationsbildes. Die dabei erforderliche Kachelgröße kann cache-optimal realisiert werden.

Bewertung

In Abbildung 4 sind die Ergebnisse verschiedener Verfahren dargestellt. Der hier prozessierte Ausschnitt (Abbildung 1) hat eine Bildgröße von 512×512 Pixel.

Allgemein gilt, je ausgeprägter der Peak im Korrelationsbild ist, um so sicherer kann eine korrekte Korrelation bestimmt werden. Es ist deutlich zu erkennen, dass die grauwertbasierte Korrelation (Abbildung 4, links) keine sichere Bestimmung der

Translationsparameter zulässt. Das Verfahren nach [3] liefert dagegen schon ein deutlich besseres Ergebnis (Abbildung 4, mitte). Allerdings wurde dieses, nicht wie in [3] beschrieben, auf Basis einer Korrelation von Einzelpixeln, sondern durch Korrelation aller Kantenpixel berechnet. Nach Meinung des Autors stellt dies somit den optimalen Grenzfall dieses Verfahrens dar.

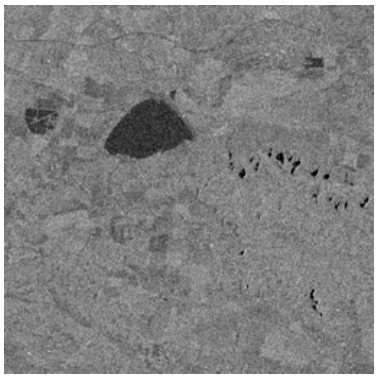
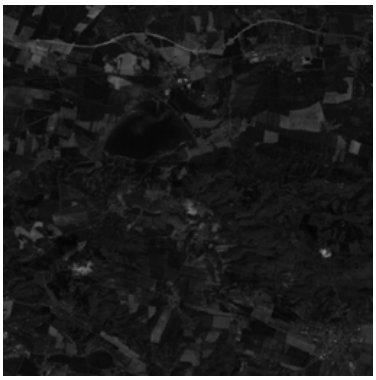
A)  B) 

Abbildung 1 Beispiel zweier zu registrierender Kanäle (Ausschnitt aus einer Szene bei Nordhausen (A: ASAR[6] VV-Polarisation B: Landsat-5, Kanal 3)

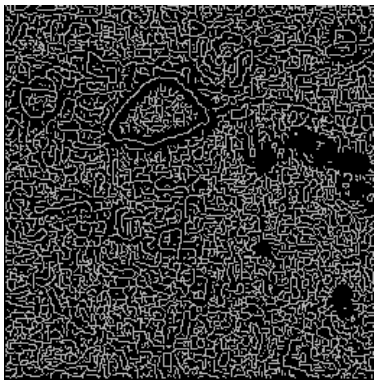
A)  B)

Abbildung 2 aus Originalbildern extrahierte Kantenbilder (links: Konturbild A, rechts: Konturbild B)

Das im Rahmen dieses Projektes entwickelte Korrelationsverfahren zeigt ein noch deutlicher ausgeprägtes Maximum (Abbildung 4, rechts), was auf eine generell sicherere Ableitung der Translationsparameter hindeutet.

Die Ausführungszeit[7] für die Bestimmung dieses Korrelationsbildes betrug weniger als 2 Sekunden. Wird die komplette Szene (korrelierter Ausschnitt: 4112×3292 Pixel) fusioniert, wird eine Ausführungszeit von ca. 45 Sekunden benötigt.

---

[6] die ENVISAT ASAR- und ERS-2-Daten wurden mit freundlicher Genehmigung der Europäischen Raumfahrt-Agentur (ESA) zur Verfügung gestellt (Category-1 Project C1P 3115)

[7] System: Pentium IV; 2,53 GHz; 1GB RAM / Kachelgröße jeweils 512x512 Pixel

Abbildung 3 Ausschnitt aus Konturbild A nach Bewertung mit Invers-Distanz-Funktion

$$D^{-1}(x,y) = \frac{1}{\left(1 + \frac{r(x,y)}{a}\right)^2}$$

Gleichung 3 Inverse-Distanz D^{-1}; mit r(x,y) als euklidischer Abstand von Pixel(x,y) zu nächsten Konturpixel ; a ist Steuer-Parameter

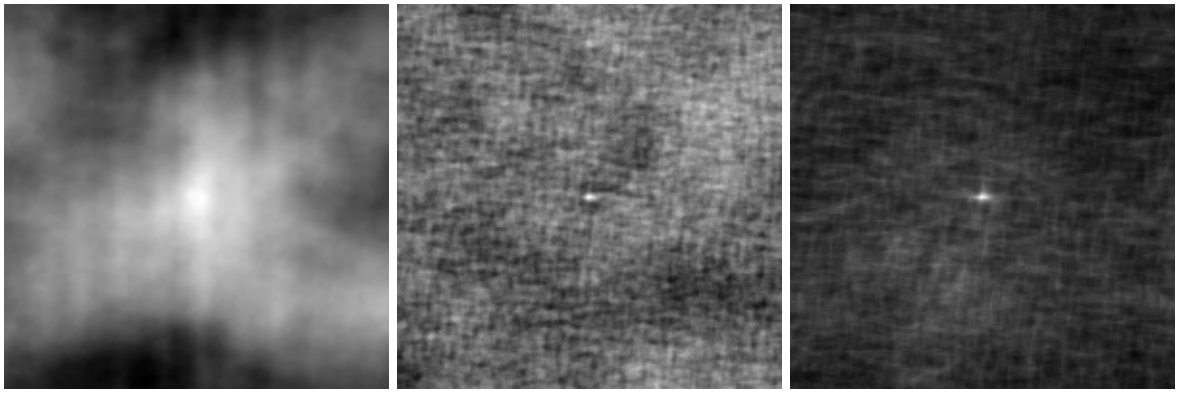


Abbildung 4 Korrelations-Ergebnisse (links: klassische Grauwert-Korrelation; mitte: Verfahren nach [3]; rechts: kombinierte Betrags- und Richtungskorrelation)

Literatur:

[1] K.-H. Franke: „Grundlagen der digitalen Bildverarbeitung und Mustererkennung“, Schriftenreihe des ZBS e.V., ISSN 1432-3346, Ilmenau, 2004

[2] Dieter Kreß, Ralf Irmer: „Angewandte Systemtheorie – Kontinuierliche und zeitdiskrete Verarbeitung“; Verlag Technik Berlin; 1989; ISBN 3-341-00693-1

[3] Axel Pinz, Manfred Prantl, Harald Ganster: „A Robust Affine Matching Algorithm Using an Exponentially Decreasing Distance Function“; Journal of Universal Computer Science, vol. 1. no. 8, pp. 614-631;1995

[4] R. Deriche, "Using Canny's criteria to derive an optimal edge detector recursively implemented," The International Journal on Computer Vision, Vol. 1, No. 2, April 1987, pp 167-187

[5] Riedel, T., C. Thiel, C. Schmullius & K.-H. Franke (2006):”Extraktion von Landbedeckungsinformationen - ein automatisierter Ansatz mittels multitemporalen C-Band Daten”; 26. Wissenschaftlich-Technische Jahrestagung der DGPF, Berlin-Adlershof, 11. -- 13 September, CD-ROM

[6] V. Heinzel, B. Waske, M. Braun, G. Menz (2005): "The potential of multitemporal and multisensoral remote sensing data for the extraction of biophysical parameters of wheat", SPIE Europe Bruges, Belgium, September, 2005, Proc. SPIE Vol. 5976, pp. 404-412

Autoren:

Torsten Koch
Karl-Heinz Franke
Zentrum für Bild- und Signalverarbeitung e.V.(ZBS), Gustav-Kirchhoff-Str. 5
98693 Ilmenau (Germany)
Phone: +49-3677-201030- 4 (Koch) / 0 (Franke)
Fax: +49-3677-2010301
E-mail: torsten.koch(at)zbs-ilmenau.de / karl-heinz.franke(at)zbs-ilmenau.de