Veröffentlichungen der DGK

Ausschuss Geodäsie der Bayerischen Akademie der Wissenschaften

Reihe C | Dissertationen | Heft Nr. 832

Thorsten Hoberg

# Conditional Random Fields zur Klassifikation multitemporaler Fernerkundungsdaten unterschiedlicher Auflösung

München 2019

Verlag der Bayerischen Akademie der Wissenschaften

ISSN 0065-5325 | ISBN 978-3-7696-5244-4

Diese Arbeit ist gleichzeitig veröffentlicht in:
Wissenschaftliche Arbeiten der Fachrichtung Geodäsie und Geoinformatik der Universität Hannover
ISSN 0174-1454, Nr. 349, Hannover 2018

Veröffentlichungen der DGK

Ausschuss Geodäsie der Bayerischen Akademie der Wissenschaften

Reihe C | Dissertationen | Heft Nr. 832

# Conditional Random Fields zur Klassifikation multitemporaler Fernerkundungsdaten unterschiedlicher Auflösung

Von der Fakultät für Bauingenieurwesen und Geodäsie

der Gottfried Wilhelm Leibniz Universität Hannover

zur Erlangung des Grades

Doktor-Ingenieur (Dr.-Ing.)

genehmigte Dissertation

Vorgelegt von

Dipl.-Ing. Thorsten Hoberg

Geboren am 23.03.1975 in Kassel

München 2019

Verlag der Bayerischen Akademie der Wissenschaften

ISSN 0065-5325 ISBN 978-3-7696-5244-4

Diese Arbeit ist gleichzeitig veröffentlicht in:
Wissenschaftliche Arbeiten der Fachrichtung Geodäsie und Geoinformatik der Universität Hannover
ISSN 0174-1454, Nr. 349, Hannover 2018

**Adresse der DGK:**

Ausschuss Geodäsie der Bayerischen Akademie der Wissenschaften (DGK)

Alfons-Goppel-Straße 11 ● D – 80 539 München

Telefon +49 – 331 – 288 1685 ● Telefax +49 – 331 – 288 1759

E-Mail post@dgk.badw.de ● http://www.dgk.badw.de

Prüfungskommission:

Vorsitzender: Prof. Dr.-Ing. Jakob Flury

Referent: Prof. Dr.-Ing. Christian Heipke

Korreferenten: Prof. Dr.-Ing. Monika Sester
Prof. Dr.-Ing. Stefan Hinz (KIT Karlsruhe)
Prof. Dr. techn. Franz Rottensteiner

Tag der mündlichen Prüfung: 14.09.2018

© 2019 Bayerische Akademie der Wissenschaften, München

Alle Rechte vorbehalten. Ohne Genehmigung der Herausgeber ist es auch nicht gestattet,
die Veröffentlichung oder Teile daraus auf photomechanischem Wege (Photokopie, Mikrokopie) zu vervielfältigen

ISSN 0065-5325 ISBN 978-3-7696-5244-4

# Zusammenfassung

Die Anzahl optischer Fernerkundungssensoren ist in den vergangenen Jahren kontinuierlich gestiegen. Insbesondere sind zunehmend Daten von hochauflösenden Satelliten und Satellitenkonstellationen mit einer geometrischen Auflösung von unter 5 Metern verfügbar. Mit der höheren Verfügbarkeit von Daten einer Region zu unterschiedlichen Zeitpunkten kann die Klassifikationsgenauigkeit von Landbedeckungs- bzw. Landnutzungsklassen sowie anderen erkennbaren Objekten der Erdoberfläche gesteigert werden. Darüber hinaus ist es möglich, Veränderungen schneller und sicherer zu erkennen. Allerdings sind die Anschaffungskosten für Daten hoher geometrischer Auflösung meist hoch und die zeitlichen Abstände, in denen ein ausgedehntes Gebiet vollständig erfasst wird, unter Umständen recht groß. Hingegen stehen Daten von Sensoren mit einer Bodenauflösung zwischen 5m und 30m oft zeitlich hochfrequent und für große Gebiete kostengünstig zur Verfügung.

In dieser Arbeit wird eine Methode zur automatischen Interpretation multitemporaler optischer Fernerkundungsdaten unterschiedlicher Auflösung vorgestellt. Die gemeinsame Auswertung dieser Daten hat mehrere Vorteile: eine hohe zeitliche Dichte verfügbarer Daten, überschaubare Kosten und die Möglichkeit, Merkmale und Objekte in unterschiedlichen Auflösungsstufen zu erfassen. Der neuartige Ansatz ermöglicht eine Steigerung der Klassifikationsgenauigkeit der Daten sämtlicher Zeitpunkte unabhängig von ihrer geometrischen Auflösung sowie die Erkennung von Veränderungen in einem Auswerteschritt. Dabei wird nicht nur erkannt, ob sich etwas verändert hat, sondern auch, wie sich etwas verändert hat. Für Daten unterschiedlicher Auflösungsstufen können unterschiedliche Klassenstrukturen definiert werden.

Die Methode basiert auf Conditional Random Fields (CRFs), einem probabilistischen Verfahren zur Mustererkennung. CRFs gehören zur Gruppe der Zufallsfelder (Random Fields). Diese zeichnen sich durch die Möglichkeit der Integration von Kontext für die Klassifikation einzelner Primitive aus. Im Unterschied zu zahlreichen anderen Arbeiten mit CRFs werden nicht ausschließlich räumliche Nachbarschaften modelliert. Zusätzlich werden unterschiedlich aufgelöste Bilder unterschiedlicher Zeitpunkte mit Übergangswahrscheinlichkeiten zwischen einzelnen Klassen verknüpft. Der CRF-Standardansatz wird hierzu um einen zusätzlichen Interaktionsterm erweitert, der zeitliche Abhängigkeiten modelliert.

Der Ansatz wurde in mehreren Experimenten für zwei großräumige Testgebiete evaluiert. Für beide Gebiete lagen mehrere Bilder unterschiedlicher Auflösung (Ikonos, RapidEye und Landsat) vor. Die alleinige Berücksichtigung räumlichen Kontexts führte für monotemporale Auswertungen im Vergleich zu einer Maximum-Likelihood-Klassifikation zu einer Verbesserung der Gesamtgenauigkeit von zumeist 5%-10%. Es wurden drei unterschiedliche Modelle für die räumlichen Interaktionen untersucht. Die Integration des zeitlichen Kontexts erhöhte die Klassifikationsgenauigkeit bei hochaufgelösten Bilddaten identischer Auflösung um weitere 1%-5%. In Kombination mit niedriger aufgelösten Daten wurde für die niedriger aufgelösten Daten eine Steigerung von teilweise über 10% erzielt. Testgebietsübergreifend wurde bei Experimenten zur Detektion von Veränderungen die Majorität der Pixel in 20 der 22 durch Simulation veränderten Flächen korrekt klassifiziert.

Schlagworte: Fernerkundung, Multitemporale Klassifikation, unterschiedliche Auflösung, Veränderungsdetektion, Zufallsfelder, Conditional Random Fields (CRF)

# Abstract

The number of optical remote sensing satellite systems is rapidly increasing. Especially high-resolution (HR) multispectral images from satellite systems at a Ground Sampling Distance (GSD) of 5 m or below, have become available at a higher rate. With the higher availability of these data, it is possible to improve the classification accuracy for landcover, landuse or other visible objects on earth surface. Moreover it is possible to analyze land-cover changes at a higher frequency than this is currently done based on a multitemporal analysis. However, multitemporal HR data is not always available and acquiring multitemporal HR data may not always be economically viable, particularly for large areas. Data having medium resolution (i.e., a GSD of 30 m) do not offer as much detail, but cover a larger area and may often be preferable from an economical point of view.

In this dissertation a method for the automatic interpretation of multitemporal optical remote sensing data having different geometrical resolutions is presented. The combined classification of HR images with data of lower resolution has several benefits: high temporal frequency of data, manageable costs and the option, to capture features and objects in different scales. This new approach not only improves the classification accuracy of all data of all individual epochs, but also allows the detection of land-cover changes in just one single step.

The new approach is based on Conditional Random Fields (CRFs), a probabilistic method for pattern recognition. CRFs belong to the family of Random Fields. In general Random Fields allow the integration of contextual knowledge for the classification of single primitives. In image analysis CRFs are frequently used for integrating spatial context in monotemporal classification. This framework is applied to multitemporal classification and change detection, taking into account interactions between images acquired at different epochs and having different geometrical resolutions. This is done by expanding the CRF model by temporal interaction terms that link neighboring epochs via transition probabilities between different classes.

Experiments were carried out for two different test sites in Germany, where Ikonos, RapidEye, and Landsat images were available. The integration of spatial context increased the overall accuracy of a monotemporal classification by 5%-10% in respect to a maximum-likelihood-classification. Three different models for the interaction potentials were evaluated in that way. The integration of temporal context increased the classification accuracy for another 1%-5% for HR images having identical spatial resolution. In combination with data of lower resolution, the classification accuracy of the images of lower resolution raised by over 10% in some experiments. Considering both test sides for experiments of change detection overall the majority of pixels in 20 of 22 changed regions (by simulation) were classified correctly.

Keywords: remote sensing, multitemporal classification, multiscale, change detection, random fields, Conditional Random Fields (CRF)

# Notation

Konvention: Variablen: *kursiv* ; Vektoren: **fett**

| | |
|---|---|
| $A_i$ | Assoziationspotential an der Stelle *i* |
| $b$ | Anzahl Merkmale |
| $C$ | Klasse |
| $d$ | Index der Merkmalsskala |
| $e$ | Kante |
| $\mathbf{E}$ | Menge aller Kanten |
| $E$ | Erwartungswert |
| $F_{all}$ | Menge aller extrahierten Merkmale |
| $F_{sel}$ | Menge aller selektierten Merkmale |
| $\mathbf{f}_i(\mathbf{y})$ | Merkmalsvektor für das Assoziationspotential der Stelle *i* |
| $\mathbf{G}$ | Graph |
| $\mathbf{h}_i(\mathbf{y})$ | Merkmalsvektor für das räumliche Interaktionspotential der Stelle *i* |
| $\mathbf{h}_i^t(\mathbf{y})$ | Merkmalsvektor für das räumliche Interaktionspotential der Stelle *i* zum Zeitpunkt *t* |
| $i$ | Primitiv |
| $I_{ij}$ | Interaktionspotential der Primitive *i* und *j* |
| $IS$ | Räumliches Interaktionspotential |
| $IT$ | Zeitliches Interaktionspotential |
| $j$ | Primitiv; räumlicher Nachbar von *i* |
| $k$ | Aufnahmezeitpunkt in der zeitlichen Nachbarschaft von *t* |
| $K_t$ | Menge der Zeitpunkte in zeitlicher Nachbarschaft von *t* |
| $l$ | Primitiv in zeitlicher Nachbarschaft zu Primitiv *i* |
| $L_i$ | Menge der Primitive in zeitlicher Nachbarschaft von *i* |
| $m$ | Komponente des Merkmalsvektors $\mathbf{h}_i^t(\mathbf{y}^t)$ |
| $M$ | Anzahl multitemporaler Bilder |
| $N_i$ | Menge der interagierenden Primitive in räumlicher Nachbarschaft von *i* |
| $PT$ | Zeitliches Potential |
| $R$ | Dimension des Merkmalsvektors $\mathbf{h}_i^t(\mathbf{y}^t)$ |
| $s$ | Auflösungsstufe |
| $S$ | Menge aller Primitive |
| $t$ | Aufnahmezeitpunkt |
| $T$ | Menge aller Aufnahmezeitpunkte |
| $\mathbf{TM}$ | Zeitliche Übergangsmatrix |
| $v$ | Knoten |
| V | Menge aller Knoten |
| $V$ | Varianz |
| $\mathbf{x}$ | Labelvektor |
| $x_i$ | Label der Stelle *i* |
| $\mathbf{y}$ | Daten |
| $\mathbf{y}_i$ | Daten der Stelle *i* |

| | |
|---|---|
| $Z$ | Partition Function |
| $\beta$ | Gewicht des räumlichen Interaktionspotentials |
| $\gamma$ | Gewicht des zeitlichen Interaktionspotentials |
| $\Delta$ | Abstand zur Berechnung der GLCM |
| $\eta$ | Glättungsparameter des räumlichen Interaktionspotentials |
| $\lambda$ | Skala der Merkmalsextraktion |
| $\boldsymbol{\mu}_{ij}^{t}(\mathbf{y}^{t})$ | Interaktionsmerkmale zwischen Primitiven *i* und *j* zum Zeitpunkt *t* |
| $\boldsymbol{\Sigma}$ | Kovarianzmatrix |

# Abkürzungen

| | |
|---|---|
| CFS | Correlation based feature selection |
| CNN | Convolutional Neural Network |
| CRF | Conditional Random Field |
| GLCM | Grey-level Co-occurrence Matrix |
| HMM | Hidden Markov Model |
| InSAR | Interferometric Synthetic Aperture Radar |
| LBP | Loopy Belief Propagation |
| LiDAR | Light Detection and Ranging |
| MRF | Markov Random Field |
| NDVI | Normalized Difference Vegetation Index |
| RNN | Recurrent Neural Network |
| RVI | Ratio Vegetation Index |
| SVM | Support Vector Machine |

# Inhaltsverzeichnis

1. EINLEITUNG ..... 13

**1.1. Motivation ..... 13**

**1.2. Zielsetzung ..... 16**

**1.3. Wissenschaftlicher Beitrag ..... 17**

**1.4. Gliederung der Arbeit ..... 18**

2. STAND DER WISSENSCHAFT ..... 20

**2.1. Klassifikation von Einzelbildern ..... 20**

**2.2. Multitemporale Klassifikation ..... 22**

2.2.1. Steigerung der Klassifikationsgenauigkeit ..... 22

2.2.2. Detektion von Veränderungen ..... 24

**2.3. Klassifikation von Bilddaten mit Zufallsfeldern ..... 26**

**2.4. Zufallsfelder in der Fernerkundung ..... 27**

**2.5. Diskussion ..... 29**

3. GRUNDLAGEN ..... 32

**3.1. Merkmale ..... 32**

3.1.1. Merkmalsextraktion ..... 32

3.1.2. Merkmalsselektion ..... 37

**3.2. Graphische Modelle für die Bildanalyse ..... 39**

3.2.1. Grundlagen der Wahrscheinlichkeitstheorie ..... 40

3.2.2. Graphische Modelle zur Modellierung von Wahrscheinlichkeiten ..... 42

3.2.3. Markov Random Fields ..... 43

3.2.4. Conditional Random Fields ..... 46

3.2.5. Inferenz ..... 47

**3.3. Bewertungskriterien ..... 50**

4. CRF ZUR KLASSIFIKATION MULTITEMPORALER BILDDATEN UNTERSCHIEDLICHER AUFLÖSUNG ..... 53

**4.1. Multitemporale Klassifikation von Bildern identischer Auflösung ..... 54**

**4.2. Multitemporale Klassifikation von Bildern unterschiedlicher Auflösung ..... 56**

**4.3. Assoziationspotential ... 57**

**4.4. Räumliches Interaktionspotential ... 58**

**4.5. Zeitliches Interaktionspotential ... 61**

5. EXPERIMENTELLE UNTERSUCHUNGEN ... 66

**5.1. Strategie ... 66**

**5.2. Daten und Klassen ... 68**

**5.3. Merkmalsextraktion ... 70**

**5.4. Merkmalsselektion ... 72**

**5.5. Maximale Merkmalsskala für das Assoziationspotential ... 75**

**5.6. Parameter des räumlichen Interaktionspotentials ... 78**

**5.7. Evaluation der Klassifikationsgenauigkeit feiner Strukturen ... 83**

**5.8. Multitemporale Klassifikation von Fernerkundungsdaten mit einheitlicher Auflösung ... 85**

**5.9. Multitemporale Klassifikation von Daten unterschiedlicher Auflösung ohne Veränderungen der Landbedeckung ... 89**

**5.10. Multitemporale Klassifikation von Daten unterschiedlicher Auflösung mit Veränderungen der Landbedeckung ... 95**

6. SCHLUSSBETRACHTUNG UND AUSBLICK ... 102

7. LITERATUR ... 106

*Knowledge without context is not knowledge;*
*it's just information.*
C. Otto Scharmer

# 1. Einleitung

## 1.1. Motivation

Fernerkundungsdaten ermöglichen den Blick aus der Luft auf Objekte der Erdoberfläche. Nicht immer ist es möglich, diese Objekte korrekt zu erkennen. Hierzu drei kleine Beispiele:

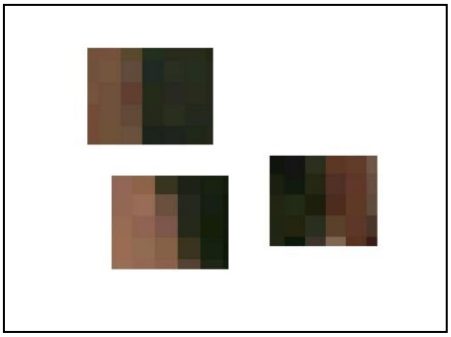


*Abbildung 1-1: Was zeigen die Bildausschnitte?*

*Abbildung 1-2: Handelt es sich bei den dunkelgrünen Flächen um Acker oder Wald?*

*Abbildung 1-3: Ist die Landbedeckung in den markierten Gebieten identisch?*

*Abbildung 1-4: Antwort zu Abbildung 1-1: Die Farbwerte der Bildausschnitte sind ähnlich. Durch Nutzung räumlichen Kontextwissens können zwei benachbarte Ackerflächen (Bild oben), ein Sportplatz (Bild unten links) sowie eine Dachfläche nebst benachbarter Grünfläche (Bild unten rechts) erkannt werden.*

*Abbildung 1-5: Antwort zu Abbildung 1-2: Einzig bei den dunkelgrünen Flächen zwischen den Siedlungsteilen handelt es sich um Wald, die übrigen Flächen sind bestellte Äcker. Durch Nutzung zeitlichen Kontextwissens unter Berücksichtigung von Daten eines zweiten Aufnahmezeitpunktes (rechtes Bild) können diese Flächen korrekt erkannt werden.*

*Abbildung 1-6: Antwort zu Abbildung 1-3: Bei den markierten Gebieten handelt es sich um Siedlungen und Ackerflächen. Die Unterscheidung ist aufgrund der niedrigen geometrischen Bildauflösung schwer möglich. Durch Nutzung von Kontextwissen aus einer höher aufgelösten Szene (rechtes Bild) können die Flächen korrekt erkannt werden.*

Die drei Beispiele verdeutlichen, dass für Menschen Kontextwissen für die Erkennbarkeit von Objekten hilfreich ist. Ist diese Beobachtung auf die automatisierte Auswertung von Fernerkundungsdaten übertragbar?

---

Aktuelle Kartierungen der Erdoberfläche dienen als Basis für vielfältige kurz-, mittel- und langfristige Planungen, z.B. im Bereich der Infrastruktur, der Ökologie oder zur Einsatzkoordination in Krisensituationen. Speziell für sehr große oder nur schwer zugängliche Gebiete werden hierzu häufig Daten von Fernerkundungssatelliten genutzt, da diese schnell und flächendeckend verfügbar sind. In den vergangenen zwei Jahrzehnten ist die Anzahl optischer Fernerkundungssensoren stetig gestiegen. Insbesondere sind zunehmend Daten von hochauflösenden Satelliten und Satellitenkonstellationen mit einer geometrischen Auflösung von unter 5 Metern, z.B. Ikonos, Quickbird, WorldView, RapidEye oder SKYSAT, um nur einige zu nennen, verfügbar. Die höhere Verfügbarkeit von Daten einer Region zu unterschiedlichen Zeitpunkten eröffnet neue Möglichkeiten der Informationsgewinnung: Verändert sich die Landschaft, kann die hochfrequente Datenerfassung für ein zeitnahes Erkennen dieser Veränderungen genutzt werden. Verändert sich die Landschaft nicht, kann durch Kombination der Daten die Klassifikationsgenauigkeit erhöht werden.

Die verschiedenen optischen Fernerkundungssensoren verfügen über unterschiedliche radiometrische und geometrische Auflösungen sowie eine unterschiedliche Anzahl erfassbarer Spektralbereiche. Geometrisch sehr hochaufgelöste Sensoren erreichen eine Bodenauflösung von unter 0.5m pro Pixel im panchromatischen Kanal und unter 1m pro Pixel in den multispektralen Kanälen. Die meisten Sensoren zeichnen neben panchromatischer Information in höherer Auflösung auch Farbinformationen im Bereich des sichtbaren Lichts sowie im Wellenlängenbereich des nahen und/oder mittleren Infrarots auf. Darüber hinaus existieren Hyperspektralsensoren, die Daten in hunderten Wellenlängenbereichen aufzeichnen. Die radiometrische Auflösung liegt im Allgemeinen bei mindestens 8bit pro Kanal.

Mit der zunehmenden Verfügbarkeit von Daten hoher geometrischer Auflösung (< 5m) ergeben sich erweiterte Möglichkeiten der Auswertung von Fernerkundungsdaten: Auch kleine Objekte, die in Daten mittlerer Auflösung nicht erkennbar sind, können erkannt und kartiert werden. Allerdings sind die Anschaffungskosten für Daten hoher geometrischer Auflösung sehr hoch und die zeitlichen Abstände, in denen ein Gebiet erfasst wird, unter Umständen recht groß. Eine Auswertung ausschließlich basierend auf diesen Daten wäre folglich zwar sehr detailliert, ist jedoch häufig aufgrund eingeschränkter Datenverfügbarkeit nicht realisierbar und insbesondere für größere Gebiete sehr kostenintensiv.

Neben den hochaufgelösten Sensoren gibt es eine Vielzahl von Sensoren mit einer Bodenauflösung zwischen 5m und 30m. Der Informationsgehalt der mittels dieser Sensoren gewonnenen Daten ist zwar geringer, allerdings ist deren Beschaffung kostengünstiger. Zudem stehen die Daten oft zeitlich hochfrequent zur Verfügung.

Methoden zur gemeinsamen Auswertung beliebiger Daten unterschiedlicher Aufnahmezeitpunkte ermöglichen es, die enthaltenen Informationen ergänzend zu nutzen. Die Entwicklung dieser Methoden galt vor einem Jahrzehnt als eine der größten Herausforderungen der Fernerkundung (Richards, 2005). Auch wenn seitdem Fortschritte erzielt wurden, ist die Aufgabe unverändert aktuell, und die Relevanz nimmt mit der Vielzahl unterschiedlicher Datenquellen weiterhin zu. Eine Kombination unterschiedlich

aufgelöster Daten zur multitemporalen Klassifikation verspricht mehrere Vorteile: Eine hohe zeitliche Dichte verfügbarer Daten, überschaubare Kosten und die Möglichkeit, Merkmale und Objekte in unterschiedlichen Auflösungsstufen zu erfassen. Zudem ist es sinnvoll, Methoden für die gemeinsame Auswertung von Daten mehrerer Sensoren zu etablieren, da der einzelne Sensor nur für einen endlichen Zeitraum aktiv ist.

Der Vorteil einer gemeinsamen Analyse multitemporaler Daten im Vergleich zu einer monotemporalen Betrachtung liegt in der Möglichkeit, zeitliches Kontextwissen zu integrieren (Abbildung 1-5 und Abbildung 1-6). Die Einbindung dieses Wissens verspricht eine höhere Klassifikationsgenauigkeit vor allem in Fällen, in denen eine Entscheidung für eine Klasse zu einem einzelnen Zeitpunkt nicht zweifelsfrei getroffen werden kann. Verbesserte Ergebnisse sind darüber hinaus durch die Integration von Modellwissen, z.B. Prognosen über wahrscheinliche Veränderungen zwischen Zeitpunkten, zu erwarten. Ein Beispiel hierfür ist, dass in Mitteleuropa die Umwandlung von Ackerflächen zu bebauten Flächen wahrscheinlicher ist als die Umwandlung von bebauten Flächen zu Ackerflächen.

Zusätzlich zur Nutzung zeitlichen Kontextwissens kann eine Klassifikation von Bilddaten durch Betrachtung räumlichen Kontexts verbessert werden (Abbildung 1-4). Es können typische nachbarschaftliche Beziehungen in die Auswertung einbezogen werden, z.B. dass Autos im Regelfall auf versiegelten Flächen stehen und mit sehr geringer Wahrscheinlichkeit komplett von Wasser umgeben sind oder dass benachbarte Primitive mit hoher Wahrscheinlichkeit identischen Klassen zugehören. Ein Ansatz zur bestmöglichen Analyse multitemporaler Daten integriert folglich sowohl zeitliche als auch räumliche Kontextinformation.

Eine derartige auf Methoden der Bildanalyse und Mustererkennung basierende Analyse multitemporaler Daten kann als Grundlage für Planungen unterschiedlicher Fachdisziplinen dienen. Potentielle Anwendungsgebiete sind sowohl in Bezug auf ihre räumliche Ausdehnung als auch den ausgewerteten wie auch den zukünftig relevanten Zeitraum breit gefächert. Beispiele sind:

- die Dokumentation von Veränderungen der Landnutzung, z.B. durch Abholzung von Waldgebieten oder durch ungeplante Ausdehnung von Siedlungen,
- der Nachweis von Umweltveränderungen, z.B. durch das Abschmelzen von Gletschern oder Waldsterben,
- die Überwachung jahreszeitliche Phänomene, z.B. des Pflanzenwachstums, oder
- die Unterstützung von Handlungsentscheidungen in Krisensituationen, z.B. infolge Tsunamis, Waldbränden oder Erdbeben.

## 1.2. Zielsetzung

Ziel dieser Arbeit war die Entwicklung einer Methode zur automatischen Interpretation multitemporaler optischer Fernerkundungsdaten unterschiedlicher Auflösung. In erster Linie sollten niedriger aufgelöste Daten unter Berücksichtigung der Interpretationsergebnisse höher aufgelöster Daten besser klassifiziert werden. Die Methode sollte darüber hinaus in der Lage sein, veränderte Gebiete zu detektieren und die Art der Veränderung zu identifizieren.

Methodisch sollte dieses Ziel durch die gleichzeitige Berücksichtigung von räumlichem und zeitlichem Kontextwissen erreicht werden. Aktuell ist der Standard bei der Auswertung von Fernerkundungsdaten das Trainieren eines lokalen Klassifikators. Dieser trifft aufgrund extrahierter Merkmale an der Position einzelner Primitive (Pixel oder Segmente) eine Entscheidung über die Zugehörigkeit zu einer Klasse. Auch wenn die Merkmale in manchen Ansätzen in lokalen Umgebungen extrahiert werden, interagieren die einzelnen Primitive meist nicht miteinander. Die maximal erreichbare Klassifikationsgenauigkeit ist hierdurch begrenzt. In vielen Szenen können einzelne Pixel hingegen bei Betrachtung des Kontexts mit höherer Sicherheit klassifiziert werden (Torralba et al., 2010).

Die im Rahmen dieser Arbeit entwickelte Methode wurde vorrangig für die Auswertung von Fernerkundungsdaten mittlerer bis hoher Auflösung entworfen und auf die in dieser Arbeit verwendeten Daten und Objektklassen optimiert. Dabei wurde die Methode so allgemein gehalten, dass sie relativ leicht auf andere Datenbestände, andere Objektarten und andere Anwendungen übertragen werden kann.

Generell können in Daten unterschiedlicher Auflösungsstufen unterschiedliche Objekte und Objektdetails identifiziert werden. Dieses Wissen wird im Idealfall für die Definition zur Klassifikation geeigneter Objektklassen genutzt. Somit ist es für die zuvor beschriebene Anwendung erforderlich, dass sich eine Klasse einer niedrigeren Auflösungsstufe in mehrere Teilklassen einer höheren Auflösungsstufe aufspalten lässt. Umgekehrt soll eine Fusion von Klassen in einer höheren Auflösungsstufe zu übergeordneten Klassen in einer niedrigeren Auflösungsstufe möglich sein.

Eine vollständige Interpretation des Gebietes zu einem frühen Zeitpunkt, z.B. repräsentiert in einem GIS, sollte nicht Vorrausetzung für die Anwendbarkeit der in dieser Arbeit entwickelten Methode sein. Die Methode ist daher vor allem auch für die zahlreichen Regionen geeignet, für die ein solcher Datenbestand nicht existiert oder nur mit geringer Zuverlässigkeit zu erzeugen ist.

## 1.3. Wissenschaftlicher Beitrag

In dieser Arbeit wird eine Methode zur simultanen Klassifikation multitemporaler optischer Fernerkundungsdaten unterschiedlicher Auflösung vorgestellt. Die Methode basiert auf Conditional Random Fields (CRF), einem probabilistischem Verfahren zur Mustererkennung. CRF gehören zur Gruppe der Zufallsfelder (Random Fields). Allgemein zeichnen sich Zufallsfelder durch die Möglichkeit der Integration von Kontextwissen für die Klassifikation einzelner Primitive aus. CRF stellen eine Erweiterung der populären Markov Random Fields (MRF) dar (Kapitel 3.2.3) und erlauben eine flexiblere Auswahl der Merkmale sowie eine Berücksichtigung von Merkmalen bei der Formulierung von Kontextwissen. Der wissenschaftliche Beitrag dieser Arbeit erstreckt sich auf folgende Aspekte:

- Im Bereich der Bildanalyse wurden Zufallsfelder bislang fast ausschließlich zur Modellierung räumlicher Abhängigkeiten eingesetzt. Einzelne Arbeiten generieren zusätzliches Wissen durch Verknüpfung unterschiedlicher Auflösungsstufen eines identischen Bildes. Die Modellierung von zeitlichen Abhängigkeiten erfolgt häufig über Markov-Ketten. Im Unterschied zu Zufallsfeldern sind diese allerdings nicht in der Lage, gleichzeitig räumliches Kontextwissen explizit zu integrieren. In dieser Arbeit werden verschiedene, unterschiedlich aufgelöste Bilder einer beliebigen Anzahl von Zeitpunkten erstmals durch Zufallsfelder verknüpft. Der Standardansatz

wird hierzu um einen zusätzlichen Interaktionsterm erweitert, der zeitliche Abhängigkeiten modelliert.

- Die Methode ermöglicht eine Steigerung der Klassifikationsgenauigkeit der Daten sämtlicher Zeitpunkte unabhängig von ihrer geometrischen Auflösung sowie die Erkennung von Veränderungen in einem Auswerteschritt. Dabei wird nicht nur erkannt, ob sich etwas verändert hat, sondern auch, wie sich etwas verändert hat. Für Daten unterschiedlicher Auflösungsstufen können unterschiedliche Klassenstrukturen definiert werden. In bisherigen Arbeiten wurden die Steigerung der Klassifikationsgenauigkeit und die Veränderungsdetektion stets getrennt behandelt.

- Der Algorithmus der entwickelten Methode beinhaltet drei unterschiedliche Terme. Der erste Term gibt die Möglichkeit der Zugehörigkeit eines Primitivs zu einer Klasse unabhängig von deren Nachbarn an. Der zweite und dritte Term modellieren das räumliche und zeitliche Kontextwissen. Die Terme können verschiedene Merkmale berücksichtigen, wobei die Merkmale in unterschiedlich großen Umgebungen, im Folgenden Skalen genannt, extrahiert werden können. Im Rahmen dieser Arbeit werden die besten Merkmale in den jeweiligen Skalen selektiert und der Einfluss der Verwendung unterschiedlicher maximaler Skalen für die einzelnen Terme auf das Klassifikationsergebnis systematisch evaluiert.

- Vergleiche zwischen MRF und CRF erfolgen in der Literatur zumeist anhand der Standardkriterien Gesamtklassifikationsgenauigkeit und Kappa-Koeffizient. Diese Kriterien sind jedoch je nach Anwendung nur beschränkt aussagekräftig. Anhand einer qualitativen Evaluation kleiner Bildstrukturen werden Differenzen der Klassifikationsergebnisse bei ausschließlicher Verwendung von Labels (MRF) zu den Ergebnissen bei der zusätzlichen Berücksichtigung von Merkmalen (CRF) für die räumlichen Interaktionen aufgezeigt.

## 1.4. Gliederung der Arbeit

Kern der Arbeit ist die Entwicklung eines neuartigen kontextbasierten Ansatzes zur Klassifikation multitemporaler Bilddaten unterschiedlicher Auflösung. Dieser wird in Kapitel 4 dargestellt und in Kapitel 5 evaluiert.

Vorab wird in Kapitel 2 der Stand der Wissenschaft aufgezeigt. Die Betrachtung erfolgt dabei aus zwei Blickwinkeln. Mit Blick auf die Anwendung werden Grundlagen der Klassifikation von Fernerkundungsdaten, insbesondere für multitemporale Auswertungen, vorgestellt. Aus methodischer Sicht werden bedeutende Arbeiten zu Zufallsfeldern mit Fokus auf Conditional Random Fields hervorgehoben.

In Kapitel 3 werden die für die Methode und deren Anwendung benötigten Grundlagen dargelegt. Die in dieser Arbeit verwendeten Merkmale sowie eine Methode zur Selektion der zur Klassifikation bestmöglich geeigneten Merkmale werden vorgestellt. Es folgt ein Überblick über graphische Modelle zur Bildanalyse, insbesondere auf den Spezialfall Conditional Random Fields. Abschließend werden die zur Auswertung der Experimente verwendeten Beurteilungskriterien beleuchtet.

Aufbauend auf den Grundlagen wird in Kapitel 4 der kontextbasierte Ansatz zur Klassifikation multitemporaler Bilddaten unterschiedlicher Auflösung entwickelt. Bei diesem auf CRF basierenden Ansatz wird für die Bestimmung der wahrscheinlichsten Konfiguration der Zuordnung von Klassen zu den einzelnen Primitiven eine Energiefunktion minimiert. Möglichkeiten der Modellierung der einzelnen Energieterme, namentlich des Assoziationspotentials sowie des räumlichen und zeitlichen Interaktionspotentials, werden im weiteren Verlauf des Kapitels beschrieben.

In Kapitel 5 folgen umfangreiche experimentelle Untersuchungen. Zunächst wird die Strategie der Experimente erläutert und die untersuchten Daten sowie die Klassendefinitionen beschrieben. In einem der Evaluation des kontextbasierten Ansatzes vorgeschalteten Auswerteschritt werden zur Klassifikation geeignete Merkmale extrahiert und eine Untermenge selektiert. Die kontextbasierten Wirkungen der einzelnen Interaktionspotentiale werden im Folgenden mit zunehmender Komplexität evaluiert und die Ergebnisse diskutiert. Zunächst liegt der Fokus ausschließlich auf dem räumlichen Interaktionspotential ohne zeitliche Interaktionen. Es folgt eine Betrachtung der multitemporalen Klassifikation von Fernerkundungsdaten mit einheitlichem Maßstab. Zuletzt erfolgt die Evaluation von multitemporalen Klassifikationen von Daten unterschiedlicher Auflösung ohne sowie mit Veränderungen.

Im letzten Kapitel dieser Arbeit werden die wesentlichen Erkenntnisse in einer Schlussbetrachtung zusammengeführt und ein Blick auf künftige Herausforderungen geworfen.

# 2. Stand der Wissenschaft

Ausgangspunkt für diese Arbeit waren zwei praktische Fragestellungen: Kann die Klassifikationsgenauigkeit von Satellitendaten durch die Kombination mehrerer Aufnahmen unterschiedlicher Sensoren zu unterschiedlichen Zeitpunkten gesteigert werden? Ist es gleichzeitig möglich, Veränderungen zu erkennen? Im ersten Teil dieses Kapitels wird der Stand der Wissenschaft zu diesen Fragestellungen aus Sicht der Anwendung betrachtet. Allgemein werden die Grundlagen der Klassifikation von Fernerkundungsdaten dargestellt, zunächst für die Auswertung von Einzelbildern, im Folgenden für multitemporale Auswertungen.

Die zur Lösung der obigen Fragestellungen entwickelte Methode basiert auf Conditional Random Fields. Im zweiten Teil dieses Kapitels wird der Stand der Wissenschaft zur Auswertung von Bilddaten mit Zufallsfeldern aufgezeigt. Zufallsfelder erfreuen sich in den letzten Jahren zunehmender Beliebtheit. Nach einem Überblick über grundlegende wegweisende Arbeiten zu Zufallsfeldern liegt der Fokus auf Arbeiten zur Klassifikation von Fernerkundungsdaten mit Zufallsfeldern.

Abschließend werden in Kapitel 2.5 die Grenzen der existierenden Arbeiten diskutiert und daraus die Anforderungen und Rahmenbedingungen für die neue Methodik entwickelt.

## 2.1. Klassifikation von Einzelbildern

Ziel einer Klassifikation ist die Zuweisung von unterschiedlichen Klassen zu den auszuwertenden Primitiven. Primitive können Pixel oder regelmäßige, z.B. quadratische, wie auch unregelmäßige Segmente sein. Die Wahl ist unter Berücksichtigung der Anwendung und der einzusetzenden Methode zu treffen. Die Definition der zu unterscheidenden Klassen hängt sowohl von den Szeneninhalten als auch von der Auflösung ab. So können z.B. bei hochaufgelösten Fernerkundungsdaten die Klassen Haus, Straße, Baum und Grünfläche unterschieden werden. Hingegen ist bei Daten mittlerer Auflösung die Unterscheidung in die übergeordneten Klassen Siedlung, Ackerland und Wald sinnvoll. Die Erkennbarkeit von Objekten ist allerdings nicht ausschließlich von deren Größe und dem räumlichen Auflösungsvermögen des Sensors abhängig. Es spielen auch die Objekteigenschaften und der Kontrast zur Umgebung eine wesentliche Rolle, so dass kein eindeutiges Verhältnis zwischen Objektgröße und Bildauflösung für die sichere Erkennbarkeit eines Objektes angegeben werden kann (Albertz, 2009).

Die Klassifikation von Fernerkundungsdaten ist ein komplexer Prozess, der u.a. die Auswahl von Trainingsgebieten, die Bildvorverarbeitung, die Merkmalsextraktion, die Wahl eines geeigneten Klassifikationsverfahrens, die Nachbearbeitung und eine Genauigkeitsbetrachtung beinhaltet (Lu und Weng, 2007; Wang et al., 2016). Einige dieser Aspekte werden für den konkreten Fall im weiteren Verlauf dieser Arbeit in Grundzügen dargestellt. Eine eingehende Betrachtung sämtlicher Aspekte übersteigt den Rahmen dieser Arbeit. Der Fokus liegt an dieser Stelle auf den Klassifikationsverfahren. Diese lassen sich nach unterschiedlichen Kriterien (Lu und Weng, 2007) einteilen. Im Folgenden werden einige dieser Kriterien aufgezeigt. Auf eine ausführliche Vorstellung einzelner Klassifikationsverfahren wird verzichtet, die Grundlagen finden sich in zahlreichen Lehrbüchern (u.a. Albertz, 2009; Richards, 2013; Mather und Tso, 2016).

Überwachte vs. unüberwachte Klassifikationsverfahren: Für eine überwachte Klassifikation müssen die zu bestimmenden Klassen zunächst definiert werden und ausreichend Trainingsgebiete für die Klassen vorhanden sein. Anhand dieser Trainingsgebiete werden die Charakteristika der einzelnen Klassen gelernt. Jedes zu klassifizierende Primitiv wird in einer überwachten Klassifikation der Klasse mit der größten Ähnlichkeit zugewiesen. Populäre Verfahren sind Maximum-Likelihood oder Neuronale Netzwerke. Hingegen sind bei unüberwachten Klassifikationen die Klassen zunächst unbekannt. Der Klassifikator gruppiert die Merkmalsvektoren der Primitive im Merkmalsraum und erzeugt auf diese Weise unterschiedliche Klassen (z.B. K-means-Clustering). Die Bedeutung dieser Klassen wird abschließend durch manuelle Interpretation oder automatische Zuordnungsverfahren zugewiesen.

Parametrische vs. nichtparametrische Klassifikationsverfahren: Bei parametrischen Verfahren liegt eine Annahme über die Verteilung der Daten zugrunde. Zumeist wird eine Normalverteilung der Daten angenommen. Die zu bestimmenden Parameter sind in diesem Fall z.B. der Mittelwert und die Kovarianzmatrix, die für eine Maximum-Likelihood-Klassifikation genutzt werden können. Ein Vergleich mehrerer parametrischer Verfahren, sowohl pixelbasiert als auch unter Einbeziehung von räumlichem Kontext, findet sich bei Hubert-Moy et al. (2001). Nichtparametrische Verfahren kommen ohne eine solche Annahme aus. Zu dieser Gruppe gehören u.a. Neuronale Netzwerke, Support Vector Machines (einen ausführlichen Überblick geben Mountrakis et al. (2011)) oder Expertensysteme. Insbesondere bei einer sehr großen Anzahl von Merkmalen, z.B. bei der Auswertung von Hyperspektraldaten, sind nichtparametrische Verfahren zu bevorzugen (Camps-Valls et al., 2014).

Pixelbasierte vs. Subpixel- vs. segmentorientierte Klassifikationsverfahren: Die Auswertung von Fernerkundungsdaten erfolgt zumeist auf Pixelebene. Nahezu sämtliche populären Klassifikationsverfahren können auf Pixelebene angewandt werden. Pixelbasierte Verfahren führen zu guten Ergebnissen, wenn die Spektralwerte einer Klasse und die daraus abgeleiteten Merkmale homogen sind. Ist dies nicht der Fall, führt dies zu verrauschten Ergebnissen. Benachbarte Pixel werden fehlerhaft unterschiedlichen Klassen zugewiesen, man spricht von einem „Salz-und-Pfeffer-Effekt“. Insbesondere bei Daten mittlerer oder niedriger Auflösung resultieren die Pixelwerte aus einer Vermischung der Reflexionen unterschiedlicher Materialien (Lu et al., 2011). Das Ziel von Subpixel-Klassifikationsverfahren ist die spektrale Entmischung. Hierdurch lassen sich die Pixelwerte als Kombination sogenannter Endmember aufspalten. Diese Endmember geben den Anteil der Materialien (und somit häufig der Objekte) an der räumlichen Position des Pixels an. Segmentorientierte Klassifikationsverfahren erfordern zunächst eine Bildsegmentierung. Diese Segmente repräsentieren im Idealfall Objekte oder werden zu Objekten zusammengefügt. Ein Nachteil dieses Ansatzes ist, dass sich Segmentierungsfehler unmittelbar auf das Klassifikationsergebnis auswirken. Alternativ können GIS-Daten zur Definition der Segmente verwendet werden (Albert et al., 2017).

Lokale vs. kontextbasierte Klassifikationsverfahren: Die Entscheidung für die Zugehörigkeit eines Primitivs zu einer Klasse erfolgt bei lokalen Klassifikationsverfahren ausschließlich basierend auf Merkmalen, die aus dem Erscheinungsbild des Primitivs abgeleitet wurden. Insbesondere bei pixelbasierten Verfahren führt dies oft zum „Salz-und-Pfeffer-Effekt“. Aus diesem Grund werden häufig Techniken zur Nachbearbeitung eingesetzt. Diese können z.B. Filtermethoden sein, die das Ergebnis glätten, jedoch oft zum Verlust wertvoller Informationen führen. Andere Ansätze nutzen Expertenwissen, um die Ergebnisse in einem regelbasierten System zu verbessern. Durch Nutzung kontextbasierter Verfahren kann dem „Salz-und-Pfeffer-Effekt“ entgegen gewirkt werden (Hubert-Moy

et al. 2001). Bei diesen Verfahren interagieren Primitive, in den meisten Fällen räumlich benachbarte Primitive, miteinander. Durch diesen Informationsaustausch kann zusätzliches Wissen berücksichtigt werden, z.B., dass benachbarte Primitive identischen Klassen zugehören oder eine Kombination bestimmter benachbarter Klassen sehr wahrscheinlich ist. Zu den kontextbasierten Verfahren zählen die in dieser Arbeit verwendeten Zufallsfelder. Ausführliche Betrachtungen zur Nutzung von Kontextwissen finden sich u.a. in Li et al. (2014), Schindler (2012) sowie Kumar und Hebert (2006).

Monoskalare vs. multiskalare Klassifikationsverfahren: Abhängig von der räumlichen Auflösung der Fernerkundungsdaten sind unterschiedliche Objekte oder deren Objektteile in den Bildern sichtbar, interpretierbar und detektierbar. Das ist die Motivation für multiskalare Klassifikationsverfahren. In der Literatur wird der Begriff multiskalare Klassifikation für zwei grundlegend unterschiedliche Ansätze verwendet. Im ersten Fall wird ein zu analysierendes Bild in unterschiedliche Skalenebenen überführt (Kumar und Hebert, 2006; Moser et al., 2009), z.B. durch eine Filterung des Originalbildes. Zwar geht damit ein Verlust an Information einher, es treten jedoch in den unterschiedlichen Skalenebenen unterschiedliche Objektcharakteristika hervor. In den jeweiligen Skalenebenen werden Merkmale extrahiert, die gemeinsam mit den in anderen Skalenebenen extrahierten Merkmalen zur Auswertung genutzt werden. Im zweiten Fall handelt es sich um Verfahren, in denen verschiedene Datenquellen, die sich in Auflösung und Informationsgehalt unterscheiden, gemeinsam ausgewertet werden. In der englischsprachigen Literatur wird hierfür auch der Begriff „multisource“ verwendet. In vielen Fällen haben die Daten unterschiedliche Aufnahmezeitpunkte, weshalb die Arbeiten gleichfalls der Gruppe der multitemporalen Klassifikationsverfahren zugehörig sind.

## 2.2. Multitemporale Klassifikation

Für eine multitemporale Klassifikation müssen mindestens zwei Bilder einer identischen Region vorliegen. Alternativ kann zu einem Zeitpunkt ein GIS-Datenbestand einbezogen werden. Multitemporale Klassifikationsverfahren verfolgen zwei wesentliche Ziele:

- Steigerung der Klassifikationsgenauigkeit
- Detektion von Veränderungen zwischen den Zeitpunkten

Die Ziele werden im Regelfall getrennt voneinander betrachtet, so dass Veröffentlichungen zur multitemporalen Klassifikation zumeist eindeutig einer Gruppe zugeordnet werden können. Liegt der Fokus auf der Steigerung der Klassifikationsgenauigkeit, müssen die Bilder identische und unveränderte Objekte enthalten. In den unterschiedlichen Bildern prägen sich diese jedoch oft mit veränderter radiometrischer Erscheinung aus. Dieser zusätzliche Informationsgehalt hilft, die Genauigkeit von Klassifikationen zu steigern. Enthalten die Bilder unterschiedliche oder veränderte Objekte, kommen Verfahren zur Detektion von Veränderungen zum Einsatz. Die Herausforderung besteht darin, sowohl das Auftreten als auch die Art der Veränderungen zu erkennen (Radke et al., 2005).

### 2.2.1. Steigerung der Klassifikationsgenauigkeit

Einzelne Objektklassen, insbesondere Vegetationsklassen, sind in Fernerkundungsdaten eines Zeitpunktes häufig schwer zu unterscheiden. Ihre Reflexionseigenschaften können ähnlich sein, die

erfassten radiometrischen Informationen somit auch. Hingegen heben sich zu anderen Zeitpunkten die Klassen stark voneinander ab (siehe zweites Beispiel der Motivation, Kapitel 1.1). Da zu unterschiedlichen Zeitpunkten häufig abweichende unterschiedliche Klassen ähnliche Eigenschaften haben, ist in vielen Fällen eine Unterscheidung sämtlicher Klassen zu nur einem Zeitpunkt schwer möglich. Durch eine Auswertung multitemporaler Daten kann dieses Problem verringert werden.

Methodisch gibt es drei populäre Ansätze (Richards, 2013):

- Sämtliche Merkmale aller Zeitpunkte eines Primitivs werden in einem Vektor zusammengeführt. Dieser Vektor wird analog zur Auswertung eines Einzelbildes zur Klassifikation genutzt.
- Die Daten der unterschiedlichen Zeitpunkte werden unabhängig voneinander mit einem oder mehreren Verfahren klassifiziert. Weichen die Klassifikationsergebnisse voneinander ab, erfolgt die Entscheidung für eine Klasse durch ein vordefiniertes Kriterium (decision maker).
- Sämtliche Informationen werden in ein wissensbasiertes System integriert.

Die Ansätze werden bevorzugt für die Klassifikation multitemporaler Daten aus einer Vegetationsperiode genutzt. Insbesondere steht die Erkennung unterschiedlicher Feldfrüchte im Fokus der Forschung. Vieira et al. (2000) nutzen Daten mehrerer Sensoren mit unterschiedlicher Auflösung. Die Trainingsdaten werden durch Punkte in einem dreidimensionalen Raum mit den Achsen Zeit, Wellenlänge und Reflexion repräsentiert. Für jede Klasse wird eine Fläche durch diese Punkte geschätzt. Die Koeffizienten der Parametrisierung der Fläche werden im Folgenden als Merkmale genutzt. Hoberg und Müller (2011) nutzen CRF zur Erkennung von Feldfrüchten in hochaufgelösten optischen Fernerkundungsdaten. Es werden zwei Ansätze beschrieben: Zum einen werden sämtliche Merkmale in einem Vektor zusammengeführt, zum anderen wird ein multitemporales CRF (Kapitel 4.1) aufgebaut, in dem die benachbarten Primitive der einzelnen Zeitpunkte interagieren. Durch die Integration räumlichen Kontextwissens liefern beide Ansätze bessere Klassifikationsgenauigkeiten als eine SVM-Klassifikation. Leite et al. (2011) nutzen Hidden Markov Models zur Erkennung von Feldfrüchten in zwölf Landsat-Aufnahmen und verdeutlichen den Vorteil einer multitemporalen Auswertung. Während in Einzelbildaufnahmen eine Gesamtgenauigkeit von ca. 60% erreicht wird, kann diese durch Auswertung von mindestens sechs Szenen auf über 90% gesteigert werden. Rußwurm und Körner (2018) unterscheiden 17 Vegetationsklassen in einer Zeitreihe von Sentinel 2 Daten mit Recurrent Neural Networks (RNN). Der Ansatz erfordert keine explizite Formulierung von Merkmalen und erlaubt die Existenz einzelner bewölkter Bilder innerhalb der Zeitreihe. RNN sind eine Weiterentwicklung der populären Convolutional Neural Networks (CNN) für die Analyse von Datensequenzen. Im Gegensatz zu CNN, in denen eine Informationsverarbeitung zwischen den Neuronen von der Eingabe- zur Ausgabeschicht erfolgt, existiert in RNN eine Rückkopplung. Interdonato et al. (2019) nutzen gleichfalls RNN zur Klassifikation von 13 Landnutzungsklassen in Sentinel 2 Daten von über 30 Zeitpunkten.

Die Anwendung dieser Methoden ist nicht auf Daten optischer Fernerkundungsdaten beschränkt und kann auf Daten beliebiger Sensoren (z.B. SAR-Daten: Waske und Braun, 2009; Tavakkoli Sabour, 2010) oder eine Kombination von Daten unterschiedlicher Sensoren übertragen werden.

Einen ausführlichen Überblick über unterschiedliche Anwendungen und Verfahren zur Klassifikation multitemporaler Fernerkundungsdaten unterschiedlicher Sensoren geben Gòmez-Chova et al. (2015). Es

zeigt sich, dass kein ideales Verfahren für sämtliche Anwendungen existiert. Das Verfahren sollte stets anhand der Aufgabenstellung und der zur Verfügung stehenden Daten gewählt werden.

### 2.2.2. Detektion von Veränderungen

Eine Auswertung von Einzelbildern kann stets nur die Landbedeckung bzw. Landnutzung zu einem Zeitpunkt beschreiben. Für zahlreiche Fragestellungen der Planung, des Monitorings oder des Krisenmanagements interessieren jedoch vorrangig die Veränderungsprozesse. Für einige Fragestellungen ist es ausreichend, die sich verändernden Gebiete zu identifizieren. Für andere Fragestellungen wird Wissen über die Art der Veränderung benötigt. Die Detektion von Veränderungen (change detection) der Landbedeckung oder Landnutzung ist immer dann möglich, wenn sich diese in Veränderungen der Radiometrie auswirken. Allerdings können radiometrische Veränderungen auch durch weitere Faktoren, z.B. unterschiedliche atmosphärische Bedingungen, unterschiedliche Sonnenstände oder unterschiedliche Vegetationsphasen hervorgerufen werden (Mas, 1999). Diese Einflüsse lassen sich in begrenztem Maß reduzieren, z.B. mit der Auswahl von Daten, die zum selben Jahreszeitpunkt aufgenommen wurden (Singh, 1989).

Die Erkennung von Veränderungen ist Thema zahlreicher wissenschaftlicher Arbeiten, die an dieser Stelle nicht einzeln vorgestellt werden können. Einen Überblick über die ersten Arbeiten zur automatischen Erkennung von Veränderungen gibt Singh (1989). Einen sehr ausführlichen qualitativen Überblick geben Lu et al. (2004). Weitere umfangreiche Zusammenstellungen geben Jianya et al. (2008) sowie Bruzzone und Bovolo (2013). Sämtliche vorgenannten Arbeiten haben den Fokus auf Fernerkundungsaufgaben. Allgemein geben Radke et al. (2005) einen ausführlichen Überblick über Methoden zur Detektion von Veränderungen aus allen Bereichen der Bildanalyse. Darüber hinaus können Veränderungen nicht nur aus dem Vergleich mehrerer Bilder, sondern auch durch Abgleich von Bildern mit GIS-Daten erkannt werden (Li, 2010).

Vor dem Hintergrund der Aufgabenstellung dieser Arbeit konzentriert sich die weitere Betrachtung auf die multitemporale Klassifikation zur Erkennung von Veränderungen der Landbedeckung bzw. Landnutzung aus optischen Fernerkundungsdaten. Diese Verfahren lassen sich in drei Kategorien einteilen (Jianya et al., 2008):

- Vergleich zweier Datenquellen (Pixel oder Merkmale)
- Vergleich extrahierter Informationen
- Simultane Analyse aller Daten in einem Modell

Beim Vergleich zweier Datenquellen werden Differenzen oder Verhältnisse der Merkmale oder abgeleiteter Indizes untersucht. Weitere Ansätze nutzen transformierte Daten, um redundante Daten zu reduzieren und den Merkmalsraum zu verkleinern, z.B. mittels Hauptachsentransformation. Die Entscheidung, ob eine Veränderung vorliegt, erfolgt mit zu definierenden Schwellwerten. Neben der Wahl eines geeigneten Schwellwertes ist eine entscheidende Schwäche dieses Ansatzes die Tatsache, dass die Radiometrie eines Primitivs verändert sein kann, die zugehörige Klasse jedoch unverändert ist (Mas, 1999). Dieses Phänomen tritt typischerweise bei Ackerflächen auf, die zu unterschiedlichen Jahreszeiten beobachtet wurden. Zudem lassen sich mit diesen Ansätzen nur Aussagen über das Vorhandensein von Veränderungen, jedoch nicht über die Art der Veränderung treffen (Lu et al., 2004).

Die zweite Gruppe sind Verfahren, die die Ergebnisse von unabhängigen vollständigen Klassifikationen der Landbedeckung oder Landnutzung der einzelnen Zeitpunkte vergleichen (Post-Klassifikation). Post-Klassifikationsverfahren sind robust gegen spektrale Variationen einzelner Klassen zu unterschiedlichen Jahreszeiten und liefern eine vollständige Matrix von Klassenübergängen. Für gute Klassifikationsergebnisse zu den einzelnen Zeitpunkten sind in Qualität und Quantität ausreichende Trainingsdaten zwingend notwendig (Lu et al., 2004). Hohe Klassifikationsgenauigkeiten zu den einzelnen Zeitpunkten sind die Grundlage für eine sichere Erkennung von Veränderungen, da sich Fehler in den Klassifikationsergebnissen unmittelbar als Fehler in der Veränderungsmatrix auswirken. In diese zweite Gruppe fallen darüber hinaus objektorientierte Verfahren, die Veränderungen einzelner extrahierter Objekte analysieren. Diese Verfahren werden vor allem in der Bildsequenzanalyse zur Bestimmung von Trajektorien genutzt.

Bis zum heutigen Zeitpunkt sind die meisten Arbeiten zur Detektion von Veränderungen in die ersten beiden Kategorien einzuordnen. Da die oben angesprochenen Schwächen die erreichbare Klassifikationsgenauigkeit limitieren, gelangen Verfahren der dritten Gruppe zunehmend in den Fokus wissenschaftlicher Betrachtungen. Der Vorteil der simultanen Analyse in einem Modell ist, dass - wie in Gruppe 1 - die originalen Beobachtungen anstelle abgeleiteter Daten für die Klassifikation genutzt werden und es - wie in Gruppe 2 - möglich ist, die Art der Veränderung zu modellieren.

Der einfachste Ansatz zur simultanen Analyse in einem Modell ist, die extrahierten Merkmale aller Zeitpunkte pro Primitiv in einen Vektor zusammenzufassen. Dieser wird mit einem beliebigen Verfahren klassifiziert. Dieses Vorgehen erfordert jedoch umfangreiche Trainingsdaten für sämtliche realistischen Klassenübergänge. Für viele Anwendungen liegt diese Voraussetzung nicht vor.

Andere Ansätze integrieren semantische Informationen, u.a. modelliert Pakzad (2001) die Klassenübergänge explizit mit Zustandsübergangsdiagrammen. Ebenso nutzen Mota et al. (2007) und Feitosa et al. (2009) Vorwissen über mögliche Änderungen von Landbedeckungsklassen, um den Suchraum zu verkleinern. Dieses Vorwissen wird in Übergangsmatrizen, die die Möglichkeiten eines Übergangs zwischen zwei Klassen ausdrücken, repräsentiert. Die vorgestellte Methode dieser Arbeiten ist auf Pixel wie auch Segmente anwendbar. Räumliche Nachbarschaften werden nicht berücksichtigt. Bei Nutzung von Segmenten wirken sich Segmentierungsfehler unmittelbar auf das Klassifikationsergebnis aus.

Die meisten Arbeiten verwenden nur ein Klassifikationsverfahren zur Veränderungsdetektion. Einige Ansätze kombinieren unterschiedliche Klassifikatoren, um die Vorzüge der jeweiligen Verfahren zu nutzen. Bruzzone et al. (2004) führen die Ergebnisse von drei multitemporalen Klassifikatoren gemäß diverser Entscheidungsregeln zusammen, ohne räumliche Nachbarschaften zu berücksichtigen. Durch die Anwendung mehrerer überwachter Klassifikatoren ist das Verfahren aufwendig. Eine signifikante Verbesserung der Ergebnisse hängt stark von der Wahl geeigneter Entscheidungsregeln ab.

Verfahren zur simultanen Auswertung multitemporaler Daten unterscheiden sich zudem im Informationsfluss zwischen den Primitiven zu den einzelnen Zeitpunkten. In Verfahren mit kaskadierendem Informationsfluss (z.B. unter Einsatz von Markov-Ketten (Feitosa et al., 2009)) werden Informationen nur in eine Richtung gegeben, in den meisten Fällen von früheren zu aktuelleren Zeitpunkten. In Verfahren mit bidirektionalem Informationsfluss interagieren die Primitive der unterschiedlichen Zeitpunkte wechselseitig. Somit kann auch die Klassifikationsgenauigkeit

zurückliegender Zeitpunkte gesteigert werden. Zufallsfelder ermöglichen einen bidirektionalen Informationsfluss, entsprechende Arbeiten werden in Kapitel 2.4. besprochen.

## 2.3. Klassifikation von Bilddaten mit Zufallsfeldern

Zahlreiche Klassifikationsverfahren wirken nur lokal auf die zu klassifizierenden Primitive ohne Betrachtung angrenzender Primitive (Kapitel 2.1). Das resultierende Klassifikationsergebnis weist oft einen „Salz-und-Pfeffer-Effekt“ aus. Ein solches Ergebnis entspricht jedoch zumeist nicht der Realität. Die meisten Objekte sind größer als die durch ein Pixel des Sensors abgedeckte Fläche. Durch Nutzung kontextbasierter Verfahren kann ein homogeneres und vor allem der Realität besser entsprechendes Ergebnis erzeugt werden (Schindler, 2012; Wang et al., 2016).

Es gibt zwei gängige Möglichkeiten, Kontext in die Auswertung zu integrieren: implizit über Merkmale oder explizit durch Modellierung der Kontextbeziehungen. Im ersten Fall werden die Bildprimitive weiterhin unabhängig voneinander klassifiziert. Kontextinformation wird über die extrahierten Merkmale integriert. Diese werden hierzu nicht nur aus dem Erscheinungsbild des Primitivs gewonnen, sondern auch aus seiner Umgebung. Im zweiten Fall werden die Kontextbeziehungen explizit in das Klassifizierungsverfahrens integriert. Hierfür bieten sich graphische Modelle (Kapitel 3.2), insbesondere Zufallsfelder, an.

Populär wurden Zufallsfelder mit den Arbeiten von Geman und Geman (1984) sowie Besag (1986). Die von ihnen vorgestellten Markov Random Fields (MRF, Kapitel 3.2.3) sind ungerichtete graphische Modelle zur probabilistischen kontextbasierten Klassifikation. Zufallsfelder bestehen aus Knoten und ungerichteten Kanten. Knoten repräsentieren die zu klassifizierenden Bildprimitive. Die Abhängigkeiten zwischen den Primitiven werden durch Kanten modelliert. Ziel der Klassifikation ist die simultane Bestimmung der wahrscheinlichsten Konfiguration aller Klassenlabels für das gesamte graphische Modell.

Bei der Anwendung von MRF ist die Interaktion benachbarter Primitive zur Berücksichtigung von Kontext auf die Klassenlabels beschränkt. Die an unterschiedlichen Bildpositionen extrahierten Merkmale werden als bedingt unabhängig betrachtet (Kapitel 3.2.3). Diese Einschränkung wird mit Conditional Random Fields aufgehoben. CRF ermöglichen zusätzlich die Modellierung von Abhängigkeiten von Merkmalen von unterschiedlichen Primitiven an beliebigen Positionen sowie die Modellierung von Interaktionen von Merkmalen und Klassenlabels (Kapitel 3.2.4).

CRF wurden ursprünglich für die Analyse von Textsequenzen entwickelt (Lafferty et al., 2001). Als erste setzten Kumar und Hebert (2003) CRF zur Klassifikation von Bilddaten ein. Ziel ihrer Arbeit ist die Erkennung von Gebäuden in beliebigen terrestrischen Szenen. Dabei handelt sich um ein binäres Klassifikationsproblem, das unter Nutzung typischer Strukturen in terrestrischen Szenen gelöst wird. He et. al (2004) nutzen ein multiskaliges CRF mit lokalen, regionalen sowie globalen Merkmalen zur Klassifikation von monotemporalen terrestrischen Szenen. Künstliche sowie natürliche Objekte in terrestrischen Szenen unter Verwendung hierarchischer CRFs klassifizieren Ladický et al. (2009) und Yang (2011). Aus den vollaufgelösten Bildern werden Bilder unterschiedlicher Auflösungsstufen generiert und aus diesen niedriger aufgelösten Bildern Merkmale abgeleitet. Die Auflösungsstufen werden mit dem hierarchischen CRF verknüpft.

In der weiteren Betrachtung liegt der Fokus auf Anwendungen aus dem Bereich der Fernerkundung.

## 2.4. Zufallsfelder in der Fernerkundung

Aufgrund der vielversprechenden Ergebnisse für vielfältige unterschiedliche Aufgabenstellungen im Bereich Computer Vision wurden MRF sowie CRF in zahlreichen Anwendungen aus dem Bereich der Fernerkundung erprobt.

CRF wurden zunächst zur Auswertung monotemporaler Fernerkundungsdaten genutzt. Zhong und Wang (2007) nutzen multiple CRF zur Erkennung urbaner Gebiete in hochaufgelösten optischen Satellitenbilddaten. In ihrem Ansatz werden mehrere unterschiedliche Merkmale extrahiert, die unabhängig voneinander mit CRF klassifiziert werden. Diese Klassifikationsergebnisse werden in einem finalen Schritt fusioniert. Lu et al. (2009) nutzen CRF für die Erstellung eines digitalen Geländemodells aus LiDAR-Daten. Roscher et al. (2010) klassifizieren Feldfrüchte und weitere Landbedeckungsklassen in Landsat-Aufnahmen. Ergebnisse, die mit drei unterschiedlichen Klassifikatoren (Maximum Likelihood, Support Vector Machines, Import Vector Machines) erzeugt wurden, werden in einem CRF fusioniert. Paisitkriangkrai et al. (2015) nutzen CRF für eine Postklassifikation einer Convolutional Neural Networks (CNN) - Klassifikation von Landbedeckungsklassen. Die Post-Klassifikation bewirkt in diesem Fall eine Glättung unter Beibehaltung der Kanten. Einen ähnlichen Ansatz verfolgen Pan und Zhao (2018), wobei zunächst einzelne Pixel zu Superpixeln zusammengefasst werden und diese mit einem CNN klassifiziert werden. Die Superpixel werden mithilfe des CRF wieder zerlegt und auf diese Weise kleine Strukturen bewahrt.

Die zuvor vorgestellten Ansätze nutzen Pixel oder kleine quadratische Segmente als Primitive. Diese spannen einen regelmäßigen Graphen (Kapitel 3.2) auf. Im Gegensatz hierzu verwendet Wegner (2011) einen aus einer Segmentierung resultierenden unregelmäßigen Graphen für die Gebäudedetektion aus optischen und InSAR-Daten. Ein solcher Ansatz ist allerdings für eine Veränderungsdetektion nicht zielführend, da keine identischen Segmentgrenzen für die einzelnen Bilder zu erwarten sind. Eine 1:1 Zuordnung von Primitiven in den Bildern der unterschiedlichen Zeitpunkte wäre somit nicht gegeben.

Eine multiskalige Auswertung einzelner Bilder oder eine Fusionierung von Daten unterschiedlicher Datenbestände bzw. unterschiedlicher Zeitpunkte kann über eine Modellierung mit mehreren Ebenen in einem Zufallsfeld erfolgen. Auf multiskaligen Markov Zufallsfeldern basierende Klassifikationsansätze präsentieren Kato et al. (1993), Wilsky (2002) sowie Choi et al. (2008). Die Idee besteht im Aufbau einer Label-Pyramide, deren gröbere Raster sich in schrittweise feinere Stufen untergliedern. Der Vorteil dieses kontextbasierten Ansatzes ist, dass in jeder Stufe neue lokale Interaktionen entstehen. Ebenfalls basierend auf einem pyramidalen Aufbau nutzen Bouman und Shapiro (1994) ein multiskaliges Zufallsfeld zur Bildsegmentierung. Die zugrundeliegende Struktur sind hierbei Quadtrees. Jeder Knoten einer Pyramidenebene interagiert mit dem zugehörigen Knoten in der gröberen Ebene in Form einer Markov-Kette. Eine Interaktion räumlich benachbarter Knoten unterbleibt. Einen ähnlichen Ansatz verfolgen Kersten et al. (2010) zur Klassifikation hochaufgelöster Fernerkundungsdaten. Im Gegensatz zu den zuvor vorgestellten Arbeiten werden jedoch basierend auf dem Informationsgehalt in den unterschiedlichen Auflösungsstufen unterschiedliche Klassen für die jeweiligen Auflösungsstufen definiert.

Unterschiedliche Klassen für die unterschiedlichen Ebenen eines multiskaligen CRFs verwenden ebenfalls Schnitzspan et al. (2008). Ziel ist die Erkennung von Objekten und deren einzelner Bestandteilen in terrestrischen Szenen. Das Modell besteht aus einem regelmäßigen Graphen in einer übergeordneten Ebene. In dieser Ebene werden komplette Objekte detektiert. Zudem existieren Nachbarschaftsbeziehungen zu weiteren Skalenebenen in einer Pyramidenstruktur. In den unteren Ebenen dieser Pyramide werden Objektteile detektiert. Das Modell gewährleistet somit eine Interaktion von Objekten und deren Teilen. Das bringt vor allem im Bereich von Verdeckungen, die in terrestrischen Szenen ein großes Problem darstellen, einen großen Gewinn. Der Ansatz wird in Yang et al. (2010) mit einer vorhergehenden Watershed-Segmentierung bzw. in Yang und Förstner (2011) mit einer Mean-Shift-Segmentierung in ein unregelmäßiges CRF überführt. Diese Ansätze eignen sich gleichfalls für die Auswertung von Fernerkundungsdaten.

Albert et al. (2017) klassifizieren simultan in einem Mehr-Ebenenmodell die Landbedeckung und Landnutzung. Es wird sowohl semantischer als auch räumlicher Kontext genutzt. Die Klassifikation der Landbedeckung erfolgt auf einer Ebene mit Super-Pixeln, wohingegen die Landnutzung auf Objektebene bestimmt wird. Diese Objekte werden als Polygone aus einer GIS-Datenbank gewonnen. Innerhalb der jeweiligen Ebenen werden die paarweisen Abhängigkeiten benachbarter Primitive genutzt. Zwischen den Ebenen werden die komplexen Abhängigkeiten von mehr als zwei Variablen mit Potentialen höherer Ordnung modelliert. Die Modellierung komplexer Abhängigkeiten durch Potentiale höherer Ordnung ist mit trad itionellen Inferenz-Methoden sehr zeitintensiv und wurde erst durch die Arbeiten von Kohli und Torr (2009) populär. Für die Klassifikation von optischen Fernerkundungsdaten wurden sie u.a. von Montoya-Zegarra et al. (2015) zur Erkennung von Straßen und Bauwerken genutzt.

Kosov et. al (2013) verwenden ein CRF mit mehreren Ebenen zur Klassifikation von Luftbildern mit zahlreichen verdeckten Objekten. Sowohl die verdeckten als auch die verdeckenden Objekte werden in einer separaten Ebene (Vordergrund und Hintergrund) modelliert. Die Knoten der in x-y-Position zugehörigen Primitive der beiden Ebenen sind durch Kanten verbunden, so dass Vordergrund- und Hintergrundprimitive interagieren können.

Zufallsfelder zur multitemporalen Auswertung von Fernerkundungsdaten wurden erstmals von Solberg et al. (1996) eingesetzt. In ihrer Methode zur simultanen Klassifikation von Fernerkundungs- und GIS-Daten zu unterschiedlichen Zeitpunkten können Daten unterschiedlicher Auflösung in den Auswerteprozess integriert werden. Die Klassenstruktur ist unabhängig von der Auflösung. Für mögliche Klassenübergänge werden Übergangswahrscheinlichkeiten genutzt. Melgani und Serpico (2003) nutzen MRF unter Berücksichtigung zeitlicher und räumlicher Nachbarschaften zur Klassifikation von Fernerkundungsdaten identischer Auflösung. Zunächst werden die Bilder einzeln mit MRF klassifiziert. In einem zweiten Schritt bezieht der Ansatz eine bidirektional wirkende Übergangsmatrix für zwei aufeinanderfolgende Zeitpunkte ein.

Die Merkmale unterschiedlicher Skalen nutzen Moser et al. (2009) in einem MRF basierten Ansatz zur Erkennung von Veränderungen in optischen Daten. Das Problem der Veränderungsdetektion wird mit einem Hypothesentest gelöst, der eine Binärkarte von Veränderungen zu zwei aufeinanderfolgenden Zeitpunkten generiert. Eine Unterscheidung der unterschiedlichen Klassen unterbleibt bei diesem unüberwachten Ansatz. Einen vergleichenden Überblick über drei multilayer Markov Random Field Methoden zur Detektion von Veränderungen geben Benedek et al. (2015). Zwei dieser Verfahren

erlauben ausschließlich die Unterscheidung in veränderte und nicht veränderte Gebiete in zwei Bildern. Das dritte Verfahren (Sziranyi und Shadaydeh, 2014) ermöglicht eine Identifikation der Klassenübergänge für mehrere Bilder in einem Post-Klassifikationsschritt. Es zeigte sich, dass mit geeigneten Merkmalen und geeigneter Parametrisierung vergleichbare Ergebnisse mit sämtlichen Ansätzen erzielt werden konnten. Die Wahl eines Verfahrens für eine konkrete Anwendung sollte somit nach praktischen Erfordernissen ausfallen. Das Training gestaltet sich für die ersten beiden Verfahren aufgrund der nicht so detaillierten Auswerteerfordernisse einfacher, das dritte Verfahren liefert hingegen den höheren Informationsgewinn.

Erste Anwendungen zur multitemporalen Klassifikation von Fernerkundungsdaten unter Verwendung von CRF stammen von Benedek und Szirányi (2009) sowie Hoberg et al. (2010). Beide Ansätze gehen von identischen Auflösungen der Bilder aus. Benedek und Szirányi (2009) detektieren Veränderungen in Luftbildern durch Kombination eines Mixed Markov Modells mit einem bedingt unabhängigen Zufallsfeld (conditionally independent random field). Die Methode unterscheidet lediglich binär in die Zustände verändert und unverändert ohne die Klassen näher zu bestimmen. Hoberg et al. (2010) erweitern ein monotemporales regelmäßiges CRF um weitere Ebenen zur Modellierung zeitlicher Interaktionen (Kapitel 4.1). Für Klassenübergänge wird eine empirisch bestimmte Übergangsmatrix genutzt. In optischen Bilddaten werden vier Landbedeckungsklassen unterschieden. Die Übertragbarkeit des Ansatzes auf Radardaten zeigen Kenduiywo et al. (2015) in ihren Arbeiten zur Klassifikation von Feldfrüchten in TerraSAR-X-Daten von zwei Zeitpunkten.

Das Problem der multitemporalen Klassifikation von Fernerkundungsdaten unterschiedlicher Auflösung mit Zufallsfeldern wurde bislang in der Literatur wenig behandelt. Die dieser Arbeit zugrunde liegende Methode wurde erstmals in Hoberg et al. (2011) sowie in Hoberg et al. (2015) komprimiert vorgestellt. Hedhli et al. (2016) verfolgen eine ähnliche Strategie, allerdings unter Verwendung von MRF. Zu den einzelnen Zeitpunkten können darüber hinaus mehrere Bilder unterschiedlicher Zeitpunkte eingebunden werden. Diese werden hierarchisch entsprechend ihrer Auflösung geordnet und über Quadtrees verknüpft. Fehlende Auflösungsebenen werden über Wavelet-Transformationen höher aufgelöster Bilder gefüllt. Die Berücksichtigung zeitlichen und räumlichen Kontexts führt in den Experimenten zu deutlichen Steigerungen im Vergleich zu Verfahren, die keinen oder nur eine Art des Kontextes integrieren.

## 2.5. Diskussion

Das Feld der Klassifikationsverfahren für Bilddaten ist sehr weit. Selbst wenn eine Einschränkung auf hochaufgelöste optische Fernerkundungsverfahren vorgenommen wird, bleibt die Anzahl der Veröffentlichungen nahezu unüberschaubar. Eine Gemeinsamkeit der Mehrzahl der Arbeiten ist, dass ausschließlich Informationen genutzt werden, die sich unmittelbar aus dem Erscheinungsbild eines zu klassifizierenden Primitivs ableiten lassen. Es wird keine Art von Kontextwissen berücksichtigt. Dabei hat sich gezeigt, dass trotz unterschiedlicher Bemühungen, die Klassifikationsgenauigkeit zu steigern, z.B. durch Integration von Modellwissen oder Subpixel-Analysen, das Potential dieser Methoden begrenzt ist.

Durch die Integration von räumlichen Kontextwissen in den Auswerteprozess können homogenere, realitätsnähere Ergebnisse erzielt werden und die Genauigkeit gesteigert werden (Wang et al., 2016; Schindler, 2012). Diese Beobachtung gilt sowohl für die Klassifikation auf Pixel- und Segmentebene als auch für Interpretationen auf Objektebene (Papadopoulos et al., 2011). Bei Betrachtungen auf Objektebene bietet sich zudem die Nutzung semantischen Kontexts an. Auf diese Weise können Relationen von Objekten modelliert werden, z.B. deuten detektierte Autos auf die Existenz einer Straße hin. Die Modellierung semantischer Beziehungen und deren Integration in den Klassifikationsprozess sind komplex und sind nicht Inhalt dieser Arbeit.

Einen ausführlichen Überblick über die Nutzung räumlichen Kontextwissens und eine Evaluation von Methoden zur Glättung für Anwendungen der Fernerkundung, u.a. einen Vergleich von lokalen Filtermethoden mit MRF und CRF, liefert Schindler (2012). Der Autor zeigt auf, dass jede Methode zur Glättung im Vergleich zu ausschließlich lokales Wissen einbeziehenden Klassifikatoren zu bevorzugen ist. Hierbei zeigen Klassifikationen mit CRF mit geringem Abstand die besten Ergebnisse.

Die Arbeiten zur Auswertung von Bilddaten mit CRF konzentrierten sich zunächst auf monotemporale terrestrische Szenen. Einige Jahre später wurden die Ansätze auf Fernerkundungsdaten übertragen. Terrestrische Szenen sind häufig einfacher auszuwerten, da Heuristiken oftmals einfach zu definieren sind und in den Auswerteprozess integriert werden können. Populäre Beispiele sind die Annahmen, dass terrestrische Aufnahmen stets so orientiert sind, dass der Himmel oben ist oder dass vertikale sowie horizontale Gebäudekanten annähernd parallel zu den jeweiligen Bildkanten verlaufen. Für Fernerkundungsdaten sind derart einfache Annahmen nicht zutreffend, da die Ausrichtung von Bodenobjekten in den wenigsten Fällen an die Himmelsrichtungen oder ähnlich klar definierte Gegebenheiten gekoppelt ist.

Ein Unterscheidungskriterium der Arbeiten zu Zufallsfeldern ist die Graphstruktur. Zumeist wird ein regelmäßiger Graph verwendet. Die Primitive repräsentieren die Pixel des Bildes und sind gleichfalls in einer Gitterstruktur angeordnet. Je nach Anwendung und auszuwertenden Daten bietet sich jedoch eine abweichende Graphstruktur an. So nutzen z.B. Wegner (2011) für die Gebäudedetektion aus optischen und InSAR-Daten sowie Niemeyer et al. (2014) für die Klassifikation von LiDAR-Daten unregelmäßige Graphen. Die Primitive, die durch die Knoten abgebildet werden, sind in diesen Arbeiten unterschiedlich. In Wegner (2011) sind die Knoten des Graphen einzelne Segmente. In Niemeyer et al. (2014) repräsentieren die Knoten einzelne Laserpunkte sowie Segmente bestehend aus Laserpunkten. Für beide Anwendungen ist das Vorgehen sinnvoll und führt jeweils zu positiven Ergebnissen. Es gibt jedoch entscheidende Unterschiede zur Anwendung in dieser Arbeit, weshalb sich die Modellierung mit einer unregelmäßigen Graphstruktur in dieser Arbeit nicht anbietet. Laserscandaten beinhalten als zusätzliche Größe eine Information zur Höhe, also einer dritten Dimension. Die originären Daten sind bereits unregelmäßig im Objektraum verteilt. Es liegt nahe, diese Verteilung in der Graphstruktur aufzunehmen. Die originären optischen Fernerkundungsdaten sind hingegen zweidimensional und regelmäßig verteilt. Die Nutzung von Segmenten für multitemporale Auswertungen ist problematisch. Bereits im monotemporalen Fall wirken sich Segmentierungsfehler unmittelbar auf das Klassifikationsergebnis aus. Erfolgt im multitemporalen Fall eine separate Segmentierung für jedes Bild, ist davon auszugehen, dass keine räumlich übereinstimmenden Segmente entstehen. In dieser Arbeit wird auf eine Segmentierung verzichtet und somit eine eindeutige Zuordnung von Primitiven identischer Position unterschiedlicher Zeitpunkte gewährleistet.

Eine Möglichkeit zur multitemporalen Klassifikation von Primitiven in einem graphischen Modell sind Markov-Ketten (Leite et al., 2011). Hierbei wird jedes Primitiv zu jedem Zeitpunkt als ein Knoten modelliert. Die möglichen Klassenübergänge eines Primitivs zwischen den Zeitpunkten können zum Beispiel durch Übergangswahrscheinlichkeiten beschrieben werden. Markov-Ketten sind jedoch in ihrer Struktur nur eindimensional, eine Berücksichtigung von räumlichen Kontext ist nicht möglich.

Ziel dieser Arbeit ist die gemeinsame Auswertung multitemporaler Daten unterschiedlicher Auflösung zur Steigerung der Klassifikationsgenauigkeit und zur Detektion von Veränderungen. Sämtliche in der Literatur beschrieben Verfahren decken nicht alle dieser Aspekte ab.

Zahlreiche Arbeiten beschäftigen sich mit multiskalaren Analysen. In diesen Arbeiten, wie auch in der vorliegenden, werden Ebenen für zusätzliche Interaktionen von Primitiven an identischer Position eingeführt. Die Bilddaten der unterschiedlichen Skalenebenen werden jedoch zumeist aus monotemporalen Aufnahmen generiert. Werden multitemporale Daten genutzt, sind diese meist von einem identischen Sensor und haben somit eine identische Auflösung (z.B. Moser et al., 2009). In wenigen Arbeiten zur multiskalaren Analyse verändert sich die Klassenstruktur entsprechend der Skalenebene. Eine unterschiedliche Klassenstruktur in Abhängigkeit von der Bodenauflösung erscheint jedoch gerechtfertigt. In unterschiedlichen Auflösungsstufen sind Objekte unterschiedlich detailliert wahrnehmbar. Einzelne Objektteile treten erst mit höherer Auflösung hervor. Somit ist es sinnvoll, übergeordnete Klassen einer niedrigeren Auflösungsstufe in mehrere untergeordnete Klassen einer höheren Auflösungsstufe aufzuspalten. Das in dieser Arbeit entwickelte Verfahren ermöglicht diesen Erkenntnissen entsprechend eine Modellierung der unterschiedlichen Klassenstrukturen.

Die meisten Arbeiten zur Detektion von Veränderungen in multitemporalen Daten basieren auf Daten identischer Auflösung, oft sogar desselben Sensors. Häufig werden jahreszeitlich identische Aufnahmezeitpunkte betrachtet. Die Analyse erfolgt in vielen Fällen für lediglich zwei Zeitpunkte. Ziel dieser Arbeit ist die volle Flexibilität in Bezug auf die Wahl des Sensors und die Anzahl der Aufnahmezeitpunkte. Diese Flexibilität kann mit Zufallsfeldern gewährleitstet werden. Die Graphstruktur kann durch Erweiterung auf mehrere Ebenen für eine Auswertung einer beliebigen Anzahl von Bildern angepasst werden.

Methodisch erfolgt die Detektion von Veränderungen in vielen Anwendungen durch Post-Klassifikationsverfahren. Es werden die einzelnen Aufnahmen zuvor unabhängig voneinander ausgewertet. In zahlreichen Arbeiten hat sich gezeigt, dass eine gemeinsame Auswertung aller Daten zu bevorzugen ist. Hierzu haben sich Übergangsmatrizen, die die Wahrscheinlichkeiten eines Klassenübergangs zwischen zwei Zeitpunkten abbilden, bewährt. Diese Vorgehensweise wird auch in dieser Arbeit verfolgt.

# 3. Grundlagen

Bevor in Kapitel 4 der kontextbasierte Ansatz zur Klassifikation multitemporaler Bilddaten unterschiedlicher Auflösung entwickelt wird, werden in diesem Kapitel die für das weitere Verständnis benötigten Grundlagen dargelegt. In Kapitel 3.1 wird auf die in dieser Arbeit verwendeten Merkmale eingegangen. Aus diesem verfügbaren großen Merkmalspool werden zur Steigerung der Performance im Weiteren nur wenige Merkmale zur Klassifikation genutzt. In Kapitel 3.1.2 wird die Methode der Merkmalsselektion beschrieben. Kapitel 3.2 führt in das Feld der graphischen Modelle zur Bildanalyse ein. Ein Spezialfall der graphischen Modelle sind Conditional Random Fields, auf die in Kapitel 3.2.4 eingegangen wird. Diese Methode ist die Basis für den in dieser Arbeit entwickelten Ansatz und wird in Kapitel 4 für die Auswertung multitemporaler Bilddaten unterschiedlicher Auflösung erweitert. Abschließend werden in Kapitel 3.3 die zur Auswertung der Experimente in Kapitel 5 verwendeten Beurteilungskriterien aufgeführt.

## 3.1. Merkmale

In der Literatur wird der Begriff Merkmal (feature) sowohl für geometrische Primitive als auch für Eigenschaften geometrischer Primitive verwendet. Dies ist oft missverständlich, die Bedeutung ergibt sich zumeist aus dem Kontext. In dieser Arbeit wird der Begriff „Merkmal“ ausschließlich für eine direkte oder abgeleitete Eigenschaft eines Primitivs verwendet.

Die Güte einer Klassifikation ist im Wesentlichen von den verfügbaren Trainingsdaten und den daraus extrahierten Merkmalen abhängig (Jain et al., 2000). Dabei wird zwischen der Merkmalsberechnung und der Merkmalsselektion unterschieden. Bei der Merkmalsberechnung werden neue Merkmale direkt oder indirekt aus den Daten abgeleitet. In einem anschließenden (optionalen) Schritt wird bei der Merkmalsselektion eine Untermenge der zur Klassifikation am besten geeigneten Merkmale ausgewählt. Im Folgenden wird ein Überblick über häufig für die Klassifikation von Fernerkundungsdaten genutzte Merkmale gegeben. Besonderer Fokus liegt dabei auf den in dieser Arbeit verwendeten Merkmalen. Im Anschluss werden Verfahren zur Merkmalsselektion diskutiert und das in dieser Arbeit genutzte Verfahren eingehend erläutert.

### 3.1.1. Merkmalsextraktion

Merkmale können die unmittelbaren radiometrischen Informationen eines Primitivs, abgeleitete Werte aus dem Erscheinungsbild des Primitivs oder abgeleitete Werte aus dem Erscheinungsbild seiner Umgebung sein. In der Literatur erfolgt bisweilen eine Unterteilung in spektrale, texturbasierte, strukturelle und geometrische Merkmale. Im Folgenden werden diese Gruppen von Merkmalen kurz vorgestellt.

#### 3.1.1.1. Spektrale Merkmale

Spektrale oder densitometrische Merkmale leiten sich direkt aus den Spektralinformationen der verschiedenen Kanäle ab. So lassen sich z.B. Gebiete mit grüner Vegetation von übrigen Gebieten trennen. Bei der Betrachtung von Segmenten lassen sich neben den Mittelwerten einzelner Kanäle z.B.

die Standardabweichung, Minimal- und Maximalwert oder die Majorität einzelner Pixelwerte bestimmen. Weitere Merkmale werden durch die Kombination verschiedener Kanäle gewonnen, z.B. das Verhältnis oder die Differenz der Pixelwerte zweier Kanäle. Für Aufgaben der Fernerkundung spielen Vegetationsindizes, die aus Informationen des roten Kanals und des nahen Infrarotkanals berechnet werden, eine wichtige Rolle. Hierbei wird die Tatsache genutzt, dass sich Vitalität im nahen Infrarotkanal besonders stark ausprägt. Somit ist es möglich, Vegetationsflächen von sonstigen Flächen, bzw. intakte Vegetation von kranker Vegetation, zu unterscheiden. Am häufigsten verwendet werden der Ratio Vegetation Index (RVI) (Jordan, 1969) sowie der Normalized Difference Vegetation Index (NDVI) (Rouse et al., 1973).

$$RVI = \frac{nir}{rot} \tag{3-1}$$

$$NDVI = \frac{nir - rot}{nir + rot} \tag{3-2}$$

#### 3.1.1.2. Texturbasierte Merkmale

Die Unterscheidung zwischen texturbasierten und strukturellen Merkmalen ist in der Literatur nicht immer eindeutig. Im Allgemeinen wird unter einer Textur ein mehr oder weniger regelmäßiges Muster in kleinräumiger Nachbarschaft verstanden. Ein solches Muster entsteht z.B. durch Ackerfurchen, die sich bei ausreichend hoher Auflösung von Fernerkundungsdaten als parallele Linien in dichtem Abstand ausprägen. Als Struktur wird die Anordnung prägender Bildelemente angesehen. Werden beide Merkmalsgruppen in einer kleinen Nachbarschaft berechnet, ist die Einteilung von Merkmalen in diese Gruppen nicht mehr eindeutig.

Texturbasierte Merkmale oder Texturmerkmale werden aus den Intensitätswerten in einer zuvor definierten Region abgeleitet. Einfache Texturmerkmale sind der Exzess oder die Schiefe eines Histogramms der Intensitätswerte der betrachteten Region. Hierbei kann jedoch nicht eindeutig auf ein räumliches Muster geschlossen werden. Um auch auf räumliche Muster schließen zu können, werden die Merkmale aus den Nachbarschaften einzelner Pixel berechnet. Dies ermöglicht eine Aussage über die räumliche Verteilung der Intensitätswerte. Die wohl am häufigsten verwendete Gruppe von Texturmerkmalen sind die *Haralick'schen Maße* (Haralick et al., 1973). Aufgrund der Relevanz für die weitere Arbeit werden sie an dieser Stelle eingehender betrachtet. Die Notation beruht auf Tönnies (2005).

Basis für die Berechnung der Merkmale ist die *grey-level co-occurrence matrix* (GLCM). Eine GLCM $P_{\alpha,\Delta}(g_1,g_2)$ gibt für eine Region die Wahrscheinlichkeit der Kombination zweier Grauwerte $g_1$ und $g_2$ von Pixeln im Abstand Δ und Winkel α zur Horizontalen des Koordinatensystems an. Für ein 8-Bit Grauwertbild ergibt sich folglich eine GLCM der Dimension 256*256.

$$P_{\alpha,\Delta}(g_1, g_2) = \frac{1}{K}\sum_{x} \delta_D\left(f(\boldsymbol{x}) - g_1\right) \cdot \delta_D(f(\boldsymbol{x} + \boldsymbol{d}) - g_2) \tag{3-3}$$

Mit

$$\boldsymbol{d} = \Delta \cdot (\sin\alpha \quad \cos\alpha) \tag{3-4}$$

Hierbei ist *K* die Anzahl Pixel $\boldsymbol{x}$ innerhalb der Region. Die Dirac-Funktion $\delta_D$ nimmt für $\delta_D(0)$ den Wert 1 und sonst den Wert 0 an. Unterschiedliche Pixelabstände Δ sowie unterschiedliche Orientierungen α resultieren in unterschiedlichen GLCM. Für einen gewählten Abstand Δ=1 werden nur unmittelbar benachbarte Pixel betrachtet. Häufig werden vier GLCM mit den Orientierungen 0°, 45°, 90° sowie 135° berechnet.

Eine unmittelbare Nutzung der GLCM für die Klassifikation führt zu einem sehr hochdimensionalen Merkmalsvektor. Es ist von daher zielführend, Texturmerkmale aus der GLCM abzuleiten. Die folgende Auswahl führt die am häufigsten verwendeten Haralick'schen Texturmerkmale auf, welche auch in dieser Arbeit genutzt werden:

$$Energie = \sum_{g_1}\sum_{g_2} P_{\alpha,\Delta}^2(g_1, g_2) \tag{3-5}$$

$$Kontrast = \sum_{g_1}\sum_{g_2} (g_1 - g_2)^2 * P_{\alpha,\Delta}^2(g_1, g_2) \tag{3-6}$$

$$Entropie = -\sum_{g_1}\sum_{g_2} P_{\alpha,\Delta}(g_1, g_2) * \log\left[P_{\alpha,\Delta}(g_1, g_2)\right] \tag{3-7}$$

$$Homogenität = \sum_{g_1}\sum_{g_2} \frac{P_{\alpha,\Delta}(g_1, g_2)}{1 + |g_1 - g_2|} \tag{3-8}$$

$$Korrelation = \sum_{g_1}\sum_{g_2} \frac{(g_1 - \mu_1)(g_2 - \mu_2)P_{\alpha,\Delta}(g_1, g_2)}{\sigma_1\sigma_2} \tag{3-9}$$

In (3-9) sind $\mu_1$ und $\mu_2$ die Mittelwerte sowie $\sigma_1$ und $\sigma_2$ die Standardabweichungen der Zeilen und Spalten der GLCM.

Alternativ zu den *Haralick'schen Maßen* werden häufig *local binary patterns* (Ojala et al., 1996) eingesetzt. In der einfachsten Form des Klassifikators wird der Grauwert jedes Pixels mit seinen acht Nachbarn verglichen. Ist der Grauwert des Nachbarpixels größer, wird dieser Stelle eine „1", ansonsten eine „0" zugewiesen. Bei Betrachtung der Werte im oder gegen den Uhrzeigersinn entsteht somit eine achtstellige Binärzahl. Aus den Binärzahlen der Pixel in einer vorher definierten Region wird im nächsten Schritt ein Histogramm berechnet, das als Merkmal für die Region dient. Erweitern lässt sich der Klassifikator u.a. durch Nutzung einer größeren Nachbarschaft (Ojala et al., 2002). Local binary patterns sind robust gegen globale Grauwertveränderungen, die z.B. aus ungleichmäßiger Beleuchtung resultieren. Zudem sind local binary patterns wenig rechenintensiv, so dass sie für Echtzeitanwendungen eingesetzt werden können.

In einer Untersuchung zur Vegetationsartenklassifikation auf Objektebene mit Verwendung unterschiedlicher Texturmerkmale erzielen Li et al. (2010) nahezu identische Klassifikationsgenauigkeiten bei Nutzung von Merkmalen der GLCM sowie von local binary patterns und Erweiterungen von local binary patterns.

Eine weitere Möglichkeit zur Erkennung von Textur in Bilddaten sind Wavelets (Mallat, 1989). Wavelets sind Basisfunktionen für eine Transformation in eine andere Repräsentation (z.B. in den Frequenzraum). Hierfür steht eine nahezu unbegrenzte Anzahl von Wavelets zur Verfügung. Wavelets können so skaliert werden, dass Textur bestmöglich in unterschiedlichen Bildauflösungen erkannt werden kann.

#### 3.1.1.3. Strukturelle Merkmale

Strukturelle Merkmale dienen der Identifikation von Strukturen, die für die jeweiligen Klassen im betrachteten Bild repräsentativ sind. Im Gegensatz zu texturbasierten Merkmalen ist es nicht erforderlich, dass sich die Abfolge von Grauwerten in einer kleinen betrachteten Region wiederholt (wie z.B. bei Ackerfurchen). Vielmehr wird betrachtet, ob sich die in einer ausgewählten Region erkannten Strukturen an anderen Stellen im Bild mit gleicher Klassenzugehörigkeit wiederholen.

Seit ihrem erfolgreichen Einsatz zur Personendetektion erfreuen sich strukturelle Merkmale basierend auf gewichteten Histogrammen orientierter Gradienten (HOG) (Dalal & Triggs, 2005) großer Popularität. Die Annahme für den Einsatz von HOG-Merkmalen ist, dass Objekte in einem Bild durch eine Verteilung der Gradientenintensitäten oder Kantenrichtungen beschrieben werden können. Hierzu werden im Bild stets nur kleine Regionen betrachtet. In jeder Region wird ein Histogramm orientierter Gradienten berechnet.

Sofern es sich um ein Multispektralbild handelt, wird dieses in einem ersten Schritt in ein Grauwertbild umgewandelt. Nachfolgend werden die horizontalen und vertikalen Ableitungen berechnet. Aus den horizontalen und vertikalen Ableitungen werden die Gradientenrichtungen und Gradientenmagnituden für ein einzelnes Pixel abgeleitet. Aus diesen Werten wird das Histogramm der orientierten Gradienten generiert. Das Histogramm wird durch Unterteilung des Wertebereichs von 0 bis 180 Grad (bzw. 0 bis 360 Grad sofern das Vorzeichen berücksichtigt wird) in mehrere Wertebereiche gleicher Größe gegliedert. Jedes Pixel liefert entsprechend seiner Gradientenrichtung einen Eintrag für den entsprechenden Wertebereich des Histogramms. Nach Berücksichtigung sämtlicher Pixel enthält jeder Wertebereich die Summe der Magnituden aller Gradienten, die eine Orientierung im Intervall des Wertebereichs aufweisen. Die Summe sämtlicher Magnituden anstelle einer Zählung der Anzahl der Einträge bewahrt den Einfluss starker Gradienten. Abbildung 3-1 zeigt die Gradientenmagnituden und das Histogramm der orientierten Gradienten für drei häufig auftretende Landbedeckungsszenarien in Fernerkundungsdaten.

Eine alternative Möglichkeit zur Identifikation identischer Strukturen in Fernerkundungsdaten sind Semivariogramme (Ruiz et al., 2011). Für die Erzeugung von Semivariogrammen werden die Differenzen der Farbwerte von Pixelpaaren in einem definierten horizontalen oder vertikalen Abstand gemessen. Wird der Abstand variiert, kann ein Graph aufgebaut werden, der die Varianz der Farbwerte der Pixelpaare (d.h. wie unterschiedlich die Pixelpaare durchschnittlich sind) als eine Funktion des Abstandes ausgibt. Auf diese Weise können die Korrelationen zwischen Pixeln in unterschiedlichen Bildregionen gemessen werden (Richards, 2013).

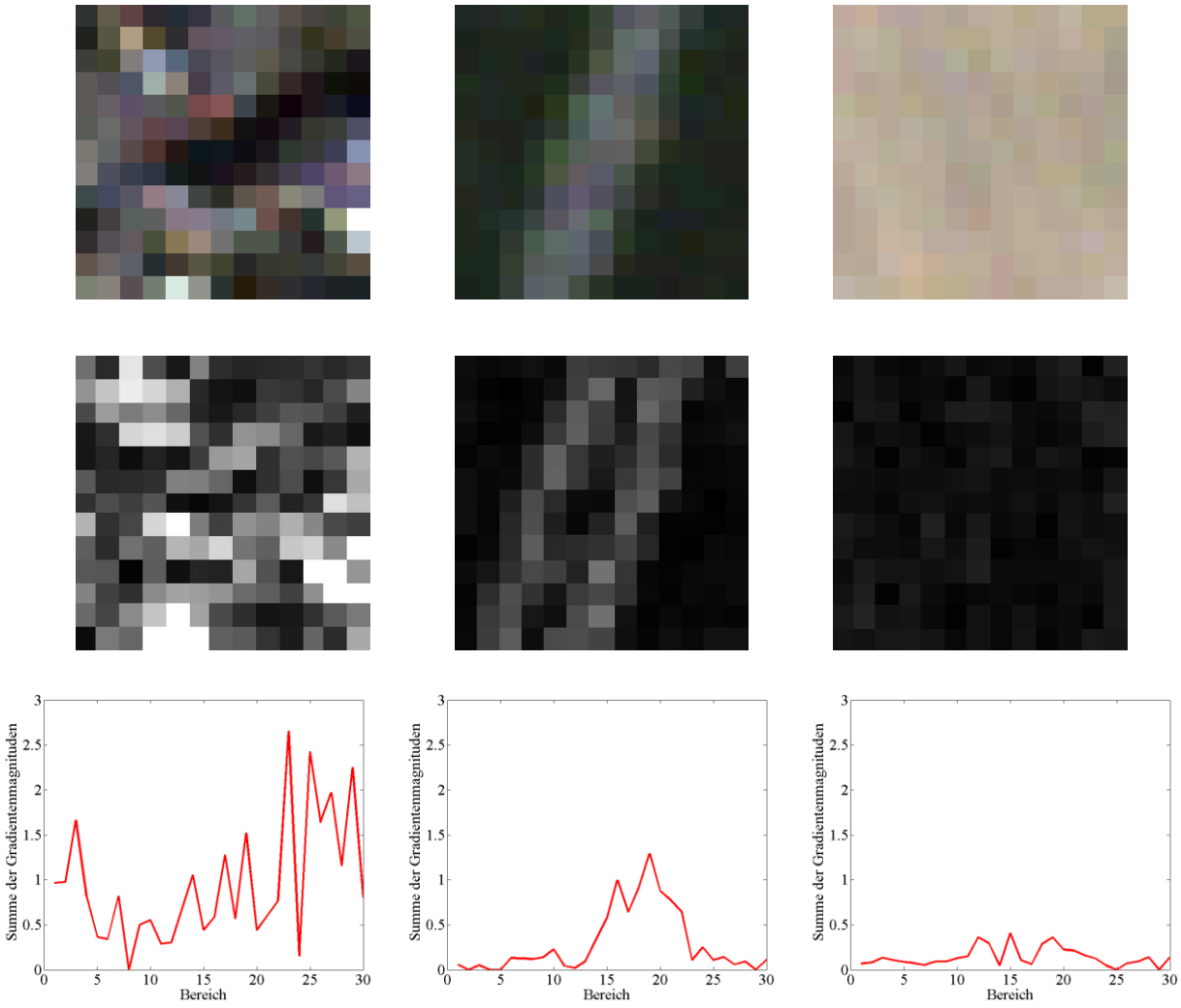



*Abbildung 3-1: Obere Reihe: Regionen (13*13 Pixel) einer multispektralen Ikonos-Szene (Bodenauflösung 30 Meter), Landbedeckung: Siedlung (links), Straße umgeben von Ackerflächen (Mitte), Ackerfläche (rechts). Mittlere Reihe: Bild der Gradientenmagnituden der Regionen; Helle Werte repräsentieren hohe Gradientenmagnituden. Untere Reihe: zugehörige Gradientenhistogramme (30 Wertebereiche mit Orientierungsintervallen von 6° Breite).*

### 3.1.1.4. Geometrische Merkmale

Zur Ableitung geometrischer Merkmale ist eine vorhergehende Segmentierung des Bildes erforderlich. Einfache Maße sind der Flächeninhalt, der Umfang oder die Ausdehnung der längsten bzw. kürzesten Achse. Zudem können auch topologische Informationen berücksichtigt werden. Geometrische Merkmale werden vor allem eingesetzt, um Objekte mit typischer Form zu detektieren. Problematisch bei der Berechnung geometrischer Merkmale ist die Abhängigkeit von der Qualität der Segmentierung; versagt die Segmentierung, sind auch die Merkmale für eine Klassifikation nicht brauchbar. Auf eine eingehende Betrachtung dieser Merkmalsgruppe wird an dieser Stelle verzichtet, da für die Experimente

in dieser Arbeit Pixel bzw. regelmäßig geformte Segmente als Primitive genutzt werden. Geometrische Merkmale sind somit für diese Arbeit nicht relevant.

### 3.1.2. Merkmalsselektion

Eine Merkmalsselektion ist eine Vorprozessierung der Daten für die Klassifikation (Refaeilzadeh et al., 2007). Das Ziel einer Merkmalsselektion ist die Auswahl einer Menge von Merkmalen aus der Gesamtheit aller Merkmale ohne die Güte der Klassifikation signifikant zu verschlechtern (Pudil et al., 1994). Mehrere Gründe sprechen für eine Reduktion von Merkmalen:

- Merkmale, die nicht in der Lage sind, Daten zuverlässig zu trennen, werden verworfen. Diese Merkmale sind für den Erfolg einer Klassifikation nicht relevant.
- Die Anzahl redundanter Merkmale wird verringert. Somit wird eine Senkung der Rechenkosten erreicht (Liu und Motoda, 2008).
- Bei geringer Anzahl von Trainingsdaten ist der Merkmalsraum zu beschränken, um dem „Fluch der Dimensionalität“ (curse of dimensionality oder Hughes phenomenon (Hughes, 1968)) zu entgehen. Hierbei ist das Verhältnis der Anzahl Merkmale zur Anzahl verfügbarer Trainingsdaten wesentlich. Zu viele Merkmale können zu einer Überanpassung führen und die Qualität des Klassifikators senken (Jain et al., 2000; Richards, 2013). In den in der vorliegenden Arbeit vorgestellten Experimenten ist die Menge der Trainingsdaten sehr groß im Verhältnis zur Anzahl der extrahierten Merkmale. Qualitätseinbußen aus diesem Grund sind nicht zu erwarten.

Die Fragestellung, ob Merkmale geeignet sind, Klassen zu trennen, kann auf einfache Weise mit Scatter-Plots visualisiert werden. Hierzu werden zwei Merkmale auf den Wertebereich [0;1] normalisiert und je nach Klassenzugehörigkeit entsprechend farbig unterschieden in einem Koordinatensystem abgebildet.

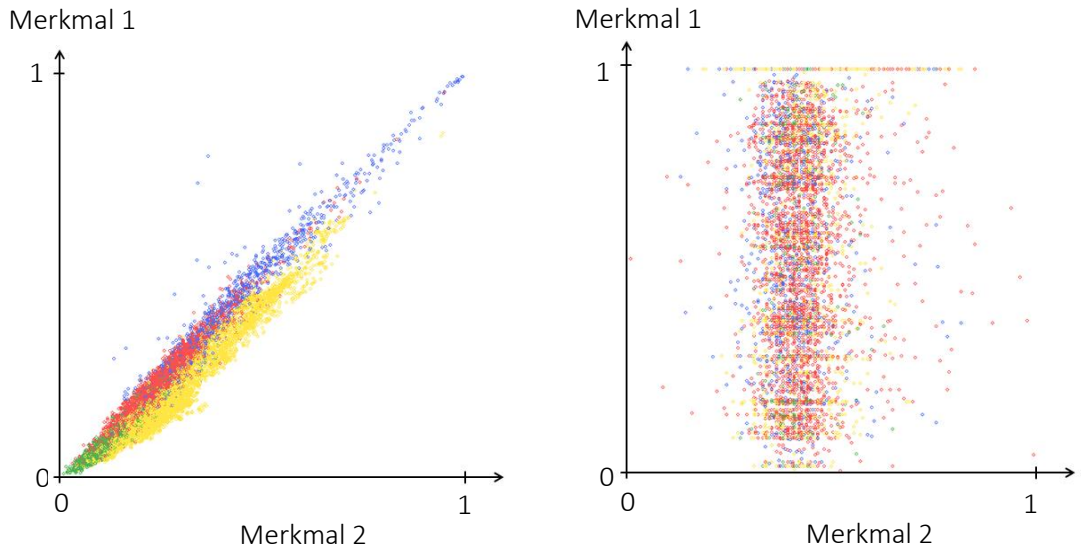



*Abbildung 3-2: Scatter Plot für 4 Klassen von zwei hoch korrelierten Merkmalen (links) und zwei schlecht trennenden Merkmalen (rechts). Die Merkmale wurden auf den Wertebereich [0;1] normalisiert.*

Die rechte Darstellung in Abbildung 3-2 zeigt den Scatter-Plot für zwei schlecht trennende Merkmale; die farbigen Punkte erscheinen willkürlich verteilt, eine Trennung der Klassen ist nur anhand dieser Merkmale nicht möglich. Eine hohe Korrelation von Merkmalen wird gleichfalls im Scatter-Plot offensichtlich (Abbildung 3-2, linke Darstellung).

Ansätze zur Merkmalsselektion werden in Wrapper und Filter unterschieden (Guyon und Elisseeff, 2003) (Liu und Motoda, 2008). Wrapper wählen die besten Merkmale in Interaktion mit dem jeweiligen Klassifikator aus. Liegt für eine Anwendung der Klassifikator fest, sind Wrapper-Methoden zu bevorzugen, da auf diese Weise optimale Merkmale ausgewählt werden. Ein Nachteil dieses Ansatzes ist, dass er einen wiederholten Aufruf des Klassifikators erfordert und somit für große Datenmengen mit vielen Merkmalen sehr rechenintensiv ist (Hall, 1999). Hingegen wird bei Filtern die Auswahl der Merkmale unabhängig vom Klassifikator getroffen. Die Auswahl basiert ausschließlich auf den Daten. Die vom Klassifikator unabhängige Lösung kann für unterschiedliche Klassifikatoren genutzt werden und eignet sich somit insbesondere für einen Vergleich unterschiedlicher Klassifikatoren. Ein weiterer wesentlicher Vorteil von Filtermethoden liegt in der deutlich höheren Geschwindigkeit. Da keine Interaktion mit dem Klassifikator gegeben ist, stellt jedoch unter Umständen die selektierte Menge eine nicht ideale Lösung für den eingesetzten Klassifikator dar. In dieser Arbeit ist die Merkmalsauswahl Grundlage für die Verwendung unterschiedlicher Klassifikatoren, weshalb eine Filtermethode zur Merkmalsselektion eingesetzt wird. Wrapper werden im Folgenden nicht weiter behandelt.

Filter lassen sich in univariate und multivariate Filter einteilen (Saeys et al., 2007). Univariat bedeutet, dass die einzelnen Merkmale unabhängig voneinander betrachtet werden. Es wird eine Anzahl von *b* Merkmalen selektiert, die einzeln betrachtet die Daten bestmöglich trennen. Bei multivariaten Ansätzen finden hingegen Abhängigkeiten zwischen den Merkmalen Berücksichtigung. Im Allgemeinen sind multivariate Ansätze zu bevorzugen. Bereits bei der Selektion von zwei Merkmalen ist in vielen Fällen die bestmögliche Kombination zweier Merkmale nicht diejenige Kombination der zwei unabhängig betrachtet am besten trennenden Merkmale (Peng et al., 2005). Der Grund hierfür ist, dass oft mehrere gut trennende Merkmale hoch korreliert sind.

In dieser Arbeit wird eine auf Korrelation basierende Merkmalsselektion (correlation based feature selection, CFS) verwendet (Hall, 1999; Hall, 2000) und deren Implementierung in der WEKA-Software (WEKA, 2012; Witten und Frank, 2005) genutzt. Diesem multivariaten Filter liegt folgende Heuristik zugrunde: „Gute Teilmengen aller Merkmale enthalten Merkmale, welche hoch korreliert mit der jeweiligen Klasse, aber untereinander möglichst unkorreliert sind“ (Hall, 2000). Diese Heuristik lässt sich wie folgt ausdrücken:

$$Merit_{fsub} = \frac{b\overline{r_{cf}}}{\sqrt{b + b(b-1)\overline{r_{ff}}}} \tag{3-10}$$

$Merit_{fsub}$ ist ein Maß für die Güte einer Teilmenge *fsub* mit *b* Merkmalen. Dabei stellt $\overline{r_{cf}}$ die durchschnittliche Korrelation der Merkmale mit den Klassen und $\overline{r_{ff}}$ die durchschnittliche Korrelation der Merkmale untereinander dar. Merkmale, die nur wenig geeignet sind, die Klassen zu trennen, tragen somit nur wenig zur Steigerung des Zählers bei. Hingegen lassen hoch korrelierte Merkmale den Nenner wachsen. In beiden Fällen wirkt sich dies negativ auf $Merit_{fsub}$ aus und die Merkmale werden nicht *fsub* zugeführt.

Als Suchstrategie wird eine Vorwärtssuche eingesetzt. Allgemein geht ein solcher Ansatz von einer leeren Merkmalsmenge aus. Es existieren zwei wesentliche Vorgehensweisen: Im ersten Fall werden die besten *b* Merkmale gewählt, wobei *b* vordefiniert wird. Im zweiten Fall wird die kleinste Untermenge an Merkmalen gewählt, die keine Einbußen der Klassifikationsgenauigkeit im Vergleich zur nächstgrößeren Merkmalsmenge bewirkt (Witten und Frank, 2005). Hierbei wird je nach Datensatz eine unterschiedliche Anzahl von Merkmalen ausgewählt. In dieser Arbeit wird eine konstante Merkmalsanzahl verwendet. Es wird zunächst das am besten trennende Merkmal gewählt. Nacheinander werden einzelne Merkmale hinzugenommen, bis die zuvor gewählte Anzahl von Merkmalen erreicht ist.

Vergleiche weiterer Filter- und Wrapper-Methoden mit der CFS-Methode in der Literatur zeigen generell keine wesentlichen Abweichungen in der Klassifikationsgenauigkeit (u.a. Pal und Foody, 2010; Ferreira und Figueiredo, 2012). Probleme treten allerdings bei sehr hochdimensionalen Daten auf, da die Berechnung in diesem Fall sehr kostenintensiv wird (Ferreira und Figueiredo, 2012; Huertas und Juárez-Ramírez, 2014). Für die vorgestellten Experimente ist dieser Punkt nicht kritisch.

Als Alternative zur Merkmalsselektion, bei der die originalen Merkmale erhalten bleiben, kann der Merkmalsraum durch eine Merkmalstransformation verkleinert werden, z.B. mittels Hauptachsentransformation. Diese ist allerdings szenenabhängig. Als Folge können die Transformationsergebnisse unterschiedlicher Szenen nicht direkt verglichen werden. Zudem müssen vor der Transformation sämtliche Merkmale berechnet werden, womit kein Zeitgewinn bei der Merkmalsextraktion erzielt werden kann. Eine weitere Möglichkeit der Datenreduktion ist der Ansatz, spektrale Informationen in abgeleitete Indikatoren zu transformieren, z.B. durch Tasseled Cap Transformation (Kauth und Thomas, 1976). Die abgeleiteten Indikatoren werden dabei als gewichtete Kombination der ursprünglichen Spektralkanäle berechnet. Diese Verfahren bieten sich jedoch nicht für die Verwendung unterschiedlicher Sensoren an, da die spezifizierte Kanalkombination für alle Bilddaten verfügbar sein muss.

## 3.2. Graphische Modelle für die Bildanalyse

Graphische Modelle ermöglichen die Bestimmung der Wahrscheinlichkeitsverteilung einer Menge von Zufallsvariablen. Sie stellen die Verbindung zwischen Wahrscheinlichkeits- und Graphentheorie dar und sind somit ein effektives Werkzeug zur statistischen Modellierung unter Verwendung beliebiger Graphstrukturen. Jeder Knoten des Graphen wird dabei als Zufallsvariable angesehen und jede Kante modelliert die statistischen Abhängigkeiten der zugehörigen Zufallsvariablen.

In der Bildanalyse werden graphische Modelle u.a. für die Modellierung von Relationen der Primitive S (d.h. Pixeln und Segmenten) eines Bildes eingesetzt. In den meisten Anwendungen werden räumlich benachbarte Primitive über Kanten miteinander verbunden. In der Folge interagieren diese benachbarten Primitive bei der Bestimmung der Wahrscheinlichkeitsverteilung. Ein solcher Graph heißt Nachbarschaftsgraph (neighbourhood graph oder region adjacency graph).

Abbildung 3-3 visualisiert die unterschiedliche Graphstruktur für zwei häufig verwendete Nachbarschaftsgraphen: 1) Die Primitive entsprechen einzelnen Pixeln oder rechtwinkligen Segmenten gleicher Größe. Das Resultat ist ein regelmäßiger Graph (linke Darstellung); 2) Die Primitive sind

unregelmäßig geformte Segmente. Folglich ist auch die Graphstruktur unregelmäßig (rechte Darstellung).

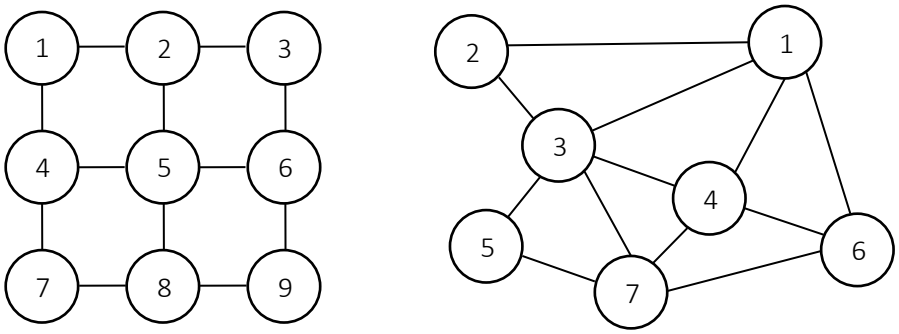



*Abbildung 3-3: Nachbarschaftsgraphen; Links: Graph mit regelmäßiger Struktur; Rechts: Graph mit unregelmäßiger Struktur*

Der Aufbau eines Graphen ist in vielen Fällen relativ unkompliziert und zeigt anschaulich die Abhängigkeiten benachbarter Primitive. Im Folgenden wird aufgezeigt, wie Graphen zur Klassifikation von Bilddaten genutzt werden können. Methoden, die diese Graphen nutzen, basieren auf der Wahrscheinlichkeitstheorie. Die Grundlagen werden in Kapitel 3.2.1 dargestellt. In Kapitel 3.2.2 wird allgemein der Zusammenhang zwischen der Graphdarstellung und der Faktorisierung der gemeinsamen Verteilung aller den Knoten entsprechenden Zufallsvariablen dargestellt. Die Kapitel 3.2.3 und 3.2.4 widmen sich Zufallsfeldern, einer Teilmenge der graphischen Modelle. Zufallsfelder bilden den Kern dieser Arbeit. Abschließend folgt in Kapitel 3.2.5 ein kurzer Überblick über Inferenz in graphischen Modellen, d.h. der Bestimmung der wahrscheinlichsten Konfiguration der Labels der einzelnen Primitive. Detaillierte Ausführungen zu graphischen Modellen finden sich u.a. in Bishop (2006) sowie Rottensteiner (2016). Darüber hinaus sind zahlreiche anschauliche Online-Tutorials verfügbar (u.a. Bishop, 2013; Koller, 2012; Larochelle, 2013).

### 3.2.1. Grundlagen der Wahrscheinlichkeitstheorie

In die Grundlagen der Wahrscheinlichkeitstheorie und der wichtigsten Begrifflichkeiten wird anhand eines kleinen praktischen Beispiels eingeführt:

*Es ist bekannt, dass in einer Region eine Landbedeckung von 70% Wald und von 30% Acker/Grünland vorliegt. In einem Luftbild wird ein grünes Pixel detektiert. Mit welcher Wahrscheinlichkeit handelt es sich um Wald?*

Die Informationen über die Landbedeckung lassen sich als Wahrscheinlichkeiten $P(x)$ formulieren, wobei $x$ eine Variable ist, die ausschließlich die Werte „Wald“ und „Acker/Grünland“ annehmen kann. Somit ist $P(x=\text{Wald}) = 70\%$, $P(x=\text{Acker/Grünland}) = 30\%$. Diese aus Vorwissen abgeleiteten Werte werden als Prior-Wahrscheinlichkeiten bezeichnet.

Zur Beantwortung der obigen Fragestellung wird allerdings noch Wissen über die Verteilung der Farbwerte innerhalb der jeweiligen Landbedeckungsklassen benötigt. Es wird angenommen, dass beide Landbedeckungsklassen ausschließlich durch grüne und braune Pixel repräsentiert werden.

| | | |
|---|---|---|
| Wald | 90 % grüne Pixel | 10 % braune Pixel |
| Acker/Grünland | 40 % grüne Pixel | 60 % braune Pixel |

Dieses Wissen lässt sich als bedingte Wahrscheinlichkeit $P(y \mid x)$ formulieren, wobei $y$ eine Variable ist, die als Realisierung die Farbwerte annehmen kann. So liegt z.B. die Wahrscheinlichkeit für ein grünes Pixel in der Klasse Wald bei 90%, $P(y=\text{grün} \mid x=\text{Wald}) = 90\%$. Die bedingten Wahrscheinlichkeiten $P(y \mid x)$ werden als Likelihood bezeichnet.

Mit diesem Wissen lässt sich die gemeinsame Wahrscheinlichkeitsverteilung für das Auftreten von zwei Zuständen der Variablen $x$ und $y$ berechnen (Tabelle 3-1), z.B. die Wahrscheinlichkeit brauner Waldpixel:

$$P(x = \text{Wald}, y = \text{braun}) = P(y = \text{braun}|x = \text{Wald})P(x = \text{Wald}) = 0{,}1 * 0{,}7 = 7\%$$

Allgemein formuliert ist dies die Produktregel (3-11).

$$P(x, y) = P(y|x)P(x) \tag{3-11}$$

Anhand Tabelle 3-1 kann auch die zweite fundamentale Regel der Wahrscheinlichkeitstheorie, die Summenregel (3-12) veranschaulicht werden. Die Wahrscheinlichkeit einer Variablen lässt sich aus der gemeinsamen Wahrscheinlichkeitsverteilung wie folgt gewinnen.

$$P(x) = \sum_{y} P(x, y) \tag{3-12}$$

| | grün | braun | $\sum_{y} P(x,y)$ |
|---|---|---|---|
| Wald | 63% | 7% | 70% |
| Acker/Grünland | 12% | 18% | 30% |
| $\sum_{x} P(x,y)$ | 75% | 25% | |

*Tabelle 3-1: Gemeinsame Wahrscheinlichkeitsverteilung sowie daraus abgeleitete Randwahrscheinlichkeiten.*

In Tabelle 3-1 ergeben sich in der letzten Spalte für die Landbedeckung die Prior-Wahrscheinlichkeiten $P(x)$. Für die Farbwerte kann die Wahrscheinlichkeit $P(y)$ für ein Auftreten im Bild, unabhängig von der Zugehörigkeit zu einer Landbedeckungsklasse, bestimmt werden (unterste Zeile). Die Wahrscheinlichkeiten $P(x)$ und $P(y)$ werden allgemein als Randwahrscheinlichkeiten bezeichnet.

Zur Beantwortung der Eingangsfrage wird die Tatsache genutzt, dass die Produktregel symmetrisch ist. Durch Umstellen ergibt sich das Theorem von Bayes (3-13):

$$P(x|y) = \frac{P(y|x)P(x)}{P(y)} \tag{3-13}$$

Mit dem Theorem von Bayes können die a-posteriori Wahrscheinlichkeiten $P(x \mid y)$ für das Auftreten der Variablen $x$ bei beobachteten $y$ durch die Likelihood $P(y \mid x)$ und die Randwahrscheinlichkeiten $P(x)$ und $P(y)$ berechnet werden.

Für das Beispiel liegt somit die Wahrscheinlichkeit, dass es sich bei einem grünen Pixel um Wald handelt, bei 84%

$$P(Wald|grün) = \frac{P(grün|Wald)P(Wald)}{P(grün)} = \frac{0.9 * 0{,}7}{0{,}75} = 84\%$$

Im Theorem von Bayes bewirkt die Division durch die Randwahrscheinlichkeit $P(y)$ eine Normalisierung auf den Wertebereich [0;1].

### 3.2.2. Graphische Modelle zur Modellierung von Wahrscheinlichkeiten

Graphische Modelle bieten die Möglichkeit, komplexe Zusammenhänge der Wahrscheinlichkeitsrechnung bildlich darzustellen. Auf diese Weise können Abhängigkeiten bzw. Unabhängigkeiten in vielen Fällen leichter erkannt werden. Es ist häufig zielführend, in einem ersten Schritt aus dem beabsichtigten Zusammenwirken einzelner Primitive den Graphen zu entwickeln und im zweiten Schritt hieraus die zugehörigen Gleichungen abzuleiten. Der Graph dient im somit als Grundlage der Faktorisierung (siehe unten, Formel (3-15)) der gemeinsamen Verteilung.

Allgemein besteht ein Graph **G** aus Knoten und Kanten. Für eine Graphdarstellung der Relation von Primitiven wird jedes Primitiv von einem Knoten $v$ repräsentiert, die Verbindung zweier Primitive ist eine Kante *e.* Die Menge aller Knoten $v$ ist **V** und die Menge aller Kanten $e$ ist **E**. **G** ist die Menge aller Knoten **V** und Kanten **E**, d.h. **G** = (**V**,**E**).

Die gemeinsame Wahrscheinlichkeit $P(x_1,x_2) = P(x_2 \mid x_1)P(x_1)$ zweier Zufallsvariablen ergibt sich aus dem folgenden Graphen:

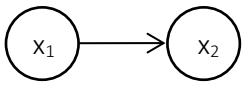



*Abbildung 3-4: Einfacher gerichteter Graph für die Berechnung der gemeinsamen Wahrscheinlichkeit zweier Zufallsvariablen.*

Hierbei handelt es sich um einen gerichteten Graphen, in dem jede Variable $x_i$ durch einen Knoten $v_i$ repräsentiert wird. Die Pfeile zeigen die Abhängigkeiten zwischen zwei Variablen, die Richtung des Pfeils zeigt von der Ursache zur Wirkung. Ist eine Kante $e_{ij}$ von Knoten $v_i$ zu Knoten $v_j$ gerichtet, werden Informationen nur von Knoten $v_i$ zu Knoten $v_j$ übermittelt. Ein Transfer in die entgegen gesetzte Richtung findet nicht statt. In dem konkreten Fall ist $x_1$ nicht kausal abhängig von anderen Variablen, hingegen ist $x_2$ abhängig von $x_1$. Diejenigen Knoten, von denen ein Pfeil ausgeht, d.h. von denen andere Knoten abhängig sind, werden als Elternknoten bezeichnet.

Der Vorteil der Graphdarstellung wird bei komplexeren probabilistischen Modellen offensichtlich:

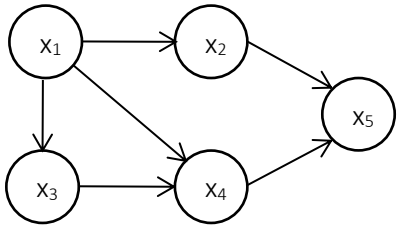



*Abbildung 3-5: Gerichteter Graph für fünf Variablen*

Aus der Struktur des Graphen in Abbildung 3-5 ergibt sich die Wahrscheinlichkeitsverteilung wie folgt:

$$P(x_1, \dots, x_5) = P(x_5|x_2, x_4)P(x_4|x_1, x_3)P(x_3|x_1)P(x_2|x_1)P(x_1) \tag{3-14}$$

Diese Zerlegung in Produkte von Faktoren, die nur noch durch Kanten verbundene Knoten aufweisen, wird als Faktorisierung bezeichnet. Dieser Regel folgende gerichtete Graphen, sogenannte Bayes-Netze, werden vor allem zur Modellierung kausaler und damit auch zeitlichen Abhängigkeiten eingesetzt. Allgemein wird für Bayes-Netze eine Verteilung von *n* Variablen mit $\pi_i$ als Menge aller Elternknoten einer Variablen *i* wie folgt faktorisiert:

$$P(x_1, \dots, x_n) = \prod_{i=1}^{n} P(x_i|\pi_i) \tag{3-15}$$

Bayes-Netze sind grundsätzlich gerichtete azyklische Graphen. Ein Zyklus existiert, wenn eine von einem Knoten $v_i$ ausgesandte Information über weitere Knoten an den Knoten $v_i$ zurückfließt. Graphische Modelle mit Zyklen werden in den folgenden Kapiteln behandelt.

### 3.2.3. Markov Random Fields

Zahlreiche Aufgaben der Bildanalyse erfordern die gegenseitige Interaktion benachbarter Primitive, um räumliches Kontextwissen zu berücksichtigen. Aus diesem Grund sind Bayes-Netze in der Bildanalyse nur begrenzt einsetzbar (eine Ausnahme ist z.B. die Objektverfolgung in Bildsequenzen). Vielmehr bietet sich der Einsatz ungerichteter graphischer Modelle an. In ungerichteten Graphen werden Informationen benachbarter Knoten in beide Richtungen übergeben, d.h. die Primitive sind gegenseitig abhängig.

Abbildung 3-6 stellt einen ungerichteten Graphen dar. In diesem Graphen ist jeder Knoten abhängig von jedem anderen Knoten, entweder in direkter Nachbarschaft oder über weitere dazwischen liegende Knoten. Darüber hinaus ist der Graph zyklisch, so fließt eine von Knoten 1 an Knoten 3 gesandte Information über Knoten 4 wieder in Knoten 1 zurück.

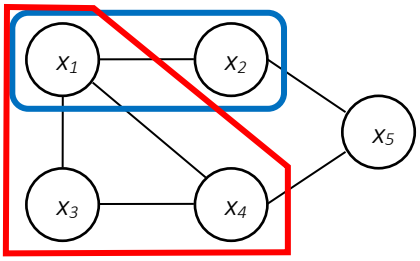



*Abbildung 3-6: Ungerichteter Graph mit zwei Beispielen für maximale Cliquen.*

Die Faktorisierung der Wahrscheinlichkeitsverteilung in ungerichteten Graphen erfolgt über die Betrachtung von Cliquen. Cliquen bestehen aus einer Menge von Knoten, die allesamt direkt über Kanten miteinander verbunden sind. Eine Clique, zu der keine weiteren Knoten hinzugefügt werden können, ohne dass eine neue, größere Clique entsteht, heißt maximale Clique *C*. In Abbildung 3-6 formen die Knoten 1,3 und 4 (roter Rahmen) eine maximale Clique, es gibt in dem Graphen keinen weiteren Knoten, der Verbindungen zu allen Knoten der Clique besitzt. Aus gleichem Grund formen auch die Knoten 1 und 2 (blauer Rahmen) eine maximale Clique.

Die in der Bildanalyse wohl populärste Anwendung von ungerichteten Graphen sind Markov Random Fields (MRF), die im Folgenden beschrieben werden. Die maximalen Cliquen eines Graphen werden durch das Hammersley-Clifford-Theorem (Hammersley & Clifford, 1971; Besag, 1974) zur Faktorisierung der Wahrscheinlichkeitsverteilung genutzt. Das Theorem gilt, wenn die paarweise Markov-Eigenschaft erfüllt ist: Jeder Knoten hängt ausschließlich von den benachbarten Knoten ab. Dies erlaubt eine Beschreibung von Beziehungen in einer lokalen Nachbarschaft anstelle der Notwendigkeit der Definition einer globalen Verteilung. Die Wahrscheinlichkeitsverteilung $P(\mathbf{x})$ ist äquivalent zu einer Gibbs-Verteilung (Geman und Geman, 1984). Hierbei ist $\mathbf{x}$ der Labelvektor, der die einzelnen Label $x_i$ der Primitive $i$ enthält. Merkmal der Gibbs-Verteilung ist, dass die Wahrscheinlichkeitsverteilung als Produkt von Potentialfunktionen $\psi_C(\mathbf{x}_C)$ über den maximalen Cliquen $C$ des Graphen ausgedrückt werden kann.

Allgemein ergibt sich für die Faktorisierung von $P(\mathbf{x})$ in ungerichteten Graphen:

$$P(\mathbf{x}) = \frac{1}{Z}\prod_C \psi_C(\mathbf{x}_C) \tag{3-16}$$

Die Potentialfunktionen können im Prinzip willkürlich gewählt werden und erlauben in vielen Fällen keine direkte probabilistische Interpretation. Der Übergang zu Wahrscheinlichkeiten erfolgt durch Normalisierung des Produktes der Potentialfunktionen auf den Wertebereich [0;1] durch Division durch $Z$. Die Normalisierungskonstante $Z$ wird auch als partition function bezeichnet.

$$Z = \sum_x \prod_C \psi_C(\mathbf{x}_C) \tag{3-17}$$

Die Bestimmung von $Z$ ist eine wesentliche Herausforderung beim Umgang mit ungerichteten Graphen, da theoretisch die Summe über alle Kombinationen sämtlicher Variablen zu berechnen ist. Aus diesem Grund erfolgt die Berechnung zumeist über Näherungsverfahren oder es wird darauf verzichtet.

Die Potentialfunktionen $\psi_C(\mathbf{x}_C)$ sind auf den positiven Wertebereich beschränkt (Bishop, 2006) und werden oft zweckmäßig als Exponentialfunktionen mit der Energiefunktion $E(\mathbf{x}_C)$ formuliert.

$$\psi_C(\mathbf{x}_C) = \exp\big(-E(\mathbf{x}_C)\big) \tag{3-18}$$

Somit wird die Wahrscheinlichkeit $P(\mathbf{x})$ umso höher, je geringer die Gesamtenergie ist. Da $e^x e^y = e^{x+y}$, ist das Produkt von Potentialen, die entsprechend Formel (3-18) beschrieben werden, äquivalent zu einer Summe der einzelnen Energien zur Basis e.

$$P(\mathbf{x}) = \frac{1}{Z} exp\left(\sum -E(\mathbf{x}_C)\right) \tag{3-19}$$

Bei der Klassifikation von Bilddaten wird ein Label $x_i$ eines Primitivs $i$ aus den Daten $\mathbf{y}_i$ abgeleitet. Darüber hinaus ist es vorteilhaft, Kontextwissen benachbarter Primitive zu nutzen. Benachbarte Primitive weisen in Bildern häufig identische Labels auf. Für eine Klassifikation von Bilddaten unter diesen Annahmen sind zwei Arten von Cliquen zu berücksichtigen (Abbildung 3-7). Die vertikalen Kanten beschreiben die Abhängigkeit der Labels $x_i$ von den zugehörigen Daten $\mathbf{y}_i$. Die übrigen Kanten zeigen die Beziehungen benachbarter Labels.

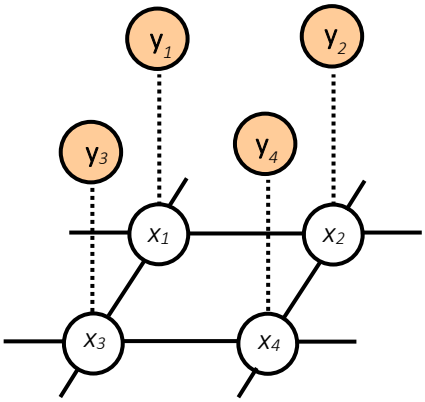



*Abbildung 3-7: Markov Random Field mit Daten $\mathbf{y}_i$ und Labels $x_i$ (ähnlich Bishop, 2006)*

Das Modell in Abbildung 3-7 ist auf maximale Cliquen von lediglich zwei benachbarten Knoten beschränkt. Für solche paarweisen Modelle kann der Energieterm als Summe über zwei Gruppen von Termen formuliert werden. Der erste Term drückt die Abhängigkeit der Labels von den Daten aus, der zweite Term modelliert die Beziehungen benachbarter Labels.

$$E(\mathbf{x}, \mathbf{y}) = \sum_{i \in S} U_i(x_i, \mathbf{y}_i) + \sum_{i \in S} \sum_{j \in N_i} B_{ij}\big(x_i, x_j\big) \tag{3-20}$$

Hierbei ist **x** der Labelvektor, **y** der Vektor sämtlicher Daten, *S* die Menge aller Primitive und *j* ein Primitiv in der Nachbarschaft $N_i$ eines betrachteten Knotens *i*. $U_i(x_i,\mathbf{y}_i)$ wird als unäres Potential oder Knotenpotential bezeichnet und kann beliebig gewählt werden. Oft werden Wahrscheinlichkeiten verwendet, so dass dieser Term als (Log-)Likelihood angesehen werden kann. $B_{ij}$ wird als binäres oder Kantenpotential bezeichnet und kann in einer Bayesschen Betrachtungsweise als Prior-Wissen interpretiert werden. Im Gegensatz zu $U_i$ ist $B_{ij}$ datenunabhängig. Als Standard hat sich für die Verwendung bei Zweiklassen-Problemen das Ising-Modell (Ising, 1925), bzw. bei Mehrklassen-Problemen das Potts-Modell (Potts, 1952) etabliert. Bei diesen Modellen erhalten zwei benachbarte Primitive, die unterschiedlichen Klassen angehören, eine Strafe (penalty). Diese Strafe erhöht die Gesamtenergie. Identische Klassen der Primitive bleiben straffrei oder werden favorisiert. Indem identische Labels interagierender Primitive bevorzugt werden, bewirkt die Anwendung der Modelle eine Glättung des Ergebnisses.

### 3.2.4. Conditional Random Fields

Conditional Random Fields (CRF) sind eng verwandt mit Markov Random Fields, weisen jedoch einige zur Klassifikation von Bilddaten vorteilhafte Unterschiede auf. Sie wurden ursprünglich für die Analyse von Textsequenzen entwickelt (Lafferty et al., 2001) und später auf den zweidimensionalen Fall für die Bildanalyse erweitert (Kumar und Hebert, 2003; Kumar und Herbert, 2006). Für paarweise Modelle ergibt sich der Energieterm erneut als Summe zweier Gruppen von Termen:

$$E(\mathbf{x},\mathbf{y}) = \sum_{i\in S} A_i(x_i,\mathbf{y}) + \sum_{i\in S}\sum_{j\in N_i} I_{ij}(x_i,x_j,\mathbf{y}) \tag{3-21}$$

Abbildung 3-8 zeigt den zugehörigen Graphen.

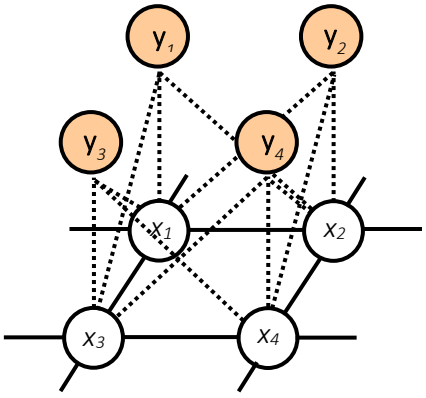



*Abbildung 3-8: Conditional Random Field (ähnlich Bishop, 2006)*

Die beiden Energieterme des CRF sind auf den ersten Blick ähnlich denen eines MRF, weisen jedoch bedeutsame Unterschiede auf.

- Der erste Term, das Knotenpotential, wird bei CRF als Assoziationspotential $A_i$ bezeichnet. Er gibt an, wie wahrscheinlich (auch wenn es sich streng genommen nicht um Wahrscheinlichkeiten handeln muss) ein Knoten *i* das Label $x_i$ für gegebene Daten **y** erhält. Im Gegensatz zu MRF ist $x_i$

theoretisch abhängig von sämtlichen Bilddaten $\mathbf{y}$, nicht nur denjenigen an der Position des betrachteten Primitivs *i*. Folglich fehlt beim Merkmalsvektor $\mathbf{y}$ der Index. Merkmale $\mathbf{y}_i$ können sich somit auch auf Label $x_j$ auswirken (Abbildung 3-8).

- Der zweite Term, das Kantenpotential, wird bei CRF als Interaktionspotential $I_{ij}$ bezeichnet. Er beschreibt, wie benachbarte Labels $x_i$ und $x_j$ interagieren. Dieser Term hängt im Unterschied zu MRF sowohl von den Labels als auch von den Daten $\mathbf{y}$ ab und erlaubt somit eine anspruchsvollere Modellierung nachbarschaftlicher Beziehungen. Auch beim Interaktionspotential ist die globale Einbeziehung sämtlicher Daten möglich.

Die Vorteile dieser Modellierung der Interaktion von Nachbarschaften werden u.a. bei der Betrachtung des für MRF häufig verwendeten Potts-Modells deutlich: Nachteil des (einfachen) Potts-Modells ist, dass die Strafe ausschließlich auf der Verschiedenheit der Labels beruht. Die zugrundeliegenden Merkmale und abgeleiteten Wahrscheinlichkeiten bleiben jedoch unberücksichtigt. Diesem Missstand kann mit dem kontrast-sensitiven Potts-Modell (Boykov et al., 2001) abgeholfen werden. Dieses Modell berücksichtigt auch die Merkmale der benachbarten Primitive. Unterschiedliche Labels benachbarter Primitive werden umso stärker bestraft, je ähnlicher deren Merkmale sind. Unterschiedliche Labels für Primitive mit stark unterschiedlichen Merkmalen werden hingegen nicht oder weniger stark bestraft. Eine detaillierte Betrachtung zur Modellierung räumlicher Interaktionspotentiale erfolgt in Kapitel 4.4.

Ein weiterer bedeutender Unterschied zwischen MRF und CRF ist die Modellierung der Wahrscheinlichkeitsverteilung. Während MRF ein generatives Verfahren sind, in dem die gemeinsame Wahrscheinlichkeit $P(\mathbf{x},\mathbf{y})$ von Labels und Daten bestimmt wird, sind CRF ein diskriminatives Verfahren, d.h. die Posterior-Wahrscheinlichkeit $P(\mathbf{x} \mid \mathbf{y})$ für die Labels $\mathbf{x}$ wird für gegebene Daten $\mathbf{y}$ direkt bestimmt. Somit gilt:

$$P(\mathbf{x}|\mathbf{y}) = \frac{1}{Z} exp\big(-E(\mathbf{x},\mathbf{y})\big) \tag{3-22}$$

Aus (3-21) und (3-22) folgt für die Wahrscheinlichkeitsverteilung eines CRF (Kumar und Hebert, 2006):

$$P(\mathbf{x}|\mathbf{y}) = \frac{1}{Z} exp\left(\sum_{i \in S} A_i(x_i,\mathbf{y}) + \sum_{i \in S}\sum_{j \in N_i} I_{ij}\big(x_i,x_j,\mathbf{y}\big)\right) \tag{3-23}$$

$$Z = \sum_{\mathbf{x}} exp\left(\sum_{i \in S} A_i(x_i,\mathbf{y}) + \sum_{i \in S}\sum_{j \in N_i} I_{ij}\big(x_i,x_j,\mathbf{y}\big)\right) \tag{3-24}$$

Die optimale Konfiguration aller Labels $\mathbf{x}$ wird wie bei MRF über eine Minimierung der Gesamtenergie gefunden.

### 3.2.5. Inferenz

In den vorhergehenden Abschnitten wurde die Modellierung von Zufallsvariablen und deren Abhängigkeiten in Graphen dargestellt. Die Abhängigkeiten der Zufallsvariablen, die zu Interaktionen

benachbarter Knoten in dem Graphischen Modell führen, wurden ganz allgemein durch Energiefunktionen beschrieben.

Abschließend erfolgt die Bestimmung des wahrscheinlichsten Labels $x_i$ (Maximum a posteriori Kriterium – MAP) für ein jedes Primitiv *i.* Das Problem, die wahrscheinlichste Klasse für ein Primitiv zu finden, wird für Zufallsfelder allerdings in der Regel nicht in einer rein lokalen Betrachtungsweise gelöst. Die Lösung erfolgt durch eine simultane Bestimmung des gemeinsamen optimalen Zustands aller Labels **x**. Dieser Prozess wird Inferenz genannt.

Der optimale Zustand aller Variablen wird durch eine globale Minimierung der Energiefunktion ermittelt. Die Minimierung von Energiefunktionen ist ein breites Forschungsfeld, ausführliche Überblicke über die unterschiedlichen Methoden geben u.a. Nocedal und Wright (2006), Bishop (2006) sowie Li (2009). Der grundsätzliche Ansatz ist hierbei in zahlreichen Verfahren identisch (Nocedal und Wright, 2006): Zunächst muss ein Startwert bzw. eine Startkonfiguration ermittelt werden. Idealerweise wird dieser Wert bzw. diese Konfiguration aus den Daten unter Berücksichtigung der Anwendung bestmöglich abgeleitet. Basierend hierauf werden iterativ bessere Lösungen gesucht. Das Kriterium für eine bessere Lösung, im Fall der Bildklassifikation eine wahrscheinlichere Konfiguration der Klassen, ist häufig eine geringere Gesamtenergie. Der Algorithmus wird gestoppt, wenn ein zuvor definiertes Abbruchkriterium, z.B. keine weitere Verbesserung oder eine zuvor definierte Anzahl Iterationen, erreicht wird.

Für ungerichtete Graphen entsprechend Kapitel 3.2.4 (diese dort beschriebenen Graphen sind für die weitere Arbeit relevant) müssen für eine eindeutige Lösung sämtliche Labelkonfigurationen getestet werden. Das ist insbesondere für Graphen mit großen Nachbarschaften einzelner Knoten und einer Vielzahl von Klassen zumeist unmöglich. Zur Berechnung werden Algorithmen verwendet, die Näherungslösungen bieten, jedoch nicht das Auffinden des globalen Minimums garantieren.

Werden ausschließlich die Interaktionen von unmittelbar benachbarten Pixeln berücksichtigt, typischerweise die 4er oder 8er Nachbarschaft in einem Bildraster, existieren für die Inferenz in ungerichteten Graphen effiziente Algorithmen (Schindler, 2012). Die prominentesten unter ihnen sind Graph Cuts (Boykov et al., 2001) und Loopy Belief Propagation (LBP) (Frey und MacKay, 1998). Beide liefern für Markov Random Fields unter Verwendung des Potts-Modells ähnliche Ergebnisse (Tappen und Freeman, 2003; Szeliski et al., 2008). Aufbauend auf einem CRF in Rasterstruktur liefert LBP nachweislich gute Ergebnisse für die Klassifikation von Bilddaten (Vishwanathan et al., 2006). Die maximalen Cliquen in der Arbeit von Vishwanathan et al. (2006) sind 2er-Cliquen, eine Analogie zur Graphstruktur in der vorliegenden Arbeit. Aufgrund dieser Ergebnisse wird in den folgenden Kapiteln LBP für die Inferenz gewählt. Ein weitergehender Vergleich unterschiedlicher Optimierungsalgorithmen steht nicht im Fokus dieser Arbeit.

LBP ist eine Weiterentwicklung von Belief Propagation (Pearl, 1982), einem etablierten Verfahren für die exakte Inferenz in Graphen mit Baumstrukturen. Belief Propagation ist ein Spezialfall des sum-product Algorithmus (Kschischang et al., 2001; Bishop, 2006) und gehört zur Gruppe der message-passing-Verfahren. Der Grundgedanke dieser Verfahren ist, dass zur Minimierung der Energie Nachrichten (messages) von Knoten über Kanten zu benachbarten Knoten übertragen werden.

Die Theorie beruht auf der Betrachtung so genannter Faktorgraphen. Ein Faktorgraph besteht aus Variablenknoten und Faktorknoten. Jeder beliebige ungerichtete Graph kann in einen Faktorgraphen

konvertiert werden. Die Variablenknoten des Faktorgraphen sind dieselben des ungerichteten Graphen. Die Faktorknoten können den maximalen Cliquen entsprechen oder zwischen je zwei benachbarten Knoten angeordnet sein und zeigen die Faktorisierung der Wahrscheinlichkeitsverteilung. Innerhalb des Faktorgraphen werden Nachrichten von den Variablenknoten an die benachbarten Faktorknoten gesendet. Diese sammeln alle eingehenden Nachrichten und senden ihrerseits Nachrichten an die benachbarten Variablenknoten. Ausführliche Erklärungen finden sich u.a. in Bishop (2006).

Interagieren Primitive (und somit Variablenknoten) ausschließlich paarweise, kann die Nachrichtenübertragung für LBP mit CRF vereinfacht ohne die explizite Betrachtung der Faktorknoten veranschaulicht werden. Ziel ist die Ermittlung der wahrscheinlichsten Klasse für jeden Variablenknoten des Graphen. Zu Beginn wird jeder Knoten mit einem Belief-Vektor entsprechend der Klassenwahrscheinlichkeiten des Assoziationspotentials initialisiert. Zwischen den Variablenknoten werden im Folgenden Nachrichten übertragen. Jede Nachricht ist dabei ein Vektor mit der Dimension der möglichen Klassen und berechnet sich entsprechend des Interaktionspotentials, welches gegebenenfalls von den Knotenlabels oder den beobachteten Merkmalen abhängt. In einem CRF mit Rasterstruktur erhält jeder Knoten, der kein Randknoten ist, vier Nachrichten. In Abbildung 3-9 erhält der Knoten 5 Nachrichten der benachbarten Knoten 2, 4, 6 und 8. Durch die eingehenden Nachrichten werden die Beliefs eines jeden Variablenknotens aktualisiert. Ausführlichere Darstellungen finden sich u.a. in Kschischang et al. (2001) sowie Yedidia et al. (2005).

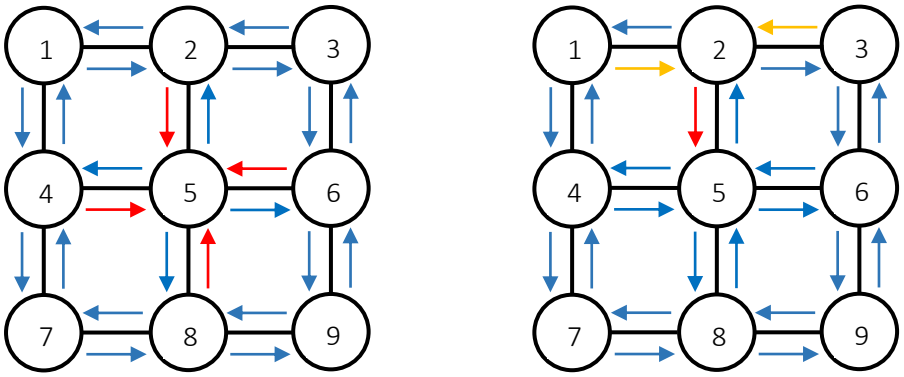



*Abbildung 3-9: Nachrichtenübertragung (message passing) in einem ungerichteten Graphen mit regelmäßiger Gitterstruktur. Links: eingehende Nachrichten für Knoten 5 (rote Pfeile). Rechts: Abhängigkeit einer Nachricht (roter Pfeil) von anderen Nachrichten (orangene Pfeile).*

Im Weiteren läuft der Prozess iterativ: Basierend auf den aktualisierten Belief-Vektoren werden neue Nachrichten zwischen den Knoten übertragen. Diese führen wiederum zu einer erneuten Aktualisierung der Beliefs der Knoten. Damit innerhalb eines Iterationsschrittes eine von einem Knoten ausgesendete Nachricht jedoch nicht sofort die eingehende Nachricht beeinflusst, basiert eine Nachricht von einem Knoten auf allen dort eingehenden Nachrichten, außer auf der Nachricht aus der entgegen gesetzten Richtung. So basiert in Abbildung 3-9 (rechte Darstellung) die Nachricht von Knoten 2 an Knoten 5 auf einer Aktualisierung des Knotens 2 ausschließlich durch die Nachrichten der Knoten 1 und 3. Wird ein definiertes Abbruchkriterium erreicht, z.B. indem der Algorithmus konvergiert, wird der Prozess beendet. Die endgültige Klasse für ein jeweiliges Primitiv ergibt sich entsprechend des maximalen Beliefs.

## 3.3. Bewertungskriterien

Die Bewertung erfasster Fernerkundungsdaten lässt sich in geometrische und thematische Genauigkeit unterteilten (Congalton und Green, 2009). Im ersten Fall werden Abweichungen von Positionen erkannter Objekte von deren wahrer Position berechnet. Im zweiten Fall werden die ermittelten Klassenzugehörigkeiten mit den Klassenzugehörigkeiten einer Referenz an den jeweiligen Positionen verglichen. In dieser Arbeit wird ausschließlich die thematische Genauigkeit untersucht, weshalb im Folgenden unter dem Begriff Genauigkeit stets die thematische Genauigkeit zu verstehen ist.

Zur Berechnung der Genauigkeit wird für jedes Primitiv die als Ergebnis einer automatischen Klassifikation ermittelte Klasse mit einer Referenz verglichen. Die in dieser Arbeit verwendeten Referenzen wurden durch manuelle Digitalisierung von Satellitenbildern gewonnen. Auch solche Referenzen sind nicht absolut fehlerfrei und bilden die wahren Klassenzugehörigkeiten nicht perfekt ab. Typische Fehlerquellen sind z.B. die ungenaue Erfassung von Gebietsgrenzen oder unvollständige Klassendefinitionen (Foody, 2010). Im Rahmen dieser Arbeit wird jedoch angenommen, dass die Anzahl von Fehlzuordnungen in der Referenz deutlich geringer ist als in den Klassifikationsergebnissen. Somit können Fehler in der Referenz für eine quantitative Evaluation vernachlässig werden.

Die am häufigsten verwendeten Qualitätsmaße lassen sich aus der Konfusionsmatrix (Fehlermatrix) ableiten (Foody, 2002). Eine Konfusionsmatrix ist eine quadratische Matrix, deren Elemente die Häufigkeit der Zuordnung klassifizierter Primitive zu den Referenzklassen aufzeigen (Tabelle 3-2). Es wird vorausgesetzt, dass die Klassenanzahl vollständig ist und jedes Primitiv eindeutig einer Klasse zugewiesen werden kann. Vor allem bei Daten mit grober Auflösung wird die Voraussetzung oft verletzt (Lu und Weng, 2007).

| $x^{klass}$ / $x^{ref}$ | $1^{klass}$ | $2^{klass}$ | $3^{klass}$ | Summe |
|---|---|---|---|---|
| $1^{ref}$ | $n_{11}$ | $n_{12}$ | $n_{13}$ | $n_{1+}$ |
| $2^{ref}$ | $n_{21}$ | $n_{22}$ | $n_{23}$ | $n_{2+}$ |
| $3^{ref}$ | $n_{31}$ | $n_{32}$ | $n_{33}$ | $n_{3+}$ |
| Summe | $n_{+1}$ | $n_{+2}$ | $n_{+3}$ | $S$ |

*Tabelle 3-2: Konfusionsmatrix.*

Auf der Hauptdiagonale der Konfusionsmatrix steht die korrekt klassifizierte Anzahl Primitive $n_{cc}$ für eine jeweilige Klasse $c$, $n_{c+}$ ist die Summe aller Primitive einer Referenzklasse, $n_{+c}$ die Summe aller Primitive, die einer Klasse zugeordnet wurden. Setzt man die Summe der Hauptdiagonalelemente in das Verhältnis zur Menge aller Primitive $S$ ergibt sich die Gesamtgenauigkeit.

$$\text{Gesamtgenauigkeit } [\%] = \frac{\sum n_{cc}}{S} * 100 \qquad \in [0; 100] \tag{3-25}$$

Korrekt klassifizierte Primitive einer Klasse werden als True Positives (TP) in Bezug auf diese Klasse bezeichnet. Als Sonderfall werden bei einem Zweiklassenproblem die korrekt klassifizierten Primitive der nicht im Fokus stehenden Klasse als True Negatives (TN) bezeichnet. Typische Zweiklassenprobleme sind die Detektion einer ausgewählten Klasse in einer Szene mit einer Vielzahl von Klassen, z.B. die Detektion von Straßen in Fernerkundungsdaten, oder die Erkennung von Veränderungen ohne weitergehende Analyse der Klassenübergänge.

Die ausschließliche Betrachtung der Gesamtgenauigkeit ist für die Beurteilung einer Klassifikation in den meisten Fällen nicht ausreichend. Vielmehr interessiert, wie gut die einzelnen Klassen detektiert werden können. Aussagekräftige Maße hierfür sind die Vollständigkeit (completeness) und die Richtigkeit (correctness) (Heipke et al., 1997). Die Vollständigkeit sowie die Richtigkeit einer Klasse können evaluiert werden, indem die Klasse gegen die Menge aller anderen Klassen getestet wird. Beide Maße lassen sich aus der Konfusionsmatrix ableiten.

Für jede Referenzklasse geben die Elemente abseits der Hauptdiagonalen der Konfusionsmatrix in der jeweiligen Zeile die Anzahl der Primitive dieser Klasse wieder, die falsch klassifiziert wurden (False Negatives, FN). Der Prozentsatz der Referenzdaten für eine Klasse, der mit dem Klassifikationsergebnis übereinstimmt, ist die Vollständigkeit.

$$\text{Vollständigkeit } [\%] = \frac{TP}{TP + FN} * 100 = \frac{n_{cc}}{n_{c+}} * 100 \qquad \in [0; 100] \tag{3-26}$$

In den Spalten zeigen die Elemente der Konfusionsmatrix abseits der Hauptdiagonalen die Anzahl der Primitive, die diesen Klassen zugeordnet werden, aber in der Referenz anderen Klassen entsprechen (False Positives, FP). Der Prozentsatz der klassifizierten Daten, der mit der Referenz übereinstimmt, ist die Richtigkeit einer Klasse.

$$\text{Richtigkeit } [\%] = \frac{TP}{TP + FP} * 100 = \frac{n_{cc}}{n_{+c}} * 100 \qquad \in [0; 100] \tag{3-27}$$

Eine Analyse der Konfusionsmatrix sollte nicht ausschließlich auf diese häufig genutzten “harten” Genauigkeitsmaße beschränkt werden. Darüber hinaus lassen sich aus der Verteilung der Klassenzugehörigkeiten häufig Muster erkennen, die spezielle Probleme der Klassifikation verdeutlichen. Ein Beispiel ist eine ungenügende Trennung zweier spektral ähnlicher Klassen. Die gewonnenen Erkenntnisse können wiederum zu einer Optimierung der Klassifikationsmethode genutzt werden, z.B. durch Nutzung weiterer, mit Fokus auf das Problem ausgewählter Merkmale.

Eine weiteres Beurteilungskriterium, das sich insbesondere für den Vergleich von Referenz und Klassifikationsergebnis mehrerer (nicht gleichverteilter) Klassen eignet, ist der Kappa Koeffizient $\kappa$ (Cohen, 1960).

$$\kappa = \frac{p_0 - p_c}{1 - p_c} \qquad \in [0; 1] \tag{3-28}$$

Der Wert $p_0$ wird als Übereinstimmung bezeichnet und ist identisch mit der Gesamtgenauigkeit.

$$p_0 = \frac{\sum n_{cc}}{S} \tag{3-29}$$

Der Wert $p_c$ ist die zufällig erwartete Übereinstimmung. Für $p_c$ werden die Produkte der Randsummen der einzelnen Klassen (Zeilensumme und Spaltensumme der Konfusionsmatrix) aufsummiert und ins Verhältnis zum Quadrat der Gesamtsumme gesetzt.

$$p_c = \frac{\sum n_{c+} n_{+c}}{S^2} \tag{3-30}$$

Bei maximaler Übereinstimmung von Klassifikationsergebnis und Referenz nimmt $\kappa$ den Wert 1 an, bei einer Übereinstimmung im Maß des Zufalls den Wert 0. Bei gleichverteilter Klassenhäufigkeit liefert der Kappa Koeffizient keinen Erkenntnisgewinn gegenüber der Betrachtung der Gesamtgenauigkeit. Ist die Häufigkeit der einzelnen Klassen jedoch stark unterschiedlich, ist der Kappa Koeffizient zumeist das aussagekräftigere Beurteilungskriterium. Dies kann anhand eines kleinen Beispiels verdeutlicht werden: Auf einer Wiese steht ein einzelner Baum, der in einer Luftbildaufnahme lediglich ein Prozent der Fläche einnimmt. Wird nun das komplette Bild als Wiese klassifiziert, beträgt die Gesamtgenauigkeit 99%, ein im Allgemeinen hervorragender Wert. Der Fakt, dass eine Klasse komplett falsch klassifiziert wurde, wird nicht deutlich. Der Kappa Koeffizient hingegen reagiert sensibel auf diesen Fehler und liefert den Wert 0.

# 4. CRF zur Klassifikation multitemporaler Bilddaten unterschiedlicher Auflösung

Das Ziel einer Klassifikation von Bilddaten ist die Bestimmung eines optimalen Labelvektors **x**. Jede Komponente $x_i$ von **x** entspricht dabei der ermittelten Klasse des Primitivs $i \in S$ für gegebene Bilddaten **y**. Hierbei ist $S$ die Menge aller Primitive. In vielen Klassifikationsverfahren basiert die Entscheidung für eine Klasse lediglich auf Informationen, die an der Position des jeweiligen Primitivs gewonnen werden (Kapitel 2). Räumliches oder zeitliches Kontextwissen wird selten berücksichtigt. Benachbarte Primitive weisen jedoch häufig Abhängigkeiten auf, die zu einer verbesserten Klassifikation beitragen können. Eine Möglichkeit zur Berücksichtigung von Kontextwissen sind CRF (Kapitel 3.2.4). Die Klassifikation monotemporaler Daten mit CRF erfolgt mit (4-1) durch Maximierung der Wahrscheinlichkeit $P(\mathbf{x}|\mathbf{y})$ (Kumar und Hebert, 2006):

$$P(\mathbf{x}|\mathbf{y}) = \frac{1}{Z} exp\left(\sum_{i \in S} A_i(x_i, \mathbf{y}) + \sum_{i \in S} \sum_{j \in N_i} I_{ij}(x_i, x_j, \mathbf{y})\right) \tag{4-1}$$

In (4-1) ist $N_i$ die Menge interagierender Primitive. In dieser Arbeit ist $N_i$ gleichbedeutend mit der räumlichen Nachbarschaft des Primitivs $i$, d.h. ein Primitiv $j$ ist ein räumlicher Nachbar von $i$. $Z$ ist die in Kapitel 3.2.4 beschriebene Normalisierungskonstante. Das Assoziationspotential $A_i$ stellt die Verbindung zwischen dem Label $x_i$ eines Primitivs $i$ und den Daten **y** her. Das Interaktionspotential $I_{ij}$ modelliert die Abhängigkeiten zwischen den Labels $x_i$ und $x_j$ der benachbarten Primitive $i$ und $j$ und den Daten **y**. Das Modell ist sehr allgemein in Bezug auf die Definition der Terme für $A_i$ und $I_{ij}$.

Wie bereits in Kapitel 3.2 erläutert, ist die Wahl der Primitive prinzipiell willkürlich. Geeignet sind sowohl Pixel als auch regelmäßig oder unregelmäßig geformte Segmente. In dieser Arbeit werden Pixel als Primitive verwendet. Sämtliche entwickelten Methoden sind jedoch bei Anpassung der Graphstruktur problemlos auf Segmente übertragbar. Wenn jedes Pixel einem Primitiv entspricht und mit sämtlichen Pixeln in der 4er-Nachbarschaft interagiert, ist der dem Zufallsfeld zugrunde liegende Graph für die Klassifikation eines Bildes ein rechteckiges, regelmäßiges Gitter. Dabei wird jedes Pixel durch einen Knoten symbolisiert und Kantenbeziehungen existieren zu allen Pixeln in einer 4er-Nachbarschaft (Abbildung 4-1).

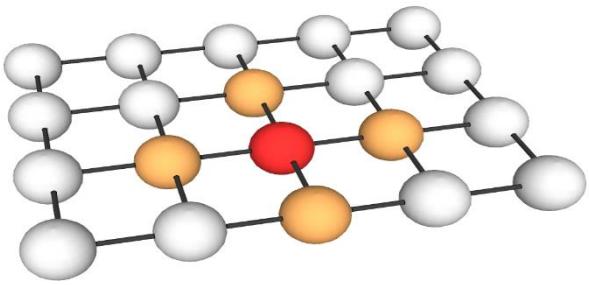


*Abbildung 4-1: Graphstruktur für die Klassifikation eines Bildes auf Pixel-Ebene – jeder Knoten entspricht einem Pixel. Roter Knoten: Betrachtes Pixel; Orange Knoten: Pixel in räumlicher Nachbarschaft zum betrachteten Pixel.*

Die Graphstruktur in Abbildung 4-1 erlaubt die Modellierung räumlicher Beziehungen von Bildinhalten. Wie die einzelnen Knoten interagieren, wird im Term $I_{ij}$ festgelegt. Je nach Definition von $I_{ij}$ kann das Klassifikationsergebnis im einfachen Fall eine Glättung des auszuwertenden Bildes sein. Jedoch sind auch komplexere Interaktionen durch die Integration von Modellwissen möglich. Der Ansatz ist zunächst allgemein und keinesfalls auf die Anwendung von Fernerkundungsdaten limitiert. Im Folgenden liegt der Fokus jedoch auf der Auswertung von Fernerkundungsdaten.

Bei einer multitemporalen Analyse liegen $M$ koregistrierte Bilder einer identischen Region zu unterschiedlichen Zeitpunkten vor. Zusätzlich zur Berücksichtigung räumlicher Interaktionen kann in diesen Daten zeitliche Kontextinformation zur Verbesserung der Klassifikationsgenauigkeit genutzt werden. Im Folgenden wird ein neuer Ansatz zur Nutzung sowohl räumlicher als auch zeitlicher Kontextinformationen basierend auf CRF entwickelt. Dies erfordert eine Erweiterung der beschriebenen CRF-Grundstruktur um eine zeitliche Komponente. Aus Sicht der Anwendung liegt die Neuartigkeit des Ansatzes in der Möglichkeit, in einer integrierten Auswertung sowohl die Klassifikationsgenauigkeit sämtlicher Bilddaten zu erhöhen als auch Veränderungen der Landbedeckung zu erkennen. In Kapitel 4.1 wird zunächst die Auswertung von Bildern identischer Auflösung behandelt. Dieser Ansatz wird in Kapitel 4.2 für die Auswertung von Bildern unterschiedlicher Auflösung erweitert. Die Bedeutung des Assoziationspotentials, des räumlichen Interaktionspotentials sowie eines neu eingeführten zeitlichen Interaktionspotentials wird in den Kapiteln 4.3 bis 4.5 vertiefend erläutert. Zudem erfolgt die Modellierung dieser Potentiale als Grundlage für die weiteren Untersuchungen dieser Arbeit.

## 4.1. Multitemporale Klassifikation von Bildern identischer Auflösung

Liegen Bilder identischer Auflösung vor, deren Pixelabgrenzungen im Objektraum deckungsgleich sind, existiert zu jedem Pixel eines Bildes genau ein korrespondierendes Pixel in sämtlichen anderen Bildern.

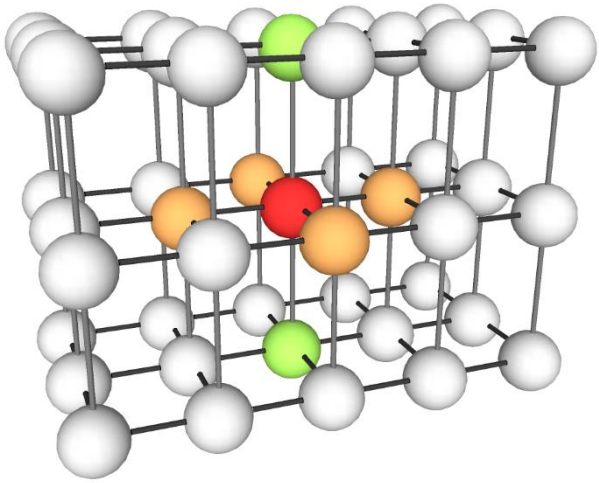


*Abbildung 4-2: Graphstruktur für die Klassifikation mehrerer Bilder mit identischer geometrischer Auflösung. Roter Knoten: Betrachtes Pixel; Orange Knoten: Pixel in räumlicher Nachbarschaft zum betrachteten Pixel; Grüne Knoten: Pixel in zeitlicher Nachbarschaft zum betrachteten Pixel.*

Werden die Bilder in der Reihenfolge ihrer Aufnahmezeitpunkte angeordnet, kann für jedes Pixel die Beziehung zum entsprechenden Pixel des vorangegangenen und des nachfolgenden Zeitpunkts modelliert werden. Abbildung 4-2 zeigt die resultierende Graphstruktur. Die horizontal angeordneten Gitterstrukturen repräsentieren die Bilder der jeweiligen Zeitpunkte. Durch die vertikalen Kanten werden zeitliche Nachbarschaften definiert. Die Graphstruktur in Abbildung 4-2 erfordert eine Erweiterung des Modells in (4-1). Hierzu gelten im Weiteren folgende Festlegungen: Die Komponenten des Datenvektors $\mathbf{y}$ sind die Datenvektoren der einzelnen Primitive $\mathbf{y}_i^t$. Dabei ist $i \in S$ und $S$ ist die Menge aller Primitive sämtlicher Bilder, d.h. $i$ entspricht nicht einer bestimmten räumlichen Position, sondern einem Primitiv in einem Bild. Primitive an der identischen Position, jedoch in unterschiedlichen Bildern erhalten folglich einen unterschiedlichen Index $i$. Der Index $t$ zeigt die Zugehörigkeit eines Primitivs $i$ zu einem Aufnahmezeitpunkt $t \in T$, wobei $T = \{1,\dots M\}$. Die Komponenten des Labelvektors $\mathbf{x}$ sind die Labels $x_i^t$ der Primitive $i$, ebenfalls mit dem Index des Aufnahmezeitpunkts $t$. Ziel ist die Bestimmung der Klasse $x_i^t$ für jedes Primitiv $i$ aus einer Menge vordefinierter Klassen. Generell können die Klassenstruktur und die Anzahl der Klassen zu den unterschiedlichen Aufnahmezeitpunkten variieren. Für Bilder identischer Auflösung ist dies jedoch zumeist nicht sinnvoll, weshalb an dieser Stelle identische Klassenstrukturen für alle Bilder angenommen werden. Um sowohl räumliche als auch zeitliche Beziehungen der Primitive modellieren zu können, wird das Modell für $P(\mathbf{x}|\mathbf{y})$ aus (4-1) wie folgt erweitert:

$$P(\mathbf{x}|\mathbf{y}) = \frac{1}{Z} exp\left(\sum_{i \in S} A\left(x_i^t, \mathbf{y}^t\right) + \sum_{i \in S}\sum_{j \in N_i} IS\left(x_i^t, x_j^t, \mathbf{y}^t\right) + \sum_{i \in S}\sum_{k \in K_t} IT^{tk}\left(x_i^t, x_l^k, \mathbf{y}^t, \mathbf{y}^k\right)\right) \tag{4-2}$$

In (4-2) ist $A$ ist das Assoziationspotential, $IS$ das räumliche Interaktionspotential, welches dem Interaktionspotential $I_{ij}$ aus (4-1) entspricht, und $IT^{tk}$ das zeitliche Interaktionspotential. Im Unterschied zu (4-1) wird in (4-2) auf den Index des betrachteten Primitivs bei den Potentialen verzichtet. Dies ist erlaubt, wenn während eines Experiments stets für sämtliche Primitive die identischen funktionalen Modelle für das Assoziationspotential $A$ und die Interaktionspotentiale $IS$ und $IT^{tk}$ gewählt werden. In diesem Fall handelt es sich um ein homogenes und isotropes Zufallsfeld. Die getroffene Annahme ist für die weitere Arbeit gültig. Die Modellierung des Assoziationspotentials für die Experimente in dieser Arbeit wird in Kapitel 4.3 beschrieben.

Das räumliche Interaktionspotential $IS$ dient der Modellierung räumlicher Beziehungen zu den jeweiligen Aufnahmezeitpunkten. Hierbei interagieren ausschließlich Labels und Merkmale eines Bildes zum Zeitpunkt $t$. Die Modellierung des räumlichen Interaktionspotentials wird in Kapitel 4.4 beschrieben. Beim zeitlichen Interaktionspotential $IT^{tk}$ repräsentieren $\mathbf{y}^t$ und $\mathbf{y}^k$ Bilder, die zu den Epochen $t$ und $k$ aufgenommen wurden. Hierbei ist $k$ der zeitliche Index für einen Aufnahmezeitpunkt in der zeitlichen Nachbarschaft von $t$. $K_t$ ist die Menge aller Zeitpunkte, mit denen ein Primitiv $i$ aus $t$ interagiert. Für jeden Zeitpunkt $k$ interagiert ein Primitiv $i$ aus $t$ mit genau einem weiteren Primitiv $l$ an der identischen räumlichen Position. Das zeitliche Interaktionspotential $IT^{tk}$ modelliert die Beziehungen der Labels und Daten aufeinanderfolgender Zeitpunkte. Je nach Wahl von $IT^{tk}$ kann der Term z.B. eine glättende Wirkung über sämtliche Zeitpunkte haben oder aufgrund von Modellannahmen ausgewählte Klassenübergänge fördern. Weitere Informationen zur Modellierung des räumlichen Interaktionspotentials $IT^{tk}$ folgen in Kapitel 4.4.

## 4.2. Multitemporale Klassifikation von Bildern unterschiedlicher Auflösung

In vielen Fällen haben die zur Verfügung stehenden Daten unterschiedliche räumliche Auflösungen. Dies kann unterschiedliche Gründe haben: Aus wirtschaftlichen Gründen wird auf die Anschaffung ausschließlich hochauflösender Daten verzichtet oder es sind keine multitemporalen Daten eines Sensors für eine Region zu den benötigten Zeitpunkten verfügbar. Dennoch ist es erstrebenswert, Informationen aus sämtlichen Bilddaten zu extrahieren, zu verknüpfen und somit die Klassifikationsgenauigkeit zu steigern.

Aufgrund der unterschiedlich hohen Bodenauflösung der Bilder entspricht ein Primitiv *i* zum Zeitpunkt *t* nicht genau einem anderen Primitiv *l* zum Zeitpunkt *k* an der identischen Position. Vielmehr hat ein Primitiv *i* eines Bildes einer niedrigeren Auflösung mehrere zeitliche Nachbarprimitive *l*, die in einem höher aufgelösten Bild die identische Fläche abdecken. Abbildung 4-3 verdeutlicht die resultierende Graphstruktur für zwei Zeitpunkte, wobei die Auflösung des ersten Zeitpunkts doppelt so hoch ist, wie jene des zweiten Zeitpunkts. Als Folge hat jeder Knoten des zweiten Zeitpunkts vier zeitliche Nachbarknoten aus dem ersten Zeitpunkt.

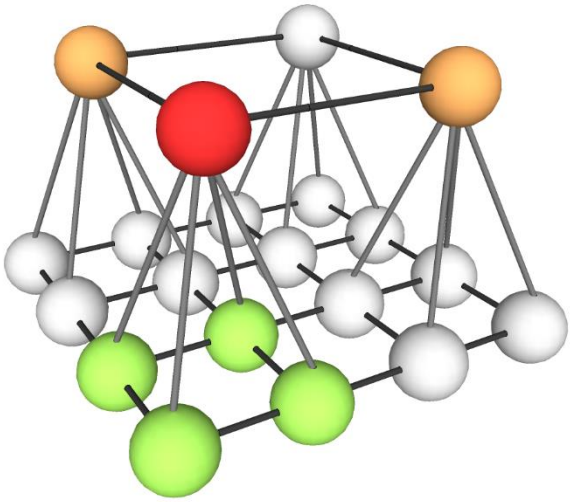


*Abbildung 4-3: Graphstruktur für die Klassifikation mehrerer Bilder mit unterschiedlicher Auflösung. Roter Knoten: Betrachtes Pixel; Orange Knoten: Pixel in räumlicher Nachbarschaft des betrachteten Knotens; Grüne Knoten: Pixel in zeitlicher Nachbarschaft des betrachteten Knotens.*

Bei vielen praktischen Anwendungen ist der Unterschied der Auflösung deutlich größer, z.B. hat bei der kombinierten Auswertung von Landsat-Daten (30m Auflösung) mit RapidEye-Daten (5m Auflösung) ein Knoten des Landsat-Bildes 36 zeitliche Nachbarn. Des Weiteren liegen die Pixelgrenzen im Objektraum zumeist nicht deckungsgleich, selbst wenn die Auflösung eines Bildes einem ganzzahligen Vielfachen der Auflösung eines zweiten Bildes entspricht. Aus diesem Grund kann auch ein Knoten in höherer Auflösung mehrere zeitliche Nachbarn in der geringeren Auflösungsstufe haben.

Unterschiedliche Auflösungen können durch eine Erweiterung von (4-2) zur Modellierung von $P(\mathbf{x}|\mathbf{y})$ berücksichtigt werden:

$$P(\mathbf{x}|\mathbf{y}) = \frac{1}{Z} exp\left(\sum_{i \in S} A\left(x_i^t, \mathbf{y}^t\right) + \sum_{i \in S} \sum_{j \in N_i} IS\left(x_i^t, x_j^t, \mathbf{y}^t\right) + \sum_{i \in S} \sum_{k \in K_t} \sum_{l \in L_i^k} IT^{tk}\left(x_i^t, x_l^k, \mathbf{y}^t, \mathbf{y}^k\right)\right) \quad (4\text{-}3)$$

In (4-3) ist $l \in L_i^k$ ein Primitiv zum Aufnahmezeitpunkt *k* in zeitlicher Nachbarschaft zu einem Primitiv *i* eines Bildes zum Aufnahmezeitpunkt *t*. $L_i^k$ ist die Menge aller Primitive *l* zu einem Zeitpunkt *k*, die zumindest teilweise die identische Fläche des Primitivs *i* aus *t* abdecken und zeitlich mit dem Primitiv *i* interagieren. Im Unterschied zu (4-2) ist die Menge aller Primitive *l* in (4-3) nicht auf lediglich ein Primitiv beschränkt.

(4-3) erlaubt den Informationstransfer zwischen beliebig vielen Bildern unterschiedlicher Auflösung. Es gibt keine Restriktionen, in welcher zeitlichen Reihenfolge die unterschiedlich aufgelösten Bilder stehen. Die Auflösung der zeitlich benachbarten Bilder kann beliebig variieren. Zudem erlaubt dieser Ansatz unterschiedliche Klassenstrukturen zu den einzelnen Zeitpunkten. Diese sind bei der Definition der zeitlichen Interaktionspotentiale $IT^{tk}$ zu berücksichtigen (Kapitel 4.5).

## 4.3. Assoziationspotential

Das Assoziationspotential kann auf unterschiedliche Weise modelliert werden. Es muss lediglich ein Maß für die Zugehörigkeit eines Primitivs *i* zu den definierten Klassen *C* für gegebene Daten $\mathbf{y}$ liefern. So erfüllt ein beliebiger lokaler Klassifikator, der ein probabilistisches Ergebnis liefert, diese Voraussetzung (Kumar und Hebert, 2006). Im Rahmen dieser Arbeit wird das Assoziationspotential $A(x_i^t, \mathbf{y}^t)$ in (4-3) allgemein durch die Wahrscheinlichkeit für eine Zugehörigkeit zu einer Klasse *C* bei gegebenen Bilddaten $\mathbf{y}^t$ zu einem Zeitpunkt *t* ausgedrückt.

$$A\left(x_i^t, \mathbf{y}^t\right) = \log\left\{P\left[x_i^t = C \middle| \mathbf{f}_i^t(\mathbf{y}^t)\right]\right\} \quad (4\text{-}4)$$

Die Bilddaten werden durch die Merkmalsvektoren $\mathbf{f}_i^t$ der Primitive *i* repräsentiert. Diese können lokal gewonnen werden oder abhängig von den kompletten Bilddaten sein, z.B. durch Merkmalsberechnung in unterschiedlichen Skalen (Kapitel 5.3). In dieser Arbeit wird ein multivariates Gauß-Modell für $P[x_i^t=C|\mathbf{f}_i^t(\mathbf{y}^t)]$ verwendet.

$$P\left[x_i^t = C \middle| \mathbf{f}_i^t(\mathbf{y}^t)\right] \propto \frac{1}{\sqrt{(2\pi)^n det\left(\mathbf{\Sigma}_{fc}^t\right)}} e^{-\frac{1}{2}\left[\mathbf{f}_i^t(\mathbf{y}^t) - \mathbf{E}_{fc}^t\right]^T \mathbf{\Sigma}_{fc}^{-1}\left[\mathbf{f}_i^t(\mathbf{y}^t) - \mathbf{E}_{fc}^t\right]} \quad (4\text{-}5)$$

Dabei sind $\mathbf{E}_{fc}^t$ und $\mathbf{\Sigma}_{fc}^t$ die Mittelwerte und die Kovarianzmatrizen der Merkmale einer Klasse *C* zum Zeitpunkt *t*. Es ist zu beachten, dass die Anzahl der möglichen Klassen zu den unterschiedlichen Zeitpunkten variieren kann. Ebenso kann sowohl die Auswahl der Merkmale als auch die Dimension der Merkmalsvektoren $\mathbf{f}_i^t(\mathbf{y}^t)$ zu den unterschiedlichen Zeitpunkten variieren. Diese Flexibilität des Modells ist notwendig, damit für die unterschiedlichen Auflösungsstufen und aufgrund unterschiedlicher

Sensoren eventuell unterschiedlicher verfügbarer Spektralinformationen für jeden Zeitpunkt aussagekräftige Merkmale ausgewählt werden können.

## 4.4. Räumliches Interaktionspotential

Das räumliche Interaktionspotential $IS(x_i^t, x_j^t, \mathbf{y}^t)$ in (4-3) ist ein Maß für den Einfluss der Daten $\mathbf{y}^t$ und für die interagierenden Labels $x_j^t$ auf die Klasse $x_i^t$ des Primitivs $i$ zum Zeitpunkt $t$. Die Daten werden durch Vektoren von Interaktionsmerkmalen $\boldsymbol{\mu}_{ij}^t(\mathbf{y}^t)$ repräsentiert. Die Interaktionsmerkmale werden für die jeweils interagierenden Primitive $i$ und $j$ unabhängig berechnet und können allgemein vom kompletten Bildinhalt zum Zeitpunkt $t$ abhängen.

In dieser Arbeit werden drei unterschiedliche Modelle für das räumliche Interaktionspotential verglichen. Das erste Modell ist das Potts-Modell (Potts, 1952). Die Wirkung hängt ausschließlich von den Labels der interagierenden Primitive ab. Das Potts-Modell „bestraft" unterschiedliche Labels benachbarter Primitive. Im Allgemeinen wird dieses Modell für MRF eingesetzt und bewirkt eine Glättung der Labels (Kapitel 3.2.3 und 3.2.4):

$$IS_1\left(x_i^t, x_j^t, \mathbf{y}^t\right) = \begin{cases} \beta & \text{für} \quad x_i^t = x_j^t \\ 0 & \text{für} \quad x_i^t \neq x_j^t \end{cases} \tag{4-6}$$

Das zweite Modell ist ein kontrast-sensitiven Potts-Modell basierend auf Shotton et al. (2009):

$$IS_2\left(x_i^t, x_j^t, \mathbf{y}^t\right) = \begin{cases} \beta * \exp\left[-\dfrac{\eta\left\|\boldsymbol{\mu}_{ij}^t(\mathbf{y}^t)\right\|^2}{R}\right] & \text{für} \quad x_i^t = x_j^t \\ 0 & \text{für} \quad x_i^t \neq x_j^t \end{cases} \tag{4-7}$$

Das dritte Modell von Hoberg et al. (2010) stellt eine Erweiterung von (4-7) dar:

$$IS_3\left(x_i^t, x_j^t, \mathbf{y}^t\right) = \begin{cases} \beta * \exp\left[-\dfrac{\eta\left\|\boldsymbol{\mu}_{ij}^t(\mathbf{y}^t)\right\|^2}{R}\right] & \text{für} \quad x_i^t = x_j^t \\ \beta * \left\{1 - \exp\left[-\dfrac{\eta\left\|\boldsymbol{\mu}_{ij}^t(\mathbf{y}^t)\right\|^2}{R}\right]\right\} & \text{für} \quad x_i^t \neq x_j^t \end{cases} \tag{4-8}$$

Der Parameter $\beta$ in (4-6), (4-7) und (4-8) ist ein Gewichtsfaktor für den Einfluss des räumlichen Interaktionspotentials auf den Klassifikationsprozess. Für $\beta$=0 entfällt der Einfluss des räumlichen Interaktionspotentials $IS(x_i^t, x_j^t, \mathbf{y}^t)$. Um die Vorteile eines CRF-Ansatzes zu nutzen, sollte $\beta$ mit dem Ziel eines ausgewogenen Einflusses sämtlicher Potentiale gewählt werden. Der Parameter $\eta$ in (4-7) und (4-8) reguliert den Glättungseffekt der räumlichen Interaktionspotentiale, welcher mit zunehmendem $\eta$ geringer wird. Bei Wahl von $\eta$=0 entsprechen die Potentiale $IS_2$ und $IS_3$ dem ausschließlich labelbasierten Potential $IS_1$, hingegen betonen hohe Werte von $\eta$ den Einfluss der Interaktionsmerkmale $\boldsymbol{\mu}_{ij}^t(\mathbf{y}^t)$.

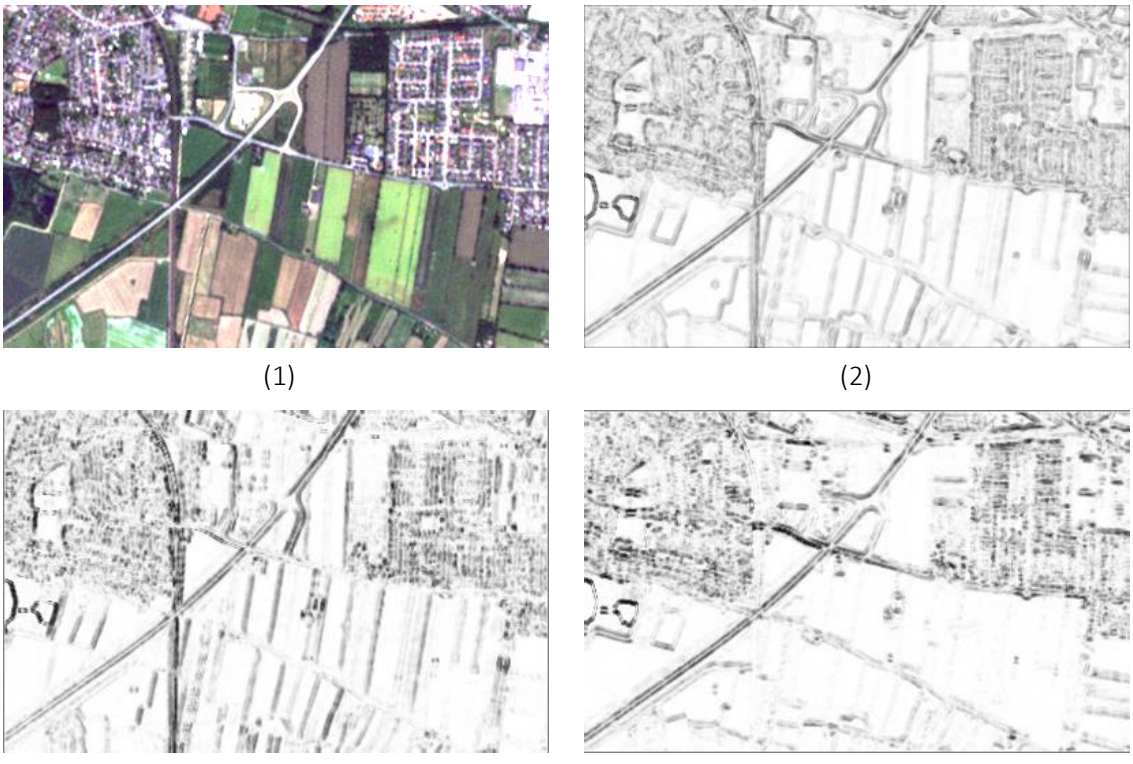
In (4-7) und (4-8) ist $||\boldsymbol{\mu}_{ij}^t(\mathbf{y}^t)||$ die Euklidische Norm von $\boldsymbol{\mu}_{ij}^t(\mathbf{y}^t)$. Für die Interaktionsmerkmale $\boldsymbol{\mu}_{ij}^t(\mathbf{y}^t)$ werden in dieser Arbeit die Differenzen der Komponenten von Merkmalsvektoren $\mathbf{h}_i^t(\mathbf{y}^t)$ und $\mathbf{h}_j^t(\mathbf{y}^t)$, die

für die Knoten berechnet werden, genutzt. Folglich ist $\boldsymbol{\mu}_{ij}^{t}(\mathbf{y}^{t}) = [\mu_{ij1}^{t}, \ldots \mu_{ijR}^{t}]^{\top}$, wobei $R$ die Dimension der Vektoren $\mathbf{h}_{i}^{t}(\mathbf{y}^{t})$ ist. Zu beachten ist, dass der Merkmalsvektor für das räumliche Interaktionspotential $\mathbf{h}_{i}^{t}(\mathbf{y}^{t})$ von dem Merkmalsvektor des Assoziationspotentials $\mathbf{f}_{i}^{t}(\mathbf{y}^{t})$ abweichen kann (Kumar und Hebert, 2006). Mit $m$ als Index werden die einzelnen Komponenten in $\mathbf{h}_{i}^{t}(\mathbf{y}^{t})$ als $h_{im}^{t}(\mathbf{y}^{t})$ ausgedrückt. Für die jeweiligen Komponenten in $\boldsymbol{\mu}_{ij}^{t}(\mathbf{y}^{t})$ ist somit:

$$\mu_{ijm}^{t} = \left| h_{im}^{t}(\mathbf{y}^{t}) - h_{jm}^{t}(\mathbf{y}^{t}) \right| \tag{4-9}$$

Die Dimension $R$ der Vektoren $\mathbf{h}_{i}^{t}(\mathbf{y}^{t})$ kann mit $t$ variieren. Die Division durch $R$ in (4-7) und (4-8) gewährleistet einen identischen Einfluss der räumlichen Interaktionspotentiale zu sämtlichen Zeitpunkten.

Ein wesentlicher Unterschied zwischen CRF und MRF ist die Möglichkeit das Interaktionspotential abhängig von den Daten und somit von den Knotenmerkmalen zu definieren. Die Wahl unterschiedlicher Modelle für das räumliche Interaktionspotential $IS(x_{i}^{t}, x_{j}^{t}, \mathbf{y}^{t})$, die im weiteren Verlauf der Arbeit vereinfacht als $IS_1$, $IS_2$ und $IS_3$ bezeichnet werden, erfolgt, um die resultierenden Differenzen durch die unterschiedliche Form der Integration von Merkmalen aufzuzeigen. Das Modell $IS_1$ ist das zumeist für MRF verwendete Modell. Der resultierende Glättungseffekt wird durch Wahl des Parameters $\beta$ festgelegt. Je größer $\beta$ ist, desto stärker werden identische Labels für interagierende Primitive präferiert.



(1) (2)

(3) (4)

*Abbildung 4-4: Räumliches Interaktionspotential unter Verwendung von $IS_2$. (1) RapidEye-Bild; (2) Summe der Kantenpotentiale je Pixel – helle Werte symbolisieren höhere Präferenz für Glättung; (3) Kantenpotentiale in horizontaler Richtung; (4) Kantenpotentiale in vertikaler Richtung.*

Wie bei $IS_1$ wird auch bei $IS_2$ bei unterschiedlichen Labels der interagierenden Primitive dem Potential der Wert null zugewiesen. Im Gegensatz zu $IS_1$ ist das Potential für identische Labels jedoch abhängig von den Interaktionsmerkmalen $\boldsymbol{\mu}_{ij}^t(\mathbf{y}^t)$. Unterscheiden sich die Merkmale der interagierenden Primitive stark, nähert sich auch $IS_2$ für $x_i^t=x_j^t$ dem Wert Null an. In diesem Fall ist der Einfluss des räumlichen Interaktionspotentials minimal und die Klassifikationsentscheidung basiert ausschließlich auf den weiteren verwendeten Potentialen. Als Folge werden Klassenübergänge interagierender Pixel bewahrt. Sind die Merkmale in $\mathbf{h}_i^t(\mathbf{y}^t)$ und $\mathbf{h}_j^t(\mathbf{y}^t)$ jedoch sehr ähnlich, steigt $IS_2$ für $x_i^t=x_j^t$ an. Die Folge ist eine Präferenz identischer Labels für die interagierenden Primitive.

Abbildung 4-4 zeigt die räumlichen Interaktionspotentiale für $IS_2$ und $x_i^t=x_j^t$ für Pixel in einer 4er-Nachbarschaft. Helle Bereiche resultieren aus ähnlichen Merkmalen benachbarter Pixel im Ausgangsbild. Dunkle Stellen indizieren große Merkmalsdifferenzen. An diesen dunklen Stellen ist die Glättungseigenschaft des räumlichen Interaktionspotentials gering und aus dem Assoziationspotential resultierende Klassenübergänge bleiben erhalten.

$IS_2$ und $IS_3$ unterscheiden sich in der Definition des Potentials für $x_i^t\neq x_j^t$. In $IS_3$ werden unterschiedliche Labels explizit präferiert, wenn die Merkmale von räumlich interagierenden Primitiven große Differenzen aufweisen. Als Folge können Klassenübergänge auch an Stellen entstehen, die aufgrund anderer Potentiale identische Labels bekommen hätten. Strukturen, die sich in den anderen Potentialen nicht ausprägen, können durch das Modell $IS_3$ somit hervorgehoben werden. Allerdings ist eine eindeutige Klassenzuweisung hierbei nicht gewährleistet, da $IS_3$ nur zwischen identischen und nicht identischen interagierenden Labels unterscheidet, aber die jeweiligen Klassen nicht explizit berücksichtigt.

Die Wirkungen der unterschiedlichen Modelle IS werden anhand von drei monotemporalen synthetischen Beispielkonfigurationen in Tabelle 4-1 verdeutlicht. Es interagieren lediglich die unmittelbar benachbarten Primitive. Hierbei besteht ein ausgewogener Einfluss des Assoziationspotentials und des räumlichen Interaktionspotentials auf das Klassifikationsergebnis. Es existieren zwei Klassen $C_1$ und $C_2$, die jeweils durch die Erwartungswerte $E_{C1}$ und $E_{C2}$ von zwei Merkmalen beschreiben werden. Beide Klassen besitzen identische Kovarianzmatrizen, d.h. $\boldsymbol{\Sigma}_{C1}=\boldsymbol{\Sigma}_{C2}$. In Beispiel (1) ähneln sich die Merkmale benachbarter Primitive stark. Basierend auf einer unabhängigen Klassifikation würden den Primitiven dennoch unterschiedliche Klassen zugewiesen. Dieser Effekt findet sich bei ausschließlicher Betrachtung des Assoziationspotentials wieder. Sämtliche räumlichen Interaktionsmodelle propagieren jedoch in diesem konkreten Fall identische Labels. Es erfolgt eine Glättung zugunsten der Klasse mit den insgesamt höheren Werten des Assoziationspotentials, in diesem Fall der Klasse $C_1$. Allein basierend auf dem Assoziationspotential werden auch in Beispiel (2) den benachbarten Primitiven unterschiedliche Labels zugewiesen. In diesem Fall sind die Merkmalsdifferenzen jedoch deutlich größer als in Beispiel (1). In dem ausschließlich labelbasierten Modell $IS_1$ wird dies nicht berücksichtigt; die Labels werden erneut geglättet. Hingegen bleiben bei $IS_2$ und $IS_3$ die Klassenübergänge erhalten. Beispiel (3) behandelt einen weniger häufig auftretenden Fall, der jedoch vor allem bei Klassen, die stark unterschiedliche Ausprägungen aufweisen können, nicht unrealistisch ist. Eine solche Klasse kann z.B. *Acker* sein, die im bestellten und nicht bestellten Zustand unterschiedliche spektrale Eigenschaften hat, oder *Siedlung*, die unterschiedliche Objekte vereint. In Beispiel (3) haben die jeweils benachbarten Primitive große Differenzen in den einzelnen Merkmalen. Mit leicht höherer Wahrscheinlichkeit wird durch das Assoziationspotential jedes einzelne Primitiv der Klasse $C_1$ zugewiesen. Da in $IS_1$ und $IS_2$ die Potentiale für $x_i^t=x_j^t$ stets größer als für $x_i^t\neq x_j^t$ sind, erhalten

unter Verwendung dieser Modelle sämtliche Primitive einheitliche Labels. Das Modell $IS_3$ bevorzugt hingegen unterschiedliche Labels der benachbarten Primitive. Auf diese Weise findet sich die in den ursprünglich in den Daten vorhandene Struktur auch im Ergebnis wieder. Werden ausschließlich die aufgeführten Primitive betrachtet, ist jedoch die Klassenzuweisung nicht eindeutig, so dass im Extremfall sämtliche Primitive falsch klassifiziert werden. Eine korrekte Zuweisung kann nur unter Einbeziehung weiterer interagierender Primitive erfolgen.

$C_1$ : $E_{C_1} = \begin{bmatrix} 0.2 \\ 0.2 \end{bmatrix}$

$C_2$ : $E_{C_2} = \begin{bmatrix} 0.8 \\ 0.8 \end{bmatrix}$

$\Sigma_{C_1} = \Sigma_{C_2}$

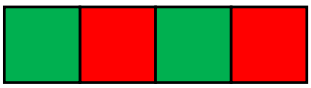
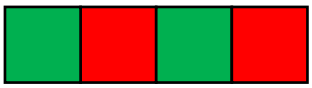
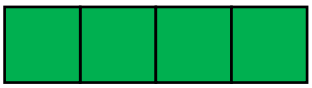
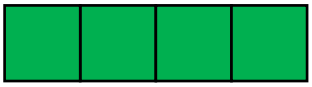
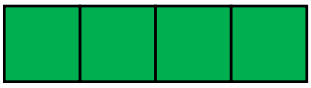
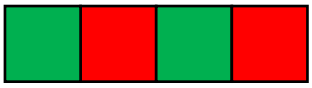
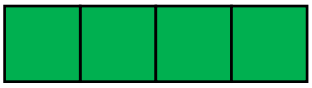
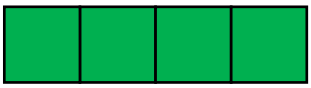
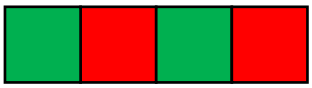
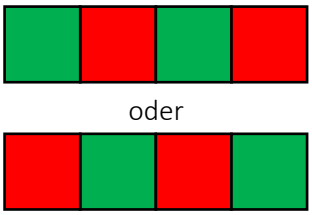
| Beispiel | (1) | (2) | (3) |
|---|---|---|---|
| Merkmale / Potentiale | 0.45 0.52 0.45 0.52<br>0.45 0.52 0.45 0.52 | 0.3 0.65 0.3 0.65<br>0.3 0.65 0.3 0.65 | 0.25 0.7 0.25 0.7<br>0.7 0.25 0.7 0.25 |
| *A* |  |  |  |
| $A + IS_1$ | | |  |
| $A + IS_2$ |  |  |  |
| $A + IS_3$ |  |  |  oder |

*Tabelle 4-1: Wirkung des räumlichen Interaktionspotentials anhand von drei synthetischen Beispielen: (1) Ähnliche Merkmale benachbarter Primitive, die bei ausschließlicher Betrachtung des Assoziationspotentials zu unterschiedlichen Labels führen; (2) Stark unterschiedliche Merkmale, unterschiedliche Klassen bleiben bei Anwendung von $IS_2$ und $IS_3$ erhalten; (3) Stark unterschiedliche Merkmale, unterschiedliche Klassen treten nur durch Anwendung von $IS_3$ hervor.*

## 4.5. Zeitliches Interaktionspotential

Das zeitliche Interaktionspotential $IT^{tk}(x_i^t, x_l^k, \mathbf{y}^t, \mathbf{y}^k)$ modelliert die Abhängigkeiten zwischen den Daten $\mathbf{y}$ und den Labels $x_i^t$ und $x_l^k$ eines Primitivs $i$ zum Zeitpunkt $t$ und eines Primitivs $l$ an der identischen Position zum Zeitpunkt $k$. Generell kann $IT^{tk}$ in ähnlicher Form wie $IS$ modelliert werden, indem zeitliche

Veränderungen der Labels unterdrückt werden, sofern Veränderungen nicht durch Differenzen in den Daten indiziert werden. Allerdings würde dies ein sehr ausgeklügeltes Modell erfordern, mit dem die Effekte durch unterschiedliche atmosphärische Bedingungen und Beleuchtungsverhältnisse der Aufnahmen, unterschiedliche Auflösung sowie saisonal bedingte Ausprägungen der Objekte wie Vegetation kompensiert werden können. Da die Entwicklung eines solchen Modells nicht im Fokus dieser Arbeit steht, wird stattdessen ein einfaches Modell für das zeitliche Interaktionspotential verwendet, welches die Abhängigkeit von den Daten vernachlässigt.

$$IT^{tk}\left(x_i^t, x_l^k, \mathbf{y}^t, \mathbf{y}^k\right) = IT^{tk}\left(x_i^t, x_l^k\right) = \frac{\gamma * \mathbf{TM}_{s(t)s(k)}\left(x_i^t, x_l^k\right)}{Q_i^k} \tag{4-10}$$

In (4-10) ist $\gamma$ ein Gewichtsparameter, der den Einfluss des zeitlichen Interaktionspotentials im Verhältnis zum Assoziationspotential und zum räumlichen Interaktionspotential bestimmt. $\mathbf{TM}_{s(t)s(k)}$ ist eine zeitliche Übergangsmatrix, die proportional einer Matrix der Übergangswahrscheinlichkeiten ist. Die Elemente von $\mathbf{TM}_{s(t)s(k)}(x_i^t, x_l^k)$ sind die zeitlichen Potentiale $PT_{C^t,C^k}$, wobei $C^t$ die Klasse zum Zeitpunkt $t$ und $C^k$ die Klasse zum Zeitpunkt $k$ beschreibt (Tabelle 4-2). Die zeitlichen Potentiale können als bedingte Wahrscheinlichkeiten $P(x_i^t = C^t \mid x_l^k = C^k)$ für die Zugehörigkeit eines Primitivs $i$ zum Zeitpunkt $t$ zu einer Klasse $C^t$ bei Zugehörigkeit des Primitivs $l$ zum Zeitpunkt $k$ zur Klasse $C^k$ angesehen werden. Der Index $s(\cdot)$ in $\mathbf{TM}_{s(t)s(k)}$ beschreibt die Auflösungsstufe der Daten zum betrachteten Zeitpunkt. In einer Auflösungsstufe werden Bilder ähnlicher geometrischer Auflösung, die eine Erkennung identischer Klassen ermöglicht, zusammengefasst. Die Menge der Zeitpunkte $K_t$, mit denen ein Primitiv $i$ zum Zeitpunkt $t$ interagiert, ist generell beliebig. In dieser Arbeit werden der jeweilige vorangehende Zeitpunkt $t$-1 und der nachfolgende Zeitpunkt $t$+1 verwendet. Ist $t$ der erste oder letzte Zeitpunkt der Sequenz, existieren nachbarschaftliche Beziehungen nur zu einem benachbarten Zeitpunkt.

Liegen Bilder unterschiedlicher Auflösung vor, kann ein Primitiv $i$ zum Zeitpunkt $t$ mehr als ein zeitlich interagierendes Primitiv $l$ zum Zeitpunkt $k$ besitzen (Abbildung 4-3). Die Menge der zeitlichen Nachbarn $L_i^k$ enthält sämtliche Primitive $l$ zum Zeitpunkt $k$, die sich räumlich mit $i$ überlappen. Die Anzahl der Elemente in $L_i^k$ ist $Q_i^k$. In (4-10) bewirkt $Q_i^k$ eine Normalisierung, um einen identischen Einfluss der Summe aller zeitlichen Interaktionspotentiale zu allen Zeitpunkten unabhängig von der Anzahl der zeitlichen Nachbarn zu gewährleisten.

Tabelle 4-2 zeigt allgemein eine Übergangsmatrix $\mathbf{TM}_{s(t)s(k)}$ für zwei Zeitpunkte $t$ und $k$ in identischer Auflösungsstufe $s$ und somit identischer Anzahl Klassen $C_s=[C_1,C_2,C_3]$. Für die weiteren Ausführungen wird stets angenommen, dass in einer Auflösungsstufe $s$ die Menge Klassen $C_s$ konstant ist. Die Wahrscheinlichkeit eines Übergangs einer Klasse $C^k$ zu einem Zeitpunkt $k$ zur Klasse $C^t$ zum Zeitpunkt $t$ ist gemäß $PT_{C^t,C^k}$ gegeben. Für den Fall, dass mit großer Wahrscheinlichkeit für alle Klassen keine Veränderungen der einzelnen Primitive zu erwarten sind, d.h. $C^k=C^t$, liegen die höchsten Potentiale $PT_{C^t,C^k}$ auf der Hauptdiagonale von $\mathbf{TM}_{s(t)s(k)}$. Die übrigen Elemente weisen geringere Werte auf und sind gemäß der Wahrscheinlichkeit des jeweiligen Klassenübergangs zu definieren. Eine Konfiguration mit den hohen Potentialen auf der Hauptdiagonale von $\mathbf{TM}_{s(t)s(k)}$ bietet sich zumeist für die Analyse von Landbedeckungen an, da diese häufig unverändert sind. Jedoch sind auch Anwendungen denkbar, in denen Klassenübergänge das wahrscheinlichere Ereignis darstellen. Dies wäre z.B. bei der Analyse von Vegetationszyklen mit den Übergängen zwischen den unterschiedlichen Wachstumsstadien denkbar. Generell bleibt festzuhalten, dass zwischen Bildern identischer Auflösungsstufe die Übergangsmatrix

zwar quadratisch, jedoch nicht symmetrisch ist. So ist z.B. in Fernerkundungsdaten die Veränderung von Ackerflächen zu Siedlungsflächen wahrscheinlicher als umgekehrt.

| $x_i^k$ / $x_i^t$ | $C_1{}^k$ | $C_2{}^k$ | $C_3{}^k$ |
|---|---|---|---|
| $C_1{}^t$ | $PT_{11}$ | $PT_{12}$ | $PT_{13}$ |
| $C_2{}^t$ | $PT_{21}$ | $PT_{22}$ | $PT_{23}$ |
| $C_3{}^t$ | $PT_{31}$ | $PT_{32}$ | $PT_{33}$ |

*Tabelle 4-2: Übergangsmatrix zwischen zwei Zeitpunkten mit Bildern identischer Auflösungsstufe und identischer Klassenstruktur.*

Sofern Bilder unterschiedlicher Auflösungsstufen *s* vorliegen, ändert sich die Erscheinung der Klassen in den Bildern. Dies kann dazu führen, dass einige Klassen nicht mehr trennbar sind, bzw. mehr Klassen unterschieden werden können. Eine Klassifikation mit den in einer anderen Auflösungsstufe definierten Klassenstrukturen ist in vielen Fällen nicht mehr zielführend. Somit ist es vorteilhaft, Klassenstrukturen als Funktion der Auflösung zu definieren. Zu sämtlichen Bildern in einer Auflösungsstufe existiert folglich eine Menge von Klassen $C_s$. Die Relationen der Klassen der verschiedenen Auflösungsstufen können in einem semantischen Netz dargestellt werden. Abbildung 4-5 zeigt ein solches Netz mit 1:*N* Relationen: Die komplette Szene wird in der niedrigsten Auflösungsstufe in zwei Oberklassen $C_1^{lowRes}$ und $C_2^{lowRes}$ unterteilt. Jede dieser Klassen teilt sich in einer höheren Auflösungsstufe wiederum in $N_i$ Unterklassen auf. Zu beachten ist, dass aufgrund der Klassenstrukturen $C_2^{highRes}$ weder die identische Klasse noch eine Unterklasse von $C_2^{lowRes}$ ist.

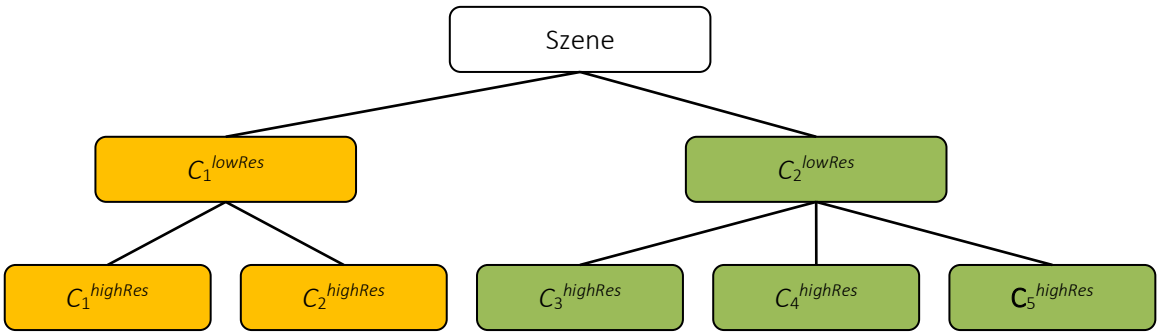



*Abbildung 4-5: Semantisches Netz mit 1:N Relationen.*

Liegt zum Zeitpunkt *t* ein Bild der niedrigeren Auflösungsstufe und zum Zeitpunkt *k* ein Bild der höheren Auflösungsstufe vor, so ist für alle Klassen $C^t=C^{lowRes}$ sowie $C^k=C^{highRes}$. Tabelle 4-3 zeigt die zugehörige Übergangsmatrix.

Die farbig unterlegten Elemente in Tabelle 4-3 verdeutlichen die Relationen aus dem semantischen Netz, wobei davon ausgegangen wird, dass keine Veränderungen der Landbedeckung zwischen den beiden Zeitpunkten existieren. Klassenübergänge kommen in diesem Fall lediglich zwischen den

Auflösungsstufen in der in Abbildung 4-5 abgebildeten Struktur vor, folglich werden nur diese (hier farbig unterlegten) Elemente mit $PT_{C^t,C^k} \neq 0$ belegt. Die 1:*N* Relation bildet sich in der Übergangsmatrix durch die farbige Unterlegung von mehreren Elementen pro Zeile, jedoch nur einem Element pro Spalte ab. Eine Oberklasse, z.B. *Siedlung*, würde in ihre Unterklassen, z.B. *Gebäude*, *Grünfläche* und *Straße*, zerfallen. Generell sind die Übergänge nicht auf die abgebildeten Relationen beschränkt. Bei Veränderungen der Landbedeckung können vielmehr sämtliche Übergänge gemäß ihrer Wahrscheinlichkeit abgebildet werden, so kann bei der multitemporalen Analyse z.B. $C_1^{lowRes}$ durchaus in $C_4^{highRes}$ übergehen.

| $x_i^t$ \ $x_j^k$ | $C_1^k$ | $C_2^k$ | $C_3^k$ | $C_4^k$ | $C_5^k$ |
|---|---|---|---|---|---|
| $C_1^t$ | $PT_{11}$ | $PT_{12}$ | $PT_{13}$ | $PT_{14}$ | $PT_{15}$ |
| $C_2^t$ | $PT_{21}$ | $PT_{22}$ | $PT_{23}$ | $PT_{24}$ | $PT_{25}$ |

*Tabelle 4-3: Übergangsmatrix zwischen zwei Zeitpunkten in unterschiedlichen Auflösungsstufen mit 1:N Relationen.*

Die Modellierung von 1:*N* Relationen ist für viele Klassenstrukturen nicht ausreichend, da eine Unterklasse Bestandteil mehrerer Oberklassen sein kann. Werden Landbedeckungen in Fernerkundungsdaten analysiert, kann z.B. *Baum* eine Unterklasse sowohl der Klasse *Siedlung* als auch der Klasse *Wald* sein. Da die Klasse *Siedlung* wiederum mehrere Unterklassen besitzt, existieren in diesem Fall *M*:*N* Relationen. Ein semantisches Netz mit drei Klassen in der niedrigen Auflösungsstufe und fünf Klassen in der höheren Auflösungsstufe und *M*:*N* Relationen ist in Abbildung 4-6 dargestellt.

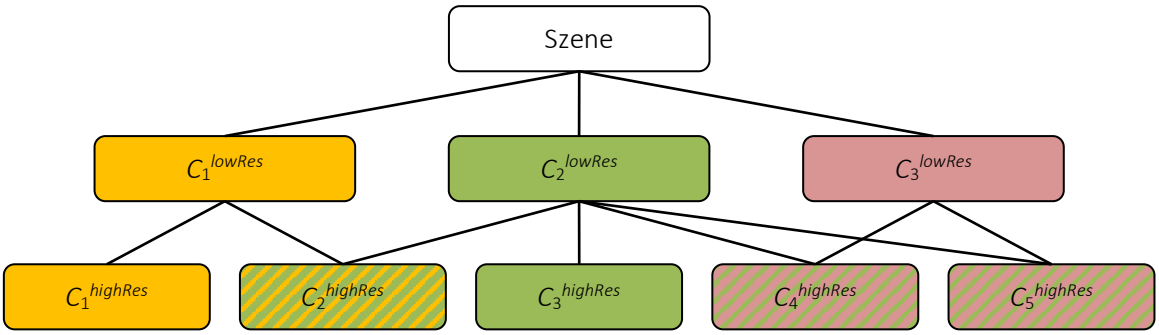



*Abbildung 4-6: Semantisches Netz mit M:N Relationen.*

Tabelle 4-4 zeigt die zugehörige Übergangsmatrix mit $C^t=C^{lowRes}$ sowie $C^k=C^{highRes}$. Die Relationen aus Abbildung 4-6 werden erneut farbig gekennzeichnet und bilden sich durch mehrere eingefärbte Elemente pro Zeile und pro Spalte ab. Selbst wenn nur diese Relationen modelliert werden, gestaltet sich die die Aufgabe, sämtliche Potentiale adäquat festzulegen, bei dieser Konstellation ungleich schwieriger, als bei der ausschließlichen Existenz von 1:*N* Relationen: Sowohl die Wahrscheinlichkeiten für die Teilung einer Oberklasse in Unterklassen als auch die Wahrscheinlichkeiten für die Zugehörigkeiten einer Unterklasse zu den verschiedenen Oberklassen müssen in den Relationen der Potentiale berücksichtigt werden.

| $x_i^t$ \ $x_l^k$ | $C_1^k$ | $C_2^k$ | $C_3^k$ | $C_4^k$ | $C_5^k$ |
|---|---|---|---|---|---|
| $C_1^t$ | $PT_{11}$ | $PT_{12}$ | $PT_{13}$ | $PT_{14}$ | $PT_{15}$ |
| $C_2^t$ | $PT_{21}$ | $PT_{22}$ | $PT_{23}$ | $PT_{24}$ | $PT_{25}$ |
| $C_3^t$ | $PT_{31}$ | $PT_{32}$ | $PT_{33}$ | $PT_{34}$ | $PT_{35}$ |

*Tabelle 4-4: Übergangsmatrix zwischen zwei Zeitpunkten in unterschiedlichen Auflösungsstufen mit M:N Relationen.*

Dieses Modell für das zeitliche Interaktionspotential ermöglicht eine Anpassung an beliebige Klassenstrukturen unterschiedlicher Auflösungsstufen mit den oben genannten Einschränkungen. Für jede Konfiguration von Auflösungsstufen [$s(t)$,$s(k)$] existiert mindestens eine Übergangsmatrix $\mathbf{TM}_{s(t)s(k)}$. Sind die Abstände zwischen den Aufnahmen bei identischer Konfiguration unterschiedlich oder sind z.B. aufgrund der Berücksichtigung jahreszeitlicher Phänomene, die Wahrscheinlichkeiten für Veränderungen abweichend, sind mehrere Übergangsmatrizen bei identischer Konfiguration erforderlich. Dabei wird impliziert, dass Kontextwissen bidirektional übergeben wird. Dies ist ein wesentlicher Unterschied zu zahlreichen Ansätzen, in denen die Information ausschließlich kaskadierend von einem Zeitpunkt zum jeweils nachfolgenden Zeitpunkt der Sequenz übergeben wird (Kapitel 2.2). Zur Informationsweitergabe in entgegen gesetzter Richtung wird die Transponierte der Übergangsmatrix genutzt: $\mathbf{TM}_{s(t)s(k)}=(\mathbf{TM}_{s(k)s(t)})^{\mathrm{T}}$. Die bidirektionale Interaktion gewährleistet, dass auch das Klassifikationsergebnis eines frühen Zeitpunkts durch die Hinzunahme späterer Zeitpunkte verbessert werden kann.

# 5. Experimentelle Untersuchungen

Zu Beginn dieses Kapitels wird die Strategie der experimentellen Untersuchungen (Kapitel 5.1) erläutert. Im Anschluss werden die für die Experimente verwendeten Daten dargestellt. In Kapitel 5.3 und 5.4 erfolgen die Merkmalsextraktion und Merkmalsselektion als Vorstufe zur Klassifikation. Die experimentellen Untersuchen zur Evaluation des Ansatzes aus Kapitel 4 folgen in den Kapiteln 5.5 bis 5.10. Die jeweiligen Ergebnisse werden direkt in den Kapiteln diskutiert.

## 5.1. Strategie

Dieses Kapitel gibt einen Überblick über die Systematik der experimentellen Untersuchungen und beinhaltet das Vorgehen zur Bestimmung der Modellparameter und zur Evaluation der Interaktionsmodelle. Da eine gemeinsame Bestimmung sämtlicher Eingangsgrößen und Parameter für ein bestmögliches Klassifikationsergebnis für die vorgestellte Applikation schwer realisierbar ist und die Wirkung einzelner Parameteränderungen verschwimmt, wird ein hierarchisches Vorgehen für deren Bestimmung präferiert (Abbildung 5-1). Zur detaillierten Beschreibung der einzelnen Schritte wird auf die jeweiligen Abschnitte in den folgenden Kapiteln verwiesen.

Zuerst wird eine Vielzahl von Merkmalen aus den Gruppen der spektralen, texturbasierten und strukturellen Merkmale extrahiert. Die Extraktion erfolgt in unterschiedlichen Skalen, d.h. die Merkmale werden in quadratischen Fenstern unterschiedlicher Größe, welche an der jeweiligen Pixelposition zentriert sind, berechnet (siehe Kapitel 5.3). Da einige dieser Merkmale hochkorreliert sind und einige Merkmale nur mäßig zur Klassifikation des Datenbestandes geeignet sind, wird eine Menge geeigneter Merkmale selektiert. Diese Selektion wird für Merkmalsmengen mit unterschiedlichen maximalen Skalen $\lambda_{max}$ durchgeführt, wobei in $\lambda_{max}$ sämtliche Merkmale der kleineren Skalen enthalten sind. Das Ergebnis dieses Prozesses ist eine Menge selektierter Merkmale $F^{\lambda max}{}_{sel}$ für jede maximale Skala (Kapitel 5.4).

Die Auswirkung der Verwendung der unterschiedlichen Merkmalsmengen $F^{\lambda max}{}_{sel}$ für das Assoziationspotential wird im nächsten Schritt qualitativ evaluiert. Generell ist es das Ziel, eine möglichst gute Klassifikationsgenauigkeit bei ausschließlicher Anwendung des Assoziationspotentials zu erreichen. Jedoch ist mit aufsteigender maximaler Skala $\lambda_{max}$ mit zunehmenden Glättungseffekten zu rechnen. Als Folge werden kleine Strukturen, die sich im Assoziationspotential nicht abbilden, auch in der vollständigen CRF-Klassifikation nicht korrekt dargestellt werden. Somit muss $\lambda_{max}$ für den Merkmalsvektor des Assoziationspotentials unter Berücksichtigung von Klassifikationsgenauigkeit und der Fähigkeit zur Erhaltung feiner Strukturen gewählt werden (Kapitel 5.5).

Mittels monotemporaler Klassifikation werden die Effekte der Variation der Parameter $\theta$ und $\eta$ sowie der maximalen Skala $\lambda_{max}$ für den Merkmalsvektor des räumlichen Interaktionspotentials evaluiert. Der Gewichtsvektor $\beta$ wird so gewählt, dass ein ausgewogener Einfluss von Assoziationspotential und räumlichem Interaktionspotential gewährleistet ist, d.h. dass keines der Potentiale bei der Klassifikationsentscheidung dominant ist. Der Parameter $\eta$ reguliert die glättende Wirkung der merkmalsbasierten räumlichen Interaktionspotentiale und wird ebenfalls unter Berücksichtigung von Klassifikationsgenauigkeit und der Fähigkeit zur Erhaltung feiner Strukturen qualitativ bestimmt. Basierend auf diesen Festlegungen wird die maximale Merkmalsskala $\lambda_{max}$ für die merkmalsbasierten

räumlichen Interaktionspotentiale durch qualitative Analyse festgelegt (Kapitel 5.6). Im Anschluss werden die verfügbaren Testszenen mit den bestimmten Konfigurationen für die unterschiedlichen Modelle des räumlichen Interaktionspotentials monotemporal klassifiziert. Besonderer Fokus liegt hierbei auf der Fähigkeit der Modelle, feine Strukturen korrekt abzubilden (Kapitel 5.7).

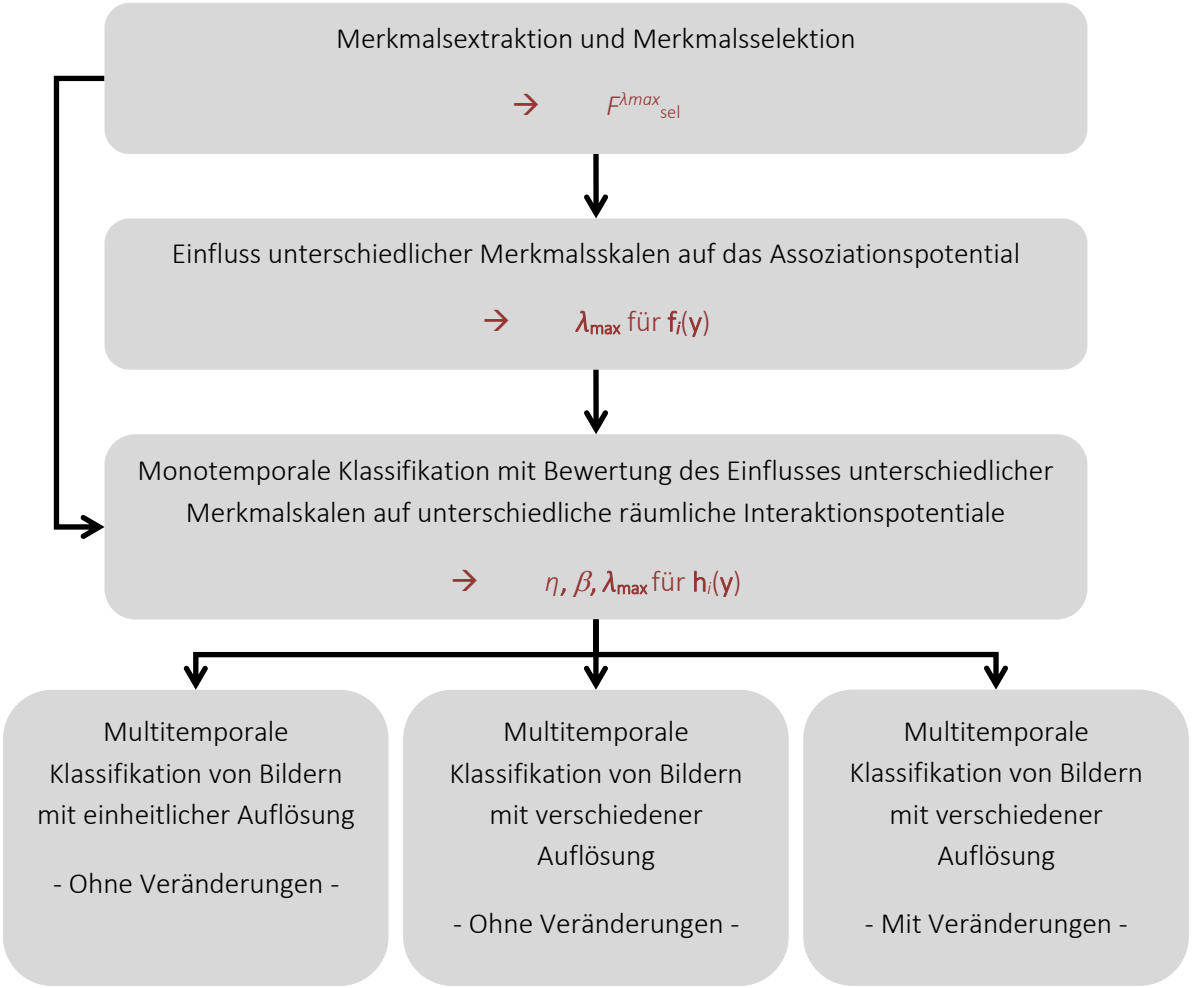



*Abbildung 5-1: Systematik der experimentellen Untersuchungen. Rot: Evaluierte Parameter und Ergebnisse der jeweiligen Untersuchung. Diese werden an den nachfolgenden Auswerteschritt übergeben.*

Aufbauend auf den monotemporalen Untersuchungen folgen Experimente mit multitemporalen Daten, zunächst mit einheitlicher Auflösung (Kapitel 5.8), dann mit unterschiedlicher Auflösung (Kapitel 5.9). In beiden Fällen ist die Landbedeckung unverändert. Im Fokus steht die Fähigkeit der Methode, falsche Klassifikationen zu einem Zeitpunkt durch Nutzung zeitlichen Kontextwissens zu minimieren sowie die Klassifikationsgenauigkeit der Daten sämtlicher Zeitpunkte zu steigern. Hierbei ist von Interesse, ob lediglich Daten niedrigerer Auflösung von der Integration zeitlichen Kontextwissens profitieren, oder ob auch die Klassifikationsgenauigkeit hoch aufgelöster Daten gesteigert werden kann. Ein bedeutender Vorteil der vorgestellten Methode zu vielen anderen Klassifikationsmethoden (Kapitel 2) ist die Fähigkeit, sowohl die Klassifikationsgenauigkeit aller Daten zu erhöhen als auch Veränderungen der Landbedeckung zu detektieren. Folglich bilden Experimente mit Daten unterschiedlicher Auflösung und

Veränderungen der Landbedeckung (Kapitel 5.10) den Abschluss der Experimente. Es wird untersucht, ob sich die theoretischen Annahmen mit Hilfe der Testdaten bestätigen. Zudem wird evaluiert, welchen Einfluss die Wahl unterschiedlicher zeitlicher Übergangsmatrizen auf das Klassifikationsergebnis hat.

## 5.2. Daten und Klassen

Im Rahmen dieser Arbeit werden Daten von drei verschiedenen Fernerkundungssatelliten mit unterschiedlicher geometrischer Auflösung genutzt: zum einen Bilder hoher Auflösung der Satelliten Ikonos und RapidEye, zum anderen Bilder mittlerer Auflösung der Landsat Satelliten. Mit diesen Bildern wird der vorgestellte Ansatz in zwei Testgebieten untersucht. Aus den multispektralen Kanälen des Ikonos Satelliten abgeleitete Bilder haben eine nominale Bodenauflösung von 4 m bei einer maximalen Bodenauflösung des Sensors von 3.28 m. Bilder der RapidEye Satelliten haben eine nominale Bodenauflösung von 5m, wobei die geometrische Auflösung der einzelnen Kanäle nur bei 6.5 m liegt. Die deutlich niedriger aufgelösten Bilder der Landsat (TM) Satelliten werden mit einer Bodenauflösung von 30m geliefert. In allen Fällen hat somit bereits eine Vorprozessierung stattgefunden.

Von sämtlichen Satelliten wird die Information der multispektralen Kanäle im Bereich des blauen, grünen, roten und nahen infraroten Lichts verwendet. Es wurden hierzu die ersten vier Kanäle der Landsat Aufnahmen und die Kanäle 1-3 und 5 der RapidEye Aufnahmen genutzt. Ikonos liefert ausschließlich die aufgeführten Spektralbereiche. Auf eine Nutzung evtl. verfügbarer höher aufgelöster panchromatischer Kanäle wird verzichtet. Unabhängig von der ursprünglichen radiometrischen Auflösung haben sämtliche Bilder eine Farbtiefe von 8bit pro Kanal. Sämtliche Daten sind georeferenziert. Radiometrische Anpassungen der unterschiedlichen Bilder oder atmosphärische Korrekturen wurden nicht angebracht. Da beide Testgebiete annähernd eben sind, mussten zudem keine topographischen Korrekturen vorgenommen werden.

Die Wahl der trennbaren Klassen hängt unmittelbar von der Bildauflösung ab. In Daten hoher Auflösung (Ikonos, RapidEye) bietet sich zur vollständigen Abdeckung der Testgebiete eine Unterscheidung der Klassen *Siedlung*, *Industrie*, *Wald* und *Landwirtschaftliche Flächen* an. Die Klassen *Siedlung* und *Industrie* unterscheiden sich insbesondere durch die Größe der Gebäude und den Anteil versiegelter Flächen. Größere Wasserflächen, die eine eigene Klasse rechtfertigen, sind im Testgebiet nicht vorhanden. *Landwirtschaftliche Flächen* enthalten Acker- und Grünflächen. Im weiteren Verlauf der Arbeit werden landwirtschaftliche Flächen kompakt als Klasse *Acker* bezeichnet. Diese Klasse weist im Jahresverlauf durch die Existenz bestellter und nicht bestellter Ackerflächen sowie unterschiedlicher Anbauarten stark unterschiedliche Erscheinungsformen auf. Eine starke Varianz der Erscheinung ist ebenfalls in der Klasse Siedlung vorhanden, da in dieser Klasse versiegelte Flächen und Vegetationsflächen vereint sind. Zudem werden außerörtliche Straßen in die Klasse Siedlung integriert. In Daten mittlerer Auflösung (Landsat) ist eine Unterscheidung der Klassen Siedlung und Industrie kaum möglich, weshalb diese zu einer gemeinsamen Klasse *bebaute Gebiete* verschmelzen.

Die erreichbaren Klassifikationsgenauigkeiten sind unter anderem abhängig von der Wahl des Trainingsgebiets (Hubert-Moy et al., 2001). Hierbei ist weniger der Umfang der Trainingsdaten, als die Auswahl repräsentativer Gebiete für sämtliche Klassen für den Erfolg einer Klassifikation entscheidend.

Die in dieser Arbeit entwickelten Methoden werden in zwei Testgebieten evaluiert, aus denen die Trainingsgebiete entnommen werden:

- Das Testgebiet Herne (Abbildung 5-2, (1)) liegt in einer industriell geprägten Region im Westen Deutschlands mit dichter Siedlungsstruktur. Der Anteil der einzelnen Klassen an der Gesamtfläche beträgt: Acker 43%, Siedlung 30%, Wald 22% und Industrie 5%. Die Referenz wurde manuell basierend auf einer Ikonos-Szene von 2005 erstellt. Diese Szene wird auch in den Experimenten genutzt. Als Vorgabe wurde definiert, dass eine Siedlung aus mindestens 4 Häusern bestehen soll und ein Waldstück mindestens eine Gruppe von Bäumen umfassen soll. Innerstädtische Wald- und Grünflächen wurden als solche referenziert, wenn diese eindeutig nicht bebauten Grundstücken zuzuordnen waren. Außerörtliche Straßen wurden erfasst, soweit sie eindeutig sichtbar waren, d.h. nur bedeutende Straßen, die eine Mindestbreitebreite von ca. 8-10 Metern (entspricht 2 Pixeln) aufweisen. Zudem wurden Straßenabschnitte, welche großflächig durch Vegetation verdeckt waren, nicht erfasst.

  Für die Experimente liegen multispektrale Ikonos Daten aus den Jahren 2005 und 2007, RapidEye Daten aus 2009 sowie Landsat Daten aus 2010 vor. Das Gebiet umfasst eine Fläche ca. 8,6*8,6 km², d.h. von 2160*2160 Pixeln in Ikonos-Daten mit einer multispektralen Auflösung von 4m und 288*288 Pixeln in Landsat-Daten (Auflösung 30m).

(1) (2)

*Abbildung 5-2: Testgebiete: (1) Herne, Sensor: Ikonos, Zeitpunkt 2005; (2) Husum, Sensor: RapidEye, Zeitpunkt 7/2009.*

- Das Testgebiet Husum (Abbildung 5-2, (2)) liegt in einer landwirtschaftlich geprägten Region Norddeutschlands. Außer der Stadt Husum gibt es nur vereinzelte kleinere und mittelgroße Siedlungen. Der Anteil der einzelnen Klassen an der Gesamtfläche beträgt: Acker 80%, Siedlung 10%, Wald 8,5% und Industrie 1,5%. Die Referenz wurde aus ATKIS-Daten (aktueller Stand: Arbeitsgemeinschaft der Vermessungsverwaltungen der Länder der Bundesrepublik Deutschland (AdV), 2017) abgeleitet. Die Referenz der außerörtlichen Straßen besteht aus sämtlichen Straßenobjekten der höchsten Funktion (ehemals FKT = 2301). Diese in ATKIS linienförmigen Objekte wurden zu flächigen Objekten mit einer Breite von 10 Metern gewandelt, d.h. jede Straße hat in der Referenz eine identische Breite. Auch wenn diese Annahme für sehr

große Straßen (Autobahnen) unzureichend ist, ist sie doch für die meisten Straßen im Rahmen der verfügbaren geometrischen Genauigkeit ausreichend.

Für dieses Testgebiet sind RapidEye Daten zu vier verschiedenen Zeitpunkten (4/2009, 7/2009, 3/2010 und 7/2010) sowie Landsat Daten von 7/2009 vorhanden. Das Gebiet umfasst eine Fläche ca. 16,8*9,6 km², d.h. von 3360*1920 Pixeln in RapidEye-Daten mit einer Auflösung von 5m und 560*320 Pixeln in Landsat-Daten (Auflösung 30m).

## 5.3. Merkmalsextraktion

Basis jeder Klassifikation ist die Wahl geeigneter Merkmale. Zur Anwendung des in Kapitel 4 vorgestellten Verfahrens müssen für jedes Primitiv die Vektoren $\mathbf{f}_i^t(\mathbf{y}^t)$ für das Assoziationspotential und $\mathbf{h}_i^t(\mathbf{y}^t)$ für das räumliche Interaktionspotential zu jedem Zeitpunkt $t$ definiert werden. Beide Vektoren sollten aus Merkmalen bestehen, die die einzelnen Klassen möglichst gut trennen. Die Auswahl der zu extrahierenden Merkmale hängt dabei sowohl von der Auflösung der Daten und den zur Verfügung stehenden Kanälen als auch von den gewählten Klassen ab. Folglich müssen für Bilddaten unterschiedlicher Auflösungsstufen unterschiedliche Merkmale gewählt werden.

Im Allgemeinen kann aus multispektralen Daten eine nahezu unüberschaubare Menge von Merkmalen extrahiert werden (Kapitel 3.1). Da die Merkmale einer Merkmalsgruppe teilweise jedoch ähnliche Charakteristika aufweisen, wird eine Vorauswahl getroffen. Dabei wird darauf geachtet, dass aus den Gruppen der spektralen, texturbasierten und strukturellen Merkmale jeweils repräsentative Merkmale enthalten sind. Eine Kombination von Merkmalen aus unterschiedlichen Gruppen führt im Allgemeinen zu einer deutlichen Steigerung der Klassifikationsgenauigkeit im Vergleich zur Nutzung von zahlreichen Merkmalen aus nur einer Gruppe (Lu und Weng, 2007). Zu diesem Ergebnis kommen auch Aksoy et al. (2010), wobei insbesondere die Berechnung multiskaliger Texturmerkmale zu einer weiteren Steigerung der Klassifikationsgenauigkeit führt. Geometrische Merkmale werden nicht berücksichtigt, da die in dieser Arbeit verwendeten Primitive Pixel sind.

Für Primitive $i$, die nur aus einzelnen Pixeln bestehen, bietet sich eine Extraktion der Merkmale nicht ausschließlich im Bildbereich des betrachteten Primitivs $i$ an, sondern zusätzlich aus einer zu definierenden Nachbarschaft. In dieser Arbeit werden sämtliche Merkmale folglich in fünf unterschiedlichen Merkmalsskalen $\lambda_d$ berechnet, wobei $d$ die jeweilige Skala anzeigt. In der Skala $\lambda_1$ werden die Merkmale lediglich aus den einzelnen Pixelwerten abgeleitet. In den weiteren Skalen werden die Merkmale in einem quadratischen Fenster, welches an der jeweiligen Pixelposition zentriert ist, berechnet. Folgende Seitenlängen der Fenster werden für die Skalen gewählt: $\lambda_2$: drei Pixel; $\lambda_3$: fünf Pixel; $\lambda_4$: neun Pixel; $\lambda_5$: 13 Pixel (Abbildung 5-3). Durch die Nutzung mehrerer Skalen wird folglich nicht nur Information an der jeweiligen Position $i$ berücksichtigt, auch benachbarte Pixel fließen in die Merkmalsberechnung ein. Somit sind bereits räumliche Kontextinformationen in den Merkmalsvektoren $\mathbf{f}_i^t(\mathbf{y}^t)$ und $\mathbf{h}_i^t(\mathbf{y}^t)$ enthalten.

Spektrale Merkmale leiten sich direkt aus den Pixelwerten der einzelnen Spektralkanäle ab. Für sämtliche Sensoren stehen die Daten von vier Kanälen zur Verfügung, so dass die Mittelwerte und Varianzen des roten ($E^d_{\mathrm{r}}$, $V^d_{\mathrm{r}}$), grünen ($E^d_{\mathrm{g}}$, $V^d_{\mathrm{g}}$), blauen ($E^d_{\mathrm{b}}$, $V^d_{\mathrm{b}}$) und des nahen infraroten ($E^d_{\mathrm{nir}}$, $V^d_{\mathrm{nir}}$) Kanals in der jeweiligen Skala $d$ direkt genutzt werden. Zudem wird die Varianz der Farbwerte ($V^d_{\mathrm{hue}}$), der Mittelwert

der Differenz des roten und grünen Kanals ($E^d_{r\text{-}g}$), des nahen infraroten und roten Kanals ($E^d_{nir\text{-}r}$) sowie des nahen infraroten und grünen Kanals ($E^d_{nir\text{-}g}$) berechnet. Durch Kombination der roten und des nahen infraroten Kanals werden der Mittelwert und die Varianz des Normalized Difference Vegetation Index ($E^d_{ndvi}$, $V^d_{ndvi}$ ; Formel (3-2)) und des Ratio Vegetation Index ($E^d_{rvi}$, $V^d_{rvi}$ ; Formel (3-1)) abgeleitet.

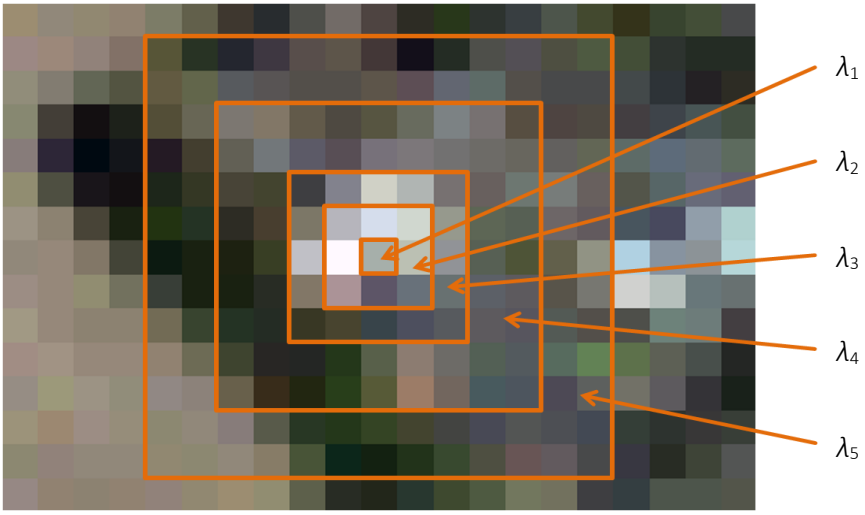



*Abbildung 5-3: Skalen der Merkmalsextraktion. Durch Wahl einer höheren Skala λ wird eine größere Nachbarschaft berücksichtigt.*

Vor allem für hoch aufgelöste Daten bietet sich neben der Erfassung radiometrischer Merkmale die Erfassung texturbasierter und struktureller Merkmale an. Aus der Gruppe der texturbasierten Merkmale werden einige der am häufigsten verwendeten Haralick'schen Texturmaße ausgewählt: Kontrast ($con^d$), Korrelation ($cor^d$), Energie ($ene^d$), Homogenität ($hom^d$) und Entropy ($ent^d$). Diese basieren auf der *grey-level co-occurrence matrix* (GLCM, Kapitel 3.1.1), die die Verteilung benachbart auftretender Werte in einem bestimmten Abstand Δ wiedergibt. In dieser Arbeit wird Δ=1 gesetzt. Um den Einfluss der Orientierung, welche bei Fernerkundungsdaten zumeist nicht aussagekräftig ist, zu beseitigen, werden die abgeleiteten Werte der vier resultierenden GLCM einer 8-Nachbarschaft mit den Orientierungen 0°, 45°, 90° sowie 135° gemittelt.

Die verwendeten strukturellen Merkmale leiten sich aus den gewichteten Histogrammen orientierter Gradienten ab (Kapitel 3.1.1). Für die Berechnung der Gradienten werden die Filtermasken [-1, 1] sowie $[-1, 1]^T$ verwendet. Dalal & Triggs (2005) zeigen, dass die Wahl der Filtermaske die Eignung der Gradientenmerkmale zur Klassifikation nur unwesentlich beeinflusst, solange die Berechnung nicht mit einer Glättung des Bildausschnittes einhergeht. Die Histogramme entstehen durch Ordnung sämtlicher Gradientenmagnituden entsprechend der Gradientenrichtungen innerhalb des jeweiligen Fensters der unterschiedlichen Skalen. Jedes Histogramm hat 30 Wertebereiche, so dass jeder Bereich einem Orientierungsintervall von 6° Breite entspricht. Jeder Wertebereich enthält die Summe der Magnituden aller Gradienten, welche eine Orientierung im Intervall des Wertebereichs aufweisen. Die Summation über alle Magnituden bewahrt den Einfluss starker Gradienten. Fünf aus dem Histogramm abgeleitete Merkmale werden in dieser Arbeit genutzt. Die Intention der Nutzung von Mittelwert $E^d_{grad}$ und Varianz $V^d_{grad}$ aller Magnituden der Wertebereiche sowie die Anzahl Wertebereiche $num^d$ mit Magnituden über der durchschnittlichen Magnitude ist die Erfassung von Ausprägungen typischer

Landbedeckungsstrukturen in den Histogrammen (vgl. Abbildung 3-1). Die Erkennung regelmäßiger künstlich erschaffener Landschaftsstrukturen wird durch Berechnung des Winkels $ang^d$ zwischen den beiden größten Maxima erwartet. Des Weiteren wird der Wert des größten Maximums $mag^d$ genutzt.

Sämtliche aufgeführten Merkmale werden in allen Skalen $\lambda_2$ bis $\lambda_5$ extrahiert. In $\lambda_1$ ist eine Ableitung der texturbasierten und strukturellen Merkmale sowie die Berechnung von Varianzen nicht möglich, so dass in dieser Skala lediglich die Farbwerte und die daraus abgeleiteten Differenzen und Vegetationsindizes verwendet werden. Insgesamt werden 113 Merkmale extrahiert, davon neun in $\lambda_1$ sowie jeweils 26 in den Skalen $\lambda_2$ bis $\lambda_5$. Alle Merkmale werden normalisiert und in einem einheitlichen Intervall [0, 10] abgebildet.

Durch die Merkmalsextraktion in Skalen enthalten die Merkmale lokale räumliche Kontextinformation. Es ist zu erwarten, dass dies generell zu einer erhöhten Klassifikationsgenauigkeit führt, jedoch eventuell auf Kosten einer zu starken Glättung. Diese Annahmen werden in den Kapiteln 5.5 und 5.6 evaluiert. Zu diesem Zweck werden Mengen der Merkmale $F^{\lambda max}{}_{all}$ definiert, die sämtliche Merkmale bis zu einer maximalen Skala $\lambda_{max}$ enthalten, d.h. $F^{\lambda 2}{}_{all}$ mit $\lambda_{max}= \lambda_2$ enthält 35 Merkmale (neun aus $\lambda_1$ sowie 26 aus $\lambda_2$), $F^{\lambda 5}{}_{all}$ sämtliche 113 Merkmale.

## 5.4. Merkmalsselektion

Die Verwendung sämtlicher Merkmale aus Kapitel 5.3 macht eine CRF-Klassifikation aus zwei Gründen sehr rechenintensiv: Zum einen müssen sämtliche Merkmale in einem ersten Schritt extrahiert werden, zum anderen resultiert ein hochdimensionaler Merkmalsvektor für die Bestimmung der Potentiale. Da jedoch viele Merkmale hoch korreliert sind oder eventuell nur geringfügig zur Verbesserung der Klassifikation beitragen, werden die jeweils am besten geeigneten Merkmale aus der Gesamtheit der Merkmale $F^{\lambda max}{}_{all}$ in einer kleinen Testszene selektiert. Nur diese selektierten Merkmale, die in $F^{\lambda max}{}_{sel}$ zusammengefasst werden, werden für die weiteren experimentellen Untersuchungen verwendet.

Die Merkmale werden mittels correlation based feature selection (CFS, Kapitel 3.1.2) selektiert. Zuerst wird das Merkmal ausgewählt, das alleine die höchste Klassifikationsgenauigkeit aufweist. Im Weiteren werden weitere Merkmale nach dem Kriterium der höchsten Korrelation mit den Klassen, jedoch einer möglichst geringen Korrelation untereinander, ausgewählt (Hall, 1999) und der Menge der selektierten Merkmale $F^{\lambda max}{}_{sel}$ zugefügt. Dabei handelt es sich um einen *greedy*-Algorithmus. *Greedy* bedeutet hier, dass Merkmale, die der Menge $F^{\lambda max}{}_{sel}$ zugefügt wurden, stets in dieser verbleiben (Guyon und Elisseeff, 2003). Der Nachteil dieses Ansatzes ist, dass es eventuell eine Kombination von Merkmalen gibt, die besser klassifiziert als die Kombination der Merkmale in $F^{\lambda max}{}_{sel}$ (Liu und Motoda, 2008). Jedoch sind *greedy*-Algorithmen sehr schnell und liefern in den meisten Fällen sehr gute Ergebnisse (Caruana und Freitag, 1994). Tabelle 5-1 gibt einen Überblick über die Reihenfolge der jeweils 20 zuerst selektierten Merkmale für $\lambda_{max}=\lambda_2$ bis $\lambda_{max}= \lambda_5$ unter Verwendung von Ikonos-Daten und den vier in dieser Arbeit verwendeten Klassen *Siedlung*, *Industrie*, *Wald* und *Acker*. Die Analyse basiert auf 512.000 zufällig ausgewählten Pixeln der Testszene Herne von 2005 unter Berücksichtigung einer repräsentativen Verteilung aller Klassen.

Es zeigt sich, dass für $\lambda_{max}=\lambda_2$ bis $\lambda_{max}=\lambda_4$ die Mehrzahl der selektierten Merkmale der höchstmöglichen Skalenstufe angehört. Jedoch werden mit steigendem $\lambda_{max}$ zunehmend auch Merkmale der niedrigen

Skalen selektiert. Zudem decken die zuerst gewählten Merkmale für alle $\lambda_{max}$ sämtliche Merkmalsgruppen ab. Dies ist wenig überraschend, da häufig Merkmale einer Merkmalsgruppe korreliert sind. Diese Beobachtung unterstreicht die Notwendigkeit der Berücksichtigung der unterschiedlichen Merkmalsgruppen für die Klassifikation.

| Merkmal | $\lambda_1$ | $\lambda_2$ | $\lambda_3$ | $\lambda_4$ | $\lambda_5$ |
|---|---|---|---|---|---|
| $E^d_r$ | | | | | |
| $V^d_r$ | | 18 | | | |
| $E^d_g$ | | 7 | 8 15 17 | 4 5 | |
| $V^d_g$ | | 16 | | | |
| $E^d_b$ | 14 13 | | 16 | 18 | 10 |
| $V^d_b$ | | 11 | | | |
| $E^d_{nir}$ | 12 18 | 5 11 17 | 2 14 13 | 5 20 | 3 |
| $V^d_{nir}$ | | 4 10 12 15 | 5 9 8 | 19 | |
| $V^d_{hue}$ | | 3 7 8 14 | 1 6 7 | 1 11 | 1 |
| $E^d_{r-g}$ | | | | | |
| $E^d_{nir-r}$ | 20 | 2 | 4 | 13 | |
| $E^d_{nir-g}$ | | | | | |
| $E^d_{ndvi}$ | | 9 | | | |
| $V^d_{ndvi}$ | | 6 14 | 9 | 10 | |
| $E^d_{rvi}$ | | | | | |
| $V^d_{rvi}$ | | | | 16 | 9 |
| $E^d_{grad}$ | | 1 19 11 12 | 3 | 2 16 | 2 |
| $V^d_{grad}$ | | 10 | | | |
| $num^d$ | | 19 | | | |
| $mag^d$ | | 15 | 15 | | |
| $ang^d$ | | | | | |
| $con^d$ | | | | | |
| $cor^d$ | | 17 20 | 17 | | 19 |
| $ene^d$ | | 13 | | | |
| $hom^d$ | | | | 7 | 6 |
| $ent^d$ | | 8 12 20 | 6 | 3 4 | 18 |

*Tabelle 5-1:* ***Rangfolge der selektierten Merkmale für unterschiedliche Skalen.*** *Rot: $\lambda_{max}=\lambda_2$; Grün: $\lambda_{max}=\lambda_3$; Blau: $\lambda_{max}=\lambda_4$: Schwarz: $\lambda_{max}=\lambda_5$.*

Abbildung 5-4 visualisiert die Punktwolken der drei am besten trennenden Merkmale für $\lambda_{max}=\lambda_3$. Die Merkmale sind relativ gering korreliert, jedoch zeigt sich die große Varianz in den Merkmalen der einzelnen Klassen. Dies lässt sich mit den unterschiedlichen Ausprägungen der einzelnen Klassen und deren zugehörigen Objekten begründen: Z.B. sind in der Klasse Siedlung versiegelte Flächen und

Vegetationsflächen enthalten. Aufgrund der großen Überlappung der Werte liegt die Klassifikationsgenauigkeit einer Maximum-Likelihood Klassifikation mit diesen drei Merkmalen lediglich bei ca. 70%.

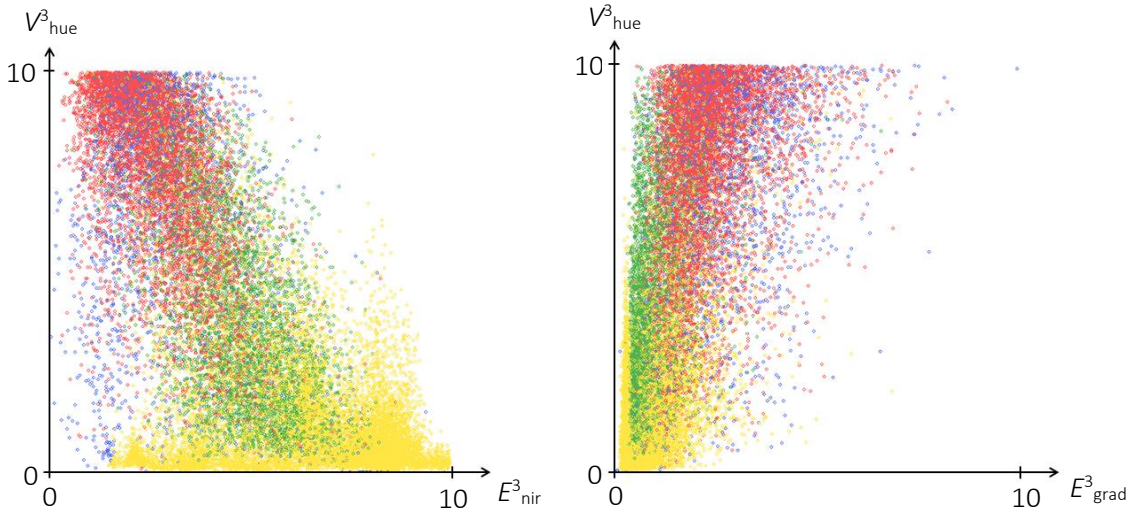



*Abbildung 5-4: Punktwolken der beiden am besten trennenden Merkmale (links) sowie des besten und des drittbesten Merkmals (rechts) für $\lambda_{max}=\lambda_3$. Rot: Siedlung; Blau: Industrie; Grün: Wald; Gelb: Acker.*

Im nächsten Schritt wird untersucht, wie viele Merkmale $F^{\lambda max}_{sel}$ enthalten sollte, um zum einen eine bestmögliche Klassifikation zu gewährleisten und zum anderen möglichst klein zu sein. Hierzu werden zufällig ausgewählte Pixel einer Ikonos-Testszene mit einer Maximum-Likelihood Klassifikation mit steigender Anzahl Merkmale in $F^{\lambda max}_{sel}$ für alle $\lambda_{max}$ ausgewertet. Die Merkmale in $F^{\lambda max}_{sel}$ ergeben sich aus deren Reihenfolge der Merkmalsselektion. Die Maximum-Likelihood Klassifikation wird gewählt, da das zugrunde liegende Gauß-Modell in der weiteren Arbeit auch für das Assoziationspotential genutzt wird. Die Trainingsstichprobe umfasst 51.200 Pixel. Getestet wird auf 460.800 Pixeln. Abbildung 5-5 zeigt die Klassifikationsfehler für unterschiedliche $F^{\lambda max}_{sel}$ für die jeweiligen $\lambda_{max}$. Für alle $\lambda_{max}$ ist die Nutzung der ersten sechs Merkmale ausreichend. Die Hinzunahme weiterer Merkmale verbessert das Klassifikationsergebnis nur geringfügig.

Diese Größenordnung der Anzahl von Merkmalen deckt sich mit zahlreichen Untersuchungen unterschiedlicher Methoden zur Merkmalsselektion bei vergleichbarer Dimension der Daten (u.a. Chandrashekar und Sahin, 2014; Mallet, 2010; Pal und Foody, 2010). Folglich werden die Anzahl der selektierten Merkmale $F^{\lambda max}_{sel}$ und somit die daraus abgeleiteten Merkmalsvektoren $\mathbf{f}_i^t(\mathbf{y}^t)$ und $\mathbf{h}_i^t(\mathbf{y}^t)$ auf sechs Merkmale begrenzt.

Es ergeben sich folgende Mengen der selektierten Merkmale $F^{\lambda max}_{sel}$.

- $F^{\lambda 2}_{sel}$: $E^2_{nir}$, $V^2_{nir}$, $V^2_{hue}$, $E^2_{nir\text{-}r}$, $V^2_{ndvi}$, $E^2_{grad}$
- $F^{\lambda 3}_{sel}$: $E^3_{nir}$, $V^3_{nir}$, $V^3_{hue}$, $E^3_{nir\text{-}r}$, $E^3_{grad}$, $ent^3$
- $F^{\lambda 4}_{sel}$: $E^4_{g}$, $E^4_{nir}$, $V^4_{hue}$, $E^4_{grad}$, $ent^4$, $V^3_{hue}$
- $F^{\lambda 5}_{sel}$: $E^5_{nir}$, $V^5_{hue}$, $E^5_{grad}$, $hom^5$, $E^4_{g}$, $ent^4$

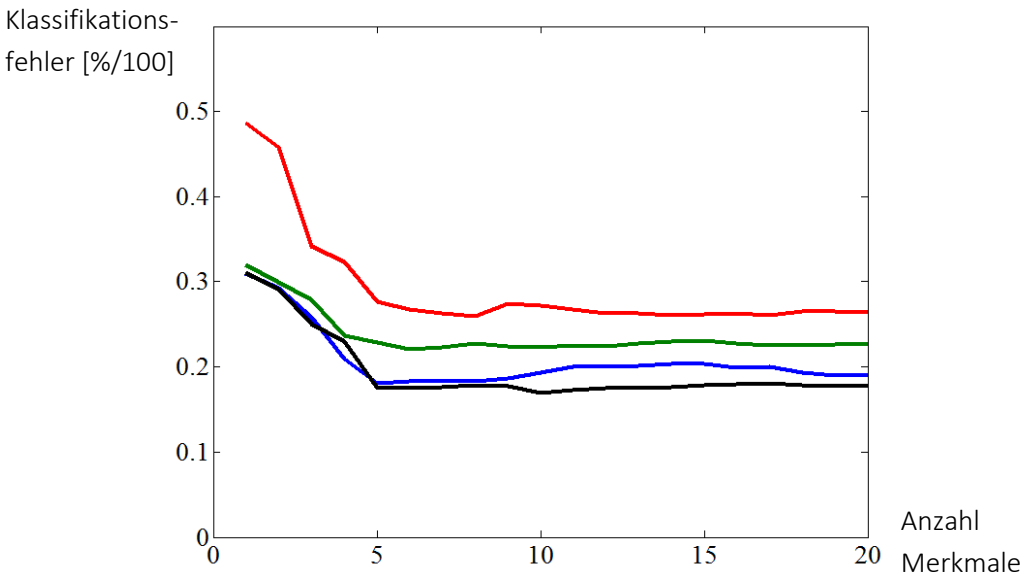



*Abbildung 5-5: Klassifikationsfehler (100% - Gesamtgenauigkeit) von Ikonos-Daten in Abhängigkeit der Anzahl der verwendeten Merkmale gemäß Merkmalsselektion. Rot: $\lambda_{max}=\lambda_2$; Grün: $\lambda_{max}=\lambda_3$; Blau: $\lambda_{max}=\lambda_4$: Schwarz: $\lambda_{max}=\lambda_5$.*

Es zeigt sich, dass in jeder Skalenstufe die Merkmale der höchstmöglichen Skala dominant sind. Des Weiteren ist bemerkenswert, dass Merkmale aus allen Merkmalsgruppen selektiert wurden. Die weiteren Untersuchungen aller hochaufgelösten Daten (Ikonos, RapidEye) erfolgen ausschließlich mit den zuvor aufgeführten Merkmalen sowie mit den folgenden Merkmalen für $F^{\lambda 1}{}_{sel}$: $E^1{}_r$, $E^1{}_g$, $E^1{}_b$, $E^1{}_{nir}$, $E^1{}_{ndvi}$, $E^1{}_{rvi}$. In Daten mittlerer Auflösung (Landsat) werden bereits pro Pixel häufig die spektralen Informationen mehrerer Klassen vermischt. Bei einer Anwendung von Skalen wird dieser Effekt verstärkt, ohne dass typische Strukturen signifikant hervortreten. Aus diesem Grund wird für Daten mittlerer Auflösung ausschließlich $F^{\lambda 1}{}_{sel}$ genutzt.

## 5.5. Maximale Merkmalsskala für das Assoziationspotential

Das Assoziationspotential beschreibt die Abhängigkeit einer Klasse von Daten an der Stelle des jeweiligen Primitivs *i* ohne explizite Modellierung nachbarschaftlicher Beziehungen. Jedoch wird durch die zuvor beschriebene Merkmalsextraktion in unterschiedlichen maximalen Skalen $\lambda_{max}$ die Nachbarschaft durchaus berücksichtigt. Je größer $\lambda_{max}$ gewählt wird, desto größer ist der zu erwartende Glättungseffekt durch das Assoziationspotential.

Die Wahl einer geeigneten maximalen Skala $\lambda_{max}$ für die Merkmale des Assoziationspotentials ist insbesondere unter Berücksichtigung der in dieser Arbeit definierten räumlichen Interaktionspotentiale wichtig. Zwar werden bei den räumlichen Interaktionspotentialen $IS_2$ und $IS_3$ sowohl Labels als auch Merkmale berücksichtigt, jedoch werden keine semantischen Nachbarschaftsbeziehungen explizit modelliert. Bilden sich Bildstrukturen nicht oder überwiegend falsch im Assoziationspotential ab, ist es nicht möglich, durch Interaktionspotentiale eine korrekte Klassenzuweisung an diesen Stellen zu gewährleisten, selbst wenn durch Wahl einer geringeren Skala $\lambda_{max}$ für das Interaktionspotential

Strukturen, die im Assoziationspotential durch große Glättung nicht hervortreten, wieder erkennbar sind. Hierbei ist eventuell unbestimmt, um welche benachbarten Klassen es sich handelt (vgl. Kapitel 4.4).

Zur qualitativen Bewertung wird der Einfluss der maximalen Skala $\lambda_{max}$ der extrahierten Merkmale auf das Klassifikationsergebnis anhand von drei Ausschnitten mit je 240*240 Pixeln aus einer Ikonos-Szene im Testgebiet Herne verdeutlicht (Abbildung 5-6 bis Abbildung 5-8). Die drei Regionen sind repräsentativ für sämtliche Testszenen, wobei in den Regionen unterschiedliche Landnutzungsstrukturen dominieren. Die Abbildungen visualisieren die Ergebnisse einer ML-Klassifikation der Regionen mit unterschiedlichen Merkmalsmengen ${F^{\lambda max}}_{sel}$ entsprechend $\lambda_{max}$ mit Angabe der in der Region erreichten Klassifikationsgenauigkeit.

In Abbildung 5-6 bis Abbildung 5-8 wird der zunehmende Glättungseffekt mit zunehmenden $\lambda_{max}$ deutlich. Nicht nur für die abgebildeten Regionen, sondern gleichfalls für die kompletten Testgebiete ist die Klassifikationsgenauigkeit für $\lambda_{max}=\lambda_4$ maximal. Allerdings bilden sich viele feine Strukturen in den Skalen $\lambda_{max}=\lambda_4$ und $\lambda_{max}=\lambda_5$ unzureichend ab. Offensichtlich wird dies z.B. in Abbildung 5-6, in der eine annähernd in Nord-Süd-Richtung verlaufende Straße durch eine Baumreihe von der naheliegenden Siedlung getrennt ist (roter Pfeil in Referenz), oder in Abbildung 5-7, in der eine Landstraße von Ackerflächen umgeben ist (roter Pfeil in Referenz).

Referenz | $\lambda_{max}=\lambda_1$ (67.0%) | $\lambda_{max}=\lambda_2$ (72.6%)

$\lambda_{max}=\lambda_3$ (77.4%) | $\lambda_{max}=\lambda_4$ (81.8%) | $\lambda_{max}=\lambda_5$ (81.4%)

*Abbildung 5-6: ML-Klassifikationsgenauigkeit von Ikonos-Daten in Abhängigkeit der verwendeten maximalen Skala $\lambda_{max}$ für die Merkmalsselektion – Region 1. Rot: Siedlung; Blau: Industrie; Grün: Wald; Gelb: Acker.*

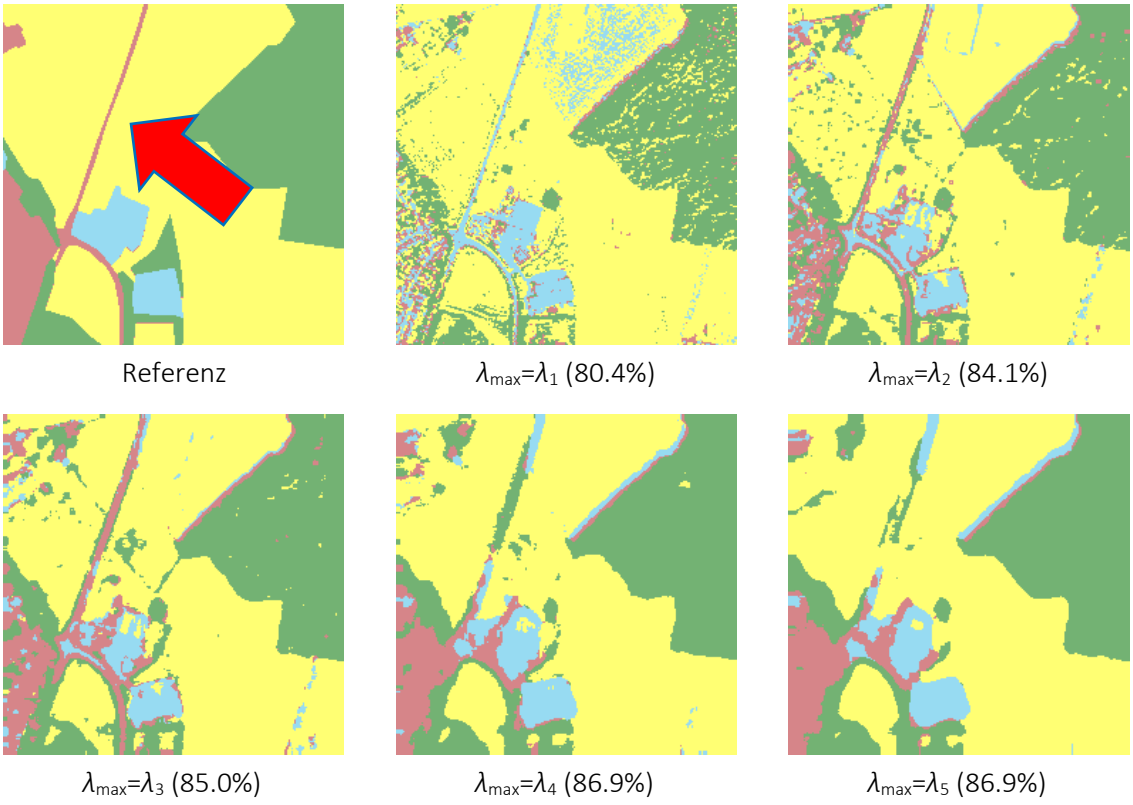



*Abbildung 5-7: ML-Klassifikationsgenauigkeit von Ikonos-Daten in Abhängigkeit der verwendeten maximalen Skala $\lambda_{max}$ für die Merkmalsselektion – Region 2.*

In beiden Beispielen in Abbildung 5-6 und Abbildung 5-7 werden die Strukturen nur bis einschließlich $\lambda_{max}=\lambda_3$ korrekt erkannt. Die Existenz dieser Strukturen ist jedoch für eine Interpretation der Szene von großer Wichtigkeit. Deshalb sind für die dieser Arbeit zugrundeliegende Anwendung $\lambda_{max}=\lambda_4$ und $\lambda_{max}=\lambda_5$ trotz einer höheren Klassifikationsgenauigkeit nicht optimal. Zudem ist durch die nachbarschaftlichen Interaktionen, die durch das räumliche Interaktionspotential an späterer Stelle erfolgen, mit zusätzlichen Glättungseffekten zu rechnen.

Hingegen reagieren die niedrigen Skalen $\lambda_{max}=\lambda_1$ und $\lambda_{max}=\lambda_2$ sehr sensibel auf unterschiedliche Ausprägungen benachbarter Primitive. Dies führt oft zu einem ungewollten Salz-und-Pfeffer-Effekt. Als Folge ist die Klassifikationsgenauigkeit unter Verwendung dieser Skalen deutlich niedriger im Vergleich zur Verwendung höherer Skalen $\lambda_{max}$. Da die in dieser Arbeit verwendeten Interaktionspotentiale nur identische und unterschiedliche benachbarte Primitive unterscheiden und keine weiteren Beziehungen unterschiedlicher Klassen explizit modelliert werden, ist eine mehrheitlich korrekte Klassenzuweisung im Assoziationspotential Bedingung für eine erfolgreiche CRF-Klassifikation. Eine Genauigkeit von nur ca. 50% - 60%, wie sie in manchen Regionen für $\lambda_{max}=\lambda_1$ und $\lambda_{max}=\lambda_2$ existiert (als Beispiel: Abbildung 5-8), ist als Grundlage für eine hohe Klassifikationsgenauigkeit durch die Integration nachbarschaftlicher Interaktionen nicht vielversprechend.

Die vorhergehenden Überlegungen verdeutlichen, dass die Entscheidung für die Wahl geeigneter Merkmale für $\mathbf{f}_i^t(\mathbf{y}^t)$ aufgrund einer qualitativen Bewertung und losgelöst von einer rein quantitativen Bewertung durch Betrachtung der Klassifikationsgenauigkeit zu treffen ist. Darauf basierend wird $\lambda_{max}=\lambda_3$

mit der Menge der Merkmale $F^{\lambda 3}{}_{sel}$ für sämtliche weiteren Untersuchungen mit hochauflösenden Daten gewählt.

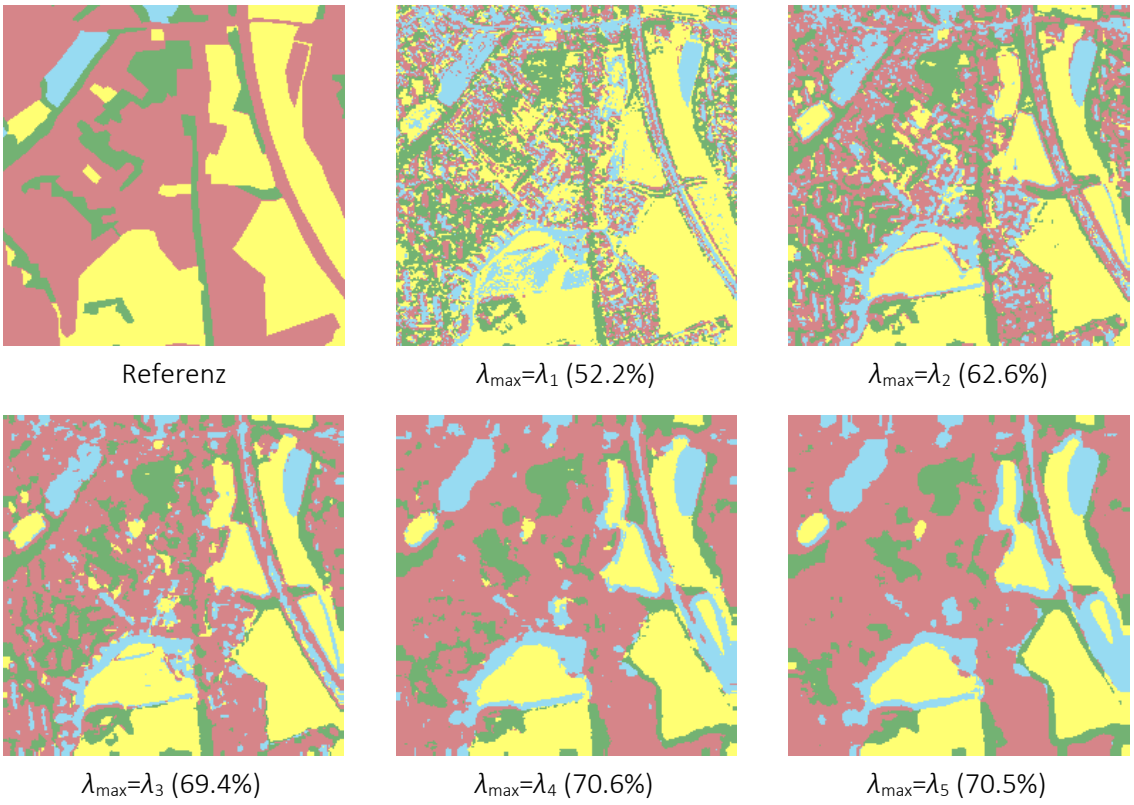



*Abbildung 5-8: ML-Klassifikationsgenauigkeit von Ikonos-Daten in Abhängigkeit der verwendeten maximalen Skala $\lambda_{max}$ für die Merkmalsselektion – Region 3.*

## 5.6. Parameter des räumlichen Interaktionspotentials

Der Einfluss der räumlichen Interaktionspotentiale $IS_1$, $IS_2$ und $IS_3$ auf das Klassifikationsergebnis ist abhängig von dem Parameter $\beta$ sowie von dem Parameter $\eta$ für $IS_2$ und $IS_3$. Der Parameter $\beta$ legt das Gewicht des räumlichen Interaktionspotentials fest, während der Glättungsparameter $\eta$ die Unterschiede in den Merkmalen für das Interaktionspotential verstärkt bzw. abschwächt (Kapitel 4.4).

Wie bei der Verwendung unterschiedlicher maximaler Merkmalsskalen $\lambda_{max}$ für das Assoziationspotential führen Experimente mit monotemporalen Daten und variierenden Parametern $\beta$ und $\eta$ in verschiedenen Testgebieten oftmals zu ähnlichen maximalen Klassifikationsgenauigkeiten, obwohl der visuelle Eindruck der Ergebnisse stark unterschiedlich ist. Wiederum werden vor allem kleine Objekte, die keinen wesentlichen Einfluss auf die Klassifikationsgenauigkeit haben, jedoch für einen Betrachter wertvolle Informationen beinhalten können, je nach Wahl des Modells *IS* für das Interaktionspotential und den Parametern $\beta$ und $\eta$ unterschiedlich gut erhalten. Aus diesem Grund wird vor den quantitativen Analysen (Kapitel 5.8 ff) eine qualitative Untersuchung durchgeführt, um die optimale Konfiguration von Parametern und maximaler Skala $\lambda_{max}$ für die räumlichen Interaktionspotentiale zu finden.

Um im Folgenden für die merkmalsbasierten Modelle eine Referenz zu haben, welche maximale Genauigkeit mit einem rein auf Labels basierendem Ansatz erzielt werden kann, wurde für $IS_1$ der Parameter $\beta$ empirisch mit dem Ziel einer maximalen Klassifikationsgenauigkeit bestimmt. Für die beiden merkmalsabhängigen Interaktionspotentiale $IS_2$ und $IS_3$ wurden die Parameter $\beta$ und $\eta$ als Kompromiss zwischen Klassifikationsgenauigkeit und visueller Beurteilung zur Erhaltung von Detailstrukturen gewählt. Abbildung 5-9 zeigt exemplarisch den Einfluss des Parameters $\eta$ auf das Modell $IS_3$ mit konstantem $\beta$. Um den Effekt zu verdeutlichen, wurde entgegen der Annahmen aus den vorhergehenden Kapiteln $\lambda_{max}=\lambda_1$ gewählt. Ein sehr klein gewählter Wert für $\eta$ führt zu einer starken Glättung. Die Abhängigkeit von den Merkmalen wird nahezu vernachlässigt. Hingegen bewirkt ein hoher Wert einen sehr geringen Glättungseffekt. Die besondere Eigenschaft von $IS_3$, „verborgene" Strukturen hervorzuheben, führt bei einem zu hohen Wert für $\eta$ sogar zu einer Verstärkung des Salz-und-Pfeffer-Effekts.

$\eta$ = 0.5 $\eta$ = 5 $\eta$ = 20 $\eta$ = 100

*Abbildung 5-9: Klassifikationsergebnisse für $IS_3$ mit $\lambda_{max}=\lambda_1$ für $\mathbf{h}_i^t(\mathbf{y}^t)$ in Abhängigkeit von $\eta$. Bilddaten: Ikonos (240*240 Pixel), $\beta = 0.7$.*

Beste Ergebnisse nach den obigen Kriterien werden für die verwendeten Testszenen mit $\beta$ = 0.9 für $IS_1$, $\beta$ = 0.7 und $\eta$ = 80 für $IS_2$ sowie $\beta$ = 0.7 und $\eta$ = 5 für $IS_3$ erzielt. Dabei ist der Parameter $\eta$ für $IS_2$ mit leichter Betonung der glättenden Wirkung gewählt, während mit der Wahl von $\eta$ für $IS_3$ etwas stärker auf die Erhaltung kleiner Strukturen fokussiert wird. Sämtliche folgenden Experimente verwenden die hier angegebenen Werte für $\beta$ und $\eta$. Beide Parameter sind ausreichend robust gegen kleine Veränderungen. Abweichungen in der Größenordnung $10^0$ für $\eta$ sowie $10^{-1}$ für $\beta$ haben nur vernachlässigbare Veränderungen des Klassifikationsergebnisses zur Folge.

Im nächsten Schritt wird die Auswirkung der verwendeten Skala $\lambda_{max}$ für den Merkmalsvektor des räumlichen Interaktionspotentials $\mathbf{h}_i^t(\mathbf{y}^t)$ auf die mit den merkmalsabhängigen Modellen $IS_2$ und $IS_3$ erzielten Ergebnisse analysiert. Diese, sowie die Ergebnisse mit merkmalsunabhängigen Modell $IS_1$, werden beispielhaft anhand der drei in Kapitel 5.5 verwendeten Regionen in Abbildung 5-10 bis Abbildung 5-12 gegenübergestellt. Die Hypothese, dass eine maximale Klassifikationsgenauigkeit nicht in allen Fällen das beste visuelle Ergebnis ergibt, wird insbesondere in Abbildung 5-11 bestätigt. Die Landstraße, die sich als schlanke Linie umgeben von Ackerflächen darstellt, ist durch die starke Glättung in $IS_1$ nahezu komplett unkenntlich (roter Pfeil), während sie in beiden merkmalsabhängigen Modellen in allen Skalen korrekt dargestellt wird. Die Existenz dieser Straße stellt für einen Betrachter eine wertvolle Information dar, die nicht durch eine geringfügig höhere Klassifikationsgenauigkeit kompensiert wird. Eine ausführlichere Betrachtung der Fähigkeiten der unterschiedlichen räumlichen Interaktionspotentiale zur Erhaltung feiner Strukturen folgt in Kapitel 5.7.

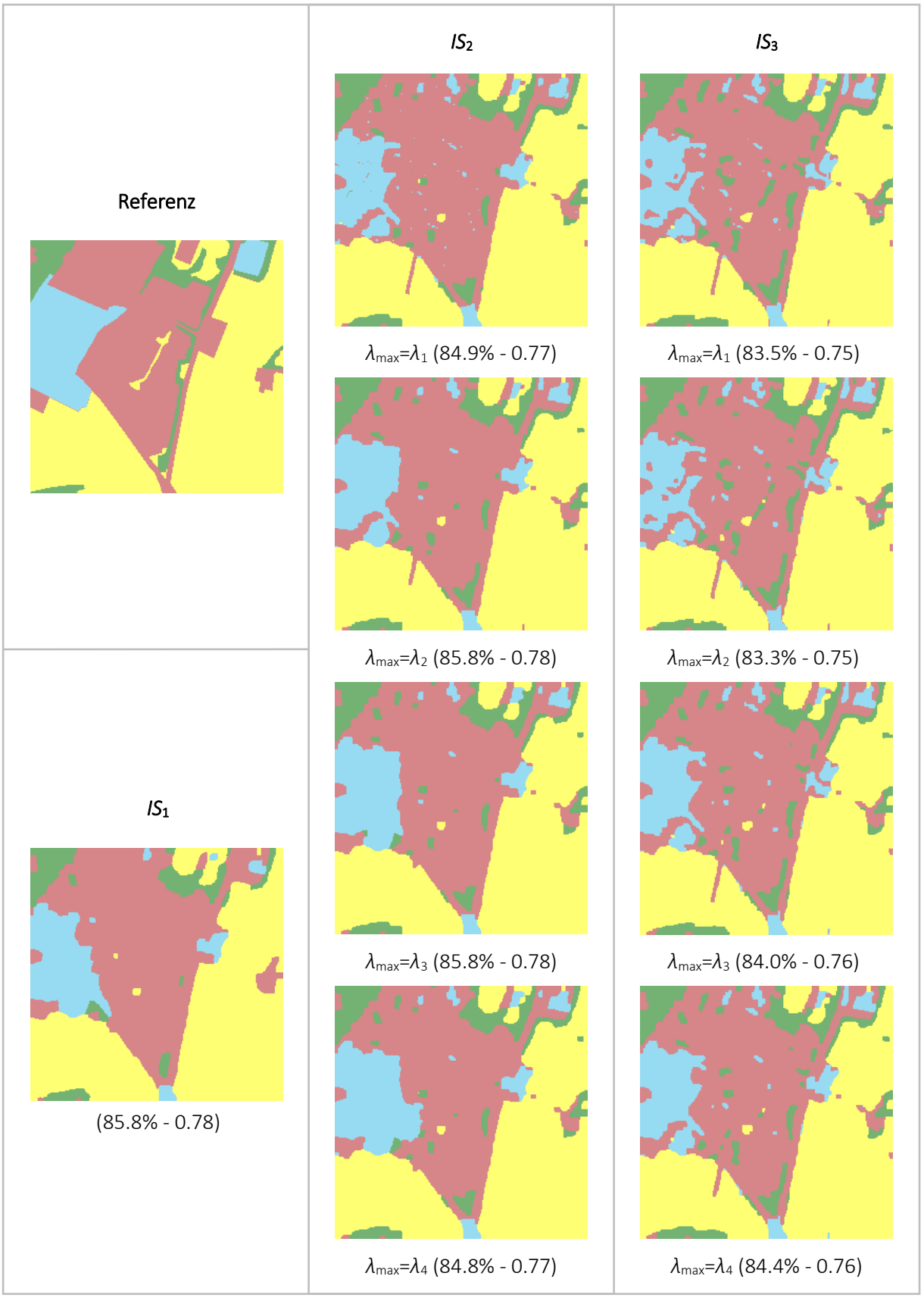



*Abbildung 5-10: Gesamtgenauigkeit und Kappa Index bei Anwendung unterschiedlicher Modelle für das räumliches Interaktionspotential unter Verwendung von Merkmalen unterschiedlicher maximaler Skalen $\lambda_{max}$ für IS2 und IS3 - Region 1.*

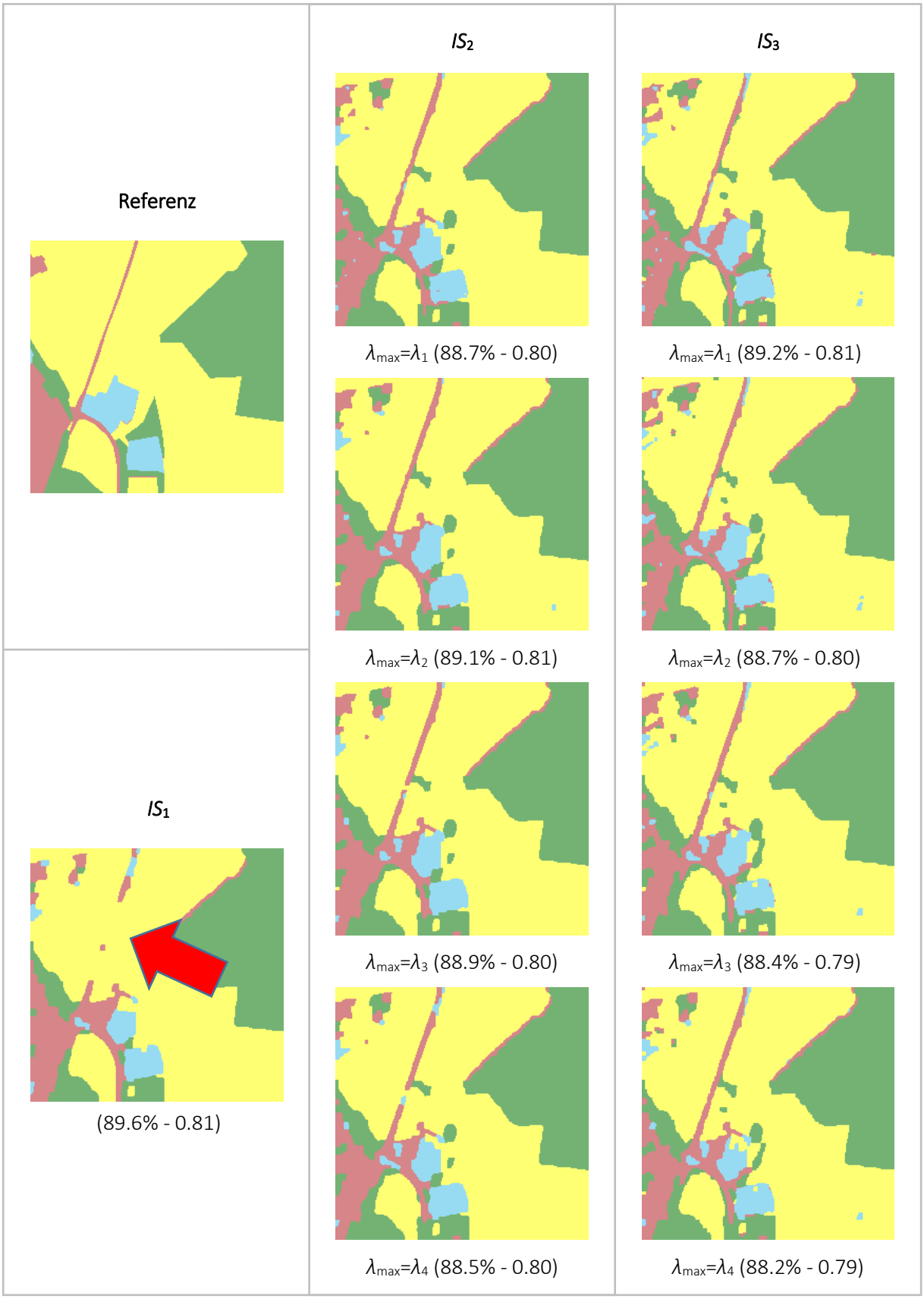



*Abbildung 5-11: Gesamtgenauigkeit und Kappa Index bei Anwendung unterschiedlicher Modelle für das räumliches Interaktionspotential unter Verwendung von Merkmalen unterschiedlicher maximaler Skalen $\lambda_{max}$ für IS2 und IS3 – Region 2.*

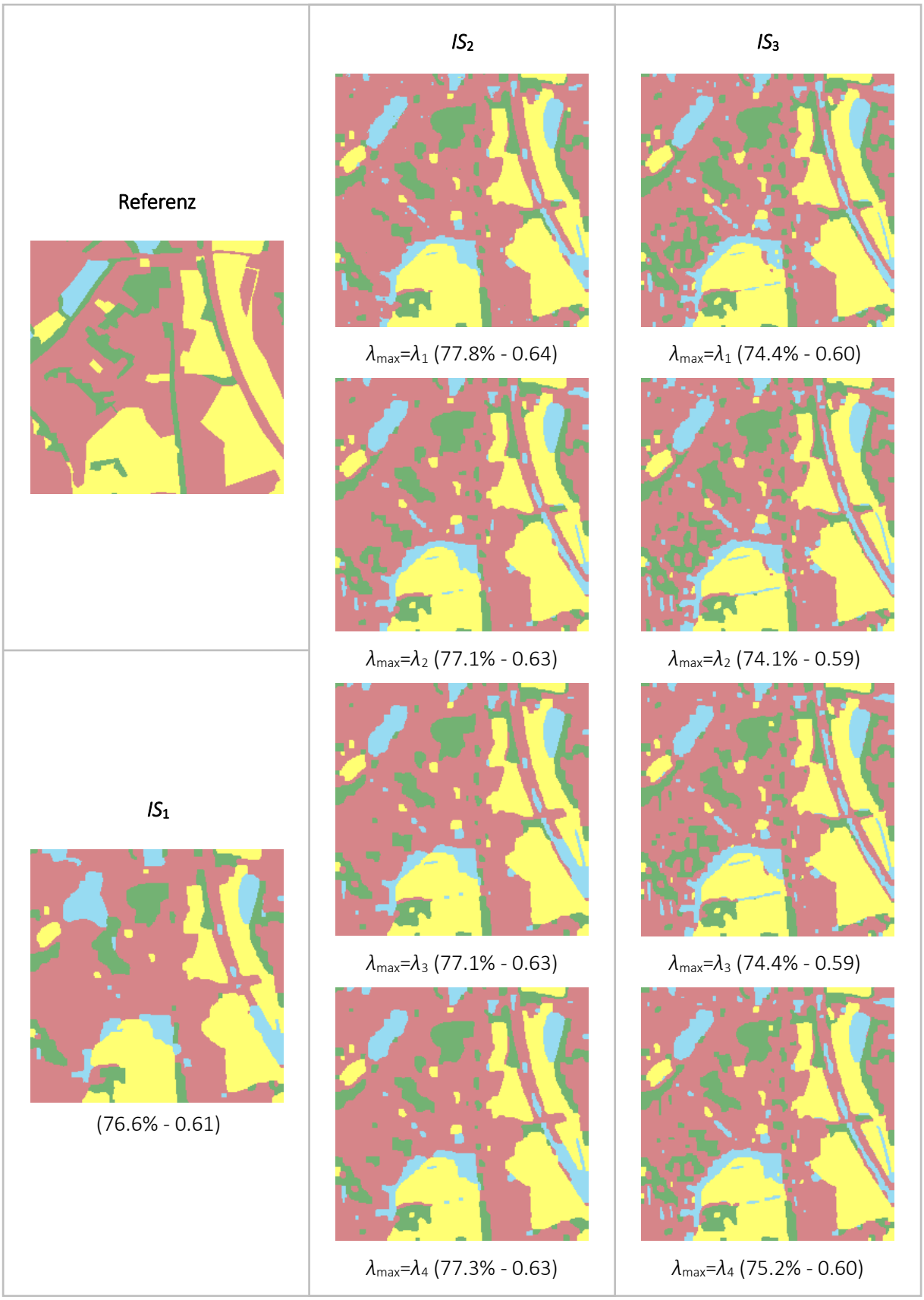
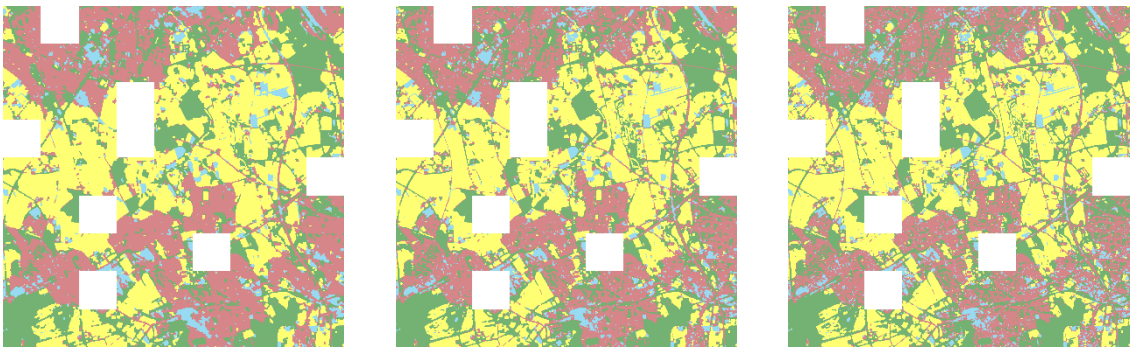



*Abbildung 5-12: Gesamtgenauigkeit und Kappa Index bei Anwendung unterschiedlicher Modelle für das räumliches Interaktionspotential unter Verwendung von Merkmalen unterschiedlicher maximaler Skalen $\lambda_{max}$ für IS2 und IS3 – Region 3.*

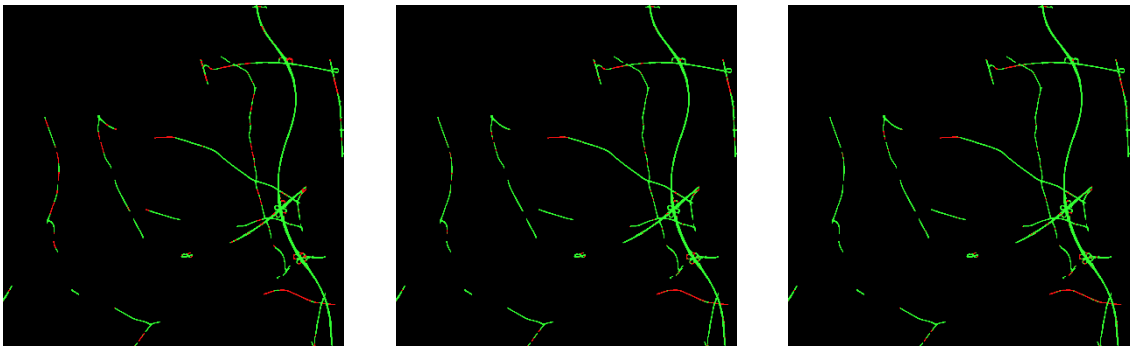
Die Verwendung von Merkmalen in unterschiedlichen maximalen Skalen $\lambda_{max}$ für $\mathbf{h}_i^t(\mathbf{y}^t)$ von $IS_2$ und $IS_3$ hat nur geringen Einfluss auf die Ergebnisse. Die Klassifikationsgenauigkeit sowie der Kappa-Koeffizient variieren nur wenig (Abbildung 5-10 bis Abbildung 5-12). Insgesamt verstärkt sich auch hier der Glättungseffekt mit zunehmendem $\lambda_{max}$. Die Differenzen sind jedoch nicht so deutlich, wie die durch Variation der anderen Parameter hervorgerufenen Unterschiede. Details werden bestmöglich mit kleinerem $\lambda_{max}$ erhalten. In $\lambda_{max}=\lambda_1$ sind dabei teilweise unerwünschte pixelige Effekte zu beobachten. Für die weiteren Untersuchungen in dieser Arbeit wird somit $\lambda_{max}=\lambda_2$ mit der Menge der Merkmale $F^{\lambda 2}_{sel}$ für den Merkmalsvektor $\mathbf{h}_i^t(\mathbf{y}^t)$ der merkmalsbasierten räumlichen Interaktionspotentiale verwendet.

## 5.7. Evaluation der Klassifikationsgenauigkeit feiner Strukturen

In den vorherigen Kapiteln wurde exemplarisch gezeigt, dass feinere Strukturen mit den merkmalsabhängigen Modellen für das räumliche Interaktionspotential besser bewahrt werden. Diese Beobachtung soll in diesem Kapitel anhand einer Evaluation bedeutender außerörtlicher Straßen für die kompletten Testgebiete verifiziert werden. Als Referenz dienen die Straßen der Klasse *Siedlung*, die nicht durch bebaute Gebiete führen, d.h. die von *Ackerflächen* oder *Wald* umgeben sind.



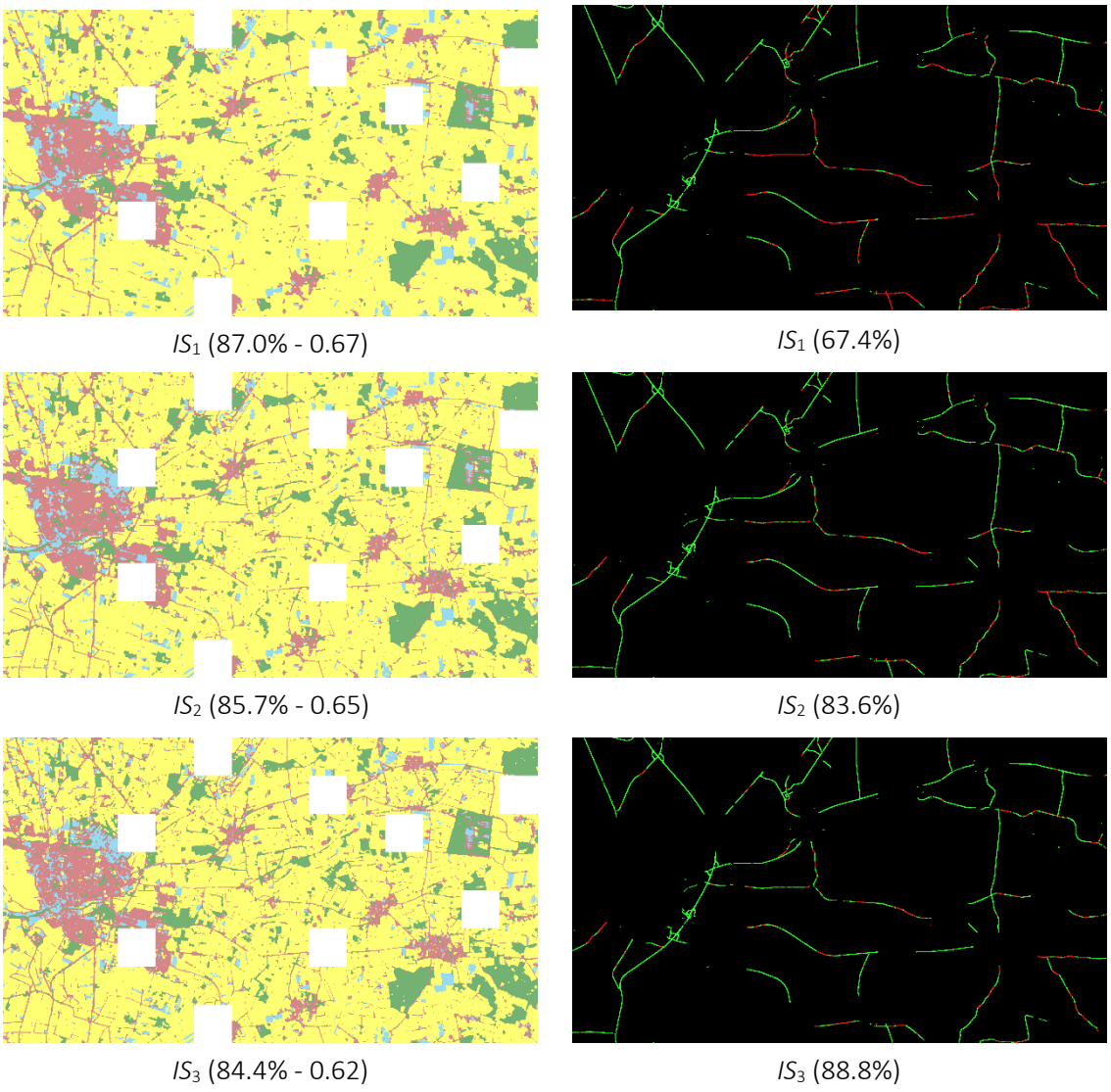
$IS_1$ (80.7% - 0.72) $IS_2$ (79.8% - 0.71) $IS_3$ (78.0% - 0.68)



$IS_1$ (89.1%) $IS_2$ (94.4%) $IS_3$ (95.5%)

*Abbildung 5-13: Oben: Monotemporales Klassifikationsergebnis (Gesamtgenauigkeit – Kappa Koeffizient) für Testszene Herne mit Ikonos-Daten (2005); Weiße Flächen: Trainingsgebiete. Unten: Klassifikation außerörtlicher Straßen; grün: korrekt erkannte Straßenabschnitte, rot: nicht erkannte Straßenabschnitt; Der Wert in Klammern gibt die Vollständigkeit an.*

In den beiden Testgebieten Herne und Husum werden in je einer hochaufgelösten Testszene jeweils ca. 10% der Fläche als Trainingsdaten ausgewählt, die restliche Fläche wird unter Verwendung der zuvor ermittelten Konfiguration mit den drei unterschiedlichen Modellen für das räumliche Interaktionspotential monotemporal klassifiziert (Abbildung 5-13, oben sowie Abbildung 5-14, links).



$IS_1$ (87.0% - 0.67) $IS_1$ (67.4%)

$IS_2$ (85.7% - 0.65) $IS_2$ (83.6%)

$IS_3$ (84.4% - 0.62) $IS_3$ (88.8%)

*Abbildung 5-14: Links: Monotemporales Klassifikationsergebnis (Gesamtgenauigkeit – Kappa Koeffizient) für Testszene Husum mit RapidEye-Daten (7/2009); Weiße Flächen: Trainingsgebiete. Rechts: Klassifikation außerörtlicher Straßen; grün: korrekt erkannte Straßenabschnitte, rot: nicht erkannte Straßenabschnitte; Der Wert in Klammern gibt die Vollständigkeit an.*

Außerörtliche Straßen werden in beiden Testgebieten zumeist korrekt als *Siedlung* klassifiziert, jedoch häufig auch als *Industriefläche*. Dies ist vor allem bei breiten Straßen der Fall, deren Merkmale denen von Industrieflächen sehr ähnlich sind. Aus diesem Grund wird für diese Analyse eine Zuordnung zur Klasse *Siedlung* oder *Industrie* als korrekt angenommen. Da eine Beurteilung der korrekt erfassten Straßenbreite in der Auflösung der zur Verfügung stehenden Daten nur unzureichend möglich ist, ist es

ausschließliches Ziel der Untersuchung, die Existenz von Straßen aufzuzeigen. Hierzu erfolgt eine Dilation der Straßenflächen im Klassifikationsergebnis. Somit wird der Einfluss von Fehlern der Straßenbreite reduziert und das Ergebnis spiegelt in guter Näherung den prozentualen Anteil der erfassten Länge aller Straßenabschnitte wider.

Für beide Testgebiete liefert das merkmalsunabhängige Modell $IS_1$ die geringfügig höchste Gesamtgenauigkeit, wobei $IS_2$ etwas besser als $IS_3$ ist. Bei Betrachtung der korrekt detektierten Straßen kehrt sich diese Reihenfolge jedoch um. Der Klassifikationsfehler für $IS_1$ ist in beiden Testszenen mindestens doppelt so hoch wie für die beiden merkmalsbasierten Modelle $IS_2$ und $IS_3$ (Abbildung 5-13, unten sowie Abbildung 5-14, rechts). Während die sehr dominanten Straßen (Autobahnen) in sämtlichen Modellen sicher erkannt werden, versagt das Modell $IS_1$ zunehmend, je kleiner die Straßen sind. In diesem Fall ist der Glättungseffekt des Interaktionspotentials größer als eine eventuelle korrekte Zuweisung zur Klasse Straße durch das Assoziationspotential. In den beiden merkmalsbasierten Modellen für das räumliche Interaktionspotential wird hingegen die Diskontinuität, die aus dem Kontrast der Straße mit ihrer Umgebung resultiert, bewahrt, bzw. sogar verstärkt.

Die insgesamt etwas höhere Genauigkeit der Straßendetektion in der Testszene Herne ist auf zwei Gründe zurück zu führen: zum einen die höhere geometrische Auflösung der verwendeten Daten, zum anderen die Art der Referenzdatenerstellung. Für die Testszene Herne wurden nur sichtbare Straßen erfasst, während die Ableitung der Referenz aus einem GIS für die Szene Husum dazu führt, dass auch durch Vegetation verdeckte Straßen in der Referenz enthalten sind.

## 5.8. Multitemporale Klassifikation von Fernerkundungsdaten mit einheitlicher Auflösung

Anhand der multitemporalen Klassifikation von Fernerkundungsdaten mit einheitlicher Auflösung wird untersucht, bis zu welchem Grad fehlerhafte Klassifikationen zu einem Zeitpunkt durch Nutzung zeitlichen Kontextwissens reduziert werden können. Hierzu stehen für das Testgebiet Husum vier RapidEye-Szenen und für das Testgebiet Herne zwei Ikonos-Szenen zur Verfügung. Wie in Kapitel 5.7 wird ca. 10% der Fläche der Testszenen zum Training, der Rest zum Testen genutzt. Da die beobachteten Veränderungen der Landbedeckung in den betrachteten Zeiträumen sehr gering sind (<0.01% der Fläche der Testszenen), wird für die Auswertung eine identische Landbedeckung für alle Bilder angenommen.

Die Testszenen werden zunächst monotemporal sowohl mit ML als auch mit CRF klassifiziert, wobei die drei unterschiedlichen Modelle für das räumliche Interaktionspotential unter Verwendung der in den Kapiteln 5.3 bis 5.6 ermittelten Konfiguration verwendet werden. Diese Klassifikationsergebnisse dienen als Referenz, um den Gewinn durch Anwendung des multitemporalen Ansatzes zu quantifizieren. Der Parameter $\gamma$, der das zeitliche Interaktionspotential im Verhältnis zu den beiden anderen Potentialen gewichtet, wird empirisch ermittelt und beträgt in allen folgenden Untersuchungen $\gamma$=1.5. Sind die räumlichen und zeitlichen Interaktionspotentiale im identischen Wertebereich definiert, sollte generell bei der gewählten Konfiguration mit vier räumlichen und zwei zeitlichen Nachbarn pro Primitiv $\gamma$ etwa doppelt so groß wie $\beta$ gewählt werden. Dies gewährleistet einen annähernd identischen Einfluss der beiden Interaktionspotentiale auf das Klassifikationsergebnis. Deutliche Verschiebungen dieses Verhältnisses führen zu einer Dominanz des räumlichen oder zeitlichen Interaktionspotentials. Gegen

kleine Veränderungen ist $\gamma$ robust. Abweichungen von 0.1 führen nur zu geringfügig abweichenden Klassifikationsergebnissen.

Die temporale Übergangsmatrix **TM** muss grundsätzlich in Abhängigkeit von den erwarteten Veränderungen erstellt werden – es wird an dieser Stelle ignoriert, dass in den Daten de facto keine Veränderungen vorliegen.

| $x_{t+1}$ / $x_t$ | *Sie* | *Ind* | *Wal* | *Ack* |
|---|---|---|---|---|
| *Sie* | 1 | 0.05 | 0.05 | 0.05 |
| *Ind* | 0.05 | 1 | 0.05 | 0.05 |
| *Wal* | 0.1 | 0.1 | 1 | 0.1 |
| *Ack* | 0.2 | 0.2 | 0.1 | 1 |

*Tabelle 5-2: Übergangsmatrix $\mathbf{TM}_{hh}$ für Daten hoher Auflösung von Zeitpunkt t zum Zeitpunkt t+1.*

Entscheidende Einflussfaktoren sind die betrachtete Region und die Zeiträume zwischen den Aufnahmen. Beide Testgebiete befinden sich in Deutschland und somit in Regionen, in denen die Landbedeckung nur wenigen Veränderungen unterworfen ist. Folglich befinden sich die höchsten Potentiale der Übergangsmatrix auf der Hauptdiagonale. Die häufigste Veränderung der Landschaft ist der Übergang von *Ackerland* (*Ack*) in *Siedlungs-* (*Sie*) oder *Industrieflächen* (*Ind*). Übergänge von bebautem Land (*Siedlung* oder *Industrie*) in *Wald* (*Wal*) oder *Ackerflächen* hingegen sind sehr unwahrscheinlich. Dieses Wissen spiegelt sich in den jeweiligen Elementen der Übergangsmatrix $\mathbf{TM}_{hh}$ für den Übergang zwischen zwei Zeitpunkten mit hoch aufgelösten Daten wieder (Tabelle 5-2). Es ist zu beachten, dass sämtliche Übergänge zugelassen werden. Der Zeitraum zwischen den Aufnahmen wird nicht explizit berücksichtigt. Es wird abweichend von den tatsächlichen Abständen impliziert, dass der Zeitraum groß genug für Veränderungen der Landbedeckung ist und zwischen den einzelnen Aufnahmen identische Zeitabstände vorliegen.

| T | monotemporal | | | | multitemporal | | |
|---|---|---|---|---|---|---|---|
| | ML | $IS_1$ | $IS_2$ | $IS_3$ | $IS_1$ | $IS_2$ | $IS_3$ |
| $t_1$ | 74.2%<br>0.63 | 80.7%<br>0.72 | 79.8%<br>0.71 | 78.0%<br>0.68 | 80.9%<br>0.72 | 81.3%<br>0.73 | 80.9%<br>0.72 |
| $t_2$ | 73.2%<br>0.61 | 78.7%<br>0.69 | 78.9%<br>0.69 | 77.3%<br>0.67 | 80.7%<br>0.72 | 80.9%<br>0.72 | 80.4%<br>0.72 |

*Tabelle 5-3: Gesamtgenauigkeit (jeweils oberer Wert) und Kappa-Koeffizient (jeweils unterer Wert) für die monotemporale und multitemporale Klassifikation von zwei Ikonos-Szenen (Testgebiet Herne).*

Die Gesamtgenauigkeiten und die Kappa-Koeffizienten für die monotemporale sowie die multitemporale Klassifikation der Szenen findet sich in Tabelle 5-3 (Testgebiet Herne) und Tabelle 5-4 (Testgebiet Husum).

| T | monotemporal | | | | multitemporal | | |
|---|---|---|---|---|---|---|---|
| | ML | $IS_1$ | $IS_2$ | $IS_3$ | $IS_1$ | $IS_2$ | $IS_3$ |
| $t_1$ | 79.9%<br>0.53 | 88.0%<br>0.67 | 86.4%<br>0.65 | 84.8%<br>0.62 | 89.5%<br>0.69 | 89.8%<br>0.72 | 89.6%<br>0.72 |
| $t_2$ | 80.6%<br>0.55 | 87.0%<br>0.67 | 85.7%<br>0.65 | 84.4%<br>0.62 | 89.9%<br>0.70 | 90.1%<br>0.73 | 89.9%<br>0.73 |
| $t_3$ | 79.8%<br>0.50 | 87.0%<br>0.63 | 86.5%<br>0.63 | 85.5%<br>0.62 | 90.1%<br>0.71 | 90.5%<br>0.73 | 90.4%<br>0.74 |
| $t_4$ | 86.1%<br>0.64 | 90.1%<br>0.72 | 89.8%<br>0.72 | 89.2%<br>0.71 | 90.1%<br>0.71 | 90.4%<br>0.73 | 90.3%<br>0.73 |

*Tabelle 5-4: **Gesamtgenauigkeit und Kappa Koeffizient für die monotemporale und multitemporale Klassifikation von vier RapidEye-Szenen (Testgebiet Husum).***

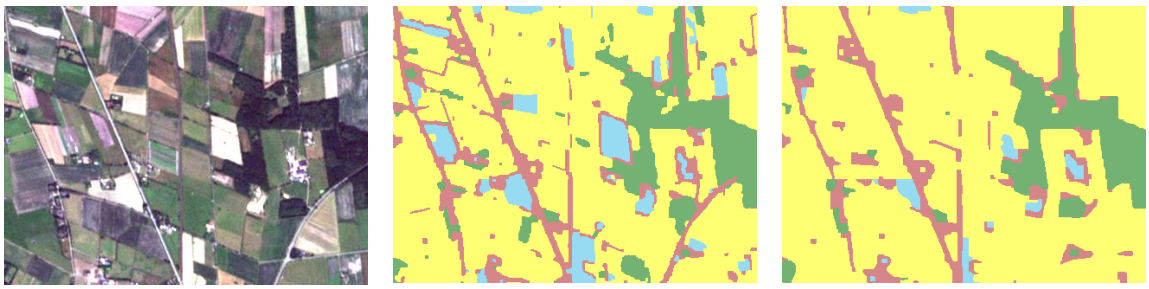
In beiden Testgebieten ist für alle Modelle *IS* eine deutliche Verbesserung durch die monotemporale CRF-Klassifikation im Vergleich zur ML-Klassifikation erkennbar, wobei $IS_1$ in Bezug auf die hier gewählten Beurteilungskriterien die besten Ergebnisse liefert. Durch die Berücksichtigung zeitlichen Kontextwissens im multitemporalen Ansatz können die Ergebnisse weiter verbessert werden. Vor allem im Testgebiet Husum, in dem vier Aufnahmen zur Verfügung stehen, wird der Klassifikationsfehler einzelner Aufnahmen um bis zu einem Drittel im Vergleich zum monotemporalen Ansatz reduziert. Alle Modelle für das räumliche Interaktionspotential führen zu ähnlichen Gesamtgenauigkeiten und Kappa-Koeffizienten. Insgesamt ist die Gesamtgenauigkeit im Testgebiet Husum deutlich höher als im Testgebiet Herne. Ein Grund hierfür ist, dass in der Referenz des Testgebiets Herne 7.2% aller Pixel mindestens ein Nachbarpixel in der 4er-Nachbarschaft mit verschiedenem Label besitzen. In der Testszene Husum beträgt der Anteil lediglich 4.2%. Diese Pixel an Klassengrenzen sind häufig Mischpixel mit den spektralen Eigenschaften unterschiedlicher Klassen und somit unsicher zu klassifizieren.

| Testgebiet | monotemporal | | | |
|---|---|---|---|---|
| | ML | $IS_1$ | $IS_2$ | $IS_3$ |
| Herne | 76,0% | 82.8% | 82.2% | 80.4% |
| Husum | 79.7% | 86.3% | 86.1% | 84.8% |

*Tabelle 5-5: Reklassifikationsgenauigkeit für die monotemporale Klassifikation der Trainingsgebiete über alle Bilder.*

Die Reklassifikationsgenauigkeit, d.h. die Genauigkeit für eine Klassifikation der Trainingsgebiete mit den aus den Gebieten gewonnen Parametern, liegt für die monotemporale Klassifikation der Trainingsgebiete über alle Bilder auf dem Niveau der Klassifikationsergebnisse der Testgebiete (Tabelle 5-5). Dies belegt, dass keine Überanpassung der Trainingsdaten vorliegt. Im Gebiet Husum liegt der Reklassifikationsfehler sogar etwas höher als der Klassifikationsfehler der kompletten Testgebiete. Dieser ungewöhnliche Sachverhalt erklärt sich durch die Tatsache, dass sämtliche Klassen in den Trainingsgebieten zu ähnlichen Anteilen vorhanden sind. Als Folge hat die Unsicherheit bei der Trennung der Klassen *Siedlung* und *Industrie*, welche nur einen geringen Anteil an der kompletten Fläche des Testgebiets haben, an dieser Stelle einen größeren Einfluss.

Im nächsten Schritt wird untersucht, welche Flächen vorrangig durch die multitemporale Auswertung profitieren. Offensichtlich wird dies bei Betrachtung der Konfusionsmatrizen zu den jeweiligen Zeitpunkten. Tabelle 5-6 zeigt exemplarisch die Konfusionsmatrix zum Zeitpunkt $t_2$ des Testgebiets Husum, die Matrizen der übrigen Zeitpunkte zeigen ähnliche Verteilungen. Insbesondere die Richtigkeit der Klasse Industrie ist sehr niedrig. Diese Tatsache ist vorrangig für den relativ geringen Kappa-Koeffizient im Verhältnis zur hohen Gesamtgenauigkeit in dem Testgebiet verantwortlich (Tabelle 5-4). Verantwortlich für die geringe Richtigkeit der Klasse Industrie sind die zahlreichen Fehlklassifikationen der Klasse Acker. Je nach Bewuchs, Beleuchtung und Aufnahmewinkel besitzen einzelne Flächen eine sehr ausgeprägte Rückstreuung und ähneln den Charakteristika industrieller Flächen. Diese Flächen können häufig durch Nutzung zeitlicher Kontextinformation korrekt klassifiziert werden (Abbildung 5-15).



RapidEye-Bild, $t_2$ | $IS_3$, $t_2$, monotemporal | $IS_3$, $t_2$, multitemporal

*Abbildung 5-15: Links: Ausschnitt aus der RapidEye-Szene zum Zeitpunkt $t_2$ im Testgebiet Husum. Mitte/rechts: Monotemporales/Multitemporales Klassifikationsergebnis, die Fehlklassifikationen von Ackerflächen als Industrieflächen sind deutlich verringert.*

In Tabelle 5-6 konnten die monotemporal fehlerhaft als Industrie klassifizierten Ackerflächen um ca. 60% verringert werden. Signifikante Verbesserungen durch die multitemporale Auswertung gibt es zudem für die Richtigkeit der Klassen Siedlung und Wald, erneut vorrangig durch Reduktion der Fehlklassifikation als Ackerfläche.

Des Weiteren zeigt Tabelle 5-6, dass die Trennung der bebauten Klassen Siedlung und Industrie durch den multitemporalen Ansatz nicht verbessert wird und zahlreiche Fehlklassifikationen der Klasse Siedlung mit allen anderen Klassen existieren. Dies ist nicht verwunderlich, da die Klasse Siedlung sowohl aus versiegelten Flächen als auch aus Vegetationsflächen besteht. Bemerkenswert ist zudem, dass die Vollständigkeit und Richtigkeit sämtlicher Klassen als Ergebnis der monotemporalen Klassifikation höher

ist als durch Anwendung der ML-Klassifikation. Diese Beobachtungen lassen sich auch in den Konfusionsmatrizen der anderen Aufnahmen der Testgebiete Husum und Herne wiederfinden.

| | $x_i$ / $x_i^{ref}$ | *Sie* | *Ind* | *Wal* | *Ack* | Vollständig |
|---|---|---|---|---|---|---|
| ML | *Sie* | 423 | 70 | 44 | 41 | **73.2%** |
| | *Ind* | 29 | 40 | 7 | 8 | **47.0%** |
| | *Wal* | 34 | 2 | 423 | 33 | **85.9%** |
| | *Ack* | 390 | 157 | 332 | 3898 | **81.6%** |
| | **Richtig** | **48.3%** | **14.8%** | **52.4%** | **97.9%** | |
| $IS_2$ monotemp | *Sie* | 458 | 60 | 28 | 33 | **79.2%** |
| | *Ind* | 27 | 44 | 7 | 7 | **51.4%** |
| | *Wal* | 32 | 2 | 426 | 33 | **86.5%** |
| | *Ack* | 289 | 152 | 178 | 4158 | **87.0%** |
| | **Richtig** | **56.9%** | **17.0%** | **66.6%** | **98.3%** | |
| $IS_2$ multitemp | Sie | 433 | 53 | 24 | 68 | **74.9%** |
| | Ind | 28 | 44 | 5 | 8 | **51.7%** |
| | Wal | 28 | 2 | 415 | 47 | **84.4%** |
| | Ack | 190 | 60 | 72 | 4454 | **93.2%** |
| | **Richtig** | **63.7%** | **27.6%** | **80.4%** | **97.3%** | |

*Tabelle 5-6: Konfusionsmatrizen der Klassifikation einer RapidEye Szene im Testgebiet Husum zum Zeitpunkt $t_2$ (7/2009). Pixelanzahl * $10^3$.*

## 5.9. Multitemporale Klassifikation von Daten unterschiedlicher Auflösung ohne Veränderungen der Landbedeckung

In diesem Kapitel wird untersucht, ob die integrierte Analyse von Daten unterschiedlicher Auflösung zu einer Steigerung der Klassifikationsgenauigkeit sämtlicher Daten führt. Dies wird anhand von drei unterschiedlichen Testkonfigurationen evaluiert. Konfiguration I besteht aus einer hoch aufgelösten Ikonos- und einer Landsat-Szene im Testgebiet Herne. Für Konfiguration II wird zwischen die beiden Aufnahmen der Konfiguration I eine RapidEye-Aufnahme eingefügt. Somit erfolgt bei dieser Konfiguration im ersten Schritt eine geringfügige Verringerung der Auflösung mit unveränderter Klassenstruktur, im nächsten Schritt eine deutlichere Verringerung der Auflösung mit veränderter Klassenstruktur. Anhand dieser beiden Konfigurationen wird untersucht, inwieweit die Hinzunahme einer zusätzlichen Aufnahme die Klassifikationsergebnisse der beiden anderen Aufnahmen beeinflusst.

Konfiguration III besteht aus zwei RapidEye-Aufnahmen und einer Landsat-Aufnahme im Testgebiet Husum. Wesentliche Unterscheidungsmerkmale zu den ersten beiden Konfigurationen sind das betrachtete Testgebiet und die Kombination zweier Aufnahmen identischer Auflösung mit einer Aufnahme in niedrigerer Auflösung.

Die Trainings- und Testgebiete werden identisch zu den vorhergehenden Experimenten definiert, d.h. ca. 10% der Fläche der Testgebiete werden zum Training verwendet. Da die in den Bildern beobachtbaren Veränderungen erneut weniger als 0.01% der Testgebiete betreffen, wird die Landbedeckung als unverändert betrachtet. Unabhängig vom realen zeitlichen Abstand der Aufnahmen wird erneut angenommen, dass die Zeitspanne zwischen den Aufnahmen ausreichend ist, um Veränderungen der Landbedeckung zu erwarten. Für zeitlich aufeinanderfolgende Aufnahmen hoher Auflösung wird dieselbe Übergangsmatrix $\mathbf{TM}_{hh}$ wie in Kapitel 5.8 gewählt (Tabelle 5-2). Da in der niedrigeren Auflösung der Landsat Bilder die Klassen *Siedlung* und *Industrie* zu einer Klasse *bebaute Gebiete* verschmelzen, muss sich die veränderte Klassenstruktur auch in der zugehörigen Übergangsmatrix $\mathbf{TM}_{hm}$ widerspiegeln (Tabelle 5-7). Diese 1:*N* Relation zeigt sich in $\mathbf{TM}_{hm}$ durch die Zuweisung hoher Potentiale für den Übergang von sowohl der Klasse *Siedlung* als auch von der Klasse *Industrie* zur Klasse *bebaute Gebiete*. Es gelten des Weiteren die identischen Annahmen über potentielle Veränderungen der Landbedeckungen. Folglich entsprechen die einzelnen Elemente in $\mathbf{TM}_{hm}$ den jeweiligen Elementen in $\mathbf{TM}_{hh}$. Der Gewichtsparameter $\gamma$ des zeitlichen Interaktionspotentials bleibt unverändert und beträgt wie in den vorangegangenen Experimenten $\gamma$=1.5.

| $x_t$ \ $x_{t+1}$ | *Beb* | *Wal* | *Ack* |
|---|---|---|---|
| *Sie* | 1 | 0.05 | 0.05 |
| *Ind* | 1 | 0.05 | 0.05 |
| *Wal* | 0.2 | 1 | 0.1 |
| *Ack* | 0.2 | 0.1 | 1 |

*Tabelle 5-7: Übergangsmatrix $\mathbf{TM}_{hm}$ für Daten hoher Auflösung vom Zeitpunkt t zum Zeitpunkt t+1 mit Daten mittlerer Auflösung.*

| T | monotemporal | | | | multitemporal | | |
|---|---|---|---|---|---|---|---|
| | ML | $IS_1$ | $IS_2$ | $IS_3$ | $IS_1$ | $IS_2$ | $IS_3$ |
| $t_1$ | 74.2%<br>0.63 | 80.7%<br>0.72 | 79.8%<br>0.71 | 78.0%<br>0.68 | 80.7%<br>0.72 | 81.2%<br>0.72 | 80.8%<br>0.72 |
| $t_2$ | 71.7%<br>0.57 | 76.0%<br>0.63 | 75.9%<br>0.63 | 76.1%<br>0.63 | 80.5%<br>0.70 | 80.5%<br>0.70 | 81.1%<br>0.71 |

*Tabelle 5-8: Gesamtgenauigkeit und Kappa Koeffizient für die monotemporale und multitemporale Klassifikation einer Ikonos- ($t_1$) und Landsat-Szene ($t_2$) im Testgebiet Herne ( Konfiguration I).*

Tabelle 5-8 zeigt die Klassifikationsergebnisse für die monotemporale und multitemporale Auswertung der Konfiguration I. Für beide Zeitpunkte führt die Berücksichtigung räumlichen Kontextwissens in der monotemporalen Klassifikation für sämtliche Modelle *IS* zu einer Steigerung von 4-6% im Vergleich zu einer ML-Klassifikation. Durch Integration zeitlichen Kontextwissens kann für die niedriger aufgelöste Landsat Szene eine weitere deutliche Steigerung erzielt werden. Die höher aufgelösten Daten profitieren ebenfalls, wenn auch in geringerem Maße, von der gemeinsamen Auswertung.

Die Ergebnisse für Konfiguration II mit der zusätzlichen RapidEye-Aufnahme zeigt Tabelle 5-9. Diese RapidEye-Aufnahme ist etwas kontrastärmer als die übrigen in dieser Arbeit verwendeten Aufnahmen. Aufgrund dieser Tatsache und der etwas geringeren geometrischen Auflösung von RapidEye im Vergleich zu Ikonos sind die monotemporalen Klassifikationsergebnisse des zugehörigen Zeitpunkts $t_2$ deutlich schlechter als die Ergebnisse der Ikonos-Aufnahme zum Zeitpunkt $t_1$. Durch die multitemporale Auswertung kann für die RapidEye-Aufnahme das Niveau der Ergebnisse der beiden anderen Aufnahmen erreicht werden. Trotz der Integration des „schlechten" Zeitpunkts $t_2$ bleiben die Ergebnisse der beiden anderen Aufnahmen im Vergleich zu Konfiguration I nahezu unverändert. Die Landsat-Szene wird geringfügig schlechter klassifiziert, die Ikonos-Szene sogar etwas besser.

| T | monotemporal | | | | multitemporal | | |
|---|---|---|---|---|---|---|---|
| | ML | $IS_1$ | $IS_2$ | $IS_3$ | $IS_1$ | $IS_2$ | $IS_3$ |
| $t_1$ | 74.2%<br>0.63 | 80.7%<br>0.72 | 79.8%<br>0.71 | 78.0%<br>0.68 | 80.8%<br>0.72 | 81.5%<br>0.73 | 81.3%<br>0.73 |
| $t_2$ | 64.4%<br>0.50 | 70.5%<br>0.58 | 68.6%<br>0.56 | 66.2%<br>0.52 | 79.8%<br>0.70 | 80.6%<br>0.72 | 80.2%<br>0.71 |
| $t_3$ | 71.7%<br>0.57 | 76.0%<br>0.63 | 75.9%<br>0.63 | 76.1%<br>0.63 | 80.1%<br>0.69 | 80.4%<br>0.70 | 80.3%<br>0.70 |

*Tabelle 5-9: **Gesamtgenauigkeit und Kappa Koeffizient für die monotemporale und multitemporale Klassifikation einer Ikonos- ($t_1$), RapidEye- ($t_2$) und Landsat-Szene($t_3$) im Testgebiet Herne (Konfiguration II).***

In Abbildung 5-16 werden die Ergebnisse der Konfiguration II exemplarisch anhand eines kleinen Szenenausschnitts gegenüber gestellt. Besonderer Fokus liegt hierbei auf zwei Objekten: (1) die außerörtliche Straße, die in der Ikonos-Aufnahme durch eine rote Ellipse gekennzeichnet ist, und (2) die beiden dunklen Ackerflächen, die in der RapidEye-Szene von einen blauen Kreis umrundet werden. Diese beiden Ackerflächen werden bei der monotemporalen Analyse der RapidEye-Szene inkorrekt als Wald klassifiziert. Die Integration des zeitlichen Kontexts bewirkt die korrekte Klassifikation dieser Flächen bei der multitemporalen Analyse. Die außerörtliche Straße wird in der ML-Klassifikation der Ikonos- und der RapidEye-Szene vorwiegend korrekt erkannt. In der Landsat-Szene ist die Straße wegen ihrer geringen Breite nicht detektierbar. Bei Verwendung der merkmalsbasierten Modelle bildet sich die Straße auch in den multitemporalen Klassifikationsergebnissen gut ab, bei Verwendung des ausschließlich labelbasierten Modells $IS_1$ verschwindet sie jedoch. Der Effekt, dass kleine Strukturen nur durch Verwendung merkmalsbasierter Modelle erhalten werden, ist für die multitemporale Auswertung ebenso gültig wie für die monotemporale Auswertung.

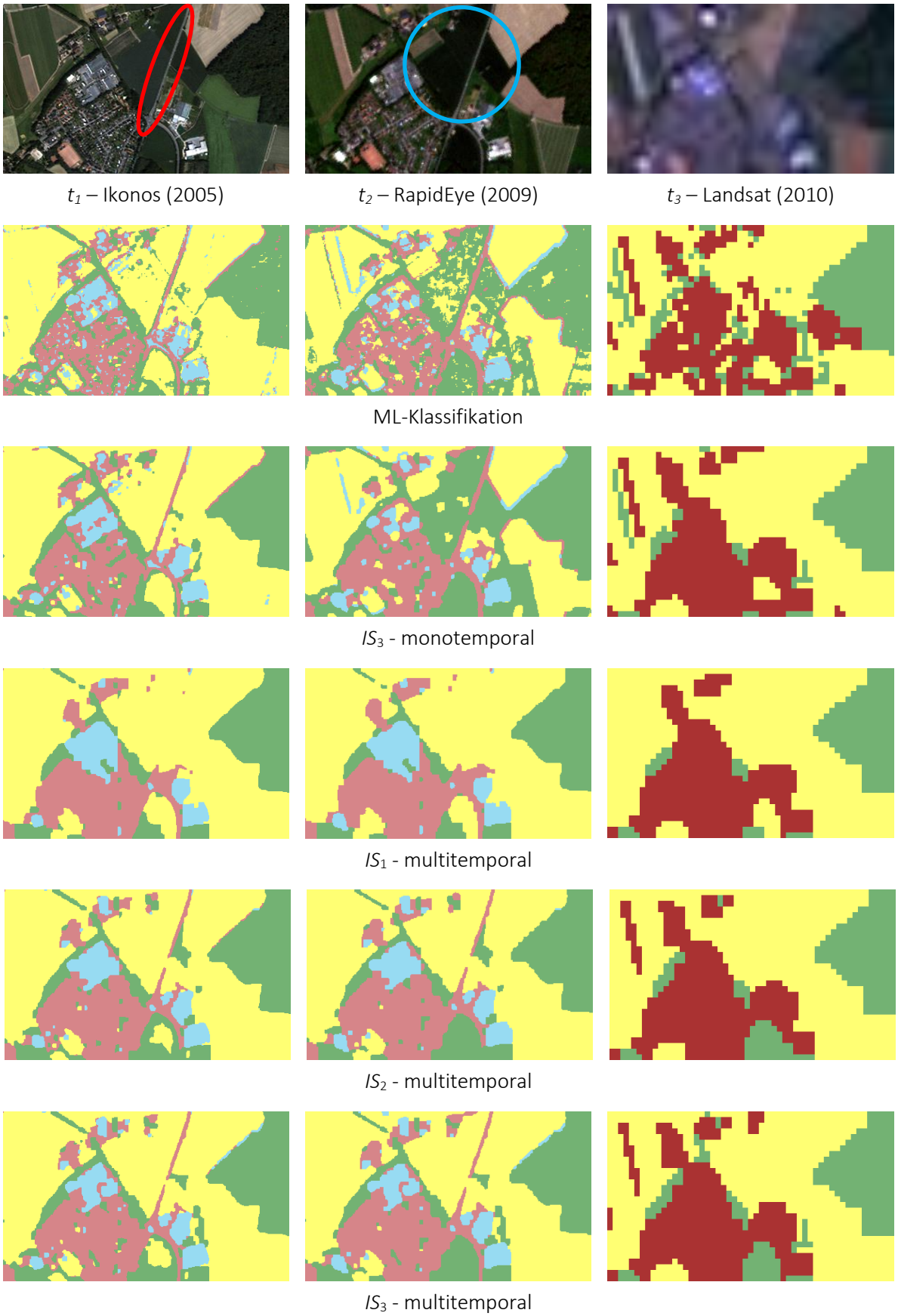



*Abbildung 5-16: Erste Zeile: Ausschnitt aus dem Testgebiet Herne zu drei Zeitpunkten; weitere Zeilen: Klassifikationsergebnisse der Aufnahmen aus der ersten Zeile unter Verwendung unterschiedlicher Methoden und Modelle.*

Für die Klassifikationsergebnisse der Konfiguration III im Testgebiet Husum (Tabelle 5-10) zeigen sich ähnliche Tendenzen wie für die Konfigurationen I und II im Testgebiet Herne. Die Verbesserungen durch die multitemporale Analyse sind sogar noch ausgeprägter. Für die Zeitpunkte mit hoch aufgelösten Daten beträgt der Gewinn der Gesamtgenauigkeit je nach Modell und Zeitpunkt bis zu 5%. Die Gesamtgenauigkeit der der Landsat-Szene wird sogar um ca. 8% im Vergleich zur monotemporalen CRF-Auswertung und um ca. 17% zur ML-Klassifikation verbessert.

| T | monotemporal | | | | multitemporal | | |
|---|---|---|---|---|---|---|---|
| | *ML* | $IS_1$ | $IS_2$ | $IS_3$ | $IS_1$ | $IS_2$ | $IS_3$ |
| $t_1$ | 79.9%<br>0.53 | 88.0%<br>0.67 | 86.4%<br>0.65 | 84.8%<br>0.62 | 89.7%<br>0.70 | 90.1%<br>0.72 | 89.7%<br>0.72 |
| $t_2$ | 80.6%<br>0.55 | 87.0%<br>0.67 | 85.7%<br>0.65 | 84.4%<br>0.62 | 89.7%<br>0.70 | 89.4%<br>0.71 | 88.9%<br>0.70 |
| $t_3$ | 71.7%<br>0.40 | 81.8%<br>0.53 | 80.2%<br>0.50 | 79.4%<br>0.50 | 89.2%<br>0.71 | 88.2%<br>0.69 | 88.7%<br>0.71 |

*Tabelle 5-10:* ***Gesamtgenauigkeit und Kappa Koeffizient für die monotemporale und multitemporale Klassifikation von zwei RapidEye-Szenen ($t_1$ und $t_2$) und einer Landsat-Szene ($t_3$) im Testgebiet Husum (Konfiguration III).***

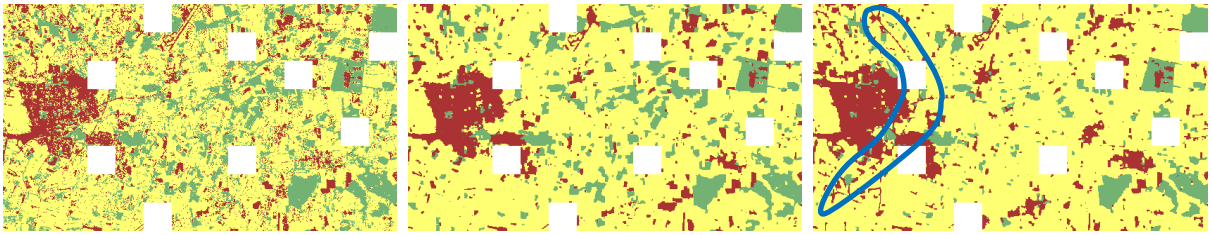
Für den Zeitpunkt $t_3$ ist zudem eine deutliche Steigerung des Kappa-Koeffizienten, vor allem zwischen monotemporaler und multitemporaler Analyse, erkennbar. Dies lässt sich anhand der zugehörigen Konfusionsmatrizen erklären (Tabelle 5-11). Dargestellt wird die Matrix für das Modell $IS_3$, die Matrizen für die anderen Modelle weichen nur geringfügig ab. Insbesondere die Richtigkeit der Klasse *Wald* ist in der monotemporalen Auswertung sehr gering (32% für ML, 43% für CRF). Dies ist auf die spektrale Ähnlichkeit bebauter Ackerflächen mit Waldflächen zurückzuführen. Mit dem multitemporalen Ansatz kann die Richtigkeit auf über 72% erhöht werden. Da die Klasse *Wald* nur einen geringen Anteil an der Gesamtfläche hat, wirkt sich diese erhebliche Verbesserung im Kappa-Koeffizient deutlich stärker aus als in der Gesamtgenauigkeit. Des Weiteren ist in Tabelle 5-11 erkennbar, dass es häufig zu Fehlklassifikationen zwischen den Klassen *Acker* und *bebautes Gebiet* kommt. Diese Fehler können durch die Berücksichtigung räumlichen und zeitlichen Kontexts um über ein Drittel im Vergleich zur ML-Klassifikation gesenkt werden. Abbildung 5-17 zeigt die zur Tabelle 5-11 gehörigen Klassifikationsergebnisse der Landsat-Szene aus Konfiguration III. Deutlich ist die korrekte Reduktion der zuvor fehlerhaft klassifizierten Waldfläche erkennbar. Bemerkenswert ist, dass die große Umgehungsstraße (innerhalb des blauen Polygons in Abbildung 5-17), die in der ML-Klassifikation zu erahnen ist, aufgrund ihrer sehr schwachen Ausprägung im monotemporalen CRF-Ergebnis nicht mehr vorhanden ist. Im multitemporalen Ergebnis tritt die Straße jedoch deutlich hervor.

In allen Konfigurationen ist bei multitemporaler Auswertung bzgl. Gesamtgenauigkeit und Kappa-Koeffizient nur ein geringer Unterschied bei Nutzung der unterschiedlichen Modelle für das räumliche Interaktionspotential zu erkennen. Wie im monotemporalen Fall sind diese beiden Kriterien auch für die multitemporale Klassifikation nur bedingt als Optimierungsfunktion zum Lernen der Parameter der

Methode geeignet. Vielmehr muss bei der Wahl des Modells erneut auch die Fähigkeit zur Erhaltung kleiner Strukturen berücksichtigt werden. Unabhängig von der Wahl des räumlichen Modells liefert die multitemporale Betrachtung mit der in dieser Arbeit entwickelten Methode eine deutliche Verbesserung der Klassifikationsergebnisse im Vergleich zur monotemporalen Analyse. Nicht nur die bei monotemporaler Betrachtung schlechter klassifizierten Zeitpunkte profitieren von der Berücksichtigung zeitlichen Kontextwissens, auch die Ergebnisse der „besseren“ Zeitpunkte werden gesteigert.

| | $x_i$ / $x_i^{ref}$ | *Beb* | *Wal* | *Ack* | **Vollständig** |
|---|---|---|---|---|---|
| *ML* | *Beb* | 128 | 32 | 99 | **49.3%** |
| | *Wal* | 2 | 112 | 11 | **89.7%** |
| | *Ack* | 117 | 204 | 943 | **74.6%** |
| | **Richtig** | **51.6%** | **32.1%** | **89.6%** | |
| *IS*$_3$ monotemp | *Beb* | 132 | 20 | 107 | **51.2%** |
| | *Wal* | 2 | 110 | 14 | **87.8%** |
| | *Ack* | 70 | 128 | 1066 | **84.3%** |
| | **Richtig** | **64.8%** | **42.6%** | **89.9%** | |
| *IS*$_3$ multitemp | *Beb* | 181 | 15 | 62 | **70.0%** |
| | *Wal* | 4 | 111 | 10 | **89.0%** |
| | *Ack* | 68 | 28 | 1169 | **92.4%** |
| | **Richtig** | **71.2%** | **72.2%** | **94.2%** | |

*Tabelle 5-11: Konfusionsmatrizen der Klassifikation der Landsat-Szene im Testgebiet Husum zum Zeitpunkt $t_3$ (7/2009). Pixelanzahl * $10^2$.*



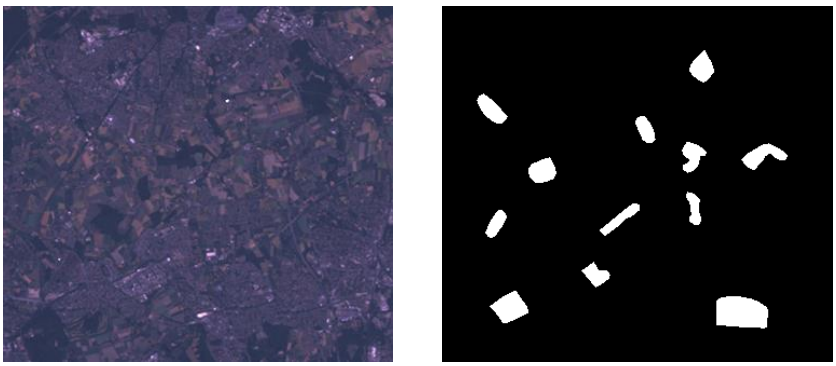
ML *IS*$_3$, monotemporal *IS*$_3$, multitemporal

*Abbildung 5-17: Klassifikationsergebnis Landsat-Szene im Testgebiet Husum. Multitemporale Klassifikation in Kombination mit zwei RapidEye-Szenen. Rot: Bebaute Gebiete; Grün: Wald; Gelb: Acker.*

## 5.10. Multitemporale Klassifikation von Daten unterschiedlicher Auflösung mit Veränderungen der Landbedeckung

Nachdem in den vorherigen Kapiteln die Landbedeckung in den einzelnen Aufnahmen stets unverändert war, wird in diesem Kapitel die Methode für die Klassifikation multitemporaler Daten mit Veränderungen eingesetzt. Es ist von Interesse, ob das Verfahren in der Lage ist, in einem Auswerteschritt sowohl die Veränderungen der Landbedeckung zu erkennen als auch die Klassifikationsgenauigkeit im Vergleich zur monotemporalen Auswertung zu steigern. Darüber hinaus wird der Effekt, den eine Veränderung der temporalen Übergangsmatrix auf die Ergebnisse hat, evaluiert.

Da in den zur Verfügung stehenden Daten beider Testgebiete die Veränderungen der Landbedeckung sehr gering sind, werden Veränderungen durch Manipulation von Landsat-Aufnahmen simuliert. Es werden ausschließlich die in Deutschland vorrangig zu erwartenden Veränderungen simuliert, d.h. Übergänge von Acker- oder Waldflächen zu bebauten Gebieten. Hierzu werden bebaute Gebiete, die außerhalb der jeweiligen Testgebiete liegen, jedoch aus derselben Aufnahme stammen, an die Position bisher unbebauter Gebiete im Testgebiet kopiert. Für das Testgebiet Herne werden 12 Veränderungsflächen simuliert (Abbildung 5-18), von denen sieben Erweiterungen existierender bebauter Flächen darstellen. Fünf Flächen simulieren das Entstehen neuer bebauter Flächen „auf der grünen Wiese“. Im Testgebiet Husum werden zehn Veränderungsflächen simuliert, von denen sich vier an bestehende bebaute Gebiete anschließen.



Manipuliertes Bild Veränderungsmaske

*Abbildung 5-18: Manipulierte Landsat-Szene mit 12 Veränderungen von Wald- oder Ackerflächen zu bebauten Gebieten entsprechend Veränderungsmaske.*

Um eine Vergleichbarkeit der Ergebnisse mit den vorherigen Experimenten zu gewährleisten, wird die Methode anhand der drei bereits in Kapitel 5.9 verwendeten Konfigurationen mit identischen Parametern evaluiert. Ebenso wird die zeitliche Übergangsmatrix $\mathbf{TM}_{hh}$ (Tabelle 5-2) für den Übergang zwischen zwei hoch aufgelösten Aufnahmen verwendet. Zusätzlich zur Verwendung der Übergangsmatrix $\mathbf{TM}_{hm}$ (im folgenden $\mathbf{TM}_{hm}^1$) zwischen den Aufnahmen hoher und mittlerer Auflösung, werden sämtliche Experimente mit einer zweiten Übergangsmatrix $\mathbf{TM}_{hm}^2$ berechnet (Tabelle 5-12). Die beiden Übergangsmatrizen unterscheiden sich einzig in den Potentialen $PT_{WalBeb}$ bzw. $PT_{AckBeb}$ für den Übergang von *Wald-* oder *Ackerfläche* zu *bebauten Gebiet*. In $\mathbf{TM}_{hm}^1$ ist $PT_{WalBeb}=PT_{AckBeb}=0.2$, während in $\mathbf{TM}_{hm}^2$ $PT_{WalBeb}=PT_{AckBeb}=0.4$ ist. Folglich ist der Übergang von *Wald-* oder Ackerflächen zu *bebautem Gebiet* in $\mathbf{TM}_{hm}^2$ wesentlich wahrscheinlicher als in $\mathbf{TM}_{hm}^1$. Dennoch ist auch in dieser Matrix eine

unveränderte Landbedeckung für sämtliche Klassen am wahrscheinlichsten. Es wird untersucht, inwieweit sich diese Unterschiede in $\mathbf{TM}_{hm}$ auf die Klassifikationsgenauigkeit und die detektierten Veränderungen auswirken.

| $x_t$ \ $x_{t+1}$ | *Beb* | *Wal* | *Ack* |
|---|---|---|---|
| *Sie* | 1 | 0.05 | 0.05 |
| *Ind* | 1 | 0.05 | 0.05 |
| *Wal* | **0.4** | 1 | 0.1 |
| *Ack* | **0.4** | 0.1 | 1 |

*Tabelle 5-12: Übergangsmatrix $TM_{hm}^2$ vom Zeitpunkt t mit hoher Auflösung zum Zeitpunkt t+1 mit mittlerer Auflösung.*

Tabelle 5-13 und Tabelle 5-14 zeigen die Klassifikationsergebnisse der Konfigurationen I und II des Testgebiets Herne jeweils unter Verwendung der Übergangsmatrizen $\mathbf{TM}_{hm}^1$ und $\mathbf{TM}_{hm}^2$. Für die hoch aufgelösten Zeitpunkte können keine signifikanten Veränderungen durch Einsatz der unterschiedlichen Matrizen festgestellt werden. Die Klassifikationsgenauigkeit der manipulierten Landsat-Daten ist zumeist geringfügig höher bei Verwendung von $\mathbf{TM}_{hm}^2$ mit $PT_{WalBeb}$=$PT_{AckBeb}$=0.4. Der Grund hierfür ist die bessere Klassifikation veränderter Flächen (siehe unten). Insgesamt unterscheiden sich die Klassifikationsergebnisse nur unwesentlich von den Werten, die mit nicht manipulierten Daten, d.h. ohne Veränderung der Landbedeckung, erzielt wurden (vgl. Tabelle 5-8 und Tabelle 5-9). Für die hoch aufgelösten Daten sind somit keine negativen Auswirkungen durch die manipulierten Landsat-Daten erkennbar.

| T | monotemporal | | | | multitemporal | | | $PT_{WalBeb}$ $PT_{AckBeb}$ |
|---|---|---|---|---|---|---|---|---|
| | *ML* | $IS_1$ | $IS_2$ | $IS_3$ | $IS_1$ | $IS_2$ | $IS_3$ | |
| $t_1$ | 74.2% 0.63 | 80.7% 0.72 | 79.8% 0.71 | 78.0% 0.68 | 80.5% 0.71 | 81.1% 0.72 | 80.6% 0.72 | 0.2 |
| | | | | | 80.9% 0.72 | 81.1% 0.72 | 80.5% 0.72 | 0.4 |
| $t_2$ | 71.3% 0.56 | 74.2% 0.60 | 75.5% 0.62 | 75.8% 0.63 | 79.5% 0.68 | 80.1% 0.69 | 80.5% 0.70 | 0.2 |
| | | | | | 80.0% 0.69 | 80.4% 0.69 | 80.6% 0.70 | 0.4 |

*Tabelle 5-13: Gesamtgenauigkeit und Kappa Koeffizient für die monotemporale und multitemporale Klassifikation einer Ikonos- ($t_1$), und der manipulierten Landsat-Szene($t_2$) im Testgebiet Herne (Konfiguration I) für unterschiedliche räumliche und zeitliche Interaktionspotentiale.*

Einen Überblick über die die Klassifikationsgenauigkeit der veränderten Flächen für die Konfigurationen I und II unter Verwendung von $\mathbf{TM}_{hm}^1$ und $\mathbf{TM}_{hm}^2$ gibt Tabelle 5-15. Mit $\mathbf{TM}_{hm}^1$ können ca. 60% aller simulierten bebauten Flächen korrekt erkannt werden. Im Vergleich zu den Ergebnissen mit $\mathbf{TM}_{hm}^1$ zeigt sich eine deutliche Steigerung von 15-20% je nach Modell und Konfiguration mit der erhöhten Übergangswahrscheinlichkeit in $\mathbf{TM}_{hm}^2$, d.h. mit $PT_{WalBeb}$=$PT_{AckBeb}$=0.4.

| T | monotemporal | | | | multitemporal | | | $PT_{WalBeb}$ $PT_{AckBeb}$ |
|---|---|---|---|---|---|---|---|---|
| | *ML* | $IS_1$ | $IS_2$ | $IS_3$ | $IS_1$ | $IS_2$ | $IS_3$ | |
| $t_1$ | 74.2% 0.63 | 80.7% 0.72 | 79.8% 0.71 | 78.0% 0.68 | 80.7% 0.72 | 81.3% 0.73 | 81.1% 0.73 | 0.2 |
| | | | | | 80.7% 0.72 | 81.3% 0.73 | 81.0% 0.72 | 0.4 |
| $t_2$ | 64.4% 0.50 | 70.5% 0.58 | 68.6% 0.56 | 66.2% 0.52 | 79.6% 0.70 | 80.3% 0.71 | 80.0% 0.71 | 0.2 |
| | | | | | 79.8% 0.71 | 80.4% 0.72 | 80.0% 0.71 | 0.4 |
| $t_3$ | 71.3% 0.56 | 74.2% 0.60 | 75.5% 0.62 | 75.8% 0.63 | 78.7% 0.67 | 79.4% 0.68 | 79.1% 0.68 | 0.2 |
| | | | | | 79.0% 0.67 | 79.3% 0.68 | 79.5% 0.68 | 0.4 |

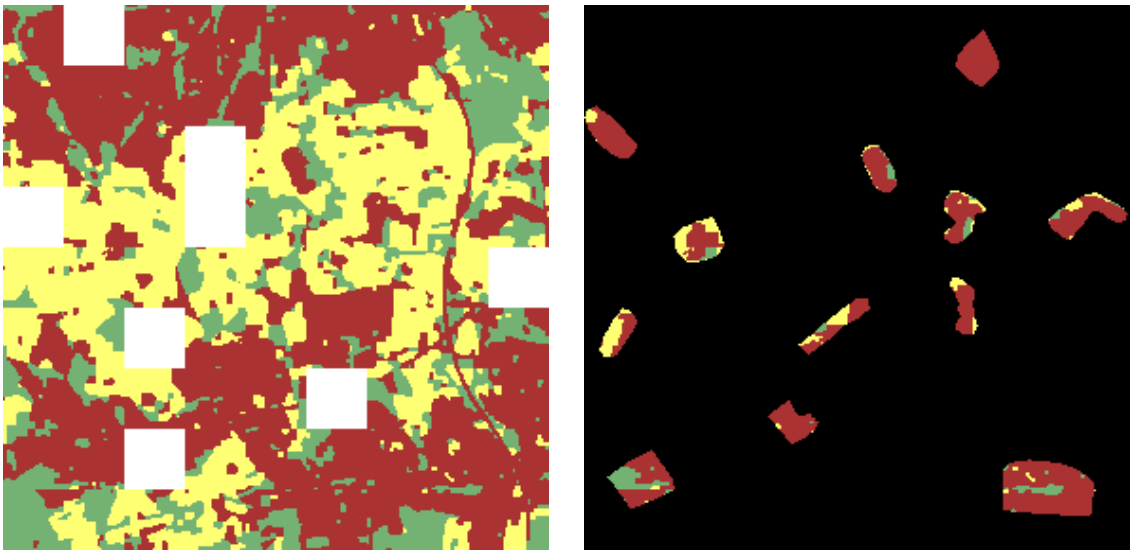
*Tabelle 5-14: **Gesamtgenauigkeit und Kappa Koeffizient für die monotemporale und multitemporale Klassifikation einer Ikonos- ($t_1$), RapidEye- ($t_2$) und der manipulierten Landsat-Szene($t_3$) im Testgebiet Herne (Konfiguration II) für unterschiedliche räumliche und zeitliche Interaktionspotentiale.***

| Konfig. | $PT_{WalBeb}$ $PT_{AckBeb}$ | multitemporal | | | monotemporal | | | |
|---|---|---|---|---|---|---|---|---|
| | | $IS_1$ | $IS_2$ | $IS_3$ | ML | $IS_1$ | $IS_2$ | $IS_3$ |
| I | 0.2 | 54% | 58% | 59% | 75% | 80% | 83% | 82% |
| | 0.4 | 73% | 73% | 75% | | | | |
| II | 0.2 | 59% | 63% | 61% | | | | |
| | 0.4 | 75% | 79% | 77% | | | | |

*Tabelle 5-15: **Korrekt erkannte Veränderungsflächen (Gesamtgenauigkeit der klassifizierten Pixel) im Testgebiet Herne (Konfiguration I und II) für unterschiedliche räumliche und zeitliche Interaktionspotentiale; rechts zum Vergleich die monotemporalen Klassifikationsgenauigkeiten für die Veränderungsflächen.***

Mit $TM_{hm}^2$ wird bis auf 4-5% der Gesamtgenauigkeit das Niveau einer monotemporalen CRF-Klassifikation für die Veränderungsflächen erreicht. Diese monotemporalen Ergebnisse werden als Referenz für die maximal erreichbare Klassifikationsgenauigkeit für die multitemporale Veränderungsanalyse angesehen, da der Informationsgehalt der Daten in diesen Flächen keine bessere Klassifikation ermöglicht. Es ist zu beachten, dass die monotemporale Klassifikationsgenauigkeit der Veränderungsflächen nur geringfügig über der Klassifikationsgenauigkeit des kompletten Testgebiets liegt. Dies dient als Indikator dafür, dass es sich bei den gewählten Veränderungsflächen um für die Klasse repräsentative Flächen handelt und nicht besonders einfach zu detektierende Manipulationen vorgenommen wurden.

(1) (2)



(3) (4)

*Abbildung 5-19: (1)-(3) Multitemporales Klassifikationsergebnis der Zeitpunkte $t_1$ bis $t_3$ für Konfiguration II (Testgebiet Herne) mit $IS_2$ und $TM_{hm}^2$. (4) Klassifikationsergebnis der manipulierten Flächen (rote Pixel wurde als bebaut klassifiziert und sind somit korrekt).*

Abbildung 5-19 zeigt exemplarisch die Klassifikationsergebnisse für die multitemporale Analyse der Konfiguration II mit $IS_2$ und $\mathbf{TM}_{hm}^2$ für die drei Aufnahmezeitpunkte. Die erkannten Veränderungsflächen in $t_3$ bewirken nur unwesentliche Veränderungen dieser Flächen zu den Zeitpunkten $t_1$ und $t_2$. Kleine Strukturen bleiben in den Ergebnissen der hoch aufgelösten Daten weiterhin gut erhalten, obwohl diese in den Daten mittlerer Auflösung zum Zeitpunkt $t_3$ nicht erkennbar sind. Die ermittelten Landbedeckungen sind in allen Aufnahmen zumeist identisch. Mit Ausnahme der Veränderungsflächen sind in Abbildung 5-19, (3) keine übermäßig häufigen Übergänge zu bebautem Land zu beobachten, d.h. die Wahl der zeitlichen Potentiale $PT_{WalBeb}$ und $PT_{AckBeb}$ in $\mathbf{TM}_{hm}^2$ ist gerechtfertigt. Abbildung 5-19, (4) fokussiert auf die manipulierten Flächen. In 11 der 12 Flächen wird die Majorität der Pixel korrekt als bebautes Gebiet klassifiziert. Mit $\mathbf{TM}_{hm}^1$ gelingt dies lediglich für neun Flächen. Insbesondere die direkt an bestehende bebaute Gebiete angrenzenden Randgebiete der manipulierten Flächen werden mit großer Sicherheit korrekt klassifiziert.

Die für die Ergebnisse der Konfigurationen I und II getroffenen Beobachtungen lassen sich auf die Konfiguration III im Testgebiet Husum übertragen. Die Klassifikationsgenauigkeit und der Kappa-Koeffizient (Tabelle 5-16) unterscheiden sich kaum von den mit nicht manipulierten Daten erzielten Werten (vgl. Tabelle 5-10). Auch die Verwendung der unterschiedlichen Übergangsmatrizen $\mathbf{TM}_{hm}^1$ und $\mathbf{TM}_{hm}^2$ führt zu keinen großen Abweichungen. Bei den korrekt erkannten Veränderungsflächen zeigt sich erneut eine deutliche Verbesserung durch die Nutzung von $\mathbf{TM}_{hm}^2$ (Tabelle 5-17). Über 70% der veränderten Pixel können mit $\mathbf{TM}_{hm}^2$ korrekt klassifiziert werden. Somit werden wiederum annähernd die mit einer monotemporalen Klassifikation in diesen Flächen erzielbaren Werte erreicht.

| T | monotemporal | | | | multitemporal | | | $PT_{WalBeb}$ $PT_{AckBeb}$ |
|---|---|---|---|---|---|---|---|---|
| | *ML* | $IS_1$ | $IS_2$ | $IS_3$ | $IS_1$ | $IS_2$ | $IS_3$ | |
| $t_1$ | 79.9% 0.53 | 88.0% 0.67 | 86.4% 0.65 | 84.8% 0.62 | 89.7% 0.70 | 90.0% 0.72 | 89.7% 0.72 | 0.2 |
| | | | | | 89.7% 0.70 | 90.0% 0.72 | 89.6% 0.72 | 0.4 |
| $t_2$ | 80.6% 0.55 | 87.0% 0.67 | 85.7% 0.65 | 84.4% 0.62 | 89.7% 0.70 | 89.2% 0.71 | 88.8% 0.70 | 0.2 |
| | | | | | 89.7% 0.70 | 89.2% 0.71 | 88.8% 0.70 | 0.4 |
| $t_3$ | 73.0% 0.45 | 81.7% 0.55 | 80.9% 0.57 | 80.6% 0.57 | 87.1% 0.68 | 86.8% 0.68 | 87.2% 0.70 | 0.2 |
| | | | | | 87.4% 0.69 | 86.1% 0.68 | 87.0% 0.70 | 0.4 |

*Tabelle 5-16: Gesamtgenauigkeit und Kappa Koeffizient für die monotemporale und multitemporale Klassifikation von zwei RapidEye-Szenen ($t_1$ und $t_2$) und der manipulierten Landsat-Szene im Testgebiet Husum (Konfiguration III) für unterschiedliche räumliche und zeitliche Interaktionspotentiale.*

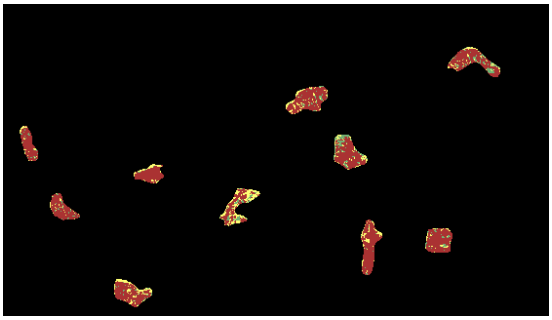
| PT | multitemporal | | | Monotemporal | | | |
|---|---|---|---|---|---|---|---|
| | $IS_1$ | $IS_2$ | $IS_3$ | *ML* | $IS_1$ | $IS_2$ | $IS_3$ |
| 0.2 | 43% | 63% | 58% | 75% | 72% | 81% | 82% |
| 0.4 | 66% | 74% | 71% | | | | |

*Tabelle 5-17: Korrekt erkannte Veränderungsflächen im Testgebiet Husum (Konfiguration III) für unterschiedliche räumliche und zeitliche Interaktionspotentiale..*

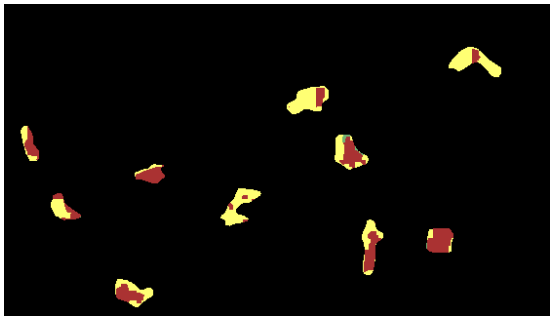
Die Ergebnisse zu Tabelle 5-17 werden in Abbildung 5-20 visualisiert. Die Steigerung der Veränderungsdetektion durch Verwendung von $\mathbf{TM}_{hm}^2$ im Vergleich zu $\mathbf{TM}_{hm}^1$ ist klar ersichtlich. Für alle räumlichen Modelle *IS* wird die Majorität der Pixel für neun der zehn Flächen mit $\mathbf{TM}_{hm}^2$ korrekt klassifiziert. Für $\mathbf{TM}_{hm}^1$ gelingt dies nur für sechs (mit $IS_1$), sieben (mit $IS_3$) bzw. acht (mit $IS_2$) Flächen.

Die in diesem Kapitel vorgestellten Untersuchungen belegen, dass die in dieser Arbeit entwickelte Methode in der Lage ist, Veränderungen der Landbedeckung mit großer Sicherheit zu erkennen und zudem gleichzeitig die Klassifikationsgenauigkeit sämtlicher Daten zu steigern. Nicht nur die Klassifikationsgenauigkeit von Daten mittlerer Auflösung wird gesteigert, auch die hoch aufgelösten Daten profitieren von der integrierten Analyse. Für beide Testgebiete werden je nach verwendetem räumlichem Modell 70-80% der veränderten Pixel korrekt erkannt. Die Majorität der Pixel wird testgebietsübergreifend in 20 der 22 manipulierten Flächen korrekt klassifiziert. Für die Veränderungsdetektion hat die Wahl des räumlichen Interaktionspotentials *IS* einen geringeren Einfluss auf die Ergebnisse als die Konfiguration der Übergangsmatrizen $\mathbf{TM}$. Insgesamt werden unter Verwendung von $IS_2$ Veränderungen am besten erkannt.

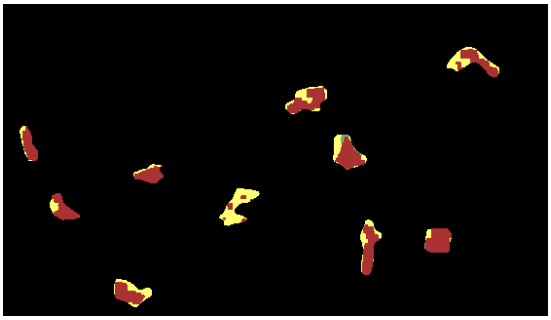
Die offensichtlichen Unterschiede aufgrund der Wahl der Übergangsmatrix $\mathbf{TM}_{hm}$ verdeutlichen die Wichtigkeit von deren Konfiguration. Die Herausforderung bei einer empirischen Bestimmung der Potentiale ist stets, Werte zu definieren, die den realen Veränderungen entsprechen. Ideal wäre es, die einzelnen Elemente *PT* der Matrix aus den Daten zu lernen. Dies erfordert jedoch einen großen Datenbestand mit zahlreichen repräsentativen Veränderungsflächen. Da ein solcher Datenbestand nur selten vorliegt, ist Expertenwissen über zu erwartende Veränderungen erforderlich. Es hat sich gezeigt, dass bei geeigneter Konfiguration der Übergangsmatrix mit der vorgestellten Methode nahezu sämtliche durch Simulation veränderten Flächen korrekt detektiert werden konnten.



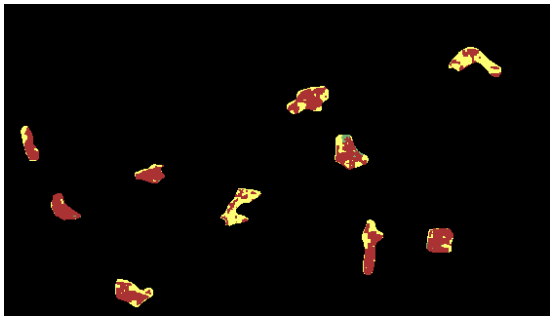
ML (75%)



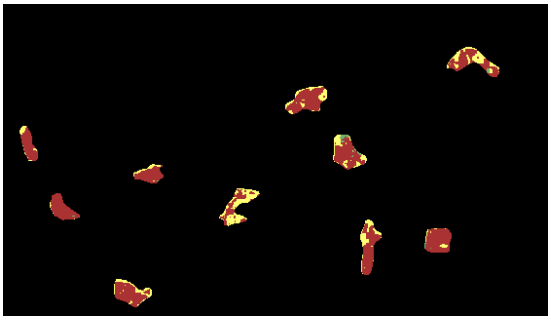
*IS*$_1$, **TM**$_{hm}^1$ (43%)



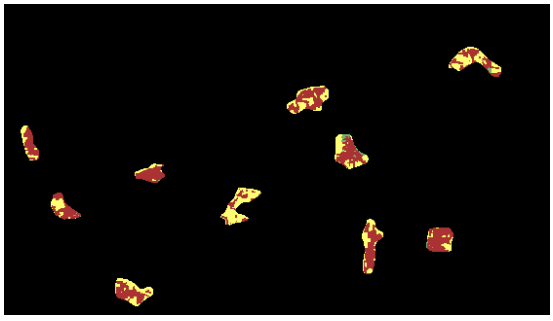
*IS*$_1$, **TM**$_{hm}^2$ (66%)



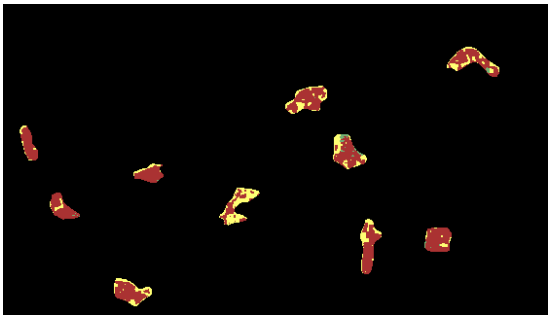
*IS*$_2$, **TM**$_{hm}^1$ (63%)



*IS*$_2$, **TM**$_{hm}^2$ (74%)



*IS*$_3$, **TM**$_{hm}^1$ (58%)



*IS*$_3$, **TM**$_{hm}^2$ (71%)

*Abbildung 5-20: Klassifikationsergebnis der manipulierten Flächen im Testgebiet Husum (Konfiguration III) für unterschiedliche räumliche und zeitliche Interaktionspotentiale.*

# 6. Schlussbetrachtung und Ausblick

In dieser Arbeit wurde ein neuer Ansatz zur gemeinsamen Analyse multitemporaler Fernerkundungsdaten unterschiedlicher Auflösung vorgestellt. Basierend auf dem Konzept der Conditional Random Fields erlaubt der Ansatz die Berücksichtigung sowohl von räumlichem als auch zeitlichem Kontext. Für die Berücksichtigung von zeitlichem Kontext werden Übergangswahrscheinlichkeiten für die Veränderung von Klassen zwischen einzelnen Zeitpunkten in Form von Übergangsmatrizen für zeitliche Interaktionen in den Ansatz integriert. Die Klassenstruktur kann entsprechend der Auflösung der Daten variiert werden. Das zugrunde liegende graphische Modell erlaubt allgemein die gemeinsame Auswertung von Daten beliebiger Zeitpunkte mit beliebiger Auflösung.

Die Innovation aus Sicht der Anwendung ist die Fähigkeit der Methode, zum einen die Klassifikationsgenauigkeit sämtlicher Daten zu steigern und zum anderen gleichzeitig Veränderungen zu detektieren und deren Klassenzugehörigkeit zu bestimmen. In bereits existierenden Ansätzen wurden diese Aufgaben getrennt voneinander gelöst. Aus methodischer Sicht liegt die Neuartigkeit in einer Erweiterung des CRF-Ansatzes zur Verknüpfung mehrerer Bilder eines identischen Szenenausschnitts. Im Gegensatz zu bestehenden Ansätzen, die bei multiskalarer Betrachtung lediglich die unterschiedlichen Auflösungsstufen eines Bildes miteinander verknüpfen, können unterschiedlich aufgelöste Bilder mit unterschiedlichem Inhalt gemeinsam analysiert werden.

Die alleinige Berücksichtigung räumlichen Kontextwissens für monotemporale Auswertungen erbrachte in den vorgestellten Experimenten für zwei großräumige Testgebiete mit mehreren Aufnahmen im Vergleich zu einer Maximum-Likelihood-Klassifikation eine Verbesserung der Gesamtgenauigkeit von zumeist 5%-10%. Für den CRF-basierten Ansatz wurden drei unterschiedliche Modelle für das räumliche Interaktionspotential getestet. Es zeigte sich, dass bei annähernd identischer Gesamtgenauigkeit der Klassifikation der visuelle Eindruck und die Interpretationsmöglichkeiten insbesondere feiner Strukturen stark unterschiedlich sind. Die Wahl geeigneter Modelle und Parameter für das räumliche Interaktionspotential sollte somit nicht ausschließlich auf den „harten“ Kriterien Gesamtgenauigkeit und Kappa-Koeffizient beruhen.

Die Integration des zeitlichen Kontextwissens erhöhte die Klassifikationsgenauigkeit bei Bilddaten identischer Auflösung um weitere 1%-5%. In Kombination mit niedriger aufgelösten Daten wurde für die niedriger aufgelösten Daten eine Steigerung von teilweise über 10% erzielt. Testgebietsübergreifend wurde bei Experimenten zur Detektion von Veränderungen die Majorität der Pixel in 20 der 22 durch Simulation veränderten Flächen korrekt klassifiziert.

Ein umfassender Vergleich mit anderen State-of-the-Art Klassifikationsverfahren ist nicht Bestandteil dieser Arbeit. Der Grund ist, dass kein anderes Verfahren bekannt ist, welches gleichzeitig die Klassifikationsgenauigkeit multitemporaler Daten unter Nutzung räumlichen Kontexts erhöht und dabei die Erkennung von Veränderungen zulässt. Sehr wohl können jedoch einzelne Aspekte des Ansatzes verglichen werden. In einer gemeinsamen Arbeit mit Raul Feitosa (Hoberg et al., 2015) verglich dieser die Ergebnisse einer multitemporalen Klassifikation mit dem vorgestellten CRF-basierten Ansatz mit den Ergebnissen einer Klassifikation mit Hidden Markov Models (HMM). Die eingesetzte Methode folgt hierbei Leite et al. (2011) und wurde verwendet, um das zeitliche Verhalten der einzelnen Pixel zu

analysieren. Das HMM ist eine Markov-Kette, wobei die Knoten den Pixeln zu den unterschiedlichen Zeitpunkten entsprechen. Jede Menge von Pixeln an einer identischen räumlichen Position, aufgenommen zu den unterschiedlichen Zeitpunkten, wurde somit einzeln ohne Berücksichtigung räumlichen Kontextwissens mit dem HMM-Ansatz klassifiziert. Der HMM-Ansatz ist in der Lage, unterschiedliche Klassenstrukturen zu berücksichtigen, erfordert jedoch eine einheitliche Auflösung der Bilddaten. Aus diesem Grund wurden die niedriger aufgelösten Landsat-Daten für die Experimente resampled. Es wurden die gleichen zeitlichen Übergangsmatrizen wie bei der CRF-Klassifikation genutzt. Die multitemporalen Klassifikationsergebnisse des HMM-Ansatzes lagen in sämtlichen Experimenten ca. 5%, im Maximum bis zu 10%, über den Ergebnissen einer monotemporalen Maximum-Likelihood-Klassifikation. Sie lagen jedoch stets ca. 4%-5% unter denen des CRF-basierten Ansatzes. Es ist davon auszugehen, dass dieser Unterschied im Wesentlichen aus der Berücksichtigung räumlichen Kontextwissens im CRF-basierten Ansatz resultiert. Die Ergebnisse der Experimente in Hoberg et al. (2015) zeigen, dass selbst bei einer reduzierten Komplexität der Anwendung, der in dieser Arbeit entwickelte CRF-basierte Ansatz zahlreichen bewährten Klassifikationsverfahren vorzuziehen ist. Darüber hinaus bietet er die zusätzlichen Optionen, Bilder unterschiedlicher Auflösungen gemeinsam zu klassifizieren und gleichzeitig Veränderungen zu detektieren. Ein Vergleich mit weiteren Klassifikationsverfahren, insbesondere Convolutional Neural Networks (CNN), sollte Bestandteil künftiger Forschungen sein.

Seit dem Jahr 2012 (Krizhevsky et al., 2012) sind CNN der bedeutendste Trend im Feld der künstlichen Intelligenz für die Analyse großer Datenmengen. CNN sind eine komplexe Ausprägung von künstlichen neuronalen Netzwerken, welche auf vernetzten künstlichen Neuronen nach biologischem Vorbild basieren. Ein Netzwerk besteht allgemein aus mehreren Schichten (layers) von Neuronen, wobei nur die hinterste Schicht, die Ausgabeschicht, sichtbar ist. Davorliegende Schichten sind nicht sichtbar (hidden layers). Die Primitive liefern Eingaben in die erste verdeckte Schicht des Netzes. Für eine Klassifikation muss in einem ersten Schritt eine initiale Topologie des Netzes in Abhängigkeit von der Klassifikationsaufgabe gewählt werden. In der folgenden Trainingsphase lernt das Netz anhand umfangreicher Trainingsdaten, so können u.a. neue Verbindungen von Neuronen entstehen oder die Gewichtungen von Neuronen verändert werden. Die Methode basiert ausschließlich auf Daten, Modellwissen wird nicht integriert. Merkmale werden aus den Daten gelernt, eine explizite Formulierung von Merkmalen durch einen Anwender ist nicht erforderlich. Künstliche neuronale Netzwerke wurden bereits in den 1980er und 1990er Jahren in der Fernerkundung eingesetzt, waren jedoch lange Zeit aufgrund hoher Rechenanforderungen auf die Nutzung von lediglich maximal zwei verborgenen Schichten limitiert. Gestiegene Rechnerleistungen ermöglichten die Verwendung von mehr als zwei verborgenen Schichten und führten zu deutlich gestiegenen Klassifikationsgenauigkeiten. Für diese „tiefen“ Netzwerke hat sich der Begriff „Deep Learning“ etabliert. Seit 2015 steigt die Anzahl der Veröffentlichungen basierend auf Deep Learning im Bereich der Fernerkundung rasant (Hu et al., 2015; Zhu et al., 2017). Waren CNN ursprünglich auf die Auswertung von Pixeln oder regelmäßigen Segmenten in monotemporalen Daten beschränkt, ermöglichen neue Ansätze die Klassifikation von Polygonen, z.B. für Landnutzungsklassen (Yang et al., 2018), oder die Auswertung von Zeitreihen (Rußwurm und Körner, 2018; Interdonato et al., 2019). Allerdings haben Deep Learning Ansätze bislang Schwächen, die Gegenstand aktueller Forschungen sind. So werden u.a. große Datenmengen zum Training benötigt, Vorwissen nicht integriert (Marcus, 2018) und Kanten schlecht erhalten (Chen et al., 2017; Marmanis et al., 2018). Der letzte Effekt entsteht, da neue Schichten des Netzwerks durch Faltung vorheriger

Schichten entstehen. Somit wird implizit Kontextwissen integriert, allerdings gehen kleine Details verloren. Marmanis et al. (2018) wirken dem Verlust durch die explizite Nutzung von Kanten entgegen. Ein anderer Ansatz ist die Kombination von CNN mit CRF um die Stärken beider Verfahren zu vereinen. Dies kann entweder durch Integration des CRF als Formulierung eines RNN in die CNN-Klassifikation (Zheng et al., 2015) oder durch Postklassifikation der CNN-Klassifikation mit einem CRF erfolgen (Paisitkriangkrai et al., 2015; Chen et al., 2017; Pan und Zhao, 2018). Volpi und Tuia (2018) bilden in einem solchen Ansatz nachbarschaftliche Beziehungen unterschiedlicher Klassen explizit in den Interaktionspotentialen des CRF ab. Bislang sind keine Arbeiten zur multitemporalen Klassifikation von Daten unterschiedlicher Auflösung und der Fähigkeit, Veränderungen zu erkennen, mit CNN bekannt. Eine Verbindung des in dieser Arbeit vorgestellten Ansatzes mit CNN ist somit erstrebenswert.

Die Qualität der Ergebnisse in dieser Arbeit hängt von der Wahl der zur Modellierung der Interaktionspotentiale erforderlichen Parameter ($\beta$, $\gamma$ und $\eta$) ab. Diese wurden in dieser Arbeit empirisch bestimmt. In vielen Applikationen werden sämtliche Parameter angelernt. Ein solcher Lernprozess erfordert eine Zielfunktion, in der die Optimierungskriterien definiert sind. Dies ist in der in dieser Arbeit beschriebenen Anwendung problematisch. Die Einführung der maximalen Klassifikationsgenauigkeit als Kriterium ist nicht sinnvoll, da in vielen Fällen diese nicht dem visuell besten Ergebnis entspricht. Insbesondere bei der Wahl einfacher Interaktionsmodelle ist die glättende Wirkung derart stark, dass feine Strukturen verschwinden. Wären sämtliche feinen Objektstrukturen in einer separaten Klasse vereinigt, würde sich der Kappa-Index als Kriterium anbieten. Durch die gewählte Klassendefinition ist dies im vorliegenden Fall jedoch ebenfalls kein geeignetes Kriterium. Aus Mangel an einer geeigneten Zielfunktion wurde deshalb auf ein Anlernen der Parameter verzichtet. Die empirische Bestimmung der einzelnen Parameter wird als nicht sonderlich kritisch angesehen, da die Parameter relativ robust gegen kleine Veränderungen sind und in wenigen Iterationsschritten bestimmt werden konnten.

Ein genereller Nachteil von Zufallsfeldern ist die Berechnungszeit, die einen Echtzeiteinsatz für viele Anwendungen bislang erschwert. Zahlreiche Anwendungen aus der Computer Vision und Photogrammetrie klassifizieren aus diesem Grund nicht auf Pixelebene sondern auf Segmentebene. Eine Möglichkeit für die Definition der Segmente sind quadratische Kacheln. Bei einer Kantenlänge von zehn Pixeln reduziert sich die Knotenanzahl für einzelne Bilder somit um den Faktor 100. Das Resultat ist ein gröberes Klassifikationsergebnis, das in vielen Fällen aber zugunsten des Geschwindigkeitsvorteils akzeptiert wird. Mit zunehmender Rechenleistung verliert dieser Aspekt allerdings zunehmend an Relevanz.

Eine weitere Möglichkeit zur Verwendung von Segmenten ist die explizite Modellierung von räumlichen Nachbarschaften. Hier erfolgt ein Übergang von Pixelebene auf Objektebene. Benachbarte Pixel einer identischen Klasse werden segmentiert. Den Segmenten wird eine Objektklasse zugewiesen. Auf diese Weise lassen sich typische Relationen von Objekten modellieren, z.B. dass Häuser in vielen Fällen parallel zur zugehörigen Straße ausgerichtet sind. Die Qualität der Ergebnisse bei diesem Vorgehen ist unmittelbar mit der Qualität der Segmentierung, die stets eine Herausforderung darstellt, verknüpft.

Das Assoziationspotential wurde in dieser Arbeit aus einem Gauß-Modell abgeleitet. Für die Klassifikation von Einzelbildern liefern andere Klassifikatoren in vielen Fällen bessere Ergebnisse. Es besteht Forschungsbedarf, in welchem Maß die Klassifikationsgenauigkeit eines CRF durch Verwendung eines anderen Klassifikators für das Assoziationspotential gesteigert werden kann. Erste

Untersuchungen liefert Schindler (2012) für die Verbesserung der Klassifikationsgenauigkeit von optischen Bilddaten mit Zufallsfeldern durch unterschiedliche Algorithmen zur Glättung. In den Experimenten erzielen CRF basierend auf einer Random-Forest-Klassifikation für das Assoziationspotential bessere Ergebnisse als CRF mit einer Maximum-Likelihood-Klassifikation für das Assoziationspotential.

Eine Herausforderung für die Klassifikation mit Zufallsfeldern stellt die Modellierung räumlicher Interaktionen dar. In vielen Fällen werden hierzu Standardansätze (z.B. Potts-Modell) verwendet oder Parameter empirisch bestimmt. In dieser Arbeit wurde die Euklidische Distanz der Merkmalsvektoren für das räumliche Interaktionspotential genutzt. Es besteht Forschungsbedarf, ob weitere Steigerungen der Klassifikationsgenauigkeit durch komplexere Modelle erzielt werden können. Alternativ zeigen Volpi und Ferrari (2015) eine Strategie basierend auf CRF und Support Vector Machines zum Lernen räumlicher Kontextbeziehungen. Diese werden aus Referenzdaten abgeleitet und repräsentieren die Präferenzen für benachbarte Klassen.

Das temporale Interaktionspotential wurde ohne Berücksichtigung von Merkmalsinformationen mit Übergangsmatrizen modelliert. Diese wurden empirisch bestimmt. In Hoberg und Müller (2011) erfolgte die Klassifikation unterschiedlicher Feldfrüchte aus multitemporalen RapidEye-Daten unter Einbeziehung von Merkmalsinformationen für das zeitliche Interaktionspotential. Dies war auf einfache Weise möglich, da die Bilder über den Zeitraum einer Vegetationsperiode verteilt waren und somit auf sämtlichen Bildern die identische Feldfrucht an der gleichen Position existierte. Auf multitemporale Daten mit Veränderungen lässt sich dieses Vorgehen nicht übertragen. Hingegen wird großes Potential für eine Ableitung der Übergangsmatrix aus Trainingsdaten gesehen. Hierzu sind allerdings sehr viele repräsentative Trainingsdaten für sämtliche realistischen Klassenübergänge notwendig. Darüber hinaus wurde der Ansatz in dieser Arbeit ausschließlich zur Modellierung von 1:*N* Relationen genutzt. Grundsätzlich können auch *M*:*N* Relationen modelliert werden. Die Einbeziehung solcher komplexeren Modellierungen sollte Bestandteil künftiger Forschungen sein.

Die Interaktionen jedes Primitivs erfolgen in dem vorgestellten Modell ausschließlich paarweise mit den unmittelbar räumlich und zeitlich benachbarten Primitiven. Es konnten auf diese Weise beachtliche Genauigkeitssteigerungen erzielt werden. Dennoch bleiben Kontextbeziehungen über größere räumliche Entfernungen oder zeitliche Abstände unberücksichtigt. Diese Kontextbeziehungen könnten durch Potentiale höherer Ordnung integriert werden. Für eine derartige Modellierung ist jedoch Vorwissen notwendig, welche Primitive interagieren sollen. Gleichfalls werden Training und Inferenz anspruchsvoller. Aktuelle Arbeiten zur Nutzung von Potentialen höherer Ordnung zeigen Ansätze zur Lösung dieser Probleme und liefern vielversprechende Ergebnisse (Albert, 2017; Niemeyer, 2017). Die Übertragbarkeit dieser Ansätze für die Klassifikation multitemporaler Daten unterschiedlicher Auflösung ist ein weiteres künftiges Forschungsfeld.

# 7. Literatur

Albert, L., Rottensteiner, F. und Heipke, C. (2017). *A higher order conditional random field model for simultaneous classification of land cover and land use*. ISPRS Journal of Photogrammetry and Remote Sensing, 130, 63-80.

Albertz, J. (2009). *Einführung in die Fernerkundung*. 4. Auflage, Wissenschaftliche Buchgesellschaft.

Aksoy, S., Akçay, H. G. und Wassenaar, T. (2010). *Automatic mapping of linear woody vegetation features in agricultural landscapes using very high resolution imagery*. IEEE Transactions on Geoscience and Remote Sensing, 48:1, 511-522.

Arbeitsgemeinschaft der Vermessungsverwaltungen der Länder der Bundesrepublik Deutschland (AdV) (2017). *AFIS-ALKIS-ATKIS-Modell.* Unter http://www.adv-online.de/AAA-Modell/ (abgerufen zuletzt September 2017).

Benedek, C. und Szirányi, T. (2009). *Change detection in optical aerial images by a multi-layer Conditional Mixed Markov Model*, IEEE Transactions on Geoscience and Remote Sensing, 47:10, 3416-3430.

Benedek, C., Shadaydeh, M., Kato, Z., Szirányi, T. und Zerubia, J. (2015). *Multilayer Markov random field models for change detection in optical remote sensing images*. ISPRS Journal of Photogrammetry and Remote Sensing, 107, 22-37.

Besag, J. (1974). *Spatial interaction and the statistical analysis of Lattice Systems*. Journal of the Royal Statistical Society. Series B (Methodological), 36:2, 192-236.

Besag, J. (1986). *On the Statistical Analysis of Dirty Pictures.* Journal of the Royal Statistical Society. Series B (Methodological), 48:3, 259-302.

Bishop, C. M. (2006). *Pattern recognition and machine learning*. 1st edition, Springer New York, USA.

Bishop, C. M. (2013). *Graphical Models – Part I-III*. Machine Learning Summer School 2013, Max Planck Institute for Intelligent Systems, Tübingen. Unter https://www.youtube.com/watch?v=ju1Grt2hdko (zuletzt abgerufen im Juni 2017).

Bouman, C. A. und Shapiro, M. (1994). *A multiscale random field model for Bayesian image segmentation*. IEEE Transactions on Image Processing, 3:2, 162-177.

Boykov, Y., Veksler, O. und Zabih, R. (2001). *Fast approximate energy minimization via graph cuts*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 23:11, 1222-1239.

Bruzzone, L. und Bovolo, F. (2013). *A novel framework for the design of change-detection systems for very-high-resolution remote sensing images*. Proceedings of the IEEE, 101:3, 609-630.

Bruzzone, L., Cossu, R. und Vernazza, G. (2004). *Detection of land-cover transitions by combining multidate classifiers*. Pattern Recognition Letters, 25:13, 1491-1500.

Camps-Valls, G., Tuia, D., Bruzzone, L. und Benediktsson, J. A. (2014). *Advances in hyperspectral image classification: Earth monitoring with statistical learning methods*. IEEE Signal Processing Magazine, 31:1, 45-54.

Caruana, R. und Freitag, D. (1994). *Greedy attribute selection*. Proceedings of the Eleventh International Conference on Machine Learning, 28—36.

Chandrashekar, G. und Sahin, F. (2014). *A survey on feature selection methods.* Computers & Electrical Engineering, 40:1, 16-28.

Chen, L. C., Papandreou, G., Kokkinos, I., Murphy, K. und Yuille, A. L. (2017). *Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 40:4, 834-848.

Choi, M. J., Chandrasekaran, V., Malioutov, D. M., Johnson, J. K. und Willsky, A. S. (2008). *Multiscale stochastic modeling for tractable inference and data assimilation.* Computer Methods in Applied Mechanics and Engineering, 197:43, 3492-3515.

Dalal, N. und Triggs, B. (2005). *Histograms of Oriented Gradients for Human Detection*. Proceedings of the IEEE Conference Computer Vision and Pattern Recognition, 886-893.

Feitosa, R. Q., Costa, G. A. O. P., Mota, G. L. A., Pakzad, K. und Costa, M. C. O. (2009). *Cascade multitemporal classification based on fuzzy Markov chains*. ISPRS Journal of Photogrammetry and Remote Sensing, 64:2, 159-170.

Ferreira, A. J., & Figueiredo, M. A. (2012). *Efficient feature selection filters for high-dimensional data*. Pattern Recognition Letters, 33:13, 1794-1804.

Foody, G.M. (2002). *Status of land cover classification accuracy assessment*, Remote Sensing of Environment, 80:1, 185-201.

Foody, G.M. (2010). *Assessing the accuracy of land cover change with imperfect ground reference data*, Remote Sensing of Environment, 114:10, 2271-2285.

Frey, B. J. und MacKay, D. J. (1998). *A revolution: Belief propagation in graphs with cycles*. Advances in neural information processing systems, 479-485.

Geman, G. und Geman, D. (1984). *Stochastic relaxation, Gibbs distribution and Bayesian restoration of images*. IEEE-Transactions on Geoscience and Remote Sensing, 6:6, 721-741.

Gómez-Chova, L., Tuia, D., Moser, G. und Camps-Valls, G. (2015). *Multimodal classification of remote sensing images: A review and future directions*. Proceedings of the IEEE, 103:9, 1560-1584.

Guyon, I. und Elisseeff, A. (2003). *An introduction to variable and feature selection*. Journal of Machine Learning Research, 3, 1157-1182.

Hall, M. A. (1999). *Correlation-based Feature Subset Selection for Machine Learning,* PhD dissertation, Department of Computer Science, University of Waikato, New Zealand.

Hall, M. A. (2000). *Correlation-based Feature Selection for Discrete and Numeric Class Machine Learning*. Seventeenth International Conference on Machine Learning, 359-366.

Hammersley, J.M. und Clifford, P. (1971). *Markov Field on Finite Graphs and Lattices*, unveröffentlicht, 26 S..

Haralick, R.M., Shanmugam, K. und Dinstein, I. (1973). *Texture features for image classification*. IEEE Transactions on Systems, Man, and Cybernetics, 3: 6, 610–622.

He, X., Zemel, R. S. und Carreira-Perpiñán, M. Á. (2004). *Multiscale conditional random fields for image labeling*. Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2.

Hedhli, I., Moser, G., Zerubia, J. und Serpico, S. B. (2016). *A new cascade model for the hierarchical joint classification of multitemporal and multiresolution remote sensing data*. IEEE Transactions on Geoscience and Remote Sensing, 54:11, 6333-6348.

Heipke, C., Mayer, H., Wiedemann, C. und Jamet, O. (1997). *Evaluation of automatic road extraction*, International Archives of Photogrammetry and Remote Sensing, 32-3:4W2, 151–160.

Hoberg, T. und Müller, S. (2011). *Multitemporal crop type classification using Conditional Random Fields and RapidEye data*. International Archives of Photogrammetry and Remote Sensing, 38-4:W19, 115-121.

Hoberg, T. und Rottensteiner, F. (2010). *Classification of settlement areas in remote sensing imagery using Conditional Random Fields*. International Archives of Photogrammetry and Remote Sensing, 38:7A, 53-58.

Hoberg, T., Rottensteiner, F. und Heipke, C. (2010). *Classification of multitemporal remote sensing data using Conditional Random Fields*: Proceedings 6. IAPR TC 7 Workshop on Pattern Recognition in Remote Sensing, 4 S.

Hoberg, T., Rottensteiner, F. und Heipke, C. (2011). *Classification of multitemporal remote sensing data of different resolution using conditional random fields*. 1st IEEE/ISPRS Workshop on Computer Vision for Remote Sensing of the Environment, IEEE ICCV Workshops, Barcelona, 235-242.

Hoberg, T., Rottensteiner, F., Feitosa, R. Q. und Heipke, C. (2015). *Conditional random fields for multitemporal and multiscale classification of optical satellite imagery*. IEEE Transactions on Geoscience and Remote Sensing, 53:2, 659-673.

Hu, F., Xia, G. S., Hu, J. und Zhang, L. (2015). *Transferring deep convolutional neural networks for the scene classification of high-resolution remote sensing imagery*. Remote Sensing, 7:11, 14680-14707.

Hubert-Moy, L., Cotonnec, A., Le Du, L., Chardin, A. und Perez, P. (2001). *A comparison of classification procedures of remotely sensed data applied on different landscape units*. Remote Sensing of Environment, 75:2, 174-187.

Huertas, C. und Juárez-Ramírez, R. (2014). *Filter feature selection performance comparison in high-dimensional data: A theoretical and empirical analysis of most popular algorithms.* 17th International Conference on Information Fusion (FUSION), 1-8.

Hughes, G. (1968). *On the mean accuracy of statistical pattern recognizers*. IEEE Transactions on Information Theory, 14:1, 55-63.

Interdonato, R., Ienco, D., Gaetano, R. und Ose, K. (2019). *DuPLO: A DUal view Point deep Learning architecture for time series classificatiOn*. ISPRS Journal of Photogrammetry and Remote Sensing, 149, 91-104.

Ising, E. (1925). *Beitrag zur Theorie des Ferromagnetismus.* Zeitschrift für Physik, 31, 253-258.

Jain, A.K., Duin, R.P.W. und Mao, J. (2000). *Statistical pattern recognition: a review*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 22:1, 4-37.

Jianya, G., Haigang, S., Guorui, M. und Qiming, Z. (2008). *A review of multi-temporal remote sensing data change detection algorithms*, International Archives of Photogrammetry and Remote Sensing, 37:B7, 757-762.

Kato, Z., Berthod, M. und Zerubia, J. (1993). *Multiscale Markov random field models for parallel image classification*, Proceedings of the 4. International Conference on Computer Vision, 253-257.

Kauth, R. J. und Thomas, G. J. (1976). *The Tasseled Cap - a graphic description of the spectral-temporal development of agricultural crops as seen by Landsat*. Proceedings of the Symposium on Machine Processing of Remotely Sensed Data, 41-51.

Kenduiywo, B. K., Bargiel, D. und Soergel, U. (2015). *Spatial-temporal conditional random fields crop classification from TerraSAR-X images.* ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2:3, 79-86.

Kersten, J., Gähler, M. und Voigt, S. (2010). *A general framework for fast and interactive classification of optical VHR satellite imagery using hierarchical and planar Markov random fields*. Photogrammetrie-Fernerkundung-Geoinformation, 2010:6, 439-449.

Kohli, P. und Torr, P. H. (2009). *Robust higher order potentials for enforcing label consistency*. International Journal of Computer Vision, 82:3, 302-324.

Koller, D. (2012). *Probabilistic Graphical Models*. OpenCourseOnline – 94 Videos, u.a. https://www.youtube.com/watch?v=WPSQfOkb1M8&index=1&list=PL50E6E80E8525B59C (Video 1 – Introduction) (zuletzt abgerufen im Juni 2017).

Kosov, S., Rottensteiner, F. und Heipke, C. (2013). *Sequential gaussian mixture models for two-level conditional random fields*. German Conference on Pattern Recognition, 153-163.

Krizhevsky, A., Sutskever, I. und Hinton, G. E. (2012). *Imagenet classification with deep convolutional neural networks*. Advances in Neural Information Processing Systems, 1097-1105.

Kschischang, F. R., Frey, B. J. und Loeliger, H. A. (2001). *Factor graphs and the sum-product algorithm*. IEEE Transactions on information theory, 47:2, 498-519.

Kumar, S. und Hebert, M. (2003). *Discriminative random fields: A discriminative framework for contextual interaction in classification*. Proceedings of the 9. IEEE International Conference on Computer Vision, 1150-1157.

Kumar, S. und Hebert, M. (2006). *Discriminative Random Fields*. International Journal of Computer Vision, 68:2, 179-201.

Ladický, L., Russell, C., Kohli, P. und Torr, P. (2009) *Associative hierarchical crfs for object class image segmentation*. Proceedings of the 12. IEEE International Conference on Computer Vision, 739–746.

Lafferty, J., McCallum, A. und Pereira, F. (2001). *Conditional random fields: Probabilistic models for segmenting and labeling sequence data*. Proceedings of the 18. International Conference on Machine Learning, 1, 282-289.

Larochelle, H. (2013). *Neural networks class*. Université de Sherbrooke – 92 Videos, u.a. https://www.youtube.com/watch?v=GF3iSJkgPbA&list=PL6Xpj9I5qXYEcOhn7TqghAJ6NAPrNmUBH&index=18 (Neural networks [3.1]: Conditional random fields – motivation) (zuletzt abgerufen im Juni 2017).

Leite, P. B. C., Feitosa, R. Q., Formaggio, A. R., da Costa, G. A. O. P., Pakzad, K. und Sanches, I. D. A. (2011). *Hidden Markov Models for crop recognition in remote sensing image sequences*. Pattern Recognition Letters, 32:1, 19-26.

Li, D. (2010). *Remotely sensed images and GIS data fusion for automatic change detection*. International Journal of Image and Data Fusion, 1:1, 99-108.

Li, M., Zang, S., Zhang, B., Li, S. und Wu, C. (2014). *A review of remote sensing image classification techniques: The role of spatio-contextual information*. European Journal of Remote Sensing, 47:1, 389-411.

Li, S. Z. (2009). *Markov random field modeling in image analysis*. Springer Science & Business Media, third edition.

Li, Z., Hayward, R., Zhang, J., Jin, H. und Walker R. (2010). *Evaluation of spectral and texture features for objet-based vegetation species classification using Support Vector Machines.* International Archives of Photogrammetry and Remote Sensing, 38:7A, 122-127.

Liu, H. und Motoda, H. (2008). *Computational methods of feature selection*. Chapman & Hall/CRC.

Lu, D. und Weng, Q. (2007). *A survey of image classification methods and techniques for improving classification performance*, International Journal of Remote Sensing, 28:5, 823-870.

Lu, D., Mausel, P., Brondizio, E. und Moran, E. (2004). *Change detection techniques*. International Journal of Remote Sensing, 25: 12, 2365-2401.

Lu, D., Moran, E. und Hetrick, S. (2011). *Detection of impervious surface change with multitemporal Landsat images in an urban–rural frontier*. ISPRS Journal of Photogrammetry and Remote Sensing, 66:3, 298-306.

Lu, W. L., Murphy, K. P., Little, J. J., Sheffer, A. und Fu, H. (2009). *A hybrid conditional random field for estimating the underlying ground surface from airborne lidar data*. IEEE Transactions on Geoscience and Remote Sensing, 47:8, 2913-2922.

Mallat, S. G. (1989). *A theory for multiresolution signal decomposition: the wavelet representation*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 11:7, 674-693.

Mallet, C. (2010). *Analyse de données lidar à retour d'onde complète pour la classification en milieu urbain.* Doctoral dissertation, PhD thesis, Télécom ParisTech, Paris, France.

Marcus, G. (2018). *Deep learning: A critical appraisal*. arXiv preprint arXiv:1801.00631, 27 S..

Marmanis, D., Schindler, K., Wegner, J. D., Galliani, S., Datcu, M. und Stilla, U. (2018). *Classification with an edge: Improving semantic image segmentation with boundary detection*. ISPRS Journal of Photogrammetry and Remote Sensing, 135, 158-172.

Mas, J.-F. (1999). Monitoring *land-cover changes: a comparison of change detection techniques*, International Journal of Remote Sensing, 20: 1, 139-152.

Mather, P. und Tso, B. (2016). *Classification methods for remotely sensed data*. CRC press.

Melgani, F. und Serpico, S. B. (2003). *A Markov Random Field approach to spatio-temporal contextual image classification*. IEEE Transactions on Geoscience and Remote Sensing, 41:11, 2478-2487.

Montoya-Zegarra, J.A., Wegner, J.D., Ladický, L. und Schindler, K. (2015). *Semantic segmentation of aerial images in urban areas with class-specific higher-order cliques*. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2-3:W4, 127–133.

Moser, G., Angiati, E., und Serpico, S. B. (2009). *A contextual multiscale unsupervised method for change detection with multitemporal remote-sensing images*. Proceedings of the 9. Conference on Intelligent Systems Design & Applications, 572-577.

Mota, G. L. A., Feitosa, R. Q., Coutinho, H. L. C., Liedtke, C.-E., Müller, S., Pakzad, K. und Meirelles, M. S. P. (2007). *Multitemporal fuzzy classification model based on class transition possibilities*. ISPRS Journal of Photogrammetry and Remote Sensing, 62:3, 186-200.

Mountrakis, G., Im, J. und Ogole, C. (2011). *Support vector machines in remote sensing: A review*. ISPRS Journal of Photogrammetry and Remote Sensing, 66:3, 247-259.

Niemeyer, J. (2017). *Verwendung von Kontext zur Klassifikation luftgestützter Laserdaten urbaner Gebiete*. Dissertation: Wissenschaftliche Arbeiten der Fachrichtung Geodäsie und Geoinformatik der Leibniz Universität Hannover, ISSN 0174-1454, Nr. 336, Hannover.

Niemeyer, J., Rottensteiner, F. und Soergel, U. (2014). *Contextual classification of lidar data and building object detection in urban areas*. ISPRS Journal of Photogrammetry and Remote Sensing, 87, 152-165.

Nocedal, J. und Wright, S. J. (2006). *Numerical optimization.* 2nd edition, Springer New York, USA.

Ojala, T., Pietikäinen, M. und Harwood, D. (1996). *A comparative study of texture measures with classification based on featured distributions*. Pattern Recognition, 29: 1, 51-59.

Ojala, T., Pietikainen, M. und Maenpaa, T. (2002). *Multiresolution gray-scale and rotation invariant texture classification with local binary patterns*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 24: 7, 971-987.

Paisitkriangkrai, S., Sherrah, J., Janney, P. und Hengel, V. D. (2015). *Effective semantic pixel labelling with convolutional networks and conditional random fields*. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 36-43.

Pan, X. und Zhao, J. (2018). *High-resolution remote sensing image classification method based on convolutional neural network and restricted conditional random field*. Remote Sensing, 10:6, 920.

Pakzad, K. (2001). *Wissensbasierte Interpretation von Vegetationsflächen aus multitemporalen Fernerkundungsdaten.* Dissertation: Wissenschaftliche Arbeiten der Fachrichtung Vermessungswesen der Universität Hannover, Nr. 243, Hannover, 104 S.

Pal, M. und Foody, G. M. (2010). *Feature selection for classification of hyperspectral data by SVM*. IEEE Transactions on Geoscience and Remote Sensing, 48:5, 2297-2307.

Papadopoulos, G.Th., Saathoff, C., Escalante, H.J., Mezaris, V., Kompatsiaris, I., und Strintzis, M.G. (2011). *A comparative study of object-level spatial context techniques for semantic image analysis*. Computer Vision and Image Understanding, 115:9, 1288-1307.

Pearl, J. (1982). *Reverend Bayes on inference engines: A distributed hierarchical approach*. Cognitive Systems Laboratory, School of Engineering and Applied Science, University of California, Los Angeles, 133-136.

Peng, H., Long, F., & Ding, C. (2005). *Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy.* IEEE Transactions on Pattern Analysis and Machine Intelligence, 27:8, 1226-1238.

Potts, R. B. (1952). *Some generalized order-disorder transformations*. Mathematical Proceedings, 48:1, 106–109.

Pudil, P., Novovičová, J. und Kittler J. (1994). *Floating search methods in feature selection*, Pattern Recognition Letters, 15:11, 1119-1125.

Radke, R.J., Andra, S., Al-Kofahi, O. und Roysam, B. (2005). *Image change detection algorithms: A systematic survey*, IEEE Transactions on Image Processing, 14:3, 294-307.

Refaeilzadeh, P., Tang, L., und Liu, H. (2007). *On comparison of feature selection algorithms.* In: Proceedings of AAAI Workshop on Evaluation Methods for Machine Learning, 5 S..

Richards, J. A. (2013). *Remote sensing digital image analysis: An introduction*. 5th edn, Springer-Verlag New York, Inc., Secaucus, NJ, USA.

Richards, J.A. (2005). *Analysis of remotely sensed data: The formative decades and the future*, IEEE Transactions on Geoscience and Remote Sensing, 43:3, 422-432.

Roscher, R., Waske, B. und Förstner, W. (2010). *Kernel discriminative random fields for land cover classification*. 6. IAPR TC 7 Workshop Pattern Recognition in Remote Sensing, 5 S..

Rottensteiner, F. (2016). *Kontextbasierte Ansätze in der Bildanalyse.* In: Heipke (Hrsg): Photogrammetrie und Fernerkundung, Springer Spektrum, 1. Auflage, 555-602.

Rouse, J.W., Haas, R.H., Schell, J.A. und Deering, D.W. (1973). *Monitoring vegetation systems in the Great Plains with ERTS*. Third ERTS S ymposium, 1, 309-317.

Ruiz, L.A., Recio, J.A., Fernández-Sarría, A. und Hermosilla, T. (2011). *A feature extraction software tool for agricultural object-based image analysis*. Computers and Electronics in Agriculture, 76:2, 284-296.

Rußwurm, M. und Körner, M. (2018). *Multi-temporal land cover classification with sequential recurrent encoders*. ISPRS International Journal of Geo-Information, 7:4, 129.

Saeys, Y., Inza, I. und Larrañaga, P. (2007). *A review of feature selection techniques in bioinformatics*, Bioinformatics, 23:19, 2507-2517.

Schindler, K. (2012). *An overview and comparison of smooth labeling methods for land-cover classification*. IEEE Transactions on Geoscience and Remote Sensing, 50:11, 4534-4545.

Schnitzspan, P., Fritz, M. und Schiele, B. (2008). *Hierarchical support vector random fields: joint training to combine local and global features*. Proceedings of the European Conference on Computer Vision, 527-540.

Shotton, J., Winn, J., Rother, C. und Criminisi, A. (2009). *TextonBoost for image understanding: Multi-class object recognition and segmentation by jointly modeling texture, layout, and context*. International Journal of Computer Vision, 81:1, 2-23.

Singh, A. (1989). *Digital change detection techniques using remotely-sensed data.* International Journal of Remote Sensing, 10:6, 989-1003.

Solberg, A. H. S., Taxt, T. und Jain, A. K. (1996). *A Markov random field model for classification of multisource satellite imagery*. IEEE Transactions on Geoscience and Remote Sensing, 34:1, 100-113.

Szeliski, R., Zabih, R., Scharstein, D., Veksler, O., Kolmogorov, V., Agarwala, A., Tappen, M. und Rother, C. (2008). *A comparative study of energy minimization methods for markov random fields with smoothness-based priors*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 30:6, 1068-1080.

Sziranyi, T. und Shadaydeh, M. (2014). *Segmentation of remote sensing images using similarity-measure-based fusion-MRF model*. IEEE Geoscience and Remote Sensing Letters, 11:9, 1544-1548.

Tappen, M. F. und Freeman, W. T. (2003). *Comparison of Graph Cuts with Belief Propagation for stereo, using Identical MRF Parameters*. Proceedings of the International Conference on Computer Vision, 2, 900-906.

Tavakkoli Sabour, S.M. (2010). *Multi-temporal classification of crops using ENVISAT ASAR data,* Dissertation: Wissenschaftliche Arbeiten der Fachrichtung Geodäsie und Geoinformatik der Leibniz Universität Hannover, Nr. 289, Hannover, 117 S..

Tönnies, K.D. (2005). *Grundlagen der Bildverarbeitung*. Pearson Studium.

Torralba, A., Murphy K. P. und Freeman, W. T. (2010). *Using the forest to see the trees: Object Recognition in context for visual object detection and localization*. Communications of the ACM, Research Highlights, 53:3, 107-114.

Vieira, C., Mather, P. und McCullagh, M. (2000). *The spectral-temporal response surface and its use in the multi-sensor, multitemporal classification of agricultural crops.* International Archives of Photogrammetry and Remote Sensing, 33:B2, 582-589.

Vishwanathan, S., Schraudolph, N. N., Schmidt, M. W. und Murphy, K. P. (2006). *Accelerated training of conditional random fields with stochastic gradient methods*. 23. International Conference on Machine Learning, 969-976.

Volpi, M. und Ferrari, V. (2015). *Semantic segmentation of urban scenes by learning local class interactions*. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 1-9.

Volpi, M. und Tuia, D. (2018). *Deep multi-task learning for a geographically-regularized semantic segmentation of aerial images*. ISPRS Journal of Photogrammetry and Remote Sensing, 144, 48-60.

Wang, L., Shi, C., Diao, C., Ji, W. und Yin, D. (2016). *A survey of methods incorporating spatial information in image classification and spectral unmixing*. International Journal of Remote Sensing, 37:16, 3870-3910.

Waske, B. and Braun M. (2009). *Classifier ensembles for land cover mapping using multitemporal SAR imagery*. ISPRS Journal of Photogrammetry and Remote Sensing, 64:5, 450–457.

Wegner, J.D. (2011). *Detection and height estimation of buildings from SAR and optical images using conditional random fields*. Dissertation: Wissenschaftliche Arbeiten der Fachrichtung Geodäsie und Geoinformatik der Leibniz Universität Hannover, Nr. 295, Hannover, 139 S..

WEKA (2012). http://sourceforge.net/projects/weka/.

Wilsky, A. S. (2002). *Multiresolution Markov models for signal and image processing*. Proceedings IEEE, 90: 8, 1396-1458.

Witten, I. H. und Frank, E. (2005). *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann, 2nd Edition.

Yang, C., Rottensteiner, F., und Heipke, C. (2018). *Classification of land cover and land use based on convolutional neural networks*. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 4:3, 251-258.

Yang, M. Y. und Förstner, W. (2011). *A hierarchical Conditional Random Field model for labeling and classifying images of man-made scenes*. 1. IEEE/ISPRS Workshop on Computer Vision for Remote Sensing of the Environment, 196-203.

Yang, M. Y., Förstner, W. und Drauschke, M. (2010). *Hierarchical conditional random field for multi-class image classification.* Proceedings of the International Conference on Computer Vision Theory and Applications (VISAPP), 464-469.

Yang, M.Y. (2011). *Hierarchical and spatial structures for interpreting images of man-made scenes using graphical models*. Dissertation: Landwirtschaftliche Fakultät, Bonn, 98 S..

Yedidia, J. S., Freeman, W. T. und Weiss, Y. (2005). *Constructing free-energy approximations and generalized belief propagation algorithms*. IEEE Transactions on Information Theory, 51:7, 2282-2312.

Zheng, S., Jayasumana, S., Romera-Paredes, B., Vineet, V., Su, Z., Du, D., Huang, C. und Torr, P. H. (2015). *Conditional random fields as recurrent neural networks*. Proceedings of the IEEE International Conference on Computer Vision, 1529-1537.

Zhong, P. und Wang, R. (2007). *A multiple conditional random fields ensemble model for urban area detection in remote sensing optical images*. IEEE Transactions on Geoscience and Remote Sensing, 45:12, 3978-3988.

Zhu, X. X., Tuia, D., Mou, L., Xia, G. S., Zhang, L., Xu, F. und Fraundorfer, F. (2017). *Deep learning in remote sensing: A comprehensive review and list of resources*. IEEE Geoscience and Remote Sensing Magazine, 5:4, 8-36.

# Danksagung

Das Verfassen einer Dissertation erfordert Inspiration, Austausch und Zeit. Zum Gelingen dieser Arbeit haben zahlreiche Menschen beigetragen, denen ich an dieser Stelle danken möchte.

An erster Stelle bedanke ich mich bei meinem Doktorvater Prof. Dr.-Ing. Christian Heipke. Er hat mir Freiräume gegeben, eigene Forschungsinteressen zu verfolgen und mich stets fördernd begleitet. Seine außerordentliche Fähigkeit, Möglichkeiten wie auch Schwachstellen zu erkennen und in Gesprächen auf das Wesentliche zu fokussieren, hat mich stets beeindruckt. Bei Prof. Dr.-Ing. Monika Sester und Prof. Dr.-Ing. Stefan Hinz bedanke ich mich für die Übernahme der Korreferate für diese Dissertation und bei Prof. Dr.-Ing. Jakob Flury für die Übernahme des Vorsitzes der Prüfungskommission. Besonderer Dank gilt Prof. Dr. techn. Franz Rottensteiner für die jahrelange intensive Zusammenarbeit am IPI und die zahlreichen anregenden Diskussionen, die wir auch nach meiner Zeit am IPI oft in dienstfernen Räumlichkeiten fortgesetzt haben. Seine stete Motivation hat wesentlich zum Erfolg dieser Dissertation beigetragen.

Allen Kolleginnen und Kollegen des IPI danke ich für sieben schöne und abwechslungsreiche Jahre. Die gute Zusammenarbeit, die angenehme Atmosphäre, die geselligen Kaffeerunden und Stammtische sowie die gemeinsamen Dienstreisen werde ich ewig in Erinnerung behalten. Insbesondere danke ich Kian Pakzad für die Verfassung des Forschungsantrages, der Grundlage für diese Dissertation war, sowie Manfred Wiggenhagen, durch den ich bereits während meines Studiums einen guten Kontakt zum IPI aufbauen konnte.

Zu guter Letzt danke ich meiner Familie für die großartige Unterstützung in allen Lebenslagen. Mit meinen Eltern Edith und Reinhold sowie meinem Bruder Hendrik durfte ich eine glückliche Kindheit verleben. Später förderten sie stets meinen (Berufs-)weg und stehen mir weiterhin mit Rat und Tat zu Seite. Meiner Frau Tina und meinen Töchtern Paula und Hanna danke ich für die unzähligen schönen Momente, in denen Sie mich glücklich machen und mir verdeutlichen, was die wirklich wichtigen Dinge im Leben sind.

# Lebenslauf

## Zur Person

| | |
|---|---|
| Name | Thorsten Hoberg (geb. Hödl) |
| geboren | 23.03.1975 in Kassel |

## Berufliche Praxis

| | |
|---|---|
| seit 01.2016 | LGLN, Regionaldirektion Hameln-Hannover, Katasteramt Hildesheim, Dezernatsleiter |
| 01.2014 – 12.2015 | Niedersächsisches Ministerium für Inneres und Sport, Referat für Vermessung, Geoinformation und Kampfmittelbeseitigung, Referent |
| 04.2013 – 12.2013 | LGLN, Landesvermessung und Geobasisinformation |
| 05.2012 – 03.2013 | LGLN, Regionaldirektion Hannover, Katasteramt Hannover |
| 03.2005 – 04.2012 | Leibniz Universität Hannover, Institut für Photogrammetrie und GeoInformation (IPI), Wissenschaftlicher Mitarbeiter |
| 08.2004 – 01.2005 | GLL Osnabrück, Katasteramt Osnabrück |

## Ausbildung

| | |
|---|---|
| 05.2002 – 06.2004 | Land Niedersachsen, Bezirksregierung Weser-Ems, Referendariat in der Fachrichtung Vermessungswesen |
| 01.2002 – 04.2002 | Universität Hannover, Institut für Photogrammetrie und GeoInformation (IPI), Freie Mitarbeit |
| 10.2001 – 12.2001 | Auslandsaufenthalt in Australien |
| 10.2000 – 08.2001 | Gemeinnützige Gesellschaft für paritätische Sozialarbeit Hannover, Zivildienst |
| 09.1998 – 01.1999 | University of Newcastle upon Tyne, Studium Geomatics |
| 09.1994 – 10.2000 | Universität Hannover, Studium Geodäsie |
| 07.1991 – 06.1994 | Gymnasium Herderschule in Kassel |
| 07.1981 – 06.1991 | Grundschule und Wilhelm-Leuchner-Gesamtschule in Niestetal-Heiligenrode |