Veröffentlichungen der DGK

Ausschuss Geodäsie der Bayerischen Akademie der Wissenschaften

Reihe C | Dissertationen | Heft Nr. 848

Alina Elisabeth Maas

Klassifikation multitemporaler Fernerkundungsdaten unter Verwendung fehlerbehafteter topographischer Daten

München 2020

Verlag der Bayerischen Akademie der Wissenschaften

ISSN 0065-5325 | ISBN 978-3-7696-5260-4

Diese Arbeit ist gleichzeitig veröffentlicht in:
Wissenschaftliche Arbeiten der Fachrichtung Geodäsie und Geoinformatik der Universität Hannover
ISSN 0174-1454, Nr. 357, Hannover 2020

Veröffentlichungen der DGK

Ausschuss Geodäsie der Bayerischen Akademie der Wissenschaften

Reihe C | Dissertationen | Heft Nr. 848

Klassifikation multitemporaler Fernerkundungsdaten unter Verwendung fehlerbehafteter topographischer Daten

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

M. Sc. Alina Elisabeth Maas

Geboren am 14.08.1990 in Hannover, Deutschland

München 2020

Verlag der Bayerischen Akademie der Wissenschaften

ISSN 0065-5325 ISBN 978-3-7696-5260-4

Diese Arbeit ist gleichzeitig veröffentlicht in:
Wissenschaftliche Arbeiten der Fachrichtung Geodäsie und Geoinformatik der Universität Hannover
ISSN 0174-1454, Nr. 357, Hannover 2020

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. Christian Heipke

Referent: Apl. Prof. Dr. techn. Franz Rottensteiner

Korreferenten: Prof. Dr.-Ing. Stefan Hinz (Karlsruher Institut für Technologie)
Prof. Dr.-Ing. Monika Sester

Tag der mündlichen Prüfung: 12.03.2020

© 2020 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 978-3-7696-5260-4

Zusammenfassung

Aktuelle topographische Daten (im Folgenden *Karten* genannt) bilden die Grundlage für verschiedene Anwendungen, wie beispielsweise die Navigation, die Gefahrenanalyse oder auch die Stadt- und Verkehrsplanung. Eine Möglichkeit zur Aktualisierung veralteter Karten besteht in der Klassifikation aktueller Fernerkundungsbilder mit überwachten Methoden. Um der zeitaufwändigen und somit kostenintensiven manuellen Generierung der Trainingslabels zu entgehen, werden die Labels der Trainingsbeispiele aus der vorhandenen Karte extrahiert. Da die Karte in der Regel nicht mehr aktuell ist, sind einige Trainingslabels fehlerhaft. Der Klassifikator muss daher in der Lage sein mit diesen Fehlern im Training umzugehen, damit die Fehler keine negativen Einflüsse auf das Klassifikationsergebnis haben. In der Literatur werden die Fehler in den Labels der Trainingdaten oft als *Label Noise* bezeichnet.

In dieser Arbeit wird ein multitemporaler Klassifikator zur Aktualisierung von veralteten Karten vorgestellt, der mit den fehlerhaften Labels der Karte umgehen kann. Hierfür wird der Random Forest Klassifikator [Breiman, 2001] angepasst, sodass dieser mit einer gewissen Anzahl an Fehlern in den Trainingslabeln trainiert werden kann. Umgesetzt wird dies durch die Integration eines komplexen Fehlermodells in die Berechnung des Informationszuwachses [Duda et al., 2001]. Dieser wird verwendet um innerhalb jedes Knotens von jedem Baum die Entscheidungsgrenze auszuwählen. Der resultierende robuste Random Forest wird in einen auf dem Conditional Random Field (CRF) [Kumar und Hebert, 2006] beruhenden iterativen Prozess integriert. Dieser Prozess berücksichtigt den räumlichen Kontext, die beobachteten Labels der veralteten Karte und die Tatsache, dass Trainingsbeispiele mit fehlerhaftem Label Segmente im Bildraum bilden. Für die gemeinsame Klassifikation von Daten mehrerer Zeitpunkte wird ein multitemporales CRF vorgestellt, welches zusätzlich zu den Labels der veralteten Karte und dem räumlichen Kontext auch temporalen Kontext berücksichtigt.

Das Verfahren wurde an fünf verschiedenen Datensätzen getestet. Zwei Datensätze bestehen aus Fernerkundungsdaten mehrerer Zeitpunkte und realem Kartenmaterial. Die anderen drei Datensätze sind monotemporal und enthalten simulierte veraltete Karten. Für die Evaluierung werden manuell erstellte Referenzen verwendet. Die Ergebnisse der neu entwickelten Methode zeigen in den meisten Experimenten eine höhere Gesamtgenauigkeit als die Ergebnisse des klassischen Random Forest Klassifikators. Das Potenzial der multitemporalen Klassifikation wurde besonders bei den Experimenten mit den hochaufgelösten Datensatz deutlich.

Schlagworte: Kartenaktualisierung, Klassifikation, Multitemporal, Fehler in den Trainingslabels

Abstract

Topographic databases (referred to as *maps*) are required for various applications such as navigation, hazard analysis or urban and traffic planning. To update outdated maps, current remote sensing images can be classified using supervised classification methods. To avoid the time-consuming and thus cost-intensive manual generation of training labels, these labels are derived from the existing map. As usually the map is not up-to-date, some training labels might be wrong. Therefore, the classifier must be able to deal with these errors during training so that the errors do not have any negative effect on the classification result. In the literature the errors in the labels of the training data are called *Label Noise*.

In this thesis a multitemporal classifier for updating outdated maps is developed, which is able to deal with errors in the traing labels. For that purpose the Random Forest classifier [Breiman, 2001] was adapted, so that it is able to cope with such errors in the training labels. This was achieved by adding a complex noise model to determine the information gain [Duda et al., 2001]. The information gain is used to select the decision boundary within each node of each tree. The resulting robust Random Forest is integrated into an iterative workflow based on Conditional Random Fields (CRFs) [Kumar und Hebert, 2006], where spatial context, the observed labels, and the fact that changes occur in segments in image space are considered. A multitemporal CRF is used for a simultanious classification of remote sensing data from different epochs. In addition to the labels of the outdated map and the spatial context, the CRF also considers temporal context.

The method was tested on five different datasets. Two datasets consist of remote sensing data of different epochs and real outdated maps. The other three datasets are monotemporal and contain simulated outdated maps. For evaluation, manually labelled reference data are used. In most experiments the results of the new method show a higher overall accuracy than the results of the standard Random Forest. The full potential of a multitemporal classification using temporal context becomes clear through the experiments on the high-resolution datasets.

Keywords: Map Update, Classification, Multitemporal, Label Noise

Nomenklatur

Allgemeine Abkürzungen

ALKIS	Amtliches Liegenschaftskatasterinformationssystem
ATKIS	Amtliches Topographisch-Kartographisches Informationssystem
DGM	Digitales Geländemodell
DOM	Digitales Oberflächenmodell
DOP	Digitales Orthophoto
LBP	Loopy Belief Propagation
LGLN	Landesamt für Geoinformation und Landesvermessung Niedersachsen
nDOM	Normalisiertes digitales Oberflächenmodell
NDVI	Normalised Difference Vegitation Index
IR	Infrarot
NIR	Nahes Infrarot
RGB	Rot-Grün-Blau
HSV	Hue-Saturation-Velocity (dt. Farbton-Sättigung-Helligkeit)
GIS	Geoinformationssystem

Klassifikatoren

CART	Klassifikations- und Regressionsbäume
RF	Random Forest
RRF	Robuster Random Forest
RRFK	Robuster Random Forest mit Kartenbeobachtung
ERFK	Ergebnisbasierter Random Forest mit Kartenbeobachtung
KRFK	Kartenbasierter Random Forest mit Kartenbeobachtung
tRF	Training des Random Forests
tRRF	Training des Robusten Random Forests
LR	Logistische Regression
RLR	Robuste Logistische Regression
RLRK	Robuste Logistische Regression mit Kartenbeobachtung
tRLR	Training der Robusten Logistischen Regression
CRF	Conditional Random Fields
KD	Direkte Klassifikation
KS	Sequenzielle Klassifikation
KM	Multitemporale Klassifikation

Liste von Symbolen

Allgemein

$n, k, a, b, i, j,$ q, m, z, b, q, o	Indizes
δ	Kronecker-Delta-Funktion
$\mathbf{I}$	Einheitsmatrix
exp	Exponentialfunktion
π	Kreiszahl (Pi)
ln	natürlicher Logarithmus
$\forall$	Für alle Elemente von
$\in$	Element von
$\cap$	Schnittmenge

Klassifikation

N	Anzahl der Bildprimitive
N_T	Anzahl der Bildprimitive für das Training
F	Anzahl der Merkmale
K	Anzahl der Klassen
$\mathbf{x}$	Merkmalsvektor
$\mathbf{d}$	Daten
C	Klassenlabel
$\widetilde{C}$	beobachtetes Klassenlabel
$\mathbb{C}$	Menge diskreter Klassenlabels
p	Wahrscheinlichkeitsverteilung

Logistische Regression

Φ	Funktion zur Merkmalsraumtransformation
$\mathbf{w}$	Parametervektor
E	Fehlerfunktion
∇E	Gradientenvektor
H	Hessematrix
τ	Iterations-Index
t	Indikatorvariable für die Klassenzugehörigkeit
σ	Regularisierungswert
y	Posterior eines zu ermittelden Klassenlabels

Robuste Logistische Regression

$\overleftarrow{\gamma}$	Transitionswahrscheinlichkeit $\overleftarrow{\gamma}_{ak} = p(\widetilde{C} = C^k \mid C = C^a)$
$\overrightarrow{\gamma}$	Transitionswahrscheinlichkeit $\overrightarrow{\gamma}_{ak} = p(C = C^a \mid \widetilde{C} = C^k)$
$\overleftarrow{\mathbf{\Gamma}}, \overrightarrow{\mathbf{\Gamma}}$	Transitionsmatrizen
S	Posterior eines beobachteten Klassenlabels

Random Forest

N_B	Anzahl an Bäumen im Random Forest
$\mathcal{T}$	Menge an Trainingsbeispielen im Random Forest
$\mathcal{T}_b$	Menge an Trainingsbeispielen des Baumes b
N_{Tb}	Anzahl an Trainingsbeispielen in $\mathcal{T}_b$
$\mathcal{T}_0$	Menge an Trainingsbeispielen des Baumes b zur Festlegung der Struktur
$\mathcal{T}_c$	Menge an Trainingsbeispielen des Baumes b für die Bestimmung der Wahrscheinlichkeit für eine Klasse in eine Blatt
$\mathcal{T}_q$	Menge an Trainingsbeispielen für den Knoten q
N_F	Anzahl an Merkmalen für die Bestimmung einer Trennfläche
N_G	Anzahl an zufälligen Trennflächen für die Bestimmung einer Entscheidungsgrenze
$\mathcal{T}_{Mj}$ mit $j \in \{1, 2\}$	Teilmenge der Trainingsbeispiele aus $\mathcal{T}_q$ nach der Trennung durch eine Trennfläche
N_{Mj}	Anzahl an Trainingsbeispiele in $\mathcal{T}_{Mj}$
$\mathcal{E}$	Entropie
$\mathcal{I}$	Informationszuwachs
κ	Blatt

Robuster Random Forest

g	Wahrscheinlichkeit für eine Änderung
$\mathcal{V}$	Wahrscheinlichkeit der Klassenzugehörigkeit
i	Index der Iteration
$\mathbf{B}$	Änderungsbild
${}^k\mathbf{B}$	Änderungsbild der Klasse C^k
s	minimale Größe eines Objektes
u	minimale Breite eines Objektes
f_a	Aktualisierungsfunktion
ζ	Konstante

Conditional Random Fields

$\mathbf{C}$	optimale Labelkonfiguration
$\hat{\mathbf{C}}$	approximierte optimalen Labelkonfiguration
Bel	Belief
$\mathcal{S}$	Menge aller Bildprimitive
$\mathcal{O}$	Menge der benachbarten Knoten
$\mathcal{L}$	Menge der zeitlich benachbarten Knoten
$\mathcal{Z}$	Partitionsfunkltion
A_d	merkmalsbezogenes Assoziationspitential
A_z	temporales Assoziationspitential
I_r	räumliches Interaktionspotential
I_z	temporales Interaktionspotential
Z	Anzahl der Zeitpunkte/ Layer
θ	Gewicht des temporalen Assoziationspitentials

kontrastsensitive Potts-Modell

β_0, β_1	Glättungsparameter
σ_D	mittlerer Distanz benachbarter Merkmalsvektoren
$\Delta\mathbf{x}$	euklidische Distanz zweier Merkmalsvektoren

Genauigkeitsmaße

GG	Gesamtgenauigkeit
VG	Vollständigkeit
KG	Korrektheit
$F1$	F1 Wert
$MF1$	Mittelwert des F1 Wert über alle Klassen

Merkmalsselektion

MI	Transinformation
$\hat{MI}$	angenäherte Transinformation
$\hat{\mathcal{E}}$	angenäherte Entropie
$\mathbf{X}$	Gruppe an Merkmalen
$\mathcal{X}$	Merkmalsraum
F_X	Anzahl an Merkmalen in $\mathbf{X}$
ξ	Anzahl der berücksichtigten Nachbarn
$\epsilon_\xi(n)$	zweifache Distanz von Trainingsbeispiel n zu seinem ξ-ten Nachbarn
Γ	Gammafunktion
Ψ	Digammafunktion
c_F	Volumen der Einheitskugel mit der Dimension F_X
N^k	Anzahl an Trainingsbeispielen mit der Klasse C^k

Inhaltsverzeichnis

1. Einleitung **13**
1.1. Motivation . . . 13
1.2. Zielsetzung und wissenschaftlicher Beitrag . . . 16
1.3. Aufbau der Arbeit . . . 18

2. Stand der Forschung **19**
2.1. Änderungsdetektion . . . 19
2.2. Reduzierung des Bedarfs an manuell erstellten Trainingslabels . . . 22
2.3. Strategien zum Umgang mit fehlerhaften Labels . . . 24
2.4. Fehlerhafte Trainingslabels im Bereich der Fernerkundung . . . 29
2.5. Merkmalsselektion . . . 31
2.6. Diskussion . . . 33

3. Grundlagen **38**
3.1. Klassifikation . . . 38
3.2. Merkmale . . . 39
3.2.1. Spektrale Merkmale . . . 40
3.2.2. Texturmerkmale: Haralick Merkmale . . . 40
3.2.3. Strukturmerkmale: gewichtetes Histogramm der Gradientenrichtungen 41
3.2.4. 3D Merkmale: normalisiertes digitales Oberflächenmodell . . . 42
3.2.5. Merkmalsselektionsalgorithmus . . . 43
3.3. Logistische Regression . . . 46
3.4. Gegenüber fehlerhaften Trainingslabels robuste logistische Regression . . . 48
3.5. Klassifikations- und Regressionsbäume . . . 50
3.6. Random Forest . . . 53
3.7. Conditional Random Fields . . . 54

4. Methodik **57**
4.1. Überblick des Klassifikationsalgorithmus . . . 57
4.1.1. Initialisierung . . . 58
4.1.2. Iterativer Prozess . . . 60
4.2. Robuster Random Forest . . . 61
4.2.1. Auswahl der Entscheidungsgrenze . . . 63
4.2.2. Wahrscheinlichkeit der Klassenzugehörigkeit . . . 64
4.2.3. Individuelle Übergangswahrscheinlichkeiten für jedes Trainingsbeispiel 65
4.2.4. Wahrscheinlichkeit in den Blättern . . . 66
4.3. Integration der Kartenlabels als Beobachtungen . . . 67
4.4. Bestimmung der Übergangsmatrix . . . 69
4.5. Änderungswahrscheinlichkeit . . . 70
4.6. Klassifikation basierend auf der robusten logistischen Regression . . . 77
4.7. Multitemporale Klassifikation . . . 79
4.7.1. Multitemporales CRF Modell . . . 80

4.7.2. Sequenzielles Training 82
4.7.3. Ablauf der multitemporalen Klassifikation 84

5. Aufbau der Experimente 85
5.1. Datensätze 85
5.1.1. Hameln 85
5.1.2. Las Vegas 89
5.1.3. Vaihingen 90
5.1.4. Herne 91
5.1.5. Husum 93
5.2. Auswahl der Trainingsbeispiele 93
5.3. Merkmale 95
5.4. Analyse der Merkmalsselektionsmethode 97
5.5. Genauigkeitsmaße 100
5.6. Aufbau und Zielsetzung der Experimente 101
5.6.1. Merkmalsselektion 101
5.6.2. Parametertests 102
5.6.3. Monotemporale Klassifikation 103
5.6.4. Multitemporale Klassifikation 108

6. Experimente 109
6.1. Merkmalsselektion 109
6.2. Einfluss der Parameter 116
6.3. Monotemporale Klassifikation 125
6.3.1. Hameln 125
6.3.2. Las Vegas 135
6.3.3. Vaihingen 145
6.3.4. Herne und Husum 156
6.3.5. Diskussion 163
6.4. Multitemporale Klassifikation 168
6.4.1. Hameln 168
6.4.2. Las Vegas 171
6.4.3. Diskussion 175

7. Fazit und Ausblick 180
7.1. Fazit 180
7.2. Ausblick 182

A. Anhang - Daten 185

B. Anhang - Klassifikationsergebnisse 191

1. Einleitung

1.1. Motivation

Topographische Daten (im Folgenden verkürzt als *Karten* bezeichnet) bilden die Grundlage für verschiedene Anwendungen, wie Navigation, Gefahrenanalyse, Stadt- und Verkehrsplanung. Neben der geometrischen Genauigkeit stellt die Aktualität ein wesentliches Qualitätsmerkmal dar. Verschiedene Ursachen, wie ein schnelles Städtewachstum, führen dazu, dass Karten schnell an Aktualität verlieren und somit fehlerhafte Informationen beinhalten. Um diesem Effekt entgegenzuwirken müssen die Karten häufig aktualisiert werden, wobei eine manuelle Bearbeitung sehr zeitaufwändig und damit kostspielig ist. Aus diesem Grund wird bereits seit längerer Zeit an Methoden zur automatischen oder halbautomatischen Aktualisierung von Karten geforscht.

Als Grundlage werden hierfür hauptsächlich Fernerkundungsdaten, wie z.B. Luft- und Satellitenbilder, verwendet. Aufgrund der schnellen technischen Entwicklung in den letzten Jahren sind Satellitenbilder in immer kürzeren Abständen und besseren Auflösungen verfügbar. Aber auch Qualität und Verfügbarkeit von Luftbildern nehmen zu. Um die Fernerkundungsbilder für die Aktualisierung von Karten nutzen zu können, müssen sie zunächst klassifiziert werden. Durch die Klassifikation wird jedem Bildprimitiv, z.B. Pixel, der Eingangsdaten eine Klasse zugeordnet. Die gewählten Klassen sind abhängig von der zu aktualisierenden Karte und bestehen daher meist aus unterschiedlichen Landbedeckungsarten, wie „Asphalt“ oder „Gras“, oder Arten der Landnutzung. Das Klassifikationsergebnis wird mit der veralteten Karte verglichen, um Änderungen über die Zeit zu detektieren und die Karte zu aktualisieren. Die Qualität der Klassifikation ist folglich ein entscheidendes Kriterium für die Qualität der neuen Karte. Wird das Klassifikationsergebnis nicht geprüft und nachbearbeitet, wie es beispielsweise in halbautomatischen Verfahren erfolgt, werden Fehler der Klassifikation direkt in die Karte übernommen.

Die Klassifikation erfolgt üblicherweise mit überwachten Methoden, da diese durch die Verwendung von Trainingsdaten leichter auf neue Datensätze übertragen werden können. Ein solcher Trainingsdatensatz besteht aus mehreren Trainingsbeispielen, bei denen neben den Sensordaten auch die Klassenlabels bekannt sind, sodass der Zusammenhang zwischen den Daten, repräsentiert durch verschiedene Merkmale, und den Klassen im Trainingsprozess erlernt werden kann. Es ist daher wichtig, dass die Trainingsdaten repräsentativ für die zu klassifizierenden Daten sind, also die gleiche Verteilung in dem von den Merkmalen aufgespannten Raum aufweisen. Die manuelle Erstellung der Klassenlabels der Trainingsdaten, auch *Trainingslabels* genannt, ist sehr zeitaufwändig und somit nicht wirtschaftlich.

Eine Alternative, die in dieser Arbeit behandelt wird, ist die Nutzung der bestehenden Karte zur Gewinnung der Trainingslabels. Den zu klassifizierenden Daten werden demnach die Labels der Karte zugeordnet, um den Trainingsdatensatz zu generieren. Aufgrund des zeitlichen Unterschiedes zwischen der Erstellung der Karte und der Aufnahme der Fernerkundungsbilder stimmen die Klassen aus der Karte mit der aktuellen Landbedeckung aus den

Fernerkundungsbildern in einigen Gebieten nicht mehr überein. Einige Trainingsdaten besitzen demnach fehlerhafte Labels. In der Literatur werden fehlerhafte Trainingslabels oft als *Label Noise* oder *Class Noise* bezeichnet [Frénay und Verleysen, 2014]. In dieser Arbeit wird jedoch davon ausgegangen, dass der Großteil jedes Gebietes unverändert bleibt und somit für jeden Objekttyp mehr Trainingsdaten mit korrekten als mit falschen Labels vorliegen.

Ein Großteil der Klassifikatoren gehen von fehlerfreien Trainingslabels aus, sodass die einfache Verwendung der Labels aus der veralteten Karte zu einer falschen Anpassung des Klassifikators und somit zu einem fehlerhaften Klassifikationsergebniss führen kann. Die Stärke und Ausprägung der Fehlklassifikation hängt dabei von dem Klassifikator, den Daten und den fehlerhaften Labels ab. Das Auftreten fehlerbehafteter Trainingslabels und deren Berücksichtigung innerhalb des Trainingsprozesses wurde bereits in vielen Arbeiten untersucht. Neben der Verwendung von Klassifikatoren, die nicht oder wenig durch fehlerhafte Trainingslabels beeinflusst werden, können Klassifikatoren auch an das Auftreten von Labelfehlern angepasst werden, z.B. indem die Fehler im Training ebenfalls modelliert werden. Bei der Beschreibung des Auftretens fehlerhafter Trainingslabels wird jedoch häufig ein zu stark vereinfachtes Modell verwendet. Beispielsweise wird angenommen, dass das Auftreten eines fehlerhaften Labels ausschließlich vom Zufall [Kalai und Servedio, 2003] oder von dem beobachteten Label und dem unbekannten wahren Label abhängt [Bootkrajang und Kabán, 2012] und nicht von den Daten. Der statistische Zusammenhang zwischen dem Auftreten der Fehler und den Labels wird überwiegend in Form von Übergangswahrscheinlichkeiten in die Trainingsprozedur integriert, welche meist zuvor empirisch bestimmt werden oder aus Expertenwissen stammen [Melgani und Serpico, 2003]. Die Modellierung mit Hilfe der Übergangswahrscheinlichkeiten ist bei der Verwendung einer Karte zur Bestimmung der Trainingslabels im Allgemeinen unzureichend. In der Darstellung fehlt beispielsweise räumlicher Kontext. Die Tatsache, dass Objekte in der Karte und somit auch die Änderungen eine gewisse Größe im Bildraum aufweisen, wird somit nicht berücksichtigt. Dies kann dazu führen, dass die Auswirkungen fehlerhafter Trainingslabels auf das Klassifikationsergebnis nicht vollständig kompensiert werden können.

Der *Random Forest* (RF) Klassifikator [Breiman, 2001] ist dafür bekannt, auch ohne Anpassung des Trainingsprozesses besser mit einem gewissen Grad an fehlerhaften Trainingslabeln umgehen zu können als andere Klassifikatoren, z.B. [Frénay und Verleysen, 2014] oder [Zhao, 2017]. Bei Vergleichen mit anderen Klassifikatoren, wie in [Folleco et al., 2008] und [Pelletier et al., 2017], zeigte der RF auch eine relativ große Robustheit gegenüber fehlerhaften Labels. Ein negativer Einfluss der Fehler auf das Ergebnis ist dennoch zu erkennen. Der Einfluss könnte zusätzlich verstärkt werden, wenn die Fehler nicht, wie z.B. in den Experimenten aus [Pelletier et al., 2017], ausschließlich vom Zufall oder den Labels abhängen, da der Bootstrap-Algorithmus innerhalb des RF in erster Linie den Einfluss zufälliger Fehler reduziert. Wie oben erwähnt ist die Verteilung der Fehler bei der Verwendung einer veralteten Karte zur Labelbestimmung deutlich komplexer. Die Tatsache, dass Änderungen eine gewisse Größe im Bildraum aufweisen, wird beispielsweise nicht berücksichtigt. Wie diese Fehler in den Trainingslabels, deren Verteilung nicht ausschließlich vom Zufall oder den Labels abhängen, in dem Training des RF berücksichtigt und kompensiert werden können, ist bisher noch ungelöst.

In dieser Arbeit wird dazu zunächst der Random Forest Klassifikator durch ein komplexes

Fehlermodell erweitert, sodass dieser auch mit fehlerhaften Trainingslabels umgehen kann, die aus der Verwendung einer veralteten Karte für die Labelbestimmung resultieren. In dem Modell werden neben Abhängigkeiten zwischen dem Auftreten der Labelfehler und dem Label der veralteten Karte auch die Daten integriert. Desweiteren wird auch berücksichtigt, dass Labelfehler in den Fernerkundungsdaten in Segmenten, also Regionen zusammenhängender Pixel, auftreten. Vorinformationen bezüglich der Änderungen sind nicht nötig, da die verwendeten Übergangswahrscheinlichkeiten und Gebiete mit potenziell fehlerhaften Trainingslabels in einem iterativen Prozess bestimmt werden.

Unabhängig von der Verwendung der Karte zur Bestimmung der Trainingslabels beinhaltet die Karte Informationen über die Position und Form von Objekten, solange diese noch in den Fernerkundungsbildern vorhanden sind. Diese Information wird meist erst nach der Klassifikation bei der eigentlichen Aktualisierung der Karte genutzt oder bei der Nachprozessierung einer Änderungsdetektion, z.B. in [Olsen et al., 2002]. Während der Klassifikation werden die Karteninformationen oft vernachlässigt, obwohl sie beispielsweise bei der Klassifikation von Objekten, bei denen sich der Klassifikator unsicher ist, helfen könnten. In dieser Arbeit wird daher das Kartenlabel als zusätzliche Beobachtung in dem Klassifikationsprozess berücksichtigt. Der statistische Zusammenhang zwischen Kartenlabels und den zu bestimmenden Labels wird, wie bei dem robusten Traininig, durch Übergangswahrscheinlichkeiten aller Klassenkombinationen modelliert. Um eine temporale Überanpassung, ähnlich wie in [Hoberg et al., 2015], zu vermeiden, wird der Einfluss des Kartenlabels auf das Klassifikationsergebnis entsprechend der Wahrscheinlichkeit, dass eine Änderung vorliegt, gewichtet.

Liegen Fernerkundungsbilder nicht nur von einem, sondern von mehreren Zeitpunkten vor, ist es sinnvoll, alle Bilder bei der Klassifikation zu berücksichtigen. Besonders bei zeitlichen Analysen kann die Klassifikation von multitemporalen Datensätzen hilfreich sein. Im Bereich der automatischen Kartenaktualisierung und der Verwendung von Kartenlabels im Trainingsprozess kommt noch der Vorteil hinzu, dass die zeitlichen Differenzen zwischen den Aufnahmezeitpunkten der Bilder einer Serie mitunter kleiner sind als jene zwischen dem Zeitpunkt der Kartenerstellung und dem Zeitpunkt der Aufnahme des aktuellsten Fernerkundungsbildes. Dementsprechend ist auch die Menge an Änderungen der Landbedeckung, die detektiert bzw. während des Trainings kompensiert werden muss, geringer. Eine Möglichkeit der multitemporalen Analyse ist daher eine sequenzielle Verarbeitung, bei der die Karte immer mit den Fernerkundungsbildern des nächstgelegenen Zeitpunktes aktualisiert wird. Statistische Zusammenhänge zwischen den Zeitpunkten gehen hierbei meist verloren, weshalb in dieser Arbeit eine neu entwickelte multitemporale Klassifikation verwendet wird, in der alle Beobachtungen, auch die Kartenlabels, simultan genutzt werden. Die Labels der veralteten Karte werden dabei als Labels eines weiteren Zeitpunktes interpretiert. Bereits bekannte multitemporale Klassifikatoren, z.B. aus [Hoberg et al., 2015], berücksichtigen hingegen ausschließlich Sensordaten. Ein weiterer Unterschied zu der multitemporalen Klassifikation in [Hoberg et al., 2015] ist, dass aufgrund des sequenziellen Trainings unter Verwendung der Kartenlabels, welche in der vorgelegten Arbeit genutzt werden, keine zusätzlichen Trainingslabels jedes Zeitpunktes benötigt werden. Übergangswahrscheinlichkeiten zwischen den Zeitpunkten gehen ebenfalls aus dem Trainingsprozess hervor.

Zusammenfassend kann festgehalten werden, dass eine zeitnahe und kostengünstige Aktualisierung von Karten in vielen Bereichen hilfreich und somit die Automatisierung dieses Prozesses von Vorteil ist. Ein Ansatz, dieser komplexen Herausforderung zu begegnen, ist die überwachte Klassifikation von aktuellen Fernerkundungsbildern. In dieser Arbeit wird daher ein Klassifikator für die Aktualisierung veralteter Karten vorgestellt. Das aufwändige manuelle Erstellen der notwendigen Trainingslabels wird durch die Verwendung der Kartenlabels umgangen. Um mit den dadurch entstandenen fehlerhaften Trainingslabels umgehen zu können, wird in dem neu entwickelten Verfahren die Trainingsprozedur des Random Forest Klassifikators durch die Integration eines komplexen Fehlermodells erweitert. Eine manuelle Erstellung von Trainingslabels ist demnach nicht mehr nötig. Für eine bessere Klassifikation von Objekten, die über die Zeit unverändert geblieben sind, fließen die Labels der Karte zusätzlich in den Klassifikationsprozess ein. Sind Fernerkundungsbilder aus mehreren Zeitpunkten bekannt, werden diese gemeinsam in einem multitemporalen Ansatz klassifiziert. Auch hier fließen die Labels aus einer veralteten Karte in die Klassifikation mit ein. Übergangswahrscheinlichkeiten zwischen den Klassen benachbarter Zeitpunkte werden während eines sequenziellen Trainings bestimmt und müssen nicht vor der Klassifikation bekannt sein. Außerdem basiert das sequenzielle Training ausschließlich auf den vorhandenen Sensordaten und den Labels der veralteten Karte. Manuell gelabelte Daten in jeglicher Form sind demnach auch in der multitemporalen Klassifikation nicht notwendig.

1.2. Zielsetzung und wissenschaftlicher Beitrag

Das Ziel der Arbeit ist die Entwicklung eines multitemporalen Klassifikators zur Aktualisierung von veralteten Karten, der mit den potenziell fehlerhaften Labels dieser Karten trainiert werden kann. Mindestens bekannt sein müssen dabei Fernerkundungsdaten des aktuellen Zeitpunktes und Kartenmaterial, mit dem jedem zu klassifizierendem Bildprimitiv ein Label zugeordnet werden kann. Einschränkungen an die Auflösung und die Klassenstruktur werden nicht vorgenommen, sodass sowohl Luftbilder mit einer Bodenpixelgröße im Dezimeter-Bereich als auch Satellitendaten mit einer Auflösung von mehreren Zehnermetern genutzt werden können. Die Objekte der gewählten Klassen aus der Karte müssen jedoch in den Sensordaten mit der gegebenen Auflösung erkennbar sein.

Der wissenschaftliche Beitrag kann folgendermaßen charakterisiert werden.

- **Die Entwicklung eines gegenüber fehlerhaften Trainingslabels robusten Klassifikators basierend auf einem neuen komplexen Fehlermodell.** Die Labels der Trainingsdaten werden aus einer veralteten Karte extrahiert, wodurch ein zeitaufwändiges manuelles Labeln vermieden wird. Die resultierenden Trainingslabels sind, aufgrund der Zeitdifferenz zwischen dem Zeitpunkt der Erstellung der Karte und der Sensordaten, fehlerbehaftet. Um Robustheit gegenüber diesen Fehlern zu erlangen, wird der Random Forest Klassifikator [Breiman, 2001] durch ein komplexes Fehlermodell erweitert. Hierbei werden jedem Trainingsbeispiel anstelle einer bestimmten Klasse alle Klassen mit einer bestimmten Wahrscheinlichkeit zugeordnet. Diese *Wahrscheinlichkeit der Klassenzugehörigkeit* hängt sowohl von den Daten als auch von dem Label aus der veralteten

Karte ab. Außerdem fließt eine Übergangswahrscheinlichkeit mit ein, in deren Bestimmung auch die Tatsache berücksichtigt wird, dass Änderungen der Karte in Segmenten im Bildraum auftreten. In einem iterativen Prozess werden die Daten klassifiziert und die Wahrscheinlichkeiten der Klassenzugehörigkeit angepasst.

- **Die Integration der Karteninformation in den Klassifikationsprozess.** Die Detektion der Änderungen erfolgt durch den Vergleich der Kartenlabels mit den resultierenden Labels der Klassifikation von aktuellen Sensordaten. Es gibt jedoch Bereiche, in denen der Klassifikator fehleranfällig ist, z.B. bei Mischpixeln, in Schattenbereichen oder wenn die Merkmalsausprägung eines Objektes von der anderer Objekte gleicher Klasse leicht abweicht. In diesen Bereichen können die Labels der veralteten Karte helfen, das korrekte Klassenlabel zu bestimmen. Die Labels der Karte werden daher als zusätzliche Beobachtung in den Klassifikationsprozess mit einbezogen. Hierfür wird der statistische Zusammenhang zwischen den Labels der Karte und den zu bestimmenden Labels modelliert und in Form eines temporalen Assoziationspotentials in ein *Conditional Random Field* (CRF) integriert. Da das Auftreten von Änderungen im Allgemeinen eher selten ist, könnte es zu einer Überanpassung kommen, wenn der statistische Zusammenhang ausschließlich durch die Übergangswahrscheinlichkeiten dargestellt wird. Aus diesem Grund wird der Einfluss des Kartenlabels auf das Klassifikationsergebnis entsprechend der Wahrscheinlichkeit, dass eine Änderung vorliegt (Änderungswahrscheinlichkeit) gewichtet.

- **Die Berücksichtigung von temporalem und lokalem Kontext mit Hilfe des Conditional Random Fields.** Mit Hilfe des CRFs wird die lokale Umgebung eines Primitivs bei dessen Klassifikation mit berücksichtigt. Neben der Abhängigkeit des Klassenlabels zu den Daten kann auch der statistische Zusammenhang zu den räumlichen und, wie bei [Hoberg et al., 2015], zeitlichen Nachbarn in die Klassifikation mit einfließen. In dieser Arbeit wird der robuste Random Forest in ein CRF integriert, um den lokalen Kontext bei der Klassifikation zu berücksichtigen. Stehen Sensordaten von mehreren Zeitpunkten zur Verfügung, kann ein multitemporales CRF verwendet werden. Der wissenschaftliche Beitrag liegt hierbei bei der Einbettung des monotemporalen CRFs in den iterativen Prozess des robusten Random Forests, der Erweiterung um ein temporales Assoziationspotential auch bei dem multitemporalen CRF und der Modellierung der temporalen Potentiale mit Hilfe der Änderungswahrscheinlichkeiten. Das Training für das multitemporale CRF basiert außerdem, wie bei dem monotemporalen CRF, auf Trainingsdaten, deren Labels aus einer veralteten Karte stammen. Auf zeitaufwändiges manuelles Labeln kann dementsprechend auch hier verzichtet werden.

- **Die Bestimmung der Übergangswahrscheinlichkeiten aus den Daten.** Die Übergangswahrscheinlichkeiten müssen nicht zuvor empirisch oder durch Expertenwissen bestimmt werden, sondern werden während der Klassifikation aus den Daten bestimmt. Hierbei werden zwei Möglichkeiten vorgestellt. Zunächst werden die Übergangswahrscheinlichkeiten innerhalb des Trainings der robusten logistischen Regression [Bootkrajang und Kabán, 2012] geschätzt. Bei der zweiten Möglichkeit werden sie nach jeder

Iteration mit dem robusten Random Forest aus dem Klassifikationsergebnis und den Labels der veralteten Karte abgeleitet.

Neben dem Klassifikationsprozess wird auch die Merkmalsselektion, wenn sie auf Trainingslabels beruht, durch die Verwendung der Labels einer veralteten Karte beeinflusst. Da in dieser Arbeit ein robuster Klassifikator entwickelt wurde, um zeitaufwändiges manuelles Labeln zu umgehen, würde die Notwendigkeit von fehlerfreien Trainingslabels für die Vorprozessierung dem Hauptziel dieser Arbeit widersprechen. Daher soll in dieser Arbeit auch gezeigt werden, dass für den gesamten Aktualisierungsprozess der Karte keine manuell gelabelten Daten benötigt werden. Aus diesem Grund wird neben der Klassifikation auch die Merkmalsselektion mit Hilfe der veralteten Karte durchgeführt. Da für die Merkmalsselektion keine neue Methode entwickelt wurde, sondern eine bereits bekannte Methode aus [Frénay et al., 2014] basierend auf der *Transinformation* (engl. *mutual information* [Shannon, 1948]) verwendet wird, ist sie kein expliziter Bestandteil des wissenschaftlichen Beitrags dieser Arbeit. Lediglich für die Evaluierung werden manuell erstellte Labels als Referenz genutzt.

Um fehlerhafte Labels auch als solche detektieren zu können, wird die Annahme getroffen, dass jede Art von Objekten zum Großteil korrekt durch die Trainingsdaten repräsentiert ist. Jede Ballung im Merkmalsraum sollte demnach mehr Trainingsbeispiele mit korrektem Label beinhalten als solche mit fehlerhaftem Label. Diese Annahme trifft in der Realität nicht immer zu, beispielsweise wenn neue Objekte, die nach der Kartenerstellung entstanden sind, andere Merkmale aufweisen, als alte Objekte der selben Klasse, die bereits in der Karte abgebildet wurden.

1.3. Aufbau der Arbeit

Die Arbeit gliedert sich wie folgt: In Kapitel 2 wird der aktuelle Stand der Forschung wiedergegeben und die vorliegende Arbeit davon abgegrenzt. In dem darauf folgenden Kapitel 3 wird zunächst die Klassifikation allgemein erläutert und die hierbei verwendete Notation eingeführt. Anschließend werden die in dieser Arbeit verwendeten Merkmale beschrieben und die Klassifikationsmethoden vorgestellt, auf denen diese Arbeit beruht. Kapitel 4 beschreibt die neu entwickelte Methodik zur Klassifikation mit Hilfe veralteter Karten. Der Aufbau der Experimente inklusive der Vorstellung der verwendeten Daten und der Evaluierungsmethoden wird in Kapitel 5 dargelegt. Die Ergebnisse der Experimente und die Evaluierung finden sich in Kapitel 6. Die Arbeit schließt mit einem Fazit und einem Ausblick in Kapitel 7.

2. Stand der Forschung

Diese Beschreibung des aktuellen Stands der Forschung beginnt in Kapitel 2.1 mit der Diskussion von verschiedenen Methoden der Änderungsdetektion. Die Erkennung von Änderungen zur Aktualisierung einer veralteten Karte erfolgt in der vorliegenden Arbeit durch den Vergleich der Karte mit zuvor klassifizierten aktuellen Sensordaten. Da hierfür ein überwachtes Klassifikationsverfahren genutzt werden soll, werden Trainingsdaten benötigt. Kapitel 2.2 legt einige Möglichkeiten dar, den manuellen Aufwand bei der Erstellung der Trainingslabels zu reduzieren. In der hier vorliegenden Arbeit werden die Labels der veralteten Karte für das Training verwendet, was dazu führt, dass einige Labels fehlerhaft sind. Da die Fehler in den Trainingsdaten das Klassifikationsergebnis ggf. beeinflussen, gibt es verschiedene Strategien mit fehlerbehafteten Trainingslabels umzugehen, welche in Kapitel 2.3 diskutiert werden. In jenem Kapitel wird auch auf den Umgang mit fehlerhaften Labels bei der Verwendung des Random Forest Klassifikators eingegangen. Wie im Bereich der Fernerkundung mit Labelfehlern umgegangen wird, zeigt Kapitel 2.4. In Kapitel 2.5 wird auf die Merkmalsselektion unter Berücksichtigung fehlerhafter Labels eingegangen. Zuletzt erfolgt in Kapitel 2.6 eine Diskussion über den zuvor dargelegten aktuellen Stand der Forschung.

2.1. Änderungsdetektion

Für die Aktualisierung von Karten mit Hilfe von aktuellen Fernerkundungsbildern werden im Allgemeinen Methoden verwendet, um Veränderungen zwischen dem Zeitpunkt der Kartenerstellung und der Sensoraufnahmen zu erkennen. Da das Thema der Änderungsdetektion in vielen Bereichen und seit längerer Zeit von Interesse ist, existieren bereits einige Veröffentlichungen, die eine Übersicht über das Thema geben, z.B. [Lu et al., 2004], [Jianya et al., 2008], [Ban, 2016] und [Liu et al., 2019]. Strukturiert ist dieses Kapitel nach den drei Arten Änderungen zu detektieren, wie sie in Jianya et al. [2008] zu finden sind.

Die erste Möglichkeit ist ein direkter Vergleich der Sensordaten zweier Zeitpunkte, was meist als *unüberwachte Änderungsdetektion* [Bruzzone und Prieto, 2000] bezeichnet wird. Hierbei werden oft Merkmale, wie das Verhältnis zweier Bänder oder andere geeignete mathematische Operatoren, verwendet, um die Sensordaten zu vergleichen, z.B. [Subudhi et al., 2014]. Diese Merkmale werden anschließend analysiert, um die Änderungsregionen zu extrahieren [Celik, 2009]. Ein großer Nachteil dieser Methoden ist, dass keine Information über die Art der Änderungen bestimmt wird, sondern lediglich zwischen *Änderung* und *keine Änderung* unterschieden wird [Lu et al., 2004]. Durch die Gruppierung der Pixel, die eine Änderung aufweisen, in verschiedene Ballungen kann die Klasse *Änderung* weiter differenziert werden [Saha et al., 2019], jedoch bleibt die Interpretation, um welche Art der Änderung es sich handelt, weiter offen. Ein weiterer Nachteil ist, dass die Methodik anfällig gegenüber unterschiedlichen Erscheinungsformen der Objekte ist, z.B. aufgrund von sich ändernden Beleuchtungsverhältnissen [Mas, 1999]. Auch die Möglichkeit, verschiedene Sensoren zu verwenden, ist bei unüberwachten Methoden im Allgemeinen nicht gegeben [Bruzzone und Prieto, 2000]. Neben

den bereits genannten Nachteilen können außerdem lediglich zwei Zeitpunkte miteinander verglichen werden.

Bruzzone und Prieto [2000] erklären, dass, obwohl diese Nachteile bekannt sind, oft auf unüberwachte Methoden zurückgegriffen wird, da Trainingsdaten für beide Zeitpunkte, z.B. aufgrund der hohen Kosten für deren Erstellung, nicht zur Verfügung stehen. Dies entspricht der Motivation der vorliegenden Arbeit, die manuelle Erstellung der Trainingslabels zu vermeiden und stattdessen auf die Labels der veralteten Karte zurückzugreifen. Eine weitere Begründung für die Verwendung dieser Art der Änderungsdetektion liefern Clement et al. [2018], die im Bereich der Ausbreitung von Hochwasser forschen. Hier ist es wichtig, nahezu in Echtzeit Änderungen zu detektieren, sodass Methoden genutzt werden, die schnell und einfach über große Gebiete anwendbar sind. Die dargelegte Arbeit hegt jedoch keine Echtzeitansprüche, sodass bei den Methoden der Aspekt der Rechenzeit nicht berücksichtigt wird.

Neben den zuvor genannten Nachteilen spricht jedoch auch die Tatsache, dass in der vorliegenden Arbeit keine Fernerkundungsdaten zum Zeitpunkt der Kartenerstellung zur Verfügung stehen, gegen die Verwendung eines direkten Vergleichs der Sensordaten.

Die zweite Möglichkeit zur Änderungsdetektion nach Jianya et al. [2008] beruht auf dem Vergleich der Ergebnisse einer unabhängig durchgeführten Klassifikation der Fernerkundungsdaten zweier Zeitpunkte, z.B. [Stefanski et al., 2014]. Obwohl auch hier lediglich zwei Zeitpunkte miteinander verglichen werden, entfallen die anderen zuvor genannten Nachteile der unüberwachten Klassifikation. Neben der aufwändigen Beschaffung von Trainingslabels, welche in der vorliegenden Arbeit durch die Verwendung der veralteten Karte entfällt, besteht ein weiterer Nachteil der Methode in der Fehlerfortpflanzung [Radoux und Defourny, 2010]. In dem Kontext der Änderungsdetektion durch den Vergleich zweier Labelbilder bedeutet dies, dass sich die Fehler der Labelbilder beider Epochen überlagern und so das Ergebnis der Änderungsdetektion eine schlechtere Qualität als die Klassifikationen selbst aufweisen kann. In der vorliegenden Arbeit ersetzt die veraltete Karte das Klassifikationsergebnis des ersten Zeitpunktes. Für die Änderungsdetektion werden somit die aktuellen Sensordaten klassifiziert und das Klassifikationsergebnis mit der Karte verglichen. Laut Jianya et al. [2008] wird in der Praxis, aufgrund der Fehleranfälligkeit der Klassifikation, der Vergleich zwischen einer Karte und einem Klassifikationsergebnis häufiger genutzt als der Vergleich zweier Bilder.

Bei der Verwendung von Neuronalen Netzen wird oft aus den Fernerkundungsdaten beider Zeitpunkte direkt die Änderung klassifiziert. Alcantarilla et al. [2018] detektieren beispielsweise mit Hilfe eines *Deep Deconvolutional Networks* (DN) Änderungen in Bildpaaren aus Street View Datensätzen unterschiedlicher Zeitpunkte, wobei auch hier Trainingsdaten benötigt werden. Bei der Änderungsdetektion mit Hyperspektralbildern unter Verwendung eines *Convolutional Neural Networks* (CNN) in [Wang et al., 2019] wird beispielsweise ein Änderungsdetektor, basierend auf der *Change Vector Analysis* (CVA) und einem Schwellwert, verwendet, um die Labels für das Training zu bestimmen.

Die dritte und allgemeinste Möglichkeit zur Änderungsdetektion nach [Jianya et al., 2008] ist die simultane Klassifikation der Bilder aus allen vorliegenden Zeitpunkten. Auch eine multitemporale Klassifikation mehrerer Zeitpunkte und somit die Analyse von längeren Zeitreihen

ist mit dieser Methode möglich. In einem probabilistischen Umfeld werden hierbei Übergangswahrscheinlichkeiten zwischen den einzelnen Zeitpunkten verwendet. Bei der Klassifikation landwirtschaftlicher Kulturen in Satellitenbildern unterschiedlicher Zeitpunkte nutzen Leite et al. [2011] diese beispielsweise innerhalb von Markov-Ketten, um die zeitlichen Abhängigkeiten zu modellieren. Die Übergangswahrscheinlichkeiten werden meist empirisch vor der Klassifikation bestimmt, z.B. in [Hoberg et al., 2015], oder basieren auf Expertenwissen, z.B. in [Melgani und Serpico, 2003]. Bruzzone et al. [1999] bestimmen diese Übergangswahrscheinlichkeiten aus zwei Bildern von unterschiedlichen Zeitpunkten mit Hilfe eines *Expectation Minimization* (EM) Algorithmus [Dempster et al., 1977]. Da keine Daten zum den Zeitpunkt der Kartenerstellung vorliegen, kann dieser Ansatz in der vorliegenden Arbeit nicht genutzt werden. Hoberg et al. [2015] modellieren sowohl temporalen Kontext als auch räumlichen Kontext innerhalb eines *Conditional Random Fields* (CRF) [Kumar und Hebert, 2006] und steigerten so die Genauigkeit des Klassifikationsergebnisses im Vergleich zu einer Klassifikation ohne räumlichen bzw. zeitlichen Kontext. Die multitemporale Klassifikation, die in der vorliegenden Arbeit vorgestellt wird, verwendet ebenfalls ein CRF, mit dem räumlicher und temporaler Kontext berücksichtigt werden. Anders als z.B. das multitemporale CRF von Hoberg et al. [2015] beinhaltet das neu entwickelte CRF auch die Kartenlabels als zusätzliche Beobachtung. Außerdem werden keine manuell erstellten Trainingslabels benötigt, da die hier genutzte sequenzielle Trainingsprozedur ausschließlich die Sensordaten, die Labels der veralteten Karte und die Klassifikationsergebnisse des vorherigen Zeitpunktes verwendet. Auch die Übergangswahrscheinlichkeiten werden in dieser Trainingsprozedur bestimmt.

Für die Aktualisierung von Karten ist es wichtig, zwischen tatsächlichen Änderungen und unveränderten Gebieten, die fälschlicherweise als Änderungen detektiert wurden, zu unterscheiden. Olsen et al. [2002] bearbeiten beispielsweise in der Nachprozessierung das Binärbild, welches die Klassen *Änderung* und *keine Änderung* beinhaltet, durch morphologische Filter für die Beseitigung kleinerer und somit potentiell falsch klassifizierter Pixelgruppen. Auch zusammenhängende Regionen (Segmente) von Pixeln, die in dem Binärbild der Klasse *Änderung* zugewiesen wurden, werden zusätzlich als *keine Änderung* klassifiziert, wenn deren Größe einen bestimmten Schwellwert nicht überschreitet, da kleinere Objekte nicht in die Karte mit aufgenommen werden sollen. Durch ein Schwellwertverfahren bezogen auf den Intensitätswert erfolgt auch die Zuordnung der Pixel der Klasse *Änderung*, die in Schattenregionen liegen, zu der Klasse *keine Änderung*. In der vorliegenden Arbeit werden Gebiete mit tatsächlichen Änderungen auf ähnliche Weise von Gebieten, die fehlerhaft als solche klassifiziert wurden, unterschieden. Ein großer Unterschied ist jedoch, dass die Bestimmung der Änderungsgebiete nicht nur mit dem finalen Klassifikationsergebnis durchgeführt wird, sondern in einem iterativen Verfahren auch aktualisiert wird und in den Trainings- und Klassifikationsprozess mit einfließt. Dies ermöglicht, dass sich sowohl das Klassifikationsergebnis als auch die Änderungsdetektion während des iterativen Prozesses verbessern und somit die Anzahl an falsch klassifizierten Pixeln abnimmt. Annahmen über die zu detektierenden Änderungen können nicht nur der Nachbearbeitung dienen, sondern auch bei der Auswahl der Trainingsdaten genutzt werden. Knudsen und Olsen [2003] entfernen beispielsweise alle Trainingspixel, die an Objektgrenzen liegen, da es sich bei diesen meist um Mischpixel handelt. Auch Schattenregionen, die durch ein Schwellwertverfahren basierend auf den drei RGB Kanälen definiert werden, werden

in [Knudsen und Olsen, 2003] nicht für das Training verwendet. In der vorliegenden Arbeit werden keine Trainingspixel aufgrund der Annahme, sie könnten die Klassifikation negativ beeinflussen, entfernt. Da in den zu klassifizierenden Daten ebenfalls Mischpixel und besonders Schattenregionen vorhanden sind, wären die Trainingsdaten nach Ausdünnung nicht mehr repräsentativ.

2.2. Reduzierung des Bedarfs an manuell erstellten Trainingslabels

Die zweite und dritte Gruppe an Verfahren zur Änderungsdetektion, die beide, abhängig von der Anzahl an Epochen mit vorhandenen Sensordaten, in der vorliegenden Arbeit genutzt werden, beinhalten die Klassifikation der Sensordaten. Bei der Klassifikation werden modellbasierte Klassifikatoren von statistischen Klassifikatoren unterschieden. Um eine große Bandbreite an Sensordaten und Klassenstrukturen verarbeiten zu können, werden in der vorliegenden Arbeit überwachte Methoden, welche zu den statistischen Klassifikatoren gehören, verwendet. Die überwachten Methoden benötigen für das Training, in dem die Parameter für die Prädiktion der Labels bestimmt werden, Trainingsdaten. Wie bereits beschrieben ist die manuelle Erstellung der Labels dieser Trainingsdaten sehr zeitaufwändig und somit unwirtschaftlich. Es gibt daher mehrere Ansätze den manuellen Aufwand zu verringern.

Bei der *Domänenadaption* (engl. *domain adaptation*) (DA) wird ein Klassifikator, der mit Hilfe von korrekten Trainingsbeispielen an einen Datensatz (Quelldomäne) angepasst wurde, auf einen neuen Datensatz (Zieldomäne) übertragen [Vogt et al., 2018]. Die gemeinsame Verteilung der Merkmale und der Labels ist in beiden Domänen zwar unterschiedlich, aber ähnlich. In der Zieldomäne sind keine oder nur vereinzelte Trainingsbeispiele gegeben. Je nachdem, was bei dem Transferlernen übertragen wird, unterscheiden Pan und Yang [2009] zwischen dem Instanzentransfer und dem Repräsentationstransfer. Beim Instanzentransfer wird ein Klassifikator schrittweise von der Quell- in die Zieldomäne übertragen, indem Trainingsbeispiele der Quelldomäne durch Trainingsbeispiele der Zieldomäne ersetzt werden. Die Labels der Trainingsbeispiele für die Zieldomäne stammen hierbei von dem aktuellen Stand des Klassifikators. Beispiele hierfür sind [Paul et al., 2016] und [Bruzzone und Marconcini, 2009]. Bei dem Repräsentationstransfer wird angenommen, dass die Unterschiede zwischen den Domänen durch Projektion beider Domänen in einen gemeinsamen Merkmalsraum abgeschwächt werden können. Ein Beispiel hierfür ist [Matasci et al., 2015]. Einen Überblick von Methoden der Domänenadaption in der Fernerkundung bieten Tuia et al. [2016]. Wird die DA mit *Deep Learning* (DL, dt. tiefes Lernen) Methoden, z.B. mit Convolutional Neural Networks (CNN), realisiert, wird von *Deep DA* gesprochen. Einen Überblick über verschiedene Deep DA Methoden bietet [Wang und Deng, 2018]. In der vorliegenden Arbeit wird allerdings davon ausgegangen, dass keine korrekt gelabelten Daten, die als Daten der Quelldomäne interpretiert werden könnten, vorliegen. Demnach kann der Klassifikator nicht oder, unter Verwendung fehlerhafter Trainingslabels, nicht korrekt in der Quelldomäne angelernt werden. Fehler beim Training in der Quelldomäne könnten durch die Domänenadaption auf die Zieldomäne übertragen werden.

Aktives Lernen (engl. *active learning*), z.B. in [Tuia et al., 2009], beschreibt einen iterativen Prozess, bei dem zunächst der Algorithmus das aussagekräftigste Trainingsbeispiel ohne bekanntes Klassenlabel bestimmt. Dieses Trainingsbeispiel wird anschließend manuell gelabelt und der Klassifikator wird mit dem neuen Trainingsbeispiel durch erneutes Training aktualisiert [Tuia et al., 2009]. Da Trainingsbeispiele, die keine Aussagekraft besitzen, nicht gelabelt werden müssen, reduzieren sich die Kosten und der Zeitaufwand für die Erstellung der Trainingslabels deutlich [Demir et al., 2010]. Beim *Batch Mode Active Learning* werden auch gleich mehrere Trainingsbeispiele pro Iteration für das manuelle Labeln von dem Algorithmus präsentiert. Besonders bei zeitaufwändigen Trainingsprozessen ist dies effizienter, als wenn der Klassifikator nach der Hinzunahme jedes einzelnen Trainingsbeispieles erneut trainiert werden müsste [Guo und Schuurmans, 2008]. Obwohl die Anzahl der Trainingsbeispiele, die manuell gelabelt werden, deutlich reduziert wird, bleibt die Notwendigkeit eines menschlichen Operateurs bestehen. In der vorliegenden Arbeit wird jedoch komplett auf manuelles Labeln verzichtet, da die Labels aus einer existierenden Karte extrahiert werden können.

Halb-überwachte Methoden verwenden sowohl Trainingsbeispiele mit bekanntem Label als auch solche ohne Label [Chapelle et al., 2003]. Um ungelabelte Trainingsbeispiele nutzen zu können, werden eine oder mehrere Annahmen getroffen. Eine mögliche Annahme ist, dass Trainingsbeispiele vermutlich derselben Klasse angehören, wenn sie innerhalb einer Ballung, also einer Region mit einer hohen Dichte an Trainingsbeispielen im Merkmalsraum, liegen [Chapelle et al., 2003]. Dies bedeutet auch, dass die Entscheidungsgrenze durch Regionen mit einer niedrigen Dichte im Merkmalsraum führt [Chapelle und Zien, 2005]. Basierend auf dieser Annahme wird in [Chapelle et al., 2003] die Verwendung ungelabelter Trainingsbeispiele in einen kernelbasiertem Klassifikator integriert, indem das Eigenspektrum der Kernel-Matrix modifiziert wird. Guillaumin et al. [2010] verwenden eine halb-überwachte Methode, um ganze Bilder einer bestimmten Klasse zuzuordnen. Für das Training stehen gelabelte und ungelabelte Bilder mit bekannten Keywords (dt. Schlüsselwörter bzw. Schlagwörter) zur Verfügung, während den zu klassifizierenden Bildern keine Keywords zugeordnet sind. Diese Keywords dienen, neben dem Bildinhalt, als Zusammenhang zwischen den gelabelten und ungelabelten Daten. Mit den gelabelten Daten wird ein *Multiple Kernel Learning* (MKL) Klassifikator trainiert, der sowohl den Bildinhalt, als auch die Keywords verwendet. Nachdem den zuvor ungelabelten Bildern mit Hilfe dieses Klassifikators ein Label zugewiesen wurde, kann ein überwachter Klassifikator, wie die *Support Vector Machines*, auf Basis aller Trainingsdaten angelernt werden. Die Keywords werden hierbei nicht berücksichtigt, sodass auch Bilder ohne Keywords klassifiziert werden können. Wie bei dem aktiven Lernen werden bei den halb-überwachten Methoden weniger gelabelte Trainingsbeispiele als im Standardfall benötigt. Dennoch wird auch hier eine gewisse Menge an Trainingslabels benötigt, was in der vorliegenden Arbeit vermieden werden soll.

Eine weitere Alternative, den Aufwand für die Erstellung der Trainingslabels zu reduzieren, ist, das manuelle Labeln durch Experten zu ersetzen. Beispielsweise werden bei der Klassifikation ganzer Bilder teilweise andere *unabhängige Algorithmen* verwendet, um die Labels der Trainingsdaten zu bestimmen [Xiao et al., 2015]. Durch *Crowd Sourcing*, bei dem anstatt Experten mehrere Laien manuell labeln, können die Kosten für die Erstellung der Trainingslabels

gesenkt werden, auch wenn deren Umfang nicht reduziert wird. In topographischen Anwendungen können auch bereits existierende Karten des Interessensgebietes verwendet werden. Hierbei ist jedoch zu beachten, dass, aufgrund der Zeitdifferenz zwischen der Kartenerstellung und den Sensoraufnahmen, die Labels ggf. fehlerbehaftet sind. In vielen Veröffentlichungen, wie [Jia et al., 2014] und [Büschenfeld, 2013] (s. Kapitel 2.4), wird daher versucht, die fehlerhaften Trainingsbeispiele zu entfernen. Eine Kombination aus Crowd Sourcing und die Verwendung existierender Kartenlabels bildet die Extraktion von Labels aus existierenden Karten, die durch Crowd Sourcing erzeugt wurden, wie Karten aus OpenStreetMap [Kaiser et al., 2017].

In der vorliegenden Arbeit wird eine existierende Karte zur Extraktion der Trainingslabels verwendet. Dabei wird davon ausgegangen, dass ein unbekannter Anteil der Labels fehlerbehaftet sein könnte. Eine gewisse Toleranz gegenüber fehlerhaften Trainingslabels sollte der Klassifikationsalgorithmus demnach aufweisen, sodass das Klassifikationsergebnis die gleiche oder eine ähnlich gute Qualität aufweist wie unter Verwendung manuell gelabelter Trainingsdaten. Verschiedene Strategien, wie Fehler in den Labels modelliert werden können und wie mit ihnen umgegangen werden kann, werden in den folgenden Unterkapiteln beleuchtet.

2.3. Strategien zum Umgang mit fehlerhaften Labels

Frénay und Verleysen [2014] unterscheiden drei mögliche Strategien, um mit fehlerhaften Labels im Training umzugehen.

Die erste Strategie basiert auf der Nutzung von Klassifikatoren, die bereits aufgrund ihrer Konzeption bzw. ihres Aufbaus zu einem gewissen Grad mit fehlerhaften Trainingslabels oder einer bestimmten Art von Labelfehlern umgehen können. Ein typisches Beispiel in diesem Kontext ist der Random Forest Klassifikator [Breiman, 2001] (Kapitel 3.6), der unter anderem durch die Verwendung der Bootstrapping-Prozedur einen gewissen Grad an fehlerhaften Trainingslabels kompensieren kann [Pelletier et al., 2017]. Die Klassifikation mit Entscheidungsbäumen ohne die Verwendung von Bootstrapping ist hingegen anfällig gegenüber fehlerbehafteter Labels im Training [Quinlan, 1986]. Bei einer hohen Anzahl an Trainingsbeispielen mit fehlerhaften Labels und einem nicht zufälligen Auftreten zeigt aber auch der *Random Forest* (RF) Klassifikator eine gewisse Anfälligkeit [Maas et al., 2016]. Folleco et al. [2008] vergleicht das Ergebnis mehrerer Klassifikatoren unter der Verwendung von Trainingsdaten mit asymmetrisch auftretenden Fehlern der Labels. Die Ergebnisse des RF Klassifikators sind zwar deutlich besser als die der anderen Klassifikatoren, ein negativer Einfluss der fehlerhaften Labels, wenn auch ein sehr geringer, ist jedoch auch bei dem RF Klassifikator zu erkennen. Pelletier et al. [2017] vergleichen ebenfalls die Klassifikationsergebnisse verschiedener Klassifikatoren unter Verwendung fehlerhafter Trainingslabels. Neben synthetischen Daten werden auch reale Fernerkundungsdaten klassifiziert, wobei als Klassen verschiedene Landbedeckungsarten genutzt werden. Die Fehler werden mit Hilfe zweier Modelle, eines basierend auf dem reinen Zufall und eines, welches ausschließlich einen Zusammenhang zwischen den Fehlern und den Labels berücksichtigt, synthetisch erstellt. Auch diese Experimente zeigen eine Reduktion der Gesamtgenauigkeit bei steigender Anzahl an fehlerhaften Labels.

Abellán et al. [2017] zeigen, dass der RF Klassifikator zur Überanpassung neigt, auch unter Verwendung von fehlerhaften Trainingslabels. Das etwas schlechtere Ergebnis des RF Klassifikators unter Verwendung von fehlerbehafteten Labels zum Training läge an den Basisklassifikatoren, also den verwendeten Entscheidungsbäumen, da diese bereits zur Überanpassung neigen würden. Aus diesem Grund verwenden Abellán et al. [2017] den *Credal Decision Tree* (CDT) [Abellán, 2013] anstatt der klassischen Entscheidungsbäume. Der hauptsächliche Unterschied der Bäume besteht in dem Kriterium zur Bestimmung der Entscheidungsgrenze in einem Knoten. Die Entscheidung innerhalb des CDT erfolgt durch den *Imprecise Info-Gain* (IIG) [Abellán und Moral, 2003], der Unsicherheitsmessungen auf Credal-Mengen nutzt, wobei Credal-Mengen geschlossene und konvexe Mengen von Wahrscheinlichkeitsverteilungen beschreiben. Experimente an mehreren Datensätzen aus dem Bereich des Maschinellen Lernens zeigen eine Verbesserung des Klassifikationsergebnisses im Vergleich zum klassischen RF Klassifikator, besonders wenn Trainingsdaten mit zufällig auftretenden Fehlern in den Labels genutzt werden [Abellán et al., 2017, 2018]. Die Fehler werden synthetisch eingefügt, indem das Label zufällig ausgewählter Trainingsbeispiele geändert wird. Hierbei werden mehrere Abstufungen bezogen auf die Anzahl an Fehlern in den Labels erzeugt, wobei maximal 30% der Trainingsbeispiele verändert werden. Wie der in der vorliegenden Arbeit vorgestellte RF Klassifikator wird auch in [Abellán et al., 2017] das Kriterium zur Bestimmung der Entscheidungsgrenze innerhalb jedes Entscheidungsbaumes angepasst, um mit Fehlern in den Trainingslabels umzugehen. Die Fehler werden jedoch nicht direkt im Training modelliert. In den Experimenten von Abellán et al. [2017] werden nur zufällig verteilte Fehler in den Trainingslabels simuliert. Ob diese Änderung auch zu einer Verbesserung führt, wenn Trainingslabels, wie in der vorliegenden Arbeit, aus einer veralteten Karte verwendet werden, müsste noch untersucht werden.

Neben dem RF Klassifikator sind auch Deep Learning Methoden in der Lage, mit fehlerhaften Trainingslabels umzugehen. Nach Rolnick et al. [2017] sind beispielsweise Tiefe Neuronale Netze in der Lage, im Training mit einer großen Anzahl an Trainingsdaten mit fehlerhaften Labels umzugehen. Sie untersuchen dies in mehreren Experimenten, bei denen ganze Bilder klassifiziert werden. Die fehlerhaften Labels werden nach unterschiedlichen Modellen simuliert, wobei die Wahl des fehlerhaften Labels entweder zufällig oder ausschließlich abhängig von dem korrekten Label erfolgt. Die Experimente zeigen, dass das Klassifikationsergebnis relativ wenig durch die Verwendung fehlerhafter Labels beeinflusst wird. Dennoch liefert das Training mit fehlerfreien Trainingslabels immer bessere Ergebnisse als mit fehlerbehafteten Daten [Rolnick et al., 2017]. Das oben genannte Modell zur Simulation der fehlerhaften Labels, bei dem die fehlerhaften Trainingsbeispiele zufällig ausgewählt werden, repräsentiert außerdem nicht die Art der Fehler, die auftreten, wenn altes Kartenmaterial zur Bestimmung der Trainingslabels verwendet wird. Laut [Damodaran et al., 2020] sind die Auswirkungen eines geringen Maßes an fehlerhaften Trainingslabels bei der Klassifikation mit neuronalen Netzen durch die Verwendung klassischer Regularisierungstechniken sehr gering. Dies gelte jedoch nicht für eine größere Menge fehlerhafter Labels. Drory et al. [2018] zeigen außerdem, dass die Fähigkeit des Netzwerks, mit fehlerhaften Labels umzugehen, von der Verteilung der fehlerhaften Labels abhängt. Ein geringer Einfluss trete bei einer Verteilung auf, bei der sich eine Vielzahl an korrekten Trainingsbeispielen in der Nachbarschaft eines fehlerbehafteten Trai-

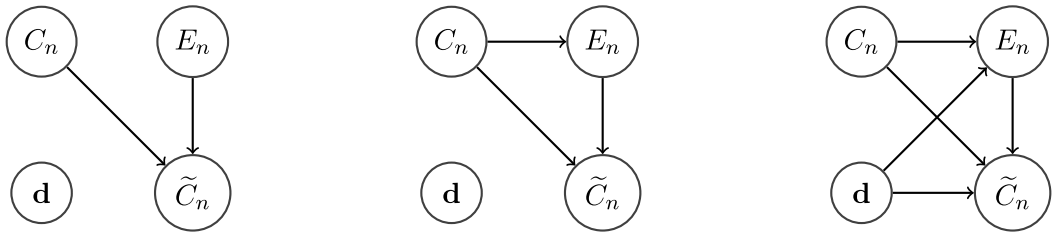
ningsbeispieles befindet, wie bei einer zufälligen Verteilung. Lokal konzentrierte Fehler können jedoch zu einer Verschlechterung des Ergebnisses führen.

In der zweiten Strategie, bekannt unter dem Begriff *Datenbereinigung* (engl. *Data Cleansing*), wird versucht, die Trainingsbeispiele mit fehlerhaften Labels zu erkennen und aus der Menge der Trainingsdaten zu entfernen. Ein Kritikpunkt an dieser Methode ist, dass oft zu viele, und demnach auch korrekte, Trainingsbeispiele entfernt werden [Frénay und Verleysen, 2014]. Da diese Vorgehensweise oft im Bereich der Fernerkundung Verwendung findet, werden entsprechende Beispiele in Kapitel 2.4 behandelt. Ein allgemeineres Beispiel für die Elimination der fehlerhaften Labels bei der Klassifikation von Objekten ist [Zhu et al., 2003], bei dem der Fokus auf der Verarbeitung großer Datensätze liegt. In den Experimenten werden Objekte anhand von zuvor definierten Attributen einer bestimmten Klasse zugeordnet. Die Detektion der Fehler basiert auf der Annahme, dass Trainingsbeispiele mit fehlerhaften Labels in einer Klassifikation, hier mit dem Baum C4.5 [Quinlan, 1993], einem anderen Label zugeordnet werden. Auch hier bleibt die Gefahr, dass Trainingsbeispiele mit korrektem Label als fehlerhaft eingestuft und somit entfernt werden.

Die dritte Strategie verwendet Trainingsmethoden, die robust gegen fehlerhafte Labels in den Trainingsdaten sind, weil das Auftreten dieser Fehler im Trainingsprozess berücksichtigt wird. *Probabilistische Verfahren* lernen meist in der Trainingsphase, simultan mit den Parametern des Klassifikators, die Parameter eines statistischen Fehlermodells. Laut Frénay und Verleysen [2014] gibt es drei verschiedene Arten, das Auftreten von fehlerhaften Labels statistisch zu modellieren, abhängig von den Faktoren, welche die Auftrittwahrscheinlichkeit eines fehlerhaften Labels beeinflussen können. Diese drei Arten werden im folgenden Absatz kurz dargestellt, bevor auf robuste Klassifikatoren, die diese Strategie verwenden, eingegangen wird.

Abb. 1 zeigt die statistischen Zusammenhänge zwischen den Daten $\mathbf{d}$, dem wahren Label C_n, dem Auftreten eines Fehlers E_n und dem potentiell fehlerhaften Label $\widetilde{C}_n$ an einem Primitiv n [Frénay und Verleysen, 2014]. Das komplett auf Zufall beruhende (engl. *Noise Completely At Random*; Abb. 1a; NCAR) Modell ist am einfachsten. Hierbei ist das Auftreten eines Fehlers in einem Label unabhängig von anderen Größen und kann somit durch eine Variable, die Auftrittwahrscheinlichkeit eines Fehlers, für alle Trainigsbeispiele repräsentiert werden. Liegt eine Abhängigkeit zwischen dem Auftreten eines Fehlers in einem Label und der Objektart vor, wird von zufälligen Fehlern (engl. *Noise At Random*; Abb. 1b; NAR) gesprochen. Ein Zusammenhang zwischen dem Auftreten eines Fehlers und den Daten wird dabei nicht modelliert. Parametrisiert wird dieses Modell z.B. durch eine asymmetrische *Übergangsmatrix*, bestehend aus *Übergangswahrscheinlichkeiten*, wie $p(\widetilde{C}|C)$, für jede Klassenkombination. Das nicht auf dem Zufall beruhende Modell (engl. *Noise Not At Random*; Abb. 1c; NNAR) beschreibt den allgemeinsten Fall, bei dem das Auftreten eines Fehlers in einem Label sowohl von der Objektart als auch von den Daten abhängt. Dieses Modell kann durch Übergangswahrscheinlichkeiten für jedes Primitiv beschrieben werden. Die Abhängigkeit zwischen dem Fehler in einem Label und den Daten kann sich auf verschiedene Arten ausprägen. Beigman und Klebanov [2009] beschreiben beispielsweise, dass fehlerhafte Labels häufig bei Trainingsbeispielen auftreten, die ein ähnliches Erscheinungsbild aufweisen wie Trainingsbeispiele einer anderen Klasse. Diese Art von Fehlern tritt besonders im Bereich der Computerlinguistik auf,

beispielsweise wenn Wörter einander phonetisch fast entsprechen [Sarma und Palmer, 2004]. In [Denoeux, 1995] wird hingegen die Distanz zu den Nachbarn im Merkmalsraum verwendet, um die Unsicherheit der Klassenlabels zu beschreiben. Es wird dabei angenommen, dass das Auftreten von Fehlern in Regionen mit einer niedrigen Dichte im Merkmalsraum wahrscheinlicher ist als in anderen.



(a) NCAR (b) NAR (c) NNAR

Abbildung 1: Statistische Zusammenhänge zwischen den Daten $\mathbf{d}$, dem wahren Label C_n, dem Auftreten eines Fehlers E_n und dem potentiell fehlerhaften Label $\widetilde{C}_n$ an einem Primitiv n für die drei Modelle *Noise Completely At Random* (NCAR), *Noise At Random* (NAR) und *Noise Not At Random* (NNAR). Der Zusammenhang zwischen den Daten $\mathbf{d}$ und dem wahren Label C_n ist für eine bessere Übersicht nicht dargestellt [Frénay und Verleysen, 2014].

Bei der Modellierung des Auftretens eines Fehlers in einem Label innerhalb eines Klassifikators wird in Mehrklassenproblemen meist ein NAR Modell angenommen, sodass für jeden Klassenübergang eine individuelle Wahrscheinlichkeit geschätzt wird. Ein Beispiel für solch ein probabilistisches Verfahren stammt von Bootkrajang und Kabán [2012], welche die logistische Regression [Bishop, 2006] erweitern und sie bei der Zuordnung ganzer Bilder zu diskreten Klassen anwenden. Bei der probabilistischen Methode von Li et al. [2007] wird die Kernel Fisher Diskriminante (KFD) [Mika et al., 1999] als Basisklassifikator verwendet. Durch ein dem EM Algorithmus vergleichbarem Verfahren werden die Parameter der Diskriminante und des NAR Fehlermodells geschätzt. Für die Evaluierung werden sowohl Daten durch das Ändern zufälliger Labels simuliert, als auch reale Daten aus dem medizinischen Bereich verwendet. Die Methode bezieht sich jedoch auf ein Zweiklassenproblem. Ein Beispiel einer pixelbasierten Klassifikation von Bildern mit Hilfe eines probabilistischen Verfahrens ist in [Lawrence und Schölkopf, 2001] zu finden. Für die Detektion von Pixeln der Klasse *Himmel* werden hier Bilder verwendet, bei denen bekannt ist, ob in ihnen der Himmel zu sehen ist oder nicht. Als Trainingsbeispiele der Klasse *kein Himmel* dienen die Pixel aus den Bildern ohne Himmel und die Trainingsbeispiele der Klasse *Himmel* bestehen aus allen Pixeln der Bilder mit Himmel. Während so die Pixel aus den Bildern ohne Himmel der korrekten Klasse zugeordnet sind, ist ein Teil der Pixel aus den Bildern mit Himmel fehlerhaft der Klasse *Himmel* zugeordnet. Um dies im Training eines auf der Kernel Fisher Diskriminante basierenden Klassifikators zu berücksichtigen, wird auch hier ein NAR Fehlermodell in die Trainingsprozedur integriert, dessen Parameter durch einen EM-ähnlichen Algorithmus bestimmt werden. Nichtprobabilistische Verfahren, beispielsweise die Methode von An und Liang [2013], basierend auf dem Prinzip des Support Vector Machine (SVM) [Vapnik, 1995] Klassifikators, erlernen im Allgemeinen keine Parameter eines Fehlermodells. Sie sind demnach ungeeignet, wenn diese Wahrscheinlichkeiten weiter verwendet werden sollen, z.B. zur Modellierung temporaler

Übergangswahrscheinlichkeiten.

Auch bei Convolutional Neural Networks (CNN) wird daran geforscht, eine gewisse Toleranz gegenüber fehlerhaften Trainingslabels zu erhalten. Eine Möglichkeit, um mit fehlerhaften Trainingslabels umzugehen, ist die Verwendung einer gegenüber fehlerhaften Trainingslabels robusten *Verlustfunktion* (engl. *loss function*) für das Training eines CNN. Diese Funktion beruht dabei meist auf einem probabilistischen Modell, welches das Auftreten eines Labelfehlers beschreibt [Veit et al., 2017]. Mnih und Hinton [2012] nutzen robuste Verlustfunktionen, um mit Hilfe von Trainingslabels, die aus einer veralteten Karte stammen, tiefe Neuronale Netze für die Detektion von Straßen zu trainieren. Dabei werden zwei Arten von Fehlern modelliert, die auf ihr Anwendungsgebiet zugeschnitten sind: *Omission noise* (dt. *Auslassungs-Fehler*), welcher das Fehlen eines Objektes in der Karte beschreibt, und *Registration noise* (dt. *Registrierungs-Fehler*), der lediglich eine ungenaue Position eines Objekts in einer Karte darstellt. Das Omission noise Modell entspricht dem NAR Fehlermodell, da das Auftreten eines Fehlers lediglich von der Objektart abhängt. Das zweite Modell beinhaltet zusätzlich einen Verschiebungs-Parameter. Ihre Methode befasst sich jedoch nur mit binären Klassifikationsproblemen.

Die in der Literatur verwendeten CNN basierten Methoden sind größtenteils nicht für die Anforderungen der vorliegenden Arbeit geeignet, da eine bestimmte Anzahl an Trainingsbeispielen mit korrekten und somit manuell bestimmten oder kontrollierten Labels benötigt wird [Rolnick et al., 2017]. Diese fehlerfreien Trainingsdaten werden entweder zur Bestimmung der Parameter des Fehlermodells, wie in [Mnih und Hinton, 2012], [Veit et al., 2017] und [Vahdat, 2017], oder zur Feinabstimmung des neuronalen Netzwerks [Maggiori et al., 2017] genutzt. In der vorliegenden Arbeit stehen solch fehlerfreie Trainingsdaten nicht zur Verfügung, da vor der Klassifikation nicht bekannt ist, welche Bereiche der Karte noch aktuell sind.

Patrini et al. [2017] und Damodaran et al. [2020] bilden hierbei eine Ausnahme, da die beschriebene Methodik ohne eine fehlerfrei gelabelte Teilmenge an Trainingsdaten auskommt. Patrini et al. [2017] stellen zwei Prozeduren vor, in denen die Verlustfunktion mit Hilfe von Übergangswahrscheinlichkeiten an das Aufkommen fehlerbehafteter Trainingslabels angepasst wird. In den Experimenten erfolgt jedoch keine pixelbasierte Klassifikation. Damodaran et al. [2020] analysieren iterativ die geometrische Struktur der Daten im Merkmalsraum für die Bestimmung der fehlerhaften Labels. Obwohl bei der Simulation der fehlerhaften Labels in den Experimenten das Aussehen der Daten von Bedeutung ist, weil ein Trainingsbeispiel einer anderen Klasse mit ähnlichem Aussehen zugeordnet wird, enthält die Fehlermodellierung keinen Zusammenhang zwischen dem Auftreten eines Labelfehlers und den Daten. Sowohl in Patrini et al. [2017] als auch in Damodaran et al. [2020] wird demnach das NAR Modell verwendet, um das Auftreten der Fehler in den Labels zu beschreiben. Werden die Trainingslabels aus einer veralteten Karte bestimmt, hängt das Auftreten von fehlerhaften Labels von der Objektart und den Daten ab, sodass das NNAR Modell greift.

An dieser Stelle soll kurz auf die Terminologie eingegangen werden. Die Wörter *robust* und *tolerant* werden in der Literatur nicht einheitlich verwendet. Frénay und Verleysen [2014]

bezeichnen Klassifikatoren der ersten Strategie, die bereits aufgrund ihres Aufbaus mit fehlerhaften Trainingslabels umgehen können, als *robust*. Klassifikatoren der dritten Strategie, in denen das Auftreten von Fehlern in den Labels im Training berücksichtigt wird, als *tolerant*. Bootkrajang und Kabán [2012] nennen den von Ihnen entwickelten Klassifikator, welcher durch die Erweiterung der logistischen Regression der dritten Kategorie zugeordnet werden kann, „Label-noise Robust Logistic Regression“. Auch in [Damodaran et al., 2020] wird *robust* im Kontext mit angepassten Verfahren genannt. In der vorliegenden Arbeit wird, aufgrund der Verwendung des Klassifikators von Bootkrajang und Kabán [2012], ein Klassifikator als *robust* bezeichnet, wenn das Auftreten von Fehlern in dem Trainingsprozess berücksichtigt und somit die dritte Strategie verwendet wird.

2.4. Fehlerhafte Trainingslabels im Bereich der Fernerkundung

Im Bereich der Fernerkundung wird meist die Strategie der Datenbereinigung verwendet, um mit fehlerbehafteten Trainingslabels umzugehen. Radoux et al. [2014] zeigen beispielsweise zwei Methoden zur Eliminierung von Ausreißern, um die Trainingslabels aus einer vorhandenen Karte zu extrahieren. Während die eine Methode Trainingsdaten in der Nähe der Grenzen zwischen den verschiedenen Landbedeckungsarten entfernt, nimmt die zweite eine Gaußverteilung der spektralen Signatur an, detektiert unter der Verwendung von statistischen Tests Ausreißer und eliminiert diese. Da die Methoden für niedrig aufgelöste Daten von 300 m Bodenpixelgröße (GSD, engl. Ground Sampling Distance) entworfen wurden, können diese Annahmen nicht einfach auf hochaufgelöste Daten angewandt werden. Eine ähnliche Methode zur Aktualisierung von Forstkarten wird in [Radoux und Defourny, 2010] vorgestellt. Die Ausreißerdetektion erfolgt hier auf Basis eines Schwellwertes, der aus den Wahrscheinlichkeitsdichten, die wiederum mit Hilfe der Kerndichteschätzung angenähert werden können, bestimmt wird. Der Nachteil dieser Methode besteht darin, dass der Schwellwert zuvor von dem Nutzer definiert werden muss und einen großen Einfluss auf das Ergebnis haben kann. Während bei zu kleinen Werten Fehler möglicherweise übersehen werden, besteht bei der Verwendung zu großer Werte die Möglichkeit, zu viele korrekte Trainingsbeispiele zu entfernen.

Jia et al. [2014] entwickeln einen Algorithmus zur Aktualisierung von Karten der Bodenbedeckung durch die Klassifikation von Landsat Bildern unter der Verwendung der Labels aus einer veralteten Karte zum Training. Bei dem Algorithmus werden alternierend die Fernerkundungsdaten klassifiziert und die Trainingsdaten, deren Label als *geändert* detektiert wurden, entfernt. In dem Algorithmus wird demnach iterativ versucht, fehlerhafte Trainingsdaten zu erkennen und zu entfernen (Datenbereinigung). Die Klassifikation erfolgt mit Hilfe eines Markov Random Fields (MRF), welches die a Posteriori Wahrscheinlichkeiten einer zuvor durchgeführten Maximum Likelihood Klassifikation verwendet. Die anschließende Änderungsdetektion erfolgt durch den Vergleich des Klassifikationsergebnisses mit dem der letzten Iteration. Unterscheiden sich die Labels eines Trainingsbeispieles wird dieses als *geändert* detektiert und aus der Menge an Trainingslabels entfernt. Die Initialisierung der Trainingslabels erfolgt mit Hilfe der veralteten Karte. Die Experimente zeigen gute Ergebnisse, wenn die Karte eine hohe

Qualität aufweist und nur wenige Änderungen der Bodenbedeckung auftreten. Bei einer größeren Fehlermenge kommt es bereits bei der Initialisierung vermehrt zu Fehlklassifikationen, sodass viele Trainingsbeispiele mit korrektem Label aus der Menge an Trainingsbeispielen entfernt und somit in der nächsten Iteration nicht weiter verwendet werden.

Der in [Büschenfeld, 2013] beschriebene Algorithmus zur Datenbereinigung, basierend auf SVM, entfernt ebenfalls Trainingsbeispiele, deren Klassifikationsergebnis sich von dem Kartenlabel unterscheiden. Durch paarweise Koppelung (engl. pairwise coupling) [Wu et al., 2004] wird zusätzlich für jedes Trainingsbeispiel die Klassifikationsunsicherheit geschätzt [Büschenfeld und Ostermann, 2012]. Trainingsbeispiele mit einer hohen Unsicherheit werden entfernt. Anwendung findet der Algorithmus bei der Klassifikation von Ikonos Bildern, wobei das Training mit Hilfe von Trainingsdaten, deren Labels aus einem Geographischen Informationssystem (GIS) stammen, erfolgt [Büschenfeld und Ostermann, 2012]. Wie bei anderen Algorithmen zur Datenbereinigung besteht auch hier das Problem, dass ggf. zu viele Trainingsbeispiele entfernt werden.

Bruzzone und Persello [2009] nutzen im Training eines kontextsensitiven halbüberwachten SVM Klassifikators die Pixel in einer lokalen Nachbarschaft eines Trainingsbeispieles, um Robustheit gegenüber fehlerhaften Trainingslabels zu erlangen. Hierfür erfolgt zunächst das Training eines klassischen SVM Klassifikators mit den gegebenen fehlerbehafteten Trainingsdaten. Mit dem Klassifikator werden anschließend Trainingsbeispiele, die noch kein Label besitzen, klassifiziert. Die neu klassifizierten und die bereits gelabelten fehlerbehafteten Trainingsbeispiele werden anschließend für das Training eines kontextbasierten SVM Klassifikators verwendet. Dessen Kostenfunktion enthält einen zusätzlichen Term, um die Auswirkungen potentiell falsch gelabelter Trainingsbeispiele, bei denen also das Klassenlabel nicht mit den Labels der Nachbarpixel übereinstimmt, zu reduzieren. Die Evaluierung erfolgt durch die Klassifikation von Ikonos und Landsat Bildern. Die Fehler in den Trainingslabels werden auf zwei verschiedene Arten simuliert: zufällig (NCAR) und klassenabhängig (NAR). Obwohl laut Bruzzone und Persello [2009] diese Methodik auch genutzt werden kann, wenn die Trainingslabels aus einer bereits existierenden Karte stammen, wird dies nicht in den Experimenten getestet oder weiter ausgeführt. Außerdem handelt es sich um ein nicht-probabilistisches Verfahren, sodass kein probabilistisches Fehlermodell, z.B. in Form von Übergangswahrscheinlichkeiten, geschätzt wird. Ferner basiert die Methode auf der Annahme, dass nur wenig gelabelte Daten vorliegen. In der vorliegenden Arbeit ist jedoch eine große Anzahl an Trainingsdaten vorhanden, die bereits ein Label besitzen. Da die Label aus der Karte stammen, wird davon ausgegangen, dass die Labelfehler räumlich korreliert sind. Die Detektion von Labelfehlern mit Hilfe des räumlichen Kontextes könnte somit problematisch sein.

Ein Großteil der Veröffentlichungen aus dem Bereich des Deep Learning befasst sich nicht mit der Klassifikation von Fernerkundungsdaten, sondern beispielsweise mit der Zuordnung ganzer Bilder zu einer bestimmten Klasse. Eine Ausnahme bildet die Veröffentlichung von Damodaran et al. [2020], in der das Auftreten fehlerhafter Trainingslabels durch eine Übergangsmatrix modelliert wird. Es wird demnach angenommen, dass das Auftreten der Fehler ausschließlich von dem Label abhängt und nicht von den Daten (NAR). Neben anderer Klassifikationsaufgaben werden in den Experimenten auch hyperspektrale Fernerkundungsbilder

pixelbasiert klassifiziert. Die Fehler in den Labels werden durch das Ändern einiger Labels in der Referenz, entsprechend eines zuvor definierten NAR Modells, simuliert. Wie bei anderen Methoden, in denen das NAR Modell verwendet wird, können Fehler in den Labels, die nicht nur von der Objektklasse abhängen, nicht korrekt widergespiegelt werden und somit das Klassifikationsergebnis unter Umständen verschlechtern.

In [Kaiser et al., 2017] erfolgt eine pixelbasierte Klassifikation unter Verwendung von Trainingsdaten, deren Labels aus Online Karten, wie OpenStreetMap, extrahiert wurden und somit auch fehlerbehaftet sind. Der verwendete CNN-Klassifikator wird leicht modifiziert, damit feinere Strukturen im Bild erhalten bleiben. Eine Anpassung, um mit fehlerhaften Trainingslabels umzugehen, erfolgt nicht. Es wird festgestellt, dass die Hinzunahme von ggf. fehlerbehafteten Trainingsdaten, deren Labels aus Online Karten stammen, zu fehlerfreien Trainingsdaten das Klassifikationsergebnis steigern kann und selbst die alleinige Nutzung der fehlerhaften Trainingsdaten zu akzeptablen Ergebnissen führt. Die Autoren erwähnen jedoch, dass dieses Ergebnis keineswegs optimal ist und zeigen, dass die Hinzunahme von korrekten Trainingsdaten ein besseres Ergebnis liefert. Die Hinzunahme von vollständig korrekten Trainingsdaten soll in der vorliegenden Arbeit jedoch vermieden werden.

Ein weiteres Beispiel für die Verwendung von Fernerkundungsdaten im Bereich des Deep Learning ist das von Maggiori et al. [2017] vorgestellte *rekurrente Neuronale Netz* (RNN, engl. *recurrent neuronal network*) zur Verbesserung vorhandener topographischer Karten. Da für die Feinabstimmung des Netzwerks ein Satz an korrekt gelabelten Trainingsdaten benötigt wird, entspricht die Methode nicht der Anforderung der vorliegenden Arbeit, vollständig auf manuelles Labeln zu verzichten.

2.5. Merkmalsselektion

Die Klassifikation von Bildern basiert auf Merkmalen, welche die Eigenschaften der Bildprimitive beschreiben. Die Wahl der Merkmale ist für das Klassifikationsergebnis entscheidend, da eine gute Trennung der Klassen nur möglich ist, wenn sich auch die Merkmale der Bildprimitive unterschiedlicher Klassen unterscheiden. Die Anzahl an Merkmalen, die aus den Sensordaten bestimmt werden können, ist jedoch meist zu groß, als dass alle in die Klassifikation mit einfließen können. Ein limitierender Faktor, besonders bei großen Datensätzen, ist unter dem Begriff *Fluch der Dimensionalität* (engl. *Curse of Dimensionality*, [Bellman, 1961]) bekannt [Frénay et al., 2014]. Dieser besagt, dass bei Hinzunahme weiterer Merkmale das Volumen des Merkmalsraumes stark ansteigt. Daran angeknüpft ist das Hughes-Phänomen [Hughes, 1968], welches besagt, dass die Hinzunahme von Merkmalen bei gleicher Anzahl an Trainingsbeispielen ab einem Punkt zu einer Verschlechterung der Klassifikationsgenauigkeit führt. Hinzu kommen Limitierungen aufgrund des Speicherplatzes und der Rechenzeit. Es ist demnach ratsam, die Anzahl an Merkmalen sinnvoll zu reduzieren, indem z.B. Merkmale, die für die Aufteilung der Klassen irrelevant sind oder nur redundante Informationen enthalten, entfernt werden. Mit Deep Learning Methoden können Merkmale auch mithilfe von Trainingsdaten erlernt werden.

Bei der Reduktion der Merkmale werden zwei Möglichkeiten unterschieden: Die Merkmalsselektion und Merkmalsextraktion. Bei der Selektion bleiben die ursprünglichen Merkmale erhalten, während bei der Extraktion neue Merkmale erstellt werden, welche die ursprünglichen Merkmale kombinieren. Der Nachteil der Merkmalsextraktion ist, dass die ursprünglichen Merkmale, die in vielen industriellen und medizinischen Anwendungen von entscheidender Bedeutung sind, verloren gehen [Frénay et al., 2014]. Die kombinierten Merkmale sind außerdem in der Regel nicht interpretierbar [Khalid et al., 2014]. Die Anzahl an Merkmalen muss zudem meist fest definiert werden. In der vorliegenden Arbeit wird daher ein Selektionsverfahren angewandt.

Im Hinblick auf die Selektionsmethoden wird zwischen Filter-, Wrapper-, eingebetteten und hybriden Methoden unterschieden [He et al., 2015]. Filtermethoden bewerten die Merkmale basierend auf intrinsischen Eigenschaften der Daten, ohne Berücksichtigung des Trainingsverfahrens einer Klassifikationsmethode. Die Vorteile liegen hierbei in den relativ kurzen Rechenzeiten sowie darin, dass in der Regel keine überwachten Klassifikatoren verwendet und somit auch keine Trainingsdaten benötigt werden. Der Nachteil ist, dass die selektierten Merkmale ggf. nicht zu der Trainingsmethode der verwendeten Klassifikationsmethode passen. Wrapper-Methoden berücksichtigen das Lernverfahren und bewerten die Merkmale abhängig von diesem. Bei eingebetteten Methoden erfolgt die Merkmalsselektion während des Lernprozesses und hybride Methoden kombinieren Filter- und Wrapper-Methoden.

Während das Auftreten von fehlerhaften Labels in den Trainingsdaten einer Klassifikation ein bekanntes Problem ist (s. Kapitel 2.3 und 2.4), ist das Thema im Bereich der Merkmalsselektion noch kaum untersucht worden. Die Notwendigkeit, sich auch in der Merkmalsselektion mit fehlerhaften Labels zu befassen, ist jedoch gegeben. Laut Frénay et al. [2014] ist anzunehmen, dass fehlerhafte Trainingslabels die Auswahl von Merkmalen mit überwachten Merkmalsselektionsverfahren verschlechtert, da diese auch das Ergebnis von überwachten Klassifikationsmethoden negativ beeinflussen.

Im Bereich der Medizin untersucht Zhang et al. [2005] den Einfluss fehlerhaft gelabelter Daten auf die Auswahl von Genen. Ist bereits eines der 30 Trainingsbeispiele falsch gelabelt, werden bereits etwa 20% der relevanten Gene nicht als solche erkannt. Den Einfluss fehlerhafter Labels auf filter- bzw. schwellwertbasierte Ranking-Methoden für die Auswahl relevanter Gene wurde in [Shanab et al., 2012] untersucht. Hierfür werden 11 verschiedene Methoden an vier Datensätzen getestet, bei denen fehlerhafte Labels in 9 verschiedenen Abstufungen simuliert werden. Die Experimente zeigen, dass die Stabilität des Rankings durch das Auftreten fehlerhafter Labels verschlechtert wird. Die Transinformation als Rankingkriterium ist laut Shanab et al. [2012] im Vergleich zu einigen anderen getesteten Methoden nicht sehr empfindlich gegenüber dem Auftreten fehlerhafter Labels. Diese starke Unempfindlichkeit ist mit ein Grund, warum eine Merkmalsselektionsmethode basierend auf der Transinformation in der vorliegenden Arbeit verwendet wird.

Frénay et al. [2014] untersuchen ebenfalls den Einfluss zufälliger fehlerhafter Labels bei der Merkmalsselektion auf Basis der Transinformation, jedoch an 11 Datensätzen für unterschiedliche Klassifikationsaufgaben. Die Autoren beobachten eine starke Beeinflussung des Klas-

sifikationsergebnisses und führen dies auf die Schätzung der Transinformation aus den k-nächsten-Nachbarn im Merkmalsraum zurück. Aus diesem Grund wird die Berechnung der Transinformation in [Frénay et al., 2014] angepasst, indem ein ausschließlich auf den Klassen basierendes Fehlermodell (NAR) integriert wird. He et al. [2015] stellen eine gegenüber fehlerhaften Trainingslabels robuste Selektionsmethode für die Verarbeitung von Omics-Daten vor. Der Begriff Omics beschreibt Teilgebiete der Biologie, wie Genomics, Transcriptomics, Metabolomics und Proteomics, die sich mit der Analyse von Gesamtheiten ähnlicher Einzelelemente befassen [Gey, 2015]. Die Methode basiert auf einem *twin boosting* Algorithmus, welcher aus zwei Stufen des Boostings, bei dem mehrere schwache Klassifikatoren zu einem starken Klassifikator kombiniert werden, besteht. Um Robustheit gegenüber fehlerhaften Trainingslabels zu erlangen, wird als schwacher Klassifikator ein robuster Klassifikator, der RCLLS (engl. *robust componentwise linear least squares*), verwendet. Dieser beinhaltet einen iterativen Algorithmus der gewichteten kleinsten Quadrate, welcher, durch die schlechtere Gewichtung von potentiell fehlerbehafteten Trainingsdaten, als eine Methode zur schwachen Datenbereinigung angesehen werden kann. Außerdem wird eine gegenüber fehlerhaften Trainingslabels robuste Verlustfunktion genutzt. Es wird dabei von fehlerhaften Trainingslabels in der Nähe der Entscheidungsgrenze im Merkmalsraum ausgegangen, was durch ein entsprechendes statistisches Modell in der Verlustfunktion beschrieben wird. Werden die Trainingslabels aus einer bereits existierenden Karte extrahiert, kann nicht davon ausgegangen werden, dass sich die Trainingsbeispiele mit fehlerhaftem Label in der Nähe der Entscheidungsgrenze im Merkmalsraum befinden. Die Auswirkungen der fehlerhaften Labels auf das Selektionsergebnis werden somit voraussichtlich nicht verringert.

2.6. Diskussion

In der Literatur finden sich viele Verfahren, um Änderungen über die Zeit zu detektieren, wobei nicht alle für die Aktualisierung veralteter Karten geeignet sind. Viele Methoden vergleichen beispielsweise die Sensordaten zweier Zeitpunkte direkt, z.B. [Subudhi et al., 2014], wodurch zwar das Auftreten einer Änderung bestimmt werden kann, nicht aber die Art der Änderung [Lu et al., 2004]. Außerdem sind Sensordaten zum Zeitpunkt der Kartenerstellung nicht immer gegeben. Eine weitere Strategie für die Änderungsdetektion ist der Vergleich der Ergebnisse unabhängig durchgeführter Klassifikationen von Fernerkundungsdaten zweier Zeitpunkte, z.B. in [Stefanski et al., 2014]. Durch die vorangegangene Klassifikation kann hier auch die Art der Änderung bestimmt werden. Fehler in den Klassifikationsergebnissen haben jedoch einen direkten Einfluss auf das Ergebnis der Änderungsdetektion. Außerdem werden, wenn eine überwachte Klassifikation verwendet wird, Trainingsdaten für jeden Zeitpunkt benötigt. Verwendung findet diese Methode insbesondere dann, wenn die Objektklassen zu einem Zeitpunkt bereits bekannt sind, z.B. in Form einer veralteten Karte. In diesem Fall werden die Fernerkundungsdaten von einem Zeitpunkt klassifiziert und das Ergebnis mit der Karte verglichen.

Stehen Daten zu mehreren Zeitpunkten zur Verfügung, kann eine simultane Klassifikation aller Zeitpunkte unter Berücksichtigung aller gegebenen Daten von Vorteil sein, da in den anderen

Strategien lediglich zwei Zeitpunkte miteinander verglichen werden können. Eine Möglichkeit, die Abhängigkeiten der Labels zwischen den Zeitpunkten probabilistisch zu modellieren, bilden hierbei Übergangswahrscheinlichkeiten. Diese werden meist empirisch vor der Klassifikation bestimmt [Hoberg et al., 2015] oder stammen aus Expertenwissen [Melgani und Serpico, 2003]. Auch aus den Sensordaten zweier Zeitpunkte kann eine Übergangswahrscheinlichkeit abgeleitet werden [Bruzzone et al., 1999]. In der vorliegenden Arbeit werden die Übergangswahrscheinlichkeiten in einem sequenziellem Training bestimmt. Wie in [Hoberg et al., 2015] wird ein multitemporales CRF verwendet, um den temporalen Zusammenhang in der Klassifikation zu berücksichtigen. Mit einer individuellen Gewichtung der Übergangswahrscheinlichkeiten bei jedem zu klassifizierendem Primitiv soll jedoch der temporalen Überanpassung, wie in [Hoberg et al., 2015] zu beobachten ist, entgegen gewirkt werden.

Wird bei der Änderungsdetektion für die Klassifikation auf überwachte Methoden zurückgegriffen, werden Daten zum Training des Klassifikators benötigt. Stammen die Labels der Trainingsdaten, wie in der vorliegenden Arbeit, aus einer veralteten Karte, sind diese, aufgrund der Zeitdifferenz zwischen der Kartenerstellung und der Aufnahme der Daten, meist fehlerbehaftet. Diese Fehler in den Labels werden in der Literatur oft *Label Noise* oder *Class Noise* genannt [Frénay und Verleysen, 2014].

Es existieren verschiedene Ansätze mit fehlerhaften Trainingslabels umzugehen. Im Bereich der Kartenaktualisierung wird oft auf Methoden der Datenbereinigung zurückgegriffen, bei denen Trainingsbeispiele mit fehlerhaften Labels vor dem Training entfernt werden sollen. Ein Nachteil dieser Methode ist jedoch, dass sie oft zu viele, und somit auch korrekte Trainingsbeispiele eliminiert [Frénay und Verleysen, 2014]. In der vorliegenden Arbeit wird daher von dem Entfernen von Trainingsbeispielen abgesehen. Stattdessen wird ein Trainingsbeispiel, anstelle zu einer bestimmten Klasse, jeder Klasse mit einer bestimmten Wahrscheinlichkeit zugeordnet. Diese Wahrscheinlichkeiten werden iterativ angepasst. Der Vorteil ist, dass ein Trainingsbeispiel, dessen Label sich potentiell von dem Label aus der Karte unterscheidet, nicht gleich herausgenommen werden muss. Es werden hingegen die Wahrscheinlichkeiten für die Zugehörigkeit zu den einzelnen Klassen angepasst. Besonders wenn sich der Algorithmus bei der Detektion einer Änderung relativ unsicher ist, kann dies durch die Verwendung der Wahrscheinlichkeiten modelliert werden, während bei der Datenbereinigung bereits eine Entscheidung für oder gegen eine Änderung getroffen werden muss.

Einige Klassifikatoren, wie der Random Forest oder Neuronale Netze, gelten bereits aufgrund ihres Aufbaus als immun gegenüber fehlerhaften Trainingslabels. Experimente zeigten jedoch, dass dies im Allgemeinen nur für eine kleine Anzahl an fehlerhaften Trainingslabels oder eine bestimmte Art an Fehlern zutrifft. Der Random Forest Klassifikator kann beispielsweise, aufgrund des Bootstrapping Prinzips, gut mit einem gewissen Grad an Fehlern umgehen [Pelletier et al., 2017]. Bei einer größeren Anzahl an Fehlern zeigt der Random Forest auch eine gewisse Anfälligkeit [Maas et al., 2016; Folleco et al., 2008; Pelletier et al., 2017]. Eine entsprechende Anpassung des Random Forest Klassifikators, mit dem Ziel Trainingslabels aus einer veralteten Karte zum Training verwenden zu können, ist noch nicht behandelt worden. In der vorliegenden Arbeit wird daher die Trainingsprozedur des Random Forest Klassifikators

durch ein komplexes Fehlermodell erweitert. Fehler, die bei der Verwendung von Kartenlabels im Training auftreten und nicht durch das Bootstraping kompensiert werden können, sollen dadurch im Training berücksichtigt werden und somit einen geringeren Einfluss auf das Klassifikationsergebnis ausüben.

Trainingsprozeduren, die das Auftreten von Fehlern in den Labels, z.B. in Form eines probabilistischen Fehlermodells, berücksichtigen, finden oft im Bereich der Computer Vision Verwendung. Probabilistische Modelle bestehen meist aus Übergangswahrscheinlichkeiten für jede Klassenkombination [Bootkrajang und Kabán, 2012; Li et al., 2007]. Da in solch einem NAR Modell [Frénay und Verleysen, 2014] lediglich ein Zusammenhang zwischen dem Auftreten der Fehler und den Objektklassen berücksichtigt werden kann, wird der Einfluss von Fehlern in den Labels, deren Auftreten auch von den Daten abhängt, ggf. nicht oder nur wenig reduziert. Stammen die Labels für das Training aus einer veralteten Karte, entstehen die Fehler durch Änderungen von Objekten oder Objektteilen und dementsprechend bilden die Fehler Segmente im Bildraum. Dies ist einer der wesentlichen Ansatzpunkte der vorliegenden Arbeit. Es wird ein komplexes Fehlermodell entwickelt, welches die Tatsache, dass Fehler in Segmenten auftreten, berücksichtigt. Dies geschieht bei der Berechnung der Wahrscheinlichkeiten für die Klassenzugehörigkeit bei jedem Trainingsbeispiel, welche wiederum in das Training des Random Forest Klassifikators einfließen.

Für die Erweiterung der Trainingsprozedur des Random Forest Klassifikators durch ein komplexes Fehlermodell wird in der vorliegenden Arbeit das Training der in dem Random Forest enthaltenen Entscheidungsbäume so angepasst, dass es mit Trainingsbeispielen, die, wie zuvor beschrieben, nicht einer bestimmten Klasse, sondern jeder Klasse mit einer bestimmten Wahrscheinlichkeit zugeordnet sind, erfolgen kann. Diese Wahrscheinlichkeit der Klassenzugehörigkeit hängt sowohl von den Daten, als auch den beobachteten Labels, welche ggf. fehlerhaft sind, ab. Die Abhängigkeit zu den beobachteten Labels wird durch Übergangswahrscheinlichkeiten $p(\widetilde{C}_n|C_n)$ modelliert, die individuell für jedes Trainingsbeispiel n mit Hilfe von Heuristiken entwickelt werden. Besonders die Tatsache, dass Trainingsbeispiele mit fehlerhaftem Label Segmente im Bildraum bilden, fließt in die Bestimmung der individuellen Übergangswahrscheinlichkeiten mit ein. Da bei der Bestimmung der Wahrscheinlichkeiten der Klassenzugehörigkeit die Daten und die individuellen Übergangswahrscheinlichkeiten berücksichtigt werden, kann die hier vorgestellte Methode zur Berücksichtigung von Fehlern in den Trainingslabels als Modellierung der Fehler entsprechend dem NNAR Modell [Frénay und Verleysen, 2014] interpretiert werden.

Abgesehen von der Bestimmung von Labels für das Training kann die Karte auch bei der Klassifikation selber genutzt werden. In Bereichen, in denen der Klassifikator fehleranfällig ist, z.B. bei Mischpixeln oder in Schattenbereichen, können die Kartenlabels helfen, das korrekte Label zu bestimmen. Nach bestem Wissen der Autorin werden bisher die Kartenlabels fast ausschließlich nach der Klassifikation genutzt, z.B. zur Detektion von Änderungen. Potentiell fehlerhaft klassifizierte Bereiche werden dann z.B. bei der Änderungsdetektion berücksichtigt [Olsen et al., 2002; Knudsen und Olsen, 2003]. Wie die Kartenlabels schon während der Klassifikation genutzt werden können, ist bisher weitgehend ungelöst. Schistad Solberg et al.

[1996] bilden eine Ausnahme, da sie die Kartenlabels als weitere Beobachtung nutzen. Hierfür werden Markov Random Fields verwendet. Der statistische Zusammenhang zwischen den Kartenlabels und den unbekannten Labels wird mithilfe von Übergangswahrscheinlichkeiten modelliert. Das Auftreten einer bestimmten Änderung der Klasse zwischen dem Zeitpunkt der Kartenerstellung und dem Zeitpunkt der Bildaufnahmen ist demnach in dem verwendeten Modell nur von dem Klassenlabel abhängig. Obwohl die Übergangswahrscheinlichkeiten mithilfe eines Parameters gewichtet werden kann, ist dieser Parameter unabhägig von dem betroffenen Pixel und somit unabhängig davon, ob eine Änderung vorliegt oder nicht. Bei der multitemporalen Klassifikation von Sensordaten mehrerer Zeitpunkte werden temporale Zusammenhänge in einigen Veröffentlichungen modelliert, z.B. durch Übergangswahrscheinlichkeiten [Hoberg et al., 2015]. Obwohl in vielen Methoden zum Umgang mit Fehlern in den Labels Übergangswahrscheinlichkeiten, welche, wie in [Schistad Solberg et al., 1996], den statistischen Zusammenhang zwischen den Labels der Karte und den unbekannten zu bestimmenden Labels beschreiben können, geschätzt werden [Bootkrajang und Kabán, 2012; Li et al., 2007; Mnih und Hinton, 2012], gehen sie nicht nach dem Training in die Klassifikation mit ein. In der vorliegenden Arbeit wird der Zusammenhang zwischen den beobachteten Labels der Karte und den zu bestimmenden Labels mit Hilfe eines erweiteren CRFs in dem Klassifikationsprozess berücksichtigt. Die Modellierung des statistischen Zusammenhanges zwischen den Kartenlabels und den zu klassifizierenden Labels erfolgt mit Hilfe von Übergangswahrscheinlichkeiten. Anders wie in [Schistad Solberg et al., 1996] ist der Einfluss des Kartenlabels so modelliert, dass er in Bereichen die sich potentiell nicht geändert haben größer ist als in Bereichen mit potentieller Änderung.

Die Neuerungen der hier vorliegenden Arbeit führen dazu, dass das Training mit den Labels einer veralteten Karte anstelle mit manuell erstellten Labels erfolgen kann und dass die Kartenlabels in der Klassifikation berücksichtigt werden. Auch in der multitemporalen Klassifikation von Sensordaten mehrerer Zeitpunkte wurde die Verwendung einer veralteten Karte zum Training und zur Klassifikation kaum behandelt. Die meisten Veröffentlichungen, wie [Bruzzone et al., 1999; Hoberg et al., 2015; Leite et al., 2011; Melgani und Serpico, 2003], fokussieren sich auf die gemeinsame Klassifikation der gegebenen Fernerkundungsdaten und die darin enthaltene Berücksichtigung von temporalen Zusammenhängen. Karten werden nicht verwendet. Die Übergangswahrscheinlichkeiten, mit denen der statistische Zusammenhang zwischen den Labels temporaler Nachbarn beschrieben wird, werden außerdem meist empirisch vor der Klassifikation bestimmt [Hoberg et al., 2015] oder basieren auf Expertenwissen [Melgani und Serpico, 2003]. In der vorliegenden Arbeit wurde für die multitemporale Klassifikation ein erweitertes CRF, basierend auf dem multitemporalen CRF aus Hoberg et al. [2015], entwickelt. Die Unterschiede zu dem CRF aus Hoberg et al. [2015] werden im Folgenden kurz dargestellt. Zum einen werden die Übergangswahrscheinlichkeiten nicht empirisch bestimmt, sondern aus den Daten und der veralteten Karte gelernt. Außerdem erfolgt eine Gewichtung bzw. Anpassung der Übergangswahrscheinlichkeiten bei jedem Pixel abhängig davon, ob sich das Pixel in einer Region mit potentiellen Änderungen befindet oder nicht. Dies soll einer temporalen Überanpassung, wie sie in [Hoberg et al., 2015] aufgetreten ist, entgegenwirken. Ein weiterer Unterschied ist, dass in dem CRF der vorliegenden Arbeit die Kartenlabels als zusätzliche Beobachtungen integriert sind. Auch werden in [Hoberg et al., 2015] korrekte Trainingslabels

für jeden Zeitpunkt benötigt. In der vorliegenden Arbeit erfolgt das Training stattdessen mit den Labels der Klassifikation des vorherigen Zeitpunktes bzw. bei der Klassifikation des ersten Zeitpunktes mit den Kartenlabels.

Für die Merkmalsselektion ist eine Selektionsmethode, die unabhängig von einem bestimmten Klassifikator ist, sinnvoll, da in der vorliegenden Arbeit unterschiedliche Methoden für die Klassifikation genutzt werden. Außerdem stammen die Labels für das Training aus einer veralteten Karte, sodass einige ggf. fehlerhaft sind. Eine Selektionsmethode, die wenig durch fehlerhafte Trainingslabels beeinflusst wird, ist demnach sinnvoll. Obwohl der Schwerpunkt der vorliegenden Arbeit auf der Klassifikation liegt, soll dennoch gezeigt werden, dass der Prozess der Klassifikation ohne manuell gelabelte Daten auskommt. Aus diesem Grund wird für die Auswahl der Merkmale auf die Selektionsmethode aus [Frénay et al., 2014] mit Hilfe der Transinformation zurückgegriffen. Die vorgeschlagene Erweiterung zur Verringerung des Einflusses fehlerhafter Labels wird nicht genutzt, da das dort enthaltene Fehlermodell nicht die Verteilung der Fehler in den Labels wiederspiegelt, die entstehen, wenn die Labels aus einer veralteten Karte extrahiert werden.

Die Darstellung der aktuellen Forschungsarbeiten zeigt, dass bei der Aktualisierung einer Karte durch die Klassifikation aktueller Fernerkundungsbilder die Labels aus der Karte zwar in manchen Arbeiten im Training verwendet werden, die dadurch entstehenden Fehler jedoch kaum analysiert werden. Meist wird versucht, die fehlerhaften Labels zu entfernen, ohne das Auftreten der Fehler konkret zu modellieren. Die Methoden, die das Auftreten von Fehlern in den Labels explizit in den Trainingsprozess integrieren, stammen meist aus dem Bereich der Computer Vision und verwenden ein vereinfachtes Fehlermodell. In diesem Fehlermodell werden lediglich Abhängigkeiten zwischen den Fehlern und den Objektklassen berücksichtigt. In der vorliegenden Arbeit wird hingegen ein komplexeres Fehlermodell in den Random Forest Klassifikator integriert, welches berücksichtigt, dass fehlerhafte Trainingslabels in Segmenten auftreten. Zusätzlich wird der statistische Zusammenhang zwischen den Labels der Karte und den zu ermittelnden Labels, der nach dem Training meist ungenutzt bleibt, in den Klassifikationsprozess mit Hilfe eines CRFs integriert, um das Klassifikationsergebnis zu verbessern. Die Übergangswahrscheinlichkeiten, mit denen der statistische Zusammenhang zwischen den zu bestimmenden Labels und den Labels der Karte und auch der Zusammenhang zwischen temporal benachbarten Labels in der multitemporalen Klassifikation beschrieben wird, müssen zuvor nicht bestimmt werden, da sie innerhalb des Klassifikationsalgorithmus aus den gegebenen Daten ermittelt werden. Manuell gelabelte Trainingsdaten werden demnach in der hier vorliegenden Arbeit nicht benötigt, weder für das Training, da hier die Labels der veralteten Karte verwendet werden, noch für die Bestimmung der Übergangswahrscheinlichkeiten.

3. Grundlagen

In diesem Kapitel werden die Methoden beschrieben, auf denen diese Arbeit hauptsächlich aufbaut. Begonnen wird in Kapitel 3.1 mit einer allgemeinen Einführung der Klassifikation von Bildern. In den Kapiteln 3.2.1 bis 3.2.4 werden anschließend die verwendeten Merkmale beschrieben, gefolgt von der Methodik zur Merkmalsselektion (Kapitel 3.2.5). Im Folgenden werden die verwendeten Klassifikationsmethoden dargestellt. Zunächst wird die logistische Regression (Kapitel 3.3) erläutert, gefolgt von der Beschreibung der gegenüber fehlerhaften Trainingslabels robusten logistischen Regression (Kapitel 3.4), den Klassifikations- und Regressionsbäumen (Kapitel 3.5) und dem Random Forest (Kapitel 3.6). Zur Einbindung von räumlichem Kontext wird das Prinzip der Conditional Random Fields in Kapitel 3.7 vorgestellt.

3.1. Klassifikation

Das Ziel der Klassifikation ist es, ein Primitiv, repräsentiert durch einen Vektor mit Merkmalen, einer diskreten Klasse zuzuweisen [Bishop, 2006]. Im Bereich der Klassifikation von Fernerkundungsdaten liegen N *Bildprimitive* vor, z.B. Pixel oder Segmente. In dieser Arbeit werden Pixel als Bildprimitiv gewählt, wobei jedes Primitiv n durch einen *Merkmalsvektor* $\mathbf{x}_n(\mathbf{d})$ mit F Merkmalen repräsentiert wird. Jeder Merkmalsvektor $\mathbf{x}_n(\mathbf{d})$ kann hierbei aus den beobachteten Sensordaten $\mathbf{d}$ des ganzen Bildes abgeleitet werden. Für die meisten Merkmale werden aber ausschließlich die beobachteten Sensordaten des Bildprimitivs $\mathbf{d}_n$ und ggf. die beobachteten Sensordaten der Bildprimitive in einer bestimmten lokalen Nachbarschaft benötigt. Ein Merkmal kann z.B. die spektrale Signatur, Textur oder Struktur von Objekten im Bild beschreiben (s. Kapitel 3.2). Der durch die Merkmale aufgespannte Raum wird auch *Merkmalsraum* genannt. Im Folgenden wird der Merkmalsvektor $\mathbf{x}_n(\mathbf{d})$ mit $\mathbf{x}_n$ abgekürzt. Wenn der Zusammenhang klar ist, wird der Index zur Vereinfachung der Notation dabei weggelassen.

Die Zugehörigkeit eines Primitivs n zu einer der K Klassen wird durch sein *Label* $C_n \in \mathbb{C} = \{C^1, ..., C^K\}$ beschrieben, wobei $\mathbb{C}$ die diskrete Menge der Klassen und C^k, mit $k \in \{1, ..., K\}$, eine bestimmte Klasse darstellt.

Bei Klassifikationsverfahren wird jedem Merkmalsvektor $\mathbf{x}_n$ ein Label C_n zugeordnet. Dies entspricht einer Unterteilung des Merkmalsraumes durch *Entscheidungsgrenzen* in mehrere Gebiete, die jeweils einer Klasse zugeordnet werden [Bishop, 2006]. Bei linearen Klassifikationsmodellen ist die Entscheidungsgrenze eine (F-1)-dimensionale Hyperebene im Merkmalsraum.

Probabilistische Verfahren verwenden Wahrscheinlichkeiten, die den statistischen Zusammenhang zwischen den Merkmalsvektoren $\mathbf{x}$ und den Labels C beschreiben. Entsprechend dem

Maximum a posteriori Kriterium wird jenes Label C gewählt, für welches die bedingte Wahrscheinlichkeit $p(C|\mathbf{x})$ maximal wird:

$$C = \arg\max_{C}(p(C|\mathbf{x})). \tag{1}$$

Bei den probabilistischen Verfahren wird zwischen generativen und diskriminativen Verfahren unterschieden. Während bei generativen Verfahren die a posteriori Verteilung über die gemeinsame Verteilung $p(C, \mathbf{x})$ modelliert wird, modellieren diskriminative Verfahren, wie die logistische Regression (Kapitel 3.3), direkt die a posteriori Wahrscheinlichkeit $p(C|\mathbf{x})$ [Bishop, 2006]. Bei nicht-probabilistischen Verfahren, wie dem Random Forest (Kapitel 3.6) wird das Label direkt und ohne die Modellierung von Wahrscheinlichkeiten zugeordnet. In einigen nicht-probabilistischen Verfahren kann jedoch das Ergebnis auch probabilistisch interpretiert werden.

Das Anlernen (engl. *Training*) eines Klassifikators, bei dem die Parameter für die Prädiktion der Labels bestimmt werden, erfolgt bei *überwachten* Klassifikationsverfahren mit Hilfe von *Trainingsdaten* [Bishop, 2006]. Diese bestehen aus N_T Bildprimitiven, auch *Trainingsbeispiele* genannt, die sowohl einen Merkmalsvektor $\mathbf{x}_n$ besitzen als auch ein bereits zugewiesenes *Trainingslabel* $\widetilde{C}_n \in \mathbb{C}$ [Bishop, 2006]. Da das Training eines Klassifikators und somit der gelernte Zusammenhang zwischen den Merkmalsvektoren und den Labels auf den Trainingsdaten basiert, ist es wichtig, dass die Trainingsdaten *repräsentativ* sind, also der gleichen Verteilung folgen wie die zu klassifizierenden Daten.

Bei Graphischen Modellen, wie z.B. dem CRF (Kapitel 3.7), wird eine optimale Labelkonfiguration $\mathbf{C}$ für das gesamte Bild abhängig von allen Daten $\mathbf{d}$ bestimmt. Anders als bei den zuvor beschriebenen Verfahren werden hier auch statistische Abhängigkeiten der Labels zwischen den verschiedenen Primitiven berücksichtigt.

3.2. Merkmale

Die Klassifikation basiert auf den Merkmalen jedes Bildprimitivs. Hierbei werden radiometrische Merkmale, die sowohl für Pixel als auch Segmente bestimmt werden können, von geometrischen Merkmalen, die nur bei Segmenten sinnvoll sind, unterschieden. Geometrische Merkmale, wie die Größe, Umfang oder Kompaktheit, werden in dieser Arbeit nicht weiter betrachtet, da eine pixelbasierte Klassifikation erfolgen soll. Beispiele für radiometrische Merkmale sind spektrale Merkmale, wie z.B. der *Normalised Difference Vegetation Index* (NDVI, dt. Normalisierter Differenzierter Vegetationsindex), Texturmerkmale wie z.B. die Haralick Merkmale [Haralick et al., 1973], und Strukturmerkmale, wie z.B. gewichtete Histogramme von Gradientenrichtungen (HOG, engl. *Histogram of Oriented Gradients*) [Dalal und Triggs, 2005]. Eine weitere Kategorie für Merkmale bilden die 3D Merkmale, welche im Allgemeinen aus einem Höhenmodell, wie z.B. dem normalisierten digitalen Oberflächenmodell, bestimmt werden. Im Folgenden werden die Merkmale, die in dieser Arbeit verwendet werden, dargestellt. Die Auflistung erhebt demnach keinen Anspruch auf Vollständigkeit.

3.2.1. Spektrale Merkmale

Spektrale Merkmale enthalten Informationen über die Reflexions- und Absorbtionseigenschaften der zu klassifizierenden Objekte. Die aufgenommenen Spektralwerte, wie Grauwerte oder Farbinformationen, geben diese Eigenschaften für bestimmte Wellenlängen wieder. Der aufgenommene spektrale Bereich ist sensorabhängig; Tabelle 1 zeigt beispielhaft die spektralen Bänder von Landsat 7 mit dem Sensor Enhanced Thematic Mapper Plus (ETM+) [Nasa, Landsat Science, 2019].

Name	Abkürzung	Spektraler Bereich
Blau	g_B	0,45 – 0,52 μm
Grün	g_G	0,53 – 0,61 μm
Rot	g_R	0,63 – 0,69 μm
Nahes Infrarot	g_{NIR}	0,78 – 0,90 μm
Infrarot	g_{IR}	1,55 – 1,75 μm

Tabelle 1: Spektrale Bänder von Landsat 7.

Viele Merkmale werden durch Funktionen der Spektralwerte bestimmt. Ein Beispiel ist die Überführung der RGB Farbkomponenten (g_R, g_G, g_B) in den HSV Farbraum [Burger und Burge, 2005]:

$$\text{Helligkeit} \quad f_V = max(g_R, g_G, g_B), \tag{2}$$

$$\text{Farbton} \quad f_H = \begin{cases} \frac{60(g_G - g_B)}{f_V - min(g_R, g_G, g_B)} & \text{für } f_V = g_R \\ \frac{120 + 60(g_B - g_R)}{f_V - min(g_R, g_G, g_B)} & \text{für } f_V = g_G \\ \frac{240 + 60(g_R - g_G)}{f_V - min(g_R, g_G, g_B)} & \text{für } f_V = g_B, \end{cases} \tag{3}$$

$$\text{Sättigung} \quad f_S = \begin{cases} \frac{f_V - min(g_R, g_G, g_B)}{f_V} & \text{für } f_V > 0 \\ 0 & \text{sonst.} \end{cases} \tag{4}$$

Ein weiteres Beispiel für ein solches Merkmal ist der NDVI [Lillesand et al., 2014]:

$$f_{NDVI} = \frac{g_{NIR} - g_R}{g_{NIR} + g_R}. \tag{5}$$

Dabei ist g_{NIR} der Grauwert im nahen infraroten Kanal. Der NDVI nimmt einen Wert zwischen -1 und 1 an und wird zur Charakterisierung von Vegetation verwendet, da der NDVI bei Vegetation im Allgemeinen einen hohen Wert besitzt.

3.2.2. Texturmerkmale: Haralick Merkmale

Texturmerkmale beschreiben charakteristische Muster innerhalb einer bestimmten Region im Bildraum, z.B. eines Segmentes oder einer lokalen Umgebung um ein Pixel. Diese Muster bestehen meist aus sich wiederholenden Grauwertabfolgen in einer bestimmten Richtung. Zur Beschreibung des Musters kann eine Co-Occurance Matrix (CM) dienen, die sich auf diese Region im Bildraum bezieht [Tönnies, 2005]. Jedes Element der Matrix beschreibt dabei die

Häufigkeit, mit der zwei Grauwerte g^a und g^b in einer bestimmten räumlichen Konfiguration auftreten. Die räumliche Konfiguration wird durch eine Distanz d und einen Winkel α repräsentiert. Für jede Kombination aus d und α entsteht eine quadratische CM der Größe $N_G \times N_G$, wobei N_G die Anzahl möglicher Grauwerte ist. Die Elemente $cm_{d,\alpha}$ werden, nach [Tönnies, 2005], dabei wie folgt bestimmt:

$$cm_{d,\alpha}(g^a, g^b) = \frac{1}{N_R} \sum_{(r,c)inR}^{N_R} \delta(g(r,c) = g^a) \cdot \delta(g(r + (d \cdot sin\alpha), c + (d \cdot cos\alpha)) = g^b).$$

N_R ist dabei die Anzahl der Pixel in der zu untersuchenden Texturregion und R ist die Menge aller Pixel in der Region; r und c geben die Reihe und Spalte an und folglich gibt $g(r, c)$ den Grauwert des Pixels der Reihe r und Spalte c an. δ ist die Kronecker-Delta-Funktion, die den Wert 1 liefert, falls das Argument wahr ist und andernfalls den Wert 0.

Auf Basis dieser Co-Occurance Matrix werden die Haralick Merkmale, unter anderem die in dieser Arbeit genutzten Merkmale Energie, Kontrast, Homogenität und Korrelation berechnet [Tönnies, 2005]:

$$\begin{array}{lll} \text{Energie:} & f_{EnerH} & = \sum_{g^a} \sum_{g^b} cm_{d,\alpha}(g^a, g^b)^2, \\ \text{Kontrast:} & f_{KontH} & = \sum_{g^a} \sum_{g^b} (g^a - g^b)^2 \cdot cm_{d,\alpha}(g^a, g^b), \\ \text{Homogenität:} & f_{HomH} & = \sum_{g^a} \sum_{g^b} \frac{cm_{d,\alpha}(g^a, g^b)}{1+|g^a - g^b|}, \\ \text{Korrelation:} & f_{KorrH} & = \sum_{g^a} \sum_{g^b} \frac{cm_{d,\alpha}(g^a, g^b) - \mu_a \mu_b}{\sigma_a^2 \sigma_b^2}. \end{array} \tag{6}$$

In Gleichung 6 bezeichnen μ_a, μ_b die Mittelwerte und σ_a, σ_b die Standardabweichung der Grauwerte g^a, g^b. Liegen mehrere Farbkanäle vor, können die Haralick Merkmale für jeden Kanal unabhängig von den anderen Kanälen bestimmt werden. Auch bereits abgeleitete Merkmale, wie die Intensität, können für die Berechnung der Haralick Merkmale verwendet werden. Für jede Kombination aus der Distanz d und dem Winkel α entsteht eine CM und somit ein Satz an Haralick Merkmalen. Um die im Bild vorkommenden Texturen ausreichend beschreiben zu können, werden ggf. mehrere Kombinationen benötigt.

3.2.3. Strukturmerkmale: gewichtetes Histogramm der Gradientenrichtungen

Strukturmerkmale beschreiben lokale Bildstrukturen, z.B. anhand einer Analyse der Grauwertvariationen. Auch hier bezieht sich das Merkmal auf eine bestimmte Region im Bildraum, z.B. ein Segment oder eine lokale Umgebung um ein Pixel. Für die Bestimmung dieser Merkmale dient meist das gewichtete Histogramm der Gradientenrichtungen (HOG) [Dalal und Triggs, 2005] als Grundlage. Zunächst werden hierfür die Grauwertgradienten gebildet und

anschließend deren Betrag und Richtung bestimmt [Zheng und Casari, 2019]:

$$\begin{aligned} \text{Gradient:} \quad & h(x,y) && = \begin{pmatrix} h_x \\ h_y \end{pmatrix} = \begin{pmatrix} \frac{\partial h}{\partial x} \\ \frac{\partial h}{\partial y} \end{pmatrix}, \\ \text{Betrag:} \quad & |h(x,y)| && = \sqrt{h_x^2 + h_y^2}, \\ \text{Richtung:} \quad & \phi && = arctan\left(\frac{h_y}{h_x}\right). \end{aligned}$$

x und y beschreiben hierbei eine Position im Bild und somit ein bestimmtes Pixel an der Stelle (x, y). Mit Hilfe der Richtungen und Beträge werden nun die Histogramme erstellt. Hierfür wird das Intervall von 0° bis 360° in gleich große Klassen aufgeteilt. Für jede dieser Klassen erfolgt die Summation der Beträge aller Gradienten mit der entsprechenden Richtung. Um Rotationsinvarianz zu erreichen, müssen die Gradientenrichtungen zuvor auf eine Hauptrichtung bezogen werden. Das resultierende Histogramm kann entweder selber als Merkmalsvektor genutzt werden oder es werden weitere Merkmale abgeleitet, wie die Anzahl relevanter Extrema (Maxima, Minima), z.B. [Albert et al., 2015].

3.2.4. 3D Merkmale: normalisiertes digitales Oberflächenmodell

3D Merkmale basieren im Allgemeinen auf Höhenmodellen, in denen für jede X,Y Position im Raum ein Höhenwert Z beschrieben wird. Bei Höhenmodellen werden grundsätzlich zwei Arten unterschieden: das *digitale Geländemodell* (DGM) und das *digitale Oberflächenmodell* (DOM). Das DGM beschreibt die Höhe der Geländeoberfläche ohne Objekte, die auf dem Gelände stehen, wie Gebäude und Vegetation. Das DOM hingegen beschreibt die Höhe der Erdoberfläche mit allen Objekten, die von einem Sensor erfasst werden, inklusive Gebäude und Vegetation. Die Höheninformation kann aus verschiedenen Quellen stammen, wie beispielsweise aus Laserscanningdaten oder aus einer stereoskopischen Bildauswertung [Landesvermessung und Geobasisinformation, 2019b,a]. Für die Klassifikation ist meist die Höhe über Grund in Form eines *normalisierten digitalen Oberflächenmodells* (nDOM) hilfreich, welches sich aus der Differenz beider Höhenmodelle ergibt [Bähr und Vögtle, 2005]:

$$nDOM = DOM - DGM. \tag{7}$$

Ist das DGM nicht bekannt, kann es näherungsweise aus einem DOM bestimmt werden. Um die Objekte über Grund aus dem DOM herauszufiltern, kann *morphologisches Opening* verwendet werden [Bähr und Vögtle, 2005]. In dieser Arbeit erfolgt dies durch die Kombination zweier *digitaler Höhenmodelle* (DHM), die durch ein Opening des DOM mit unterschiedlich großen Strukturelementen entstanden sind. In Abb. 2 ist die Vorgehensweise zur Kombination beider DGM skizziert.

Das erste Opening filtert mit Hilfe eines größeren Strukturelementes der Größe S^{gross} die Objekte über dem Gelände aus dem DOM heraus. Aufgrund der großen Ausdehnung des Strukturelementes können jedoch auch Strukturen des Geländes entfernt werden. Das zweite Opening erfolgt mit Hilfe eines kleineren Strukturelements der Größe S^{klein}. Hierbei bleibt

die Struktur des Geländes größtenteils erhalten, jedoch werden kleinere Objekte herausgefiltert und größere bleiben bestehen. Um die Vorteile beider resultierenden Höhenmodelle (DHM^{gross} bzw. DHM^{klein}) zu nutzen, werden die Werte an jedem Pixel n verglichen. Ist der Wert von DHM^{klein} deutlich größer (größer als ein Schwellwert S^{Diff}) als der Wert von DHM^{gross}, handelt es sich vermutlich um ein größeres Objekt über dem Gelände, welches in DHM^{klein} nicht herausgefiltert wurde, sodass der Wert von DHM^{gross} für das kombinierte DGM ($DGM^{kombiniert}$) genutzt wird. Andernfalls handelt es sich eher um Geländestrukturen, die in DHM^{gross} fäschlicherweise fehlen, sodass der Wert von DHM^{klein} weiterhin verwendet wird:

$$DGM_n^{kombiniert} = \begin{cases} DHM_n^{gross} & \text{für } DHM_n^{klein} - DHM_n^{gross} > S^{Diff} \\ DHM_n^{klein} & \text{für } DHM_n^{klein} - DHM_n^{gross} \leq S^{Diff}. \end{cases} \tag{8}$$

Zusätzliche Glättungsoperationen mit einem Medianfilter der Größe S^{Median} vor und einem Gaußfilter der Standardabweichung $\sigma^{GaussDGM}$ nach dem Opening verringern den Einfluss von Artefakten, die durch das Opening erzeugt werden.

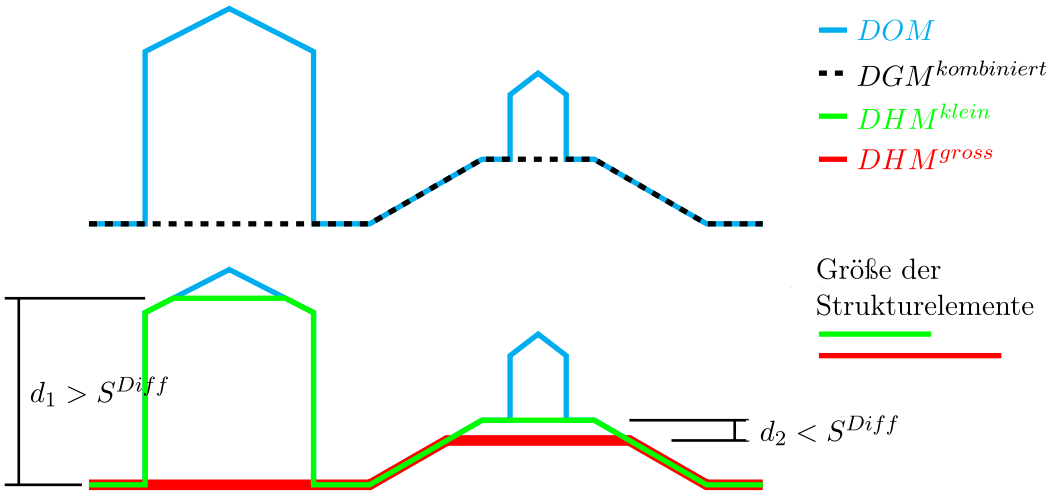


Abbildung 2: Erstellung eines DGMs aus einem DOM mit Hilfe zweier Opening-Operationen mit unterschiedlich großen Strukturelementen.

3.2.5. Merkmalsselektionsalgorithmus

Die Wahl der Merkmale hat einen großen Einfluss auf das Klassifikationsergebnis. Damit die Primitive anhand ihrer Merkmale einer Klasse zugeordnet werden können, müssen sich die Merkmale von Bildprimitiven einer Klasse von denen anderer Klassen unterscheiden. Im Idealfall bilden Primitive einer gemeinsamen Klasse Ballungen im Merkmalsraum, die sich nicht mit Ballungen weiterer Klassen überschneiden. Die Anzahl an möglichen Merkmalen, die aus den Sensordaten bestimmt werden können, ist in der Regel relativ groß, wobei nicht alle Merkmale für die Aufteilung der Klassen hilfreich sind. Besonders bei großen Datensätzen ist es außerdem sinnvoll, nicht zu viele Merkmale zu nutzen, da ab einem gewissen Punkt die Hinzunahme weiterer Merkmale bei gleich bleibender Anzahl an Trainingsdaten zu einer Verschlechterung des Klassifikationsergebnisses führt (Hughes-Phänomen) [Hughes, 1968]. Außerdem ist die Anzahl an Merkmalen meist auch durch die gegebenen Speichermöglichkeiten und durch

ein Zeitlimit für die Prozessierung begrenzt. Aus diesem Grund wird versucht mit Hilfe von Selektionsverfahren jene Merkmale aus einer größeren Menge von Merkmalen zu selektieren, die für die Klassifikation relevant sind und keine redundanten Informationen enthalten.

Die in dieser Arbeit verwendete Selektionsmethode basiert auf der *Transinformation* (engl. *mutual information*, MI) [Shannon, 1948] und wurde in [Frénay et al., 2014] präsentiert. Die in [Frénay et al., 2014] vorgestellte Erweiterung für die Selektion unter Berücksichtigung fehlerhafter Labels wird nicht verwendet. Die ausführliche Erläuterung, warum stattdessen der ursprüngliche Algorithmus verwendet wird, findet sich in Kapitel 5.4.

Die Transinformation ist ein Maß für die Menge an Informationen, die eine Zufallsvariable über eine andere enthält [Cover und Thomas, 2012]. Im Bereich der Merkmalsselektion kann so die Relevanz einer bestimmten Gruppe an Merkmalen $\mathbf{X}$ für die Prädiktion der Klassenlabels C gemessen werden [Frénay et al., 2014]. Für die Bestimmung der MI werden demnach Trainingsbeispiele benötigt, die sowohl einen Merkmalsvektor als auch ein Klassenlabel aufweisen. Die MI basiert auf der Entropie $\mathcal{E}$; ein Maß für die Unsicherheit einer Zufallsvariablen [Vergara und Estévez, 2014]. Hat jedes Ereignis in etwa die gleiche Wahrscheinlichkeit, weist die Entropie einen hohen Wert auf, während unterschiedliche Auftrittswahrscheinlichkeiten zu einem niedrigen Wert führen.

Die Transinformation MI zwischen $\mathbf{X}$ und C kann aus der Entropie $\mathcal{E}(\mathbf{X})$ der Merkmale und der bedingten Entropie $\mathcal{E}(\mathbf{X}|C)$ der Merkmale bei gegebenen Klassenlabels C bestimmt werden [Frénay und Verleysen, 2014]:

$$MI(\mathbf{X}, C) = \mathcal{E}(\mathbf{X}) - \mathcal{E}(\mathbf{X}|C). \tag{9}$$

Die Entropie $\mathcal{E}(\mathbf{X})$ kann formal mit Hilfe der Wahrscheinlichkeitsdichtefunktion p_x von $\mathbf{X}$ berechnet werden:

$$\mathcal{E}(\mathbf{X}) = -\int_{\mathcal{X}} p_x(\mathbf{X}) ln(p_x(\mathbf{X})) dX, \tag{10}$$

wobei $\mathcal{X}$ den Merkmalsraum beschreibt. Im Allgemeinen ist p_x jedoch unbekannt, sodass $\mathcal{E}(\mathbf{X})$ durch $\hat{\mathcal{E}}(\mathbf{X})$ angenähert werden muss. Frénay et al. [2014] nutzen dafür die Schätzung von [Kozachenko und Leonenko, 1987]:

$$\hat{\mathcal{E}}(\mathbf{X}) = -\Psi(\xi) + \Psi(N) + ln(c_F) + \frac{F_X}{N} \sum_{n=1}^{N} ln(\epsilon_\xi(n)). \tag{11}$$

F_X ist hierbei die Anzahl an Merkmalen in der Gruppe an Merkmalen $\mathbf{X}$, N die Anzahl an Trainingsbeispielen und $\epsilon_\xi(n)$ die zweifache Distanz von Trainingsbeispiel n zu seinem ξ-ten nächsten Nachbarn im Merkmalsraum. ξ beschreibt die Anzahl der berücksichtigten Nachbarn und ist der einzige Parameter, der vor der Selektion festgesetzt werden muss. An dieser Stelle sei kurz erwähnt, dass in vielen Veröffentlichungen der Parameter ξ mit k deklariert wird, welcher in dieser Arbeit jedoch für den Klassenindex genutzt wird. Ψ ist die Digammafunktion, die der ersten Ableitung der Gammafunktion Γ entspricht [Bishop, 2006]:

$$\Psi(\xi) \equiv \frac{d}{d\xi} ln(\Gamma(\xi)) \quad \text{bzw.} \quad \Psi(N) \equiv \frac{d}{dN} ln(\Gamma(N)). \tag{12}$$

c_F beschreibt das Volumen der Einheitskugel mit der Dimension F_X und ist wie folgt bestimmt:

$$c_F = \frac{2\pi^{\frac{F_X}{2}}}{\Gamma(\frac{F_X}{2})}. \tag{13}$$

Die bedingte Entropie $\hat{\mathcal{E}}(\mathbf{X}|C)$ ist in [Frénay et al., 2014] definiert als

$$\hat{\mathcal{E}}(\mathbf{X}|C) = \sum_{k=1}^{K}(p_C(k) \cdot \hat{\mathcal{E}}(\mathbf{X}|C = C^k)), \tag{14}$$

wobei $p_C(k) = \frac{N^k}{N}$ die Wahrscheinlichkeit für $C = C^k$ ist und mit Hilfe der Anzahl an Trainingsbeispielen N und der Anzahl an Trainingsbeispielen N^k mit der Klasse C^k angenähert wird. $\hat{\mathcal{E}}(\mathbf{X}|C = C^k)$ ist die Entropie von $\mathbf{X}$ (Gleichung 11), jedoch beschränkt auf die Trainingsbeispiele mit dem Label C^k.

Die MI kann folglich angenähert werden durch:

$$\hat{MI}(\mathbf{X}, C) = \hat{\mathcal{E}}(\mathbf{X}) - \hat{\mathcal{E}}(\mathbf{X}|C). \tag{15}$$

Ein hoher Wert der MI zeigt an, dass die eine Zufallsvariable viel Information über die andere Zufallsvariable enthält. Dies deutet auf eine starke statistische Abhängigkeit zwischen den Zufallsvariablen, in diesem Fall $\mathbf{X}$ und C, hin. Das Ziel der Merkmalsselektion liegt demnach darin, eine Merkmalsauswahl zu finden, die einen hohen Wert für MI aufweist.

In [Frénay et al., 2014] wird für die Merkmalsselektion eine einfache Rückwärtssuche (engl. backward search) vorgeschlagen. Hierbei wird mit einer großen Menge an Merkmalen gestartet, aus der iterativ Merkmale entfernt werden. Diese Methode weise laut Frénay et al. [2014] eine höhere Stabilität im Bezug zu den selektierten Merkmalen auf als eine fortlaufende Suche, bei der mit einem Merkmal gestartet wird und iterativ Merkmale hinzugefügt werden.

Bei dieser Suche gibt es in jeder Iteration τ $F_\tau = F_X - \tau + 1$ Merkmale, die die Gruppe an Merkmalen $\mathbf{X}_\tau$ bilden mit $\tau \in (1, ..., F_X - 1)$. Durch die Herausnahme von jeweils einem Merkmal f aus dem Pool von F_τ Merkmalen lassen sich insgesamt F_τ Transinformationswerte MI_{-f} bestimmen. Der höchste Wert dieser Transinformationen ${}^\tau MI_{-f}^{max}$ zeigt das Merkmal f_τ, welches am wenigsten Informationen beisteuert und dementsprechend in der Iteration τ entfernt wird. Das Entfernen der einzelnen Merkmale kann nun entweder bis zur letzten Iteration $\tau = F_X - 1$ weitergeführt werden, oder es wird zuvor durch ein Abbruchkriterium, wie eine vorgegebene Höchstanzahl an Merkmalen, unterbrochen. Werden alle Iterationen durchgeführt, kann der Verlauf der jeweils größten Transinformation pro Iteration ${}^\tau MI_{-f}^{max}$ auch graphisch abhängig von der Anzahl an Merkmalen dargestellt werden. Ausgewählt werden die Merkmale, die zu dem maximalen Wert MI^{max} aller Transinformationen ${}^\tau MI_{-f}^{max}$ führen. Die genaue Anzahl an Merkmalen kann daher variieren.

3.3. Logistische Regression

Die *logistische Regression* (LR) ist ein probabilistischer diskriminativer Klassifikator, welcher direkt die a posteriori Wahrscheinlichkeit $p(C_n = C^k|\mathbf{x}_n)$ für das Klassenlabel C_n bei gegebenem Merkmalsvektor $\mathbf{x}_n$ modelliert. Dieser Überblick orientiert sich an [Bishop, 2006].

Da es sich bei der LR um ein lineares Klassifikationsmodell handelt, können nur lineare Grenzen im Merkmalsraum zur Unterscheidung der Klassen verwendet werden. Um auch nicht lineare Entscheidungsgrenzen zu erzeugen, kann der Merkmalsvektor $\mathbf{x}_n$ durch eine Merkmalstransformation $\Phi(\mathbf{x}_n)$ in einen höherdimensionalen Raum überführt werden. Das Training und die Klassifikation erfolgen dann mit dem transformierten Vektor $\Phi(\mathbf{x}_n)$ anstatt mit dem ursprünglichen Merkmalsvektor. Die resultierende lineare Entscheidungsgrenze in dem höherdimensionalen Raum entspricht einer nicht linearen Grenze in dem ursprünglichen Merkmalsraum. Die Elemente des erweiterten Merkmalsvektors $\Phi(\mathbf{x}_n)$ werden mit prinzipiell beliebigen Funktionen aus $\mathbf{x}_n$ bestimmt. Ein Beispiel ist die quadratische Erweiterung, bei der $\Phi(\mathbf{x}_n)$ die ursprünglichen Merkmale $\mathbf{x}_n$, sowie Quadrate gemischter Produkte der Merkmale enthält. Das lineare Polynom zur Beschreibung der Entscheidungsgrenze im Merkmalsraum enthält ein konstantes Glied w_0. Um die Notation zu vereinfachen, wird ein konstantes Merkmal $\Phi_1(\mathbf{x}_n) = 1$ hinzugefügt.

Die a posteriori Wahrscheinlichkeit $p(C_n = C^k|\mathbf{x}_n)$ wird bei der Verwendung mehrerer Klassen mit der Softmax-Funktion modelliert:

$$p(C_n = C^k|\mathbf{x}_n) = \frac{exp\Big(\mathbf{w}_k^T \cdot \mathbf{\Phi}(\mathbf{x}_n)\Big)}{\sum_{j=1}^{K} exp\Big(\mathbf{w}_j^T \cdot \mathbf{\Phi}(\mathbf{x}_n)\Big)}, \tag{16}$$

wobei der Vektor $\mathbf{w}_k$ die Gewichtskoeffizienten einer bestimmten Klasse C^k beinhaltet, welche wiederum indirekt die Entscheidungsgrenze im transformierten Merkmalsraum bestimmt. Da die Summe aller a posteriori Wahrscheinlichkeiten 1 ergeben muss, sind die Parametervektoren nicht unabhängig voneinander. Berücksichtigt wird dies durch die Festsetzung von jedem Element in $\mathbf{w}_1$ zu 0. Der im Training zu bestimmende Vektor an Parametern ist demnach $\mathbf{w} = (\mathbf{w}_2^T, ..., \mathbf{w}_K^T)^T$.

Während des Trainingprozesses müssen nun die Elemente dieses Parametervektors $\mathbf{w}$ bestimmt werden. Hierfür werden N_T Trainingsbeispiele, jeweils bestehend aus einem Merkmalsvektor $\mathbf{x}_n$ und einem Klassenlabel $\widetilde{C}_n$, genutzt. Um die wahrscheinlichsten Werte von $\mathbf{w}$ zu bestimmen, wird die a posteriori Wahrscheinlichkeit von $\mathbf{w}$ bei gegebenen Trainingsdaten optimiert:

$$p(\mathbf{w}|\mathbf{x}_1, ..., \mathbf{x}_{N_T}, \widetilde{C}_1, ..., \widetilde{C}_{N_T}) \propto p(\mathbf{w}) \cdot \prod_{n,k} p(\widetilde{C}_n = C^k|\mathbf{x}_n, \mathbf{w}) \rightarrow max. \tag{17}$$

Um eine Überanpassung an die Trainingsdaten zu vermeiden, die sich unter anderem durch eine sehr große Steigung der Softmax-Funktion äußert, wird hier der Regularisierungsterm $p(\mathbf{w})$ eingeführt. Dieser basiert auf der Normalverteilung mit dem Erwartungswert $\mathbf{0}$ und der

Kovarianzmatrix $\sigma \cdot \mathbf{I}$, wobei $\mathbf{I}$ die Einheitsmatrix darstellt. Die a posteriori Wahrscheinlichkeit $p(\widetilde{C}_n = C^k|\mathbf{x}_n, \mathbf{w})$ entspricht $p(C_n = C^k|\mathbf{x}_n)$ aus Gleichung 16.

Anstatt $p(\mathbf{w}|\mathbf{x}_1, ..., \mathbf{x}_{N_T}, \widetilde{C}_1, ..., \widetilde{C}_{N_T})$ zu maximieren, wird das Minimum des negativen Logarithmus $E(\mathbf{w}) = -ln\big(p(\mathbf{w}|\mathbf{x}_1, ..., \mathbf{x}_{N_T}, \widetilde{C}_1, ..., \widetilde{C}_{N_T})\big)$ ermittelt:

$$E(\mathbf{w}) = -\sum_{n=1}^{N_T} \cdot \sum_{k=1}^{K} t_{nk} ln(p(\widetilde{C}_n = C^k|\mathbf{x}_n, \mathbf{w})) + \frac{\mathbf{w}^T\mathbf{w}}{2\sigma^2} \rightarrow min. \tag{18}$$

In Gleichung 18 entspricht t_{nk} einer Indikatorvariablen, die den Wert 1 annimmt wenn $\widetilde{C}_n = C^k$ gilt und andernfalls den Wert 0.

Aufgrund der Nichtlinearität der Funktion $E(\mathbf{w})$ kann das Minimum nicht direkt bestimmt werden, sondern nur iterativ, etwa mit dem Newton-Raphson Verfahren. Da $E(\mathbf{w})$ konkav ist und somit nur ein einziges Minimum existiert, konvergiert z.B. das Newton-Raphson Verfahren garantiert zum globalen Minimum [Bishop, 2006]. Hierbei wird, ausgehend von einem Initialwert $\mathbf{w}^0$, der Parametervektor $\mathbf{w}^\tau$ in Iteration τ aus dem Parametervektor $\mathbf{w}^{\tau-1}$ der vorherigen Iteration, dem Gradientenvektor $\nabla E(\mathbf{w}^{\tau-1})$ und der Hessematrix $\mathbf{H} = \nabla\nabla E(\mathbf{w}^{\tau-1})$ ermittelt:

$$\mathbf{w}^\tau = \mathbf{w}^{\tau-1} - \mathbf{H}^{-1}\nabla E. \tag{19}$$

Im Folgenden werden der Gradientenvektor und die Hessematrix detaillierter dargestellt. Die Wahrscheinlichkeit $p(\widetilde{C}_n = C^j|\mathbf{x}_n)$ wird mit dem Symbol $\widetilde{y}_{nj}$ dargestellt. Wird, wie bei der klassischen logistischen Regression, von fehlerfreien, repräsentativen Trainingsdaten ausgegangen, entspricht $\widetilde{y}_{nj}$ der a posteriori Wahrscheinlichkeit $y_{nj} = p(C_n = C^j|\mathbf{x}_n)$ aus Gleichung 16. Der Gradientenvektor $\nabla E(\mathbf{w})$ besteht aus den Gradientenvektoren für alle Teilkomponenten von $\mathbf{w}$:

$$\nabla E(\mathbf{w}) = (\nabla_{w_2} E(\mathbf{w}))^T, ..., \nabla_{w_K} E(\mathbf{w}))^T)^T. \tag{20}$$

$\nabla_{w_j} E(\mathbf{w})$ bezeichnet hierbei die erste Ableitung von $E(\mathbf{w})$ nach dem Parameter $\mathbf{w}_j$:

$$\nabla_{w_j} E(\mathbf{w}) = \sum_{n=1}^{N_T} (\widetilde{y}_{nj} - t_{nj}) \cdot \mathbf{\Phi}(\mathbf{x}_n) + \frac{1}{\sigma^2} \cdot \mathbf{w}. \tag{21}$$

Die Hessematrix $\mathbf{H} = \nabla\nabla E(\mathbf{w})$ setzt sich aus Blöcken $\mathbf{H}_{ij} = \nabla_{w_i}\nabla_{w_j} E(\mathbf{w})$ der Größe $F \times F$ zusammen, wobei jeder Block der zweiten Ableitung nach den Parametern $\mathbf{w}_i$ und $\mathbf{w}_j$ entspricht:

$$\nabla_{w_i}\nabla_{w_j} E = \sum_{n=1}^{N_T} \widetilde{y}_{ni}(I_{ij} - \widetilde{y}_{nj})\mathbf{\Phi}(\mathbf{x}_n)\mathbf{\Phi}(\mathbf{x}_n)^T + \frac{\delta(i=j)}{\sigma^2}\mathbf{I}. \tag{22}$$

I_{ij} bezeichnet das Element aus der Reihe i und der Spalte j der Einheitsmatrix und $\delta(\cdot)$ ist die Kronecker-Delta-Funktion, die den Wert 1 liefert, fallt das Argument wahr ist und andernfalls den Wert 0.

Sind die einzelnen Elemente von $\mathbf{w}$ nach der Trainingsprozedur bekannt, kann für ein zu klassifizierendes Primitiv die a posteriori Wahrscheinlichkeit mit Hilfe der Gleichung 16 bestimmt werden.

3.4. Gegenüber fehlerhaften Trainingslabels robuste logistische Regression

In allen Klassifikationsverfahren wird davon ausgegangen, dass das Trainingslabel $\widetilde{C}_n$ eines Primitivs dem Klassenlabel C_n entspricht, welches für den entsprechenden Merkmalsvektor $\mathbf{x}_n$ prädiziert werden soll. Im Rahmen der vorliegenden Arbeit trifft diese Annahme nicht für alle Primitive zu, weil die Trainingslabels $\widetilde{C}$ aus einer veralteten Karte stammen. Der verwendete Klassifikator muss demnach mit Fehlern in den Trainingslabels umgehen können. Aus diesem Grund wird in diesem Kapitel eine Erweiterung der LR von Bootkrajang und Kabán [2012] vorgestellt, bei der fehlerhafte Trainingslabels beim Training berücksichtigt werden. Diese Klassifikationsmethode wird in dieser Arbeit verkürzt *robuste logistische Regression* (RLR) genannt.

RLR ist ein probabilistischer Klassifikator, bei dem das Auftreten fehlerhafter Labels im Training entsprechend einem NAR Fehlermodell modelliert wird. Das NAR Modell berücksichtigt ausschließlich statistische Abhängigkeiten der Fehler vom Objekttyp, jedoch keine von den Merkmalen (s. Kapitel 2.3). In diesem konkreten Klassifikator erfolgt die Beschreibung des Auftretens eines Labelfehlers abhängig vom Objekttyp durch die Wahrscheinlichkeit, dass für ein Primitiv ein Label C^k zum Training vorliegt, während das abgebildete Objekt das Label C^a aufweist. Da es $K \times K$ Kombinationen der K Labels gibt, lassen sich diese *Übergangswahrscheinlichkeiten* $\overleftarrow{\gamma}_{ak} = p(\widetilde{C} = C^k | C = C^a)$ in einer *Übergangsmatrix* $\overleftarrow{\mathbf{\Gamma}}$ zusammenfassen, wobei $\overleftarrow{\mathbf{\Gamma}}(a,k) = \overleftarrow{\gamma}_{ak}$.[1]

Im Trainingsprozess wird das Auftreten von fehlerhaften Trainingslabels berücksichtigt, indem die Wahrscheinlichkeit $p(\widetilde{C}_n = C^k | \mathbf{x}_n, \mathbf{w})$ aus Gleichung 18 als Randverteilung des beobachteten Labels $\widetilde{C}_n$ über alle Werte, die das unbekannte Klassenlabel C_n annehmen kann, bestimmt wird:

$$p(\widetilde{C}_n = C^k | \mathbf{x}_n, \mathbf{w}) \; = \sum_{a=1}^{K} p(\widetilde{C} = C^k | C = C^a)\, p(C_n = C^a | \mathbf{x}_n, \mathbf{w}). \tag{23}$$

Im Folgenden wird die Wahrscheinlichkeit $p(\widetilde{C}_n = C^k | \mathbf{x}_n, \mathbf{w})$ mit dem Symbol S_{nk} abgekürzt.

Wie bei der klassischen logistischen Regression werden im Training die Parametervektoren $\nabla E(\mathbf{w}) = (\nabla_{w_2} E(\mathbf{w}))^T, ..., \nabla_{w_K} E(\mathbf{w}))^T)^T$ bestimmt, indem die Fehlerfunktion in Gleichung 18 minimiert wird. Bei Berücksichtigung der Fehler in den Labels entsprechend Gleichung 23 resultiert die zu minimierende Funktion

$$E(\mathbf{w}, \overleftarrow{\mathbf{\Gamma}}) = -\sum_{n=1}^{N_T} \sum_{k=1}^{K} \big(t_{nk} \cdot ln(S_{nk})\big) + \frac{\mathbf{w}^T \mathbf{w}}{2\sigma^2}. \tag{24}$$

Für die Minimierung bezüglich des Parametervektors $\mathbf{w}$ kann auch hier die Newton-Raphson Methode genutzt werden, wobei sich der Gradient und die Hessematrix aufgrund der verän-

[1] Der Pfeil auf dem Symbol für die Übergangswahrscheinlichkeiten $\overleftarrow{\gamma}_{ak}$ bzw. der Übergangsmatrix $\overleftarrow{\mathbf{\Gamma}}$ gibt die Richtung des Übergangs an. Während $\overleftarrow{\gamma}_{ak} = p(\widetilde{C} = C^k | C = C^a)$ die Wahrscheinlichkeit für die Beobachtung des Labels C^k, obwohl das Objekt das Label C^a aufweist, darstellt, gibt $\overrightarrow{\gamma}_{ak} = p(C = C^a | \widetilde{C} = C^k)$ die Wahrscheinlichkeit, dass ein Objekt das Label C^a besitzt, obwohl das Label C^k beobachtet wurde, an.

derten Zielfunktion von der LR unterscheiden. Der Grandient $\nabla E = [\nabla_{w_2} E^T, ..., \nabla_{w_K} E^T]^T$ von $E(\mathbf{w}, \overleftarrow{\mathbf{\Gamma}})$ setzt sich dabei aus folgenden Komponenten zusammen:

$$\nabla_{w_j} E = \sum_{n=1}^{N_T} (y_{nj} - \bar{t}_{nj}) \mathbf{\Phi}(\mathbf{x}_n) + \frac{1}{\sigma^2} \mathbf{w}. \tag{25}$$

In Gleichung 25 ist y_{nj} die a posteriori Wahrscheinlichkeit $p(C_n = C^j | \mathbf{x}_n)$ und $\bar{t}_{nj} = y_{nj} \sum_{k=1}^{K} \left(\overleftarrow{\gamma}_{jk} \cdot \frac{t_{nk}}{S_{nk}} \right)$.

Die Hessematrix $\mathbf{H}$ besteht aus $(K-1) \times (K-1)$ Blöcken $\mathbf{H}_{ij} = \nabla_{w_i} \nabla_{w_j} E$:

$$\begin{aligned} \nabla_{w_i} \nabla_{w_j} E = & \sum_{n=1}^{N_T} \left(y_{ni} y_{nj} \xi + I_{ij} (y_{nj} - \bar{t}_{nj}) \right) \mathbf{\Phi}(\mathbf{x}_n) \mathbf{\Phi}(\mathbf{x}_n)^T \\ & + \frac{\delta(i=j)}{\sigma^2} \mathbf{I}, \end{aligned}$$

wobei $\xi = \sum_{k=1}^{k} \left(\overleftarrow{\gamma}_{jk} \cdot \overleftarrow{\gamma}_{ik} \cdot \frac{t_{nk}}{S_{nk}^2} \right)$ ist, I_{ij} dem Element (i, j) der Einheitsmatrix entspricht und $\delta(\cdot)$ die Kronecker-Delta-Funktion beschreibt, die den Wert 1 liefert, falls das Argument wahr ist und andernfalls den Wert 0.

Anders als in Gleichung 18 muss jedoch nicht nur der Parametervektor $\mathbf{w}$ ermittelt werden, sondern auch die Übergangsmatrix $\overleftarrow{\mathbf{\Gamma}}$. Dies erfolgt mithilfe eines iterativen Verfahrens, in dem der Parametervektor und die Übergangsmatrix abwechselnd aktualisiert werden. Zunächst werden mit Initialwerten der Transitionsmatrix $\overleftarrow{\mathbf{\Gamma}}^0$ die Parameter in $\mathbf{w}$ mit der Newton-Raphson Methode entsprechend Gleichung 19 neu bestimmt. Mit diesen Parametern kann wiederum $\overleftarrow{\mathbf{\Gamma}}$ aktualisiert werden, wobei

$$\overleftarrow{\gamma}_{jk}^{\tau} = \frac{1}{c} \cdot \overleftarrow{\gamma}_{jk}^{\tau-1} \cdot \sum_{n=1}^{N} (t_{nk} \cdot \frac{y_{nj}}{S_{nk}^{\tau-1}}), \tag{26}$$

mit $c = \sum_{l=1}^{K} \left(\overleftarrow{\gamma}_{jl}^{\tau-1} \cdot \sum_{n=1}^{N} (t_{nl} \cdot \frac{y_{nj}}{S_{nl}^{\tau-1}}) \right)$ und $S_{nk}^{\tau-1} = \sum_{a=1}^{K} \left(\overleftarrow{\gamma}_{ak}^{\tau-1} \cdot y_{na} \right)$. Danach werden mit den neuen Übergangswahrscheinlichkeiten $\overleftarrow{\gamma}^{\tau}$ erneut die Parameter in $\mathbf{w}$ aktualisiert, mit denen wiederum die Übergangswahrscheinlichkeiten angepasst werden. Diese Iteration wird fortgeführt bis ein Abbruchkriterium erfüllt ist, wie beispielsweise, dass eine maximale Anzahl an Iterationen erreicht wird oder dass die Änderung der Parameter unter einem zuvor definierten Schwellwert liegt.

Die a posteriori Wahrscheinlichkeiten für einen zu klassifizierenden Merkmalsvektor $\mathbf{x}_n$ können nun wie bei der LR mit der Gleichung 16 bestimmt werden, da die geschätzten Parameter in $\mathbf{w}$ im Zusammenhang zu dem unbekannten Label C_n stehen und nicht zu dem fehlerbehafteten Label $\widetilde{C}_n$ [Bootkrajang und Kabán, 2012].

Das von Bootkrajang und Kabán [2012] vorgestellte Verfahren weist durch das alternierende Aktualisieren des Parametervektors$\mathbf{w}$ und der Übergangsmatrix $\overleftarrow{\mathbf{\Gamma}}$ Ähnlichkeiten zu einem EM Algorithmus [Dempster et al., 1977] auf. Die zu bestimmenden Parameter müssen

vor dem Training initialisiert werden. Neben den Initialwerten müssen die Art der Merkmalstransformation, der Parameter σ des Regularisierungsterms und ein Abbruchkriterium definiert werden. Aus dem Training resultieren sowohl die Parameter $\mathbf{w}$, mit denen direkt die a posteriori Wahrscheinlichkeiten $p(C_n = C^k|\mathbf{x}_n)$ bestimmt werden können, als auch die Übergangswahrscheinlichkeiten $\overleftarrow{\gamma}$, die beispielsweise für die Integration der beobachteten Labels in den Klassifikationsprozess (s. Kapitel 4.3) oder im multitemporalen Kontext (s. Kapitel 4.7) verwendet werden können. Abhängigkeiten zwischen dem Auftreten der Fehler in den Trainingslabels und den Daten werden jedoch nicht modelliert.

3.5. Klassifikations- und Regressionsbäume

Klassifikations- und Regressionsbäume (engl. *Classification and Regression Trees*; CART) [Breiman et al., 1984] sind randomisierte Entscheidungsbäume, die mit Hilfe von Trainingsdaten erlernt werden. Ein Entscheidungsbaum ist ein hierarchischer Graph, beginnend mit einem Wurzelknoten. Jeder *Knoten* entspricht einem Test, der den Merkmalsraum in zwei Bereiche unterteilt, M_1 und M_2. Typischerweise wird dies durch eine lineare Grenze im Merkmalsraum realisiert. Diese zwei Bereiche werden in dem Entscheidungsbaum durch zwei Nachfolgeknoten, auch *Kinder* genannt, repräsentiert. Ist keine weitere Aufspaltung nötig, handelt es sich bei dem Knoten um ein *Blatt*, welches keiner Grenze, sondern einer Wahrscheinlichkeitsverteilung für die Klasse C oder gleich einer Klasse C entspricht. Diese Art von Entscheidungsbaum, bei dem jeder Knoten genau zwei oder keine Kinder besitzt, wird als binärer Baum bezeichnet. Die maximale Anzahl an Knoten, die passiert werden, um von dem Wurzelknoten zu einem Blatt zu gelangen, bestimmt die *Tiefe* des Baumes.

Für das Training eines CART wird die Menge an N_{Tb} Trainingsbeispielen $\mathcal{T}_b = \{(\mathbf{x}, \widetilde{C})_n\}_{n=1,\ldots,N_{Tb}}$ zunächst in zwei Teilmengen aufgespalten. Die erste Teilmenge $\mathcal{T}_0$ wird dafür verwendet, die Struktur des Baumes festzulegen und die Parameter der Tests in jedem Knoten zu ermitteln, während die zweite Teilmenge $\mathcal{T}_c$ für die Zuweisung einer Klasse zu einem Blatt bzw. die Bestimmung der Wahrscheinlichkeitsverteilung für die Klassen in einem Blatt genutzt wird.

Das Training des Baumes erfolgt durch die rekursive Aufspaltung der Trainingsbeispiele; Gestartet wird mit $\mathcal{T}_0$. An jedem Knoten q werden N_F Merkmale, wobei standardmäßig $N_F = \sqrt{F}$ gilt, ausgewählt. Unter ausschließlicher Berücksichtigung dieser N_F Merkmale werden N_G Trennflächen zufällig erzeugt. Die Trennfläche, welche die Daten am besten trennt, wird als Entscheidungsgrenze des Knotens q verwendet. Die Bewertung einer Trennfläche erfolgt nach einem zuvor definierten Qualitätskriterium, wie dem Gini-Index [Breiman, 2001] oder dem Informationszuwachs [Duda et al., 2001]. Der Informationszuwachs basiert auf den beiden Histogrammen, die bei der Aufspaltung der Trainingsbeispiele $\mathcal{T}_q$ des Knotens q in $\mathcal{T}_{M1}$ und $\mathcal{T}_{M2}$ entstehen. Für jede Teilmenge $\mathcal{T}_{Mj}$, wobei $j \in \{1, 2\}$, mit N_{Mj} Trainingsbeispielen wird ein Histogramm der Klassen erstellt und die normalisierten Histogrammeinträge als Wahrschein-

lichkeiten $p_j(C^k)$ für die jeweilige Klasse C^k interpretiert:

$$p_j(C^k) = \frac{1}{N_{Mj}} \sum_{m \in \mathcal{T}_{Mj}} \delta(\widetilde{C}_m = C^k). \tag{27}$$

wobei $\delta(\widetilde{C}_m = C^k)$ den Wert 1 liefert, wenn $\widetilde{C}_m = C^k$ gilt, und andernfalls den Wert 0 annimmt. Mit den Wahrscheinlichkeiten für alle K Klassen lässt sich für jede Teilmenge j die Entropie des Histogramms bestimmen:

$$\mathcal{E}_j = -\sum_{k=1}^{K} (p_j(C^k) \cdot log_2(p_j(C^k))). \tag{28}$$

Die Entropie $\mathcal{E}$ ist ein Maß für die Unsicherheit einer Zufallsvariablen und wurde bereits in Kapitel 3.2.5 bei der Merkmalsselektion eingeführt.

Der Informationszuwachs resultiert aus den gewichteten negativen Entropien beider Histogramme:

$$\mathcal{I} = -\frac{N_{M1}}{N_{M1} + N_{M2}} \cdot \mathcal{E}_1 + -\frac{N_{M2}}{N_{M1} + N_{M2}} \cdot \mathcal{E}_2. \tag{29}$$

Ein hoher Wert des Informationszuwachses $\mathcal{I}$ wird als Indikator für eine gute Trennung der Trainingsdaten interpretiert, sodass die Trennfläche mit dem größten $\mathcal{I}$ als Entscheidungsgrenze des Knoten genutzt wird. Mit der Entscheidungsgrenze werden die Trainingsbeispiele in zwei Teilmengen aufgespalten, die jeweils einen neuen Knoten bilden. Rekursiv können so die Trainingsbeispiele in immer kleinere Teilmengen aufgeteilt werden. Die rekursive Aufteilung endet, sobald ein Abbruchkriterium erreicht ist. Es entsteht somit ein Blatt anstelle eines neuen Knotens. Eine Möglichkeit für das Abbruchkriterium ist, dass die Rekursion endet, wenn nur noch Trainingsbeispiele einer Klasse in das Blatt fallen. Da dies jedoch zur Überanpassung führen kann, wird als Abbruchkriterium auch eine maximale Tiefe T_{max} des Baumes, eine minimale Anzahl an Trainingsbeispielen N_{min} für einen Knoten oder ein Minimalwert für den Informationszuwachs genutzt.

Für die Zuweisung einer Klasse bzw. einer Wahrscheinlichkeitsverteilung zu einem Blatt durchlaufen nun die Trainingsbeispiele in $\mathcal{T}_c$ den Baum bis zu einem Blatt. Für jedes Blatt κ wird mit Hilfe der Trainingsbeispiele, die das Blatt erreichen, ein normiertes Histogramm bestimmt. Hierfür werden zunächst die Trainingsbeispiele einer bestimmten Klasse in κ gezählt:

$$V_{b\kappa}^k = \sum_{n \in \kappa} \delta(\widetilde{C}_n = C^k). \tag{30}$$

Anschließend werden alle $V_{b\kappa}^k$ über alle Trainingsbeispiele in κ normiert. Wie zuvor im Training bei Gleichung 27 kann auch hier der Histogrammeintrag als Wahrscheinlichkeit $p_{b\kappa}(C^k)$ für eine Klasse C^k interpretiert werden. Alternativ kann auch eine direkte Zuordnung der Klasse mit der größten Wahrscheinlichkeit zu dem Blatt erfolgen.

Bei der Klassifikation eines neuen Primitivs n mit dem Merkmalsvektor $\mathbf{x}_n$ durchläuft dieser ebenfalls den Baum, wobei in jedem Knoten der Merkmalsvektor getestet und in Abhängigkeit von dem Testergebnis in dem Knoten an das linke oder rechte Kind weiter gereicht wird. Die

Weiterleitung des Merkmalsvektors endet, sobald dieser auf ein Blatt trifft. Dort wird ihm nun direkt das zuvor bestimmte Label des Blattes zugewiesen oder die Wahrscheinlichkeiten $p_b(C_n = C^k|\mathbf{x}_n) = p_{b\kappa}(C^k)$ für jede Klasse C^k.

Da das Training vom Zufall abhängt, können die Ergebnisse des Trainings selbst bei Nutzung der gleichen Trainingsdaten variieren. Als Beispiel sind in Abb. 3 zwei mögliche Aufteilungen eines zweidimensionalen Merkmalsraumes mit vier Klassen zu sehen. Im Allgemeinen können sich bei einem erneuten unabhängigen Training anhand des gleichen Datensatzes nicht nur die Parameter der Tests unterscheiden, wie in diesem Beispiel, sondern auch der Aufbau des Entscheidungsbaumes. Durch das unterschiedliche Ergebnis des Trainings kann auch das Klassifikationsergebnis variieren, wie z.B. in Abb. 3 bei der Klassifikation des schwarz dargestellten Punktes.

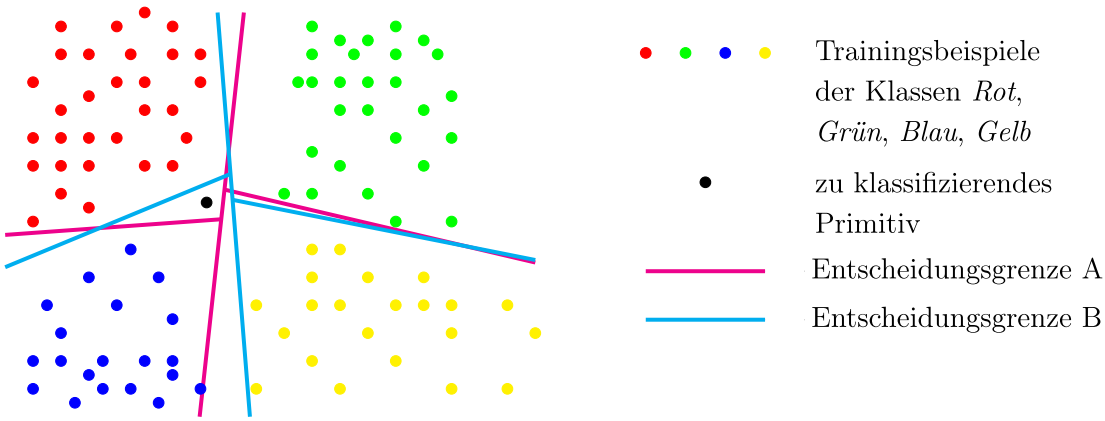


Abbildung 3: Zwei mögliche Aufteilungen eines 2D Merkmalsraumes mit Hilfe des CART Algorithmuss (A und B) in die Klassen *Rot*, *Grün*, *Blau* und *Gelb*. Das unklassifizierte Primitiv (schwarzer Punkt) wird, abhängig von dem gewähltem Baum, entweder der Klasse *Rot* oder *Blau* zugeordnet.

CART sind binäre Entscheidungsbäume, deren Aufbau und Parameter mit Hilfe von Trainingsdaten erlernt werden. Vorteile des CART-Algorithmus sind, dass keine Annahmen über die Verteilung der Merkmale und Klassen getroffen werden müssen und dass der Algorithmus auch mit teilweise fehlerhaften oder unvollständigen Daten umgehen kann [Hänsch und Hellwich, 2017]. Auch bei Klassifikationsaufgaben in denen mehr als zwei Klassen unterschieden werden sollen (Mehrklassenprobleme) kann der CART ohne weitere Anpassungen verwendet werden. Ein großer Nachteil bei der Verwendung von Entscheidungsbäumen ist deren starke Neigung zur Überanpassung, welche durch die hohe Modellkomplexität und einfache Abbruchkriterien entstehen kann [Süße und Rodner, 2014]. Eine Möglichkeit dieser Überanpassung entgegenzuwirken ist das sogenannte *Pruning*, bei dem der Baum zurückgeschnitten wird [Süße und Rodner, 2014]. Eine weitere Möglichkeit, die Überanpassung zu verringern, ist eine weitere Randomisierung durch die Verwendung von Bootstrap-Datensätzen [Süße und Rodner, 2014]. Hierbei werden mehrere Bäume an zufällig ausgewählten Trainingsdaten erlernt und anschließend kombiniert. Der so entstehende Random Forest wird im folgenden Kapitel 3.6 erläutert.

3.6. Random Forest

Der *Random Forest* (dt. *Zufallswald*; RF) [Breiman, 2001] ist ein diskriminativer nichtprobabilistischer Klassifikator, der aus N_B CART (s. Kapitel 3.5) besteht. Dementsprechend kann auch hier das Klassifikationsergebnis probabilistisch interpretiert und für eine weiterführende Prozessierung, z.B. innerhalb eines CRF (Kapitel 3.7), genutzt werden.

Wie zuvor beschrieben besteht bei dem Training von Entscheidungsbäumen die Gefahr einer Überanpassung an den Trainingsdatensatz. Dem wird im RF entgegengewirkt, indem mehrere Entscheidungsbäume trainiert und anschließend für die Klassifikation kombiniert werden. Um unabhängige Entscheidungsbäume zu trainieren und somit die Diversität der Bäume sicherzustellen, erfolgt das Training jedes CART mit einem Bootstrap-Datensatz $\mathcal{T}_b$ [Süße und Rodner, 2014]. Bei der Generierung des Bootstrap-Datensatzes werden N_{Tb} zufällig gewählte Trainingsbeispiele aus der vorhandenen Menge an Trainingsbeispielen $\mathcal{T}$ mit Zurücklegen gezogen. Die Ziehung der Bootstrap-Datensätze erfolgt aus der gleichen Menge an Trainingsbeispielen $\mathcal{T}$ und unabhängig voneinander. Ein Trainingsbeispiel kann somit sowohl mehrmals in einem Bootstrap-Datensatz $\mathcal{T}_b$ als auch in mehreren Bootstrap-Datensätzen $\mathcal{T}_b$ vorkommen. Die Nutzung von Bootstrap-Datensätzen zur Klassifikation wird auch *Bagging* (**b**ootstrap **agg**regat**ing**) genannt [Breiman, 1996].

Auf jedem Bootstrap-Datensatz wird nun ein Entscheidungsbaum b, wie in Kapitel 3.5 beschrieben, angelernt. Die Klassifikation eines Primitivs n mit dem Merkmalsvektor $\mathbf{x}_n$ erfolgt durch die Kombination der Ergebnisse aller Bäume des RF. Hierfür wird das zu klassifizierende Primitiv durch jeden Entscheidungsbaum geschleust. Wie in Kapitel 3.5 beschrieben, liefert jeder Baum b eine a posteriori Wahrscheinlichkeit $p_b(C_n = C^k|\mathbf{x}_n)$ für alle Klassen C^k. Diese Wahrscheinlichkeiten werden zu einer Gesamtwahrscheinlichkeit $p(C_n = C^k|\mathbf{x}_n)$ kombiniert [Hänsch und Hellwich, 2017]:

$$p(C_n = C^k|\mathbf{x}_n) = \frac{1}{N_B} \sum_{b=1}^{N_B} p_b(C_n = C^k|\mathbf{x}_n). \tag{31}$$

Das Klassenlabel kann nun nach dem Maximum a posteriori Kriterium (s. Kapitel 3.1) aus der a posteriori Wahrscheinlichkeit abgeleitet werden.

Alternativ kann für jeden Baum b individuell, entsprechend der Wahrscheinlichkeitsverteilung in dem Blatt, eine Entscheidung für ein Klassenlabel C^b getroffen werden. Die Gesamtwahrscheinlichkeit für eine Klasse C^k ergibt sich aus der Summe aller Bäume, deren Entscheidung das gleiche Klassenlabel aufweist $C^k = C^b$, dividiert durch die Anzahl der Bäume N_B [Hänsch und Hellwich, 2017]:

$$p(C_n = C^k|\mathbf{x}_n) = \frac{1}{N_B} \sum_{b=1}^{N_B} \delta(C^b = C^k). \tag{32}$$

Dies entspricht einer Abstimmung nach dem Mehrheitsprinzip, bei dem, anders als bei der Kombination der a posteriori Wahrscheinlichkeiten aus jedem Baum, die Unsicherheit bei der Schätzung des Klassenlabels innerhalb eines Baumes bei der Klassifikation nicht berücksichtigt wird [Hänsch und Hellwich, 2017].

Durch das Bootstrapping kann der RF bereits mit einem gewissen Grad an fehlerhaften Trainingslabels umgehen [Pelletier et al., 2017], was besonders von Vorteil ist, wenn die Trainingslabels aus einer veralteten Karte stammen und somit fehlerbehaftet sind. Auch gegenüber Ausreißern und Rauschen in den Daten ist der RF relativ robust [Breiman, 2001]. Laut Hänsch und Hellwich [2017] gäbe es für die Einstellung der Parameter gute Heuristiken und daher sei der RF, im Gegensatz zu neuronalen Netzen, auch von Laien gut nutzbar. Außerdem kann das Lernen gut parallelisiert werden, wodurch der Klassifikator relativ recheneffizient ist [Hänsch und Hellwich, 2017]. Positiv für eine weitere Verarbeitung der Ergebnisse in einem probabilistischen Kontext ist, dass das Klassifikationsergebnis probabilistisch interpretiert werden kann. In dieser Arbeit werden die resultierenden Gesamtwahrscheinlichkeiten $p(C_n = C^k|\mathbf{x}_n)$ beispielsweise innerhalb eines CRF weiter genutzt (s. Kapitel 3.7).

3.7. Conditional Random Fields

Conditional Random Fields (dt. *Bedingte Zufallsfelder*; CRF; [Kumar und Hebert, 2006]) sind ungerichtete graphische Modelle, die zur kontextbasierten Klassifikation genutzt werden können. Sie bestehen aus *Knoten*, welche Zufallsvariablen repräsentieren, und *Kanten*, welche die statistische Abhängigkeit zwischen jeweils zwei Knoten beschreiben. Die Knoten können sowohl beobachtet sein, wie z.B. die Merkmalsvektoren $\mathbf{x}$, oder unbeobachtet, wie die Klassenlabels der Bildprimitive C. Die Integration von Kontext erfolgt durch die Verbindung der unbeobachteten Labels benachbarter Primitive durch Kanten.

In der Klassifikation wird nun die optimale Labelkonfiguration $\mathbf{C} = (C_1, ..., C_N)^T$ des gesamten Bildes abhängig von allen beobachteten Daten $\mathbf{d}$ simultan bestimmt. Da CRF zu den diskriminativen Klassifikationsverfahren gehören, wird die a posteriori Wahrscheinlichkeit $p(\mathbf{C}|\mathbf{d})$ direkt modelliert [Kumar und Hebert, 2006]:

$$p(\mathbf{C}|\mathbf{d}) = \frac{1}{\mathcal{Z}}\Big(\prod_{n\in\mathcal{S}} \phi(C_n, \mathbf{d}) \cdot \prod_{n\in\mathcal{S}} \prod_{m\in\mathcal{O}_n} \psi(C_n, C_m, \mathbf{d})\Big). \tag{33}$$

$\mathcal{S}$ ist hierbei die Menge aller Bildprimitive und $\mathcal{O}_n$ die Menge der benachbarten Knoten von Primitiv n. $\phi(C_n, \mathbf{d})$ wird als *Assoziationspotential* oder *unäres Potential* bezeichnet und $\psi(C_n, C_m, \mathbf{d})$ als *Interaktionspotential* oder *binäres Potential*. Die Partitionsfunktion $\mathcal{Z}$ ist eine Normalisierungskonstante, um die Potentiale in Wahrscheinlichkeiten umzuwandeln. Häufig wird auch, z.B. aus numerischen Gründen, die Berechnung mit Summen statt Produkten verwendet, wobei jede Darstellungsform in die jeweils andere konvertierbar ist:

$$p(\mathbf{C}|\mathbf{d}) = \frac{1}{\mathcal{Z}} exp\Big(\sum_{n\in\mathcal{S}} A(C_n, \mathbf{d}) + \sum_{n\in\mathcal{S}} \sum_{m\in\mathcal{O}_n} I(C_n, C_m, \mathbf{d})\Big). \tag{34}$$

Hierbei gilt $A(C_n, \mathbf{d}) = ln(\phi(C_n, \mathbf{d}))$ und $I(C_n, C_m, \mathbf{d}) = ln(\psi(C_n, C_m, \mathbf{d}))$. Im Folgenden werden nur die logarithmierten Potentiale A und I verwendet, die zur Vereinfachung auch, wie in der Literatur üblich, als Assoziationspotential und Interaktionspotential bezeichnet werden.

Das Assoziationspotential $A(C_n, \mathbf{d})$ verknüpft die beobachteten Daten $\mathbf{d}$ mit dem Label C_n des Primitivs n. Der Zusammenhang zu den beobachteten Merkmalen eines Primitivs n wird dabei oft durch einen Merkmalsvektor $\mathbf{x}_n(\mathbf{d})$ beschrieben, der eine Funktion der Merkmale aller oder bestimmter Primitive des Bildes sein kann. Für die Modellierung des Assoziationspotentials kann ein beliebiger diskriminativer Klassifikator verwendet werden, z.B. die in Kapitel 3.3 bis 3.6 vorgestellten Klassifikatoren. Ein mögliches Modell für das Assoziationspotential mit Hilfe der a posteriori Wahrscheinlichkeit des Klassifikators ist:

$$A(C_n, \mathbf{d}) = ln\big(p(C_n|\mathbf{x}_n(\mathbf{d}))\big). \tag{35}$$

Das Interaktionspotential $I(C_n, C_m, \mathbf{d})$ beschreibt den statistischen Zusammenhang zweier benachbarter Labels C_n und C_m unter der Berücksichtigung der beobachteten Daten $\mathbf{d}$. Eine mögliche Art, die Interaktion zu beschreiben, ist das kontrastsensitive Potts-Modell, dessen Verwendung zu einer datenabhängigen Glättung des Labelbildes führt [Boykov et al., 2001]:

$$I(C_n, C_m, \mathbf{d}) = \delta(C_n, C_m) \cdot \beta_0 \Big(\beta_1 + (1 - \beta_1) \cdot e^{-\frac{\Delta\mathbf{x}^2}{2\sigma_D^2}}\Big). \tag{36}$$

In Gleichung 36 ist $\Delta\mathbf{x} = \|\mathbf{x}_n(\mathbf{d}) - \mathbf{x}_m(\mathbf{d})\|$ die euklidische Distanz zweier Merkmalsvektoren $\mathbf{x}_n(\mathbf{d})$ und $\mathbf{x}_m(\mathbf{d})$, σ_D die mittlere Distanz und $\delta(\cdot)$ die Kronecker-Delta-Funktion. Zur Berechnung von $\Delta\mathbf{x}$ können auch nur die Bilddaten direkt verwendet werden, anstatt der gesamten Merkmalsvektoren. Dies ist besonders sinnvoll, wenn für die Bestimmung der Merkmale die Daten geglättet wurden, da in diesem Fall die euklidische Distanz relativ klein ausfallen würde.

Die Stärke der Glättung und der Einfluss des datenabhängigen Terms werden durch die Parameter β_0 bzw. β_1 beschrieben. Die Idee des kontrastsensitiven Potts-Modells ist, dass eine starke Glättung nur erfolgt, wenn die Daten homogen sind. Ist der Abstand $\Delta\mathbf{x}$ sehr klein und somit die Daten homogen, wird mit β_0 geglättet: Ist $\Delta\mathbf{x}$ sehr groß, erfolgt die Glättung mit $\beta_0 \cdot \beta_1$, was wegen $\beta_1 \in [0, 1]$ kleiner als β_0 ist.

Die Struktur des Graphen für eine klassische kontextbasierte monotemporale pixelbasierte Klassifikation mit einem CRF ist in Abb. 4 zu sehen. Die Kanten basieren auf einer Vierernachbarschaft jedes einzelnen Pixels.

Bei der *Inferenz* wird die optimale Labelkonfiguration $\hat{\mathbf{C}}$ bestimmt, wobei die Wahrscheinlichkeit $p(\mathbf{C}|\mathbf{d})$ der Klassenlabels bei gegebenen Daten zu maximieren ist:

$$\hat{\mathbf{C}} = \arg\max_{\mathbf{C}} \big(p(\mathbf{C}|\mathbf{d})\big). \tag{37}$$

Da eine exakte Bestimmung im Allgemeinen nicht möglich ist, wird für die Inferenz meist eine approximative Methode verwendet. Ein Beispiel ist die *Loopy Belief Propagation* (LBP) [Frey und MacKay, 1998], ein *Message Passing* (dt. *Nachrichten übertragender*) Algorithmus für zyklische Graphen. Bei der LBP wird die Randverteilung der Klassen an den Knoten iterativ angepasst, bis ein Abbruchkriterium erreicht ist, z.B. dass die Änderungen unter einem vorgegebenen Schwellwert liegen. Die Randverteilung der Klassen an einem Knoten n wird als *Belief* bezeichnet. Die Aktualisierung des Beliefs $Bel(C_n)$ beruht auf zwei Arten von

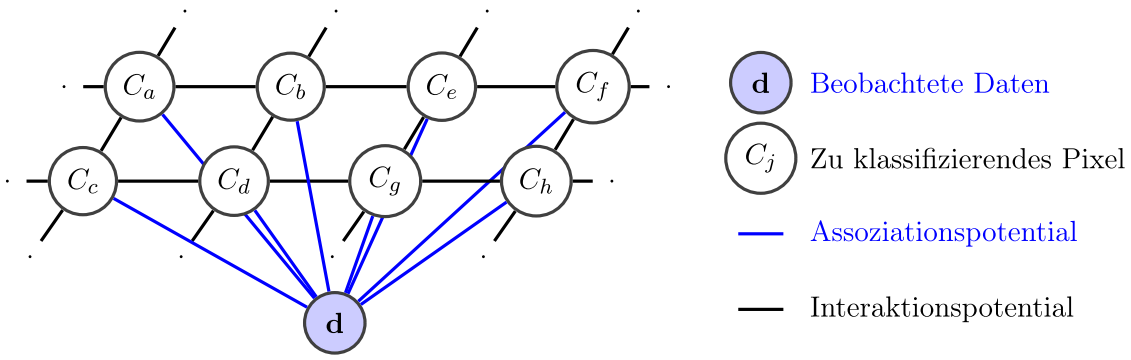


Abbildung 4: Graphenstruktur des klassischen CRF für die Klassifikation eines Bildes aus einem Zeitpunkt nach [Förstner, 2013].

Nachrichten: Der Nachricht von einem Knoten n zu einer Kante (n, m) und der Nachricht von einer Kante (n, m) zu einem Knoten n. Die LBP konvergiert nicht immer und, falls sie konvergiert, ist eine optimale Lösung nicht garantiert. Dennoch liefert das Verfahren meist gute Lösungen [Szeliski et al., 2008].

CRF stellen eine gute Möglichkeit dar, statistische Abhängigkeiten zwischen Klassenlabels von neben einander liegenden Bildprimitiven in die Klassifikation zu integrieren und somit lokalen Kontext zu berücksichtigen. Der Einfluss von fehlerhaften Trainingslabels hängt von der Modellierung der Potentiale ab. Die Parameter des in dieser Arbeit gewählten kontrastsensitiven Potts-Modells für das Interaktionspotential werden beispielsweise nicht durch Training bestimmt. Das Assoziationspotential kann mit Hilfe eines beliebigen diskriminativen Klassifikators modelliert werden, sodass auch ein Klassifikator, der Fehler in den Trainingslabels berücksichtigt, wie die robuste logistische Regression (Kapitel 3.4), verwendet werden kann.

4. Methodik

In diesem Kapitel wird die neue Methode zur Klassifikation von Fernerkundungsbildern mit Hilfe einer veralteten Karte dargestellt. Zunächst erfolgt in Kapitel 4.1 ein Überblick über den iterativen Algorithmus zur Klassifikation. Das Training und die Klassifikation des robusten Random Forest, mit dem der Zusammenhang zwischen den Daten und den unbekannten Labels in dem Algorithmus modelliert wird, zeigt Kapitel 4.2. Darauf folgend wird in Kapitel 4.3 die Integration der vorhandenen Labels der veralteten Karte als Beobachtungen in den Klassifikationsprozess erläutert. Die Bestimmung der Übergangswahrscheinlichkeiten, die sowohl für das Training des robusten Random Forest als auch für die Berücksichtigung der beobachteten Labels in der Klassifikation mit dem CRF benötigt werden, wird in Kapitel 4.4 beschrieben. Kapitel 4.5 beschäftigt sich mit der iterativen Aktualisierung der Änderungswahrscheinlichkeiten, welche ebenfalls für das Training des robusten Random Forest benötigt werden, und der Gewichte zur Regulierung des Einflusses der beobachteten Labels auf das Klassifikationsergebnis. Alternativ zu der Verwendung des robusten Random Forest in dem iterativen Klassifikationsprozess wird in Kapitel 4.6 ein Algorithmus vorgestellt, welcher die robuste logistische Regression zur Modellierung des Zusammenhanges zwischen den Labels und den Daten verwendet. Die Gewichtung der beobachteten Labels in dem CRF wird weiterhin iterativ angepasst. Zuletzt wird in Kapitel 4.7 ein multitemporaler Klassifikationsansatz beschrieben.

4.1. Überblick des Klassifikationsalgorithmus

In diesem Kapitel wird der Arbeitsablauf für eine monotemporale Klassifikation vorgestellt, welche sowohl den Einfluss fehlerhafter Labels im Training anhand eines Fehlermodells reduziert, als auch diese Kartenlabels als Beobachtungen mit in den Klassifikationsprozess integriert. Alle einzelnen Elemente des iterativen Prozesses werden anschließend im Detail erläutert. In diesem Kapitel erfolgt zunächst eine Darstellung des Zusammenhangs der Elemente und des Gesamtablaufes.

Gegeben sind die Daten $\mathbf{d}$, aus denen für jedes Pixel n ein Merkmalsvektor $\mathbf{x}_n(\mathbf{d})$ abgeleitet wird. Für eine bessere Übersicht wird $\mathbf{x}_n(\mathbf{d})$ mit $\mathbf{x}_n$ abgekürzt. Außerdem ist ein Label $\widetilde{C}_n$ aus einer vorhandenen Karte bekannt. Das Ziel der Klassifikation ist es, jedem Pixel ein diskretes Label C_n zuzuweisen. Da die Karte ggf. veraltet ist, kann sich das Label C_n bei einigen Pixeln von dem Label der Karte $\widetilde{C}_n$ unterscheiden.

Der hier vorgestellte Klassifikationsprozess basiert auf dem in dieser Arbeit entwickelten *robusten Random Forest* (RRF, Kapitel 4.2), ist iterativ und baut auf folgenden Annahmen auf:

- Änderungen treten in Segmenten im Bildraum auf.
- Pixel, deren klassifiziertes Label sich in mehreren Iterationsschritten von dem Label der veralteten Karte unterscheidet, sind mit einer hohen Wahrscheinlichkeit fehlerhaft in der Karte.

Jeder Iterationsschritt des Klassifikators kann in drei Abschnitte unterteilt werden:

a. Das Training des RRF und anschließend die Bestimmung der a posteriori Wahrscheinlichkeit $p(C_n|\mathbf{x}_n)$ für jeden beobachteten Merkmalsvektor $\mathbf{x}_n$.

b. Die kontextbasierte Klassifikation basierend auf dem CRF. An dieser Stelle wird nicht nur der räumliche Kontext berücksichtigt, sondern auch der statistische Zusammenhang zwischen den zu klassifizierenden unbekannten Labels C_n und den beobachteten Labels der Karte $\widetilde{C}_n$ in Form von Übergangswahrscheinlichkeiten $p(C|\widetilde{C})$. Dies wird durch ein zusätzliches temporales Assoziationspotential A_z im CRF realisiert. Durch eine Gewichtung θ_n des temporalen Assoziationspotentials jedes Pixels n wird der Einfluss dieses Potentials auf das Klassifikationsergebnis modelliert.

c. Die Bestimmung der Wahrscheinlichkeit g_n, dass eine Änderung des Labels vorliegt, für jedes Primitiv n. Die Änderungswahrscheinlichkeit g_n wird, wie die Gewichtung θ_n, in jeder Iteration aktualisiert. Für die Aktualisierung von g_n und θ_n werden die Labels aus der Klassifikation mit den Labels der veralteten Karte verglichen und so Regionen mit potentiellen Änderungen extrahiert.

Abgekürzt wird dieser Algorithmus in der Arbeit mit dem Akronym *RRFK* (**R**obuster **R**andom **F**orest mit **K**artenbeobachtung).

Zum Training werden die gegebenen Merkmalsvektoren $\mathbf{x}_n$ mit den beobachteten Labels $\widetilde{C}_n$ verwendet. Demnach wird mit den gleichen Merkmalsvektoren trainiert und klassifiziert. Es ist auch möglich, nur mit einer Teilmenge der gegebenen Merkmalsvektoren zu trainieren, solange diese repräsentativ für das ganze zu klassifizierende Gebiet sind.

Minimalerweise müssen dementsprechend für das Interessensgebiet bildbasierte Sensordaten vorliegen, wie beispielsweise ein Orthophoto, und die Labels einer ggf. veralteten Karte. Dabei muss, beispielsweise durch eine korrekte Georeferenzierung der Sensordaten und der Karte, gewährleistet sein, dass jedem Primitiv bzw. Pixel der Daten das entsprechende Label aus der veralteten Karte eindeutig und korrekt zugeordnet werden kann.

4.1.1. Initialisierung

Der Initialisierungsprozess ist in Abb. 5 dargestellt.

Für die Initialisierung wird der Random Forest (RF, Kapitel 3.6) verwendet, da für das Training des RRF die Beliefs $Bel(C_n|\mathbf{d})$ der vorherigen Iteration, die Übergangswahrscheinlichkeit $p(\widetilde{C}|C)$ und die Änderungswahrscheinlichkeit g_n verwendet werden, welche zunächst jedoch unbekannt sind. Der RF wird zunächst mit den Trainingsdaten trainiert (Abb. 5, türkis, tRF) und anschließend wird mit dem trainierten Klassifikator die a posteriori Wahrscheinlichkeit $p(C_n|\mathbf{x}_n)$ für jeden Merkmalsvektor bestimmt (Abb. 5, türkis, RF).

Die a posteriori Wahrscheinlichkeiten $p(C_n|\mathbf{x}_n)$ werden für die Modellierung des Assoziationspotentials eines CRF verwendet (Kapitel 3.7) (Abb. 5, schwarz). Für das kontrastsensitive Potts-Modell, welches für das Interaktionspotential verwendet wird, werden die Daten in Form

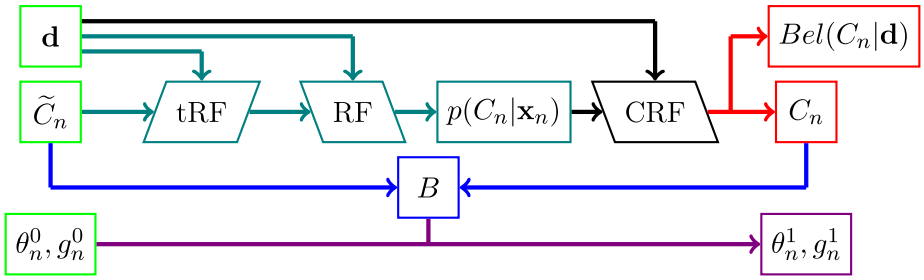


Abbildung 5: Initialisierungsprozess für den **RRFK** basierend auf dem RF. Rechteckige Boxen: Variablen. Parallelogramme: Prozesse. Parallelogramm mit spitzem Winkel oben rechts: Trainingsprozess. Die Farben in der Abbildung werden im Text aufgegriffen.

der Merkmalsvektoren benötigt (Abb. 5, schwarz). Aus der Klassifikation mit Hilfe des CRF resultieren die Klassenlabels $\mathbf{C} = (C_1, ..., C_N)^T$ des ganzen Bildes und die Beliefs $Bel(C_n|\mathbf{d})$ für jedes Primitiv bzw. Pixel (Abb. 5, rot).

Die Labels aus der Klassifikation werden anschließend mit den Labels der veralteten Karte verglichen. Aus dem Vergleich lässt sich ein binäres Änderungsbild B erstellen, welches für jedes Pixel n anzeigt, ob eine Änderung des Labels vorliegt ($C_n \neq \widetilde{C}_n$) (Abb. 5, blau). Eine solche Änderung des Labels kann sowohl darauf zurück zu führen sein, dass eine Änderung vorliegt und der Klassifikator das Pixel korrekt einer neuen Klasse zugeordnet hat, als auch darauf, dass der Klassifikator das Pixel einer falschen Klasse zugeordnet hat, obwohl gar keine Änderung vorliegt. Mit Hilfe von Heuristiken werden vermeintliche Änderungen, die vermutlich auf Fehlklassifikationen beruhen, von den korrekten Änderungen unterschieden. Diese vermutlich fehlerhaften Änderungen werden nicht in dem Änderungsbild dargestellt.

Das Änderungsbild B wird verwendet, um die Wahrscheinlichkeit für eine Änderung g_n und die Gewichtung θ_n zu aktualisieren (Abb. 5, lila). Für die Änderungswahrscheinlichkeit wird der Startwert $g_n^0 = \frac{1}{K}$ gewählt. Das Gewicht des temporalen Assoziationspotentials A_z wird mit $\theta_n^0 = 0 \ \forall n$ initialisiert, damit das Klassenlabel zunächst keinen Einfluss auf das Klassifikationsergebnis hat. Die Änderungen werden dadurch besser erkannt, als wenn das veraltete Klassenlabel bereits berücksichtigt wird.

Zusammengefasst entspricht die Initialisierung einer Klassifikation mit Hilfe des CRF unter Verwendung des RF zur Modellierung des Zusammenhanges zwischen den Daten und den Labels. Anschließend erfolgt eine Änderungsdetektion, basierend auf dem Vergleich des Labelbildes der Klassifikation mit den Labels einer verwendeten Karte. Obwohl die Labels für das Training aus einer veralteten Karte stammen, werden in der Initialisierung Fehler in den Labels nicht berücksichtigt. Es wird daher angenommen, dass das resultierende Labelbild und auch die Änderungsdetektion fehlerhaft sind. Aus diesem Grund werden, anders als bei der Datenbereinigung, keine Trainingsbeispiele entfernt oder umgelabelt. Stattdessen wird die Änderungswahrscheinlichkeit g_n, welche im Training der Folgeiteration genutzt wird, und die Gewichtung θ_n, welche den Einfluss des beobachteten Labels aus der Karte auf das Ergebnis der Klassifikation modelliert, angepasst.

4.1.2. Iterativer Prozess

Abb. 6 zeigt den Ablauf des iterativen Prozesses nach der Initialisierung.

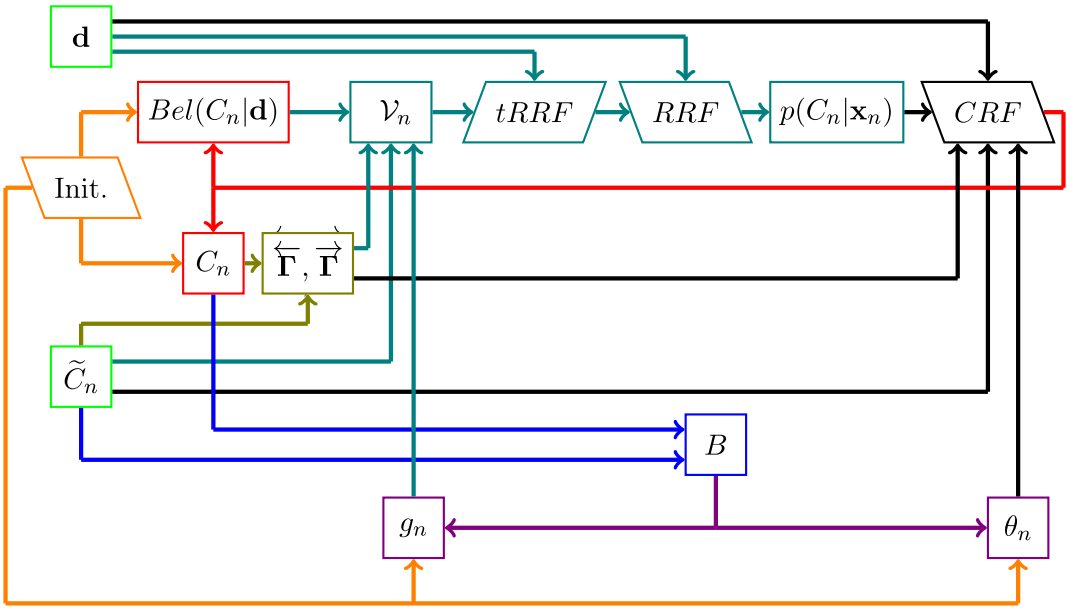


Abbildung 6: Iterativer Prozessablauf des **RRFK**. Rechteckige Boxen: Variablen. Parallelogramme: Prozesse. Parallelogramm mit spitzem Winkel oben rechts: Trainingsprozess. Init: Initialisierungsprozess. Die Farben in der Abbildung werden im Text aufgegriffen.

Zunächst werden mit den Labels der Karte und den Labels aus der Klassifikation die Übergangswahrscheinlichkeiten $\overleftarrow{\gamma} = p(\widetilde{C}|C)$ und $\overrightarrow{\gamma} = p(C|\widetilde{C})$ bestimmt, welche die Übergangsmatrizen $\overleftarrow{\mathbf{\Gamma}}$ und $\overrightarrow{\mathbf{\Gamma}}$ bilden (Abb. 6, gelbgrün, Abschnitt 4.4, Gleichungen 52 und 53).

Mit Hilfe der Änderungswahrscheinlichkeiten g_n, der zuvor berechneten Übergangsmatrix $\overleftarrow{\mathbf{\Gamma}}$ und den Beliefs $Bel(C_n|\mathbf{d})$ der vorherigen Iteration wird für jedes Pixel n die Wahrscheinlichkeit der Klassenzugehörigkeit $\mathcal{V}_n^a = p(C_n = C^a|\mathbf{x}_n, \widetilde{C}_n)$ bezüglich jeder Klasse a berechnet (Kapitel 4.2.2, Geichung 43). $\mathcal{V}_n^k$ beschreibt die Wahrscheinlichkeit, dass ein Pixel n zu der Klasse k gehört. Diese Wahrscheinlichkeiten der Klassenzugehörigkeit $\mathcal{V}_n = p(C_n|\mathbf{x}_n, \widetilde{C}_n)$ werden aus den Wahrscheinlichkeiten $p(C_n|\mathbf{x}_n)$ und $p(\widetilde{C}_n|C_n)$ probabilistisch abgeleitet (Kapitel 4.2.2). Anstatt der unbekannten Wahrscheinlichkeiten $p(C_n|\mathbf{x}_n)$ werden die Beliefs $Bel(C_n|\mathbf{d})$ der vorherigen Iteration verwendet. Die individuellen Übergangswahrscheinlichkeiten $p(\widetilde{C}_n|C_n)$ werden aus den Änderungswahrscheinlichkeiten g_n und den Übergangswahrscheinlichkeiten $\overleftarrow{\gamma}$ angenähert. Diese Modellierung der Wahrscheinlichkeiten der Klassenzugehörigkeit hebt sich dabei von bestehenden Arbeiten ab, da die Wahrscheinlichkeit für eine Klasse C hier von den Daten und den Kartenlabels abhängt, anstatt nur von einer der Größen.

Nach der Modellierung von $\mathcal{V}_n$ folgt das Training des RRF (Abb. 6, türkis, tRRF). Dieses Training verwendet ausschließlich die Wahrscheinlichkeiten der Klassenzugehörigkeit $\mathcal{V}_n$ und die Merkmalsvektoren $\mathbf{x}_n$ aus den Daten $\mathbf{d}$ (Kapitel 4.2.1, 4.2.4). Die beobachteten Labels $\widetilde{C}_n$ fließen nur indirekt über $\mathcal{V}_n$ in das Training mit ein. Nach dem Training kann für jeden

Merkmalsvektor $\mathbf{x}_n$ die a posteriori Wahrscheinlichkeit $p(C_n|\mathbf{x}_n)$ ermittelt werden (Abb. 6, türkis, RRF).

Die Modifikation des RF Klassifikators, dass dieser mit Trainingsbeispielen trainiert werden kann, die anstelle eines einzigen Labels Wahrscheinlichkeiten für die Klassenzugehörigkeit *jeder* Klasse besitzen, ist eine weitere Innovation dieser Arbeit. Der Vorteil hierbei liegt darin, dass Unsicherheiten bezüglich des Labels der Trainingsbeispiele mit modelliert werden können, anstatt annehmen zu müssen, dass die Labels korrekt sind.

Die a posteriori Wahrscheinlichkeiten $p(C_n|\mathbf{x}_n)$ werden anschließend für die Modellierung des statistischen Zusammenhangs zwischen den zu bestimmenden Labels und den Daten innerhalb eines CRF verwendet (Abb. 6, schwarz). Zusätzlich wird in dem CRF auch der Zusammenhang zwischen den Kartenlabels und den zu bestimmenden Labels durch ein temporales Assoziationspotential berücksichtigt. In anderen Veröffentlichungen wurden temporale Zusammenhänge bereits mit Hilfe eines CRF modelliert [Hoberg et al., 2015], jedoch sind dort die Labels beider Zeitpunkte unbekannt, während in dieser Arbeit die Labels aus der veralteten Karte bereits beobachtet sind. Ein weiterer Unterschied zu [Hoberg et al., 2015] liegt in der Bestimmung der verwendeten Übergangswahrscheinlichkeiten $\overrightarrow{\gamma}$. In dieser Arbeit werden sie am Anfang jeder Iteration aus den Ergebnissen der vorherigen Iteration und den Kartenlabels abgeleitet. In Hoberg et al. [2015] hingegen werden sie vor der Klassifikation empirisch bestimmt. Außerdem wird der Einfluss der Kartenlabels in der hier vorgestellten Arbeit mit θ_n gewichtet. In Hoberg et al. [2015] werden die Kartenlabels nicht verwendet.

Aus der Inferenz des CRF resultieren sowohl die Labels des gesamten Bildes $\mathbf{C}$ als auch die Beliefs $Bel(C_n = C^k|\mathbf{d})$ jeder Klasse k für jedes Pixel n (Abb. 6, rot). Wie bei der Initialisierung werden nun die Labels C_n mit den Labels $\widetilde{C}_n$ der Karte verglichen, um das Binärbild B, welches die Änderungen beinhaltet, zu erstellen. Um Labelunterschiede, die aufgrund einer fehlerhaften Klassifikation entstanden sind, von tatsächlichen Änderungen zu unterscheiden, werden Heuristiken angewendet. Hierbei wird auch berücksichtigt, dass tatsächliche Änderungen Segmente im Bildraum bilden. Anschließend werden die Wahrscheinlichkeiten für eine Änderung g_n und die Gewichte θ_n mithilfe des Änderungsbildes B aktualisiert (Abb. 6, lila).

Nun kann die nächste Iteration mit der Bestimmung der Übergangswahrscheinlichkeiten $\overleftarrow{\gamma}$ und $\overrightarrow{\gamma}$, der Wahrscheinlichkeiten der Klassenzugehörigkeit $\mathcal{V}_n$ und dem anschließenden Training des RRF beginnen. Der iterative Prozess wird beendet, sobald ein Abbruchkriterium erreicht ist. Dies kann beispielsweise durch eine maximale Anzahl an Iterationen N_{it} oder eine maximale Anzahl an Pixeln, deren Änderungswahrscheinlichkeit g_n oder Gewicht θ_n sich ändern, definiert werden.

4.2. Robuster Random Forest

Der Random Forest (Kapitel 3.6) kann durch die Verwendung von Bootstrap-Datensätzen beim Training einen gewissen Grad an Fehlern in den Trainingslabels kompensieren, besonders, wenn das Auftreten der Fehler unabhängig von den Daten ist. Andernfalls hängt die Robustheit des RF von der Robustheit des Trainingsalgorithmus der CART (Kapitel 3.5) ab.

Die Entscheidungsgrenze, welche die Daten im Merkmalsraum innerhalb eines CART trennt, wird durch den Vergleich von N_G zufällig gewählten Trennflächen bestimmt. Der Vergleich basiert auf dem Informationszuwachs $\mathcal{I}$ (s. Gleichung 29), welcher wiederum mit Hilfe der Entropie $\mathcal{E}$ (s. Gleichung 28) berechnet wird. Wird die Auswahl der Entscheidungsgrenze aus den zufälligen Trennflächen durch die fehlerhaften Trainingslabels beeinflusst, verringert sich auch die Robustheit des RF. Ein Beispiel, in dem die Wahl der Entscheidungsgrenze durch die fehlerhaften Labels negativ beeinflusst wird, ist in Abb. 7 zu sehen. Dort sind die Trainingsbeispiele von zwei Klassen (*Orange*, *Blau*) in einem zweidimensionalen Merkmalsraum dargestellt, wobei einige der blauen Trainingsbeispiele fehlerhaft als *Orange* gekennzeichnet sind (Abb. 7, orange mit blauem Rand). Von den beiden zufällig generierten Trennlinien, die in Abb. 7 dargestellt werden, erreicht die rote Trennlinie den höchsten Informationszuwachs (Gleichung 29) und wird somit vom Algorithmus als Entscheidungsgrenze ausgewählt. Es ist jedoch erkennbar, dass die grüne Trennlinie, die ebenfalls getestet wird und zu einem geringeren Informationszuwachs führt, die wahren Labels besser trennt.

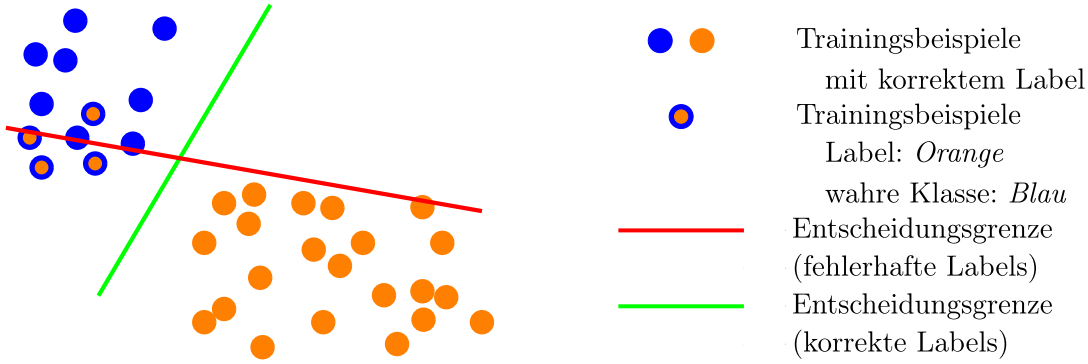


Abbildung 7: Beispiel für die Auswirkung von fehlerhaften Trainingsbeispielen auf die Klassifikation von zwei Klassen (*Blau* und *Orange*) in einem zweidimensionalen Merkmalsraum. Die rote Entscheidungsgrenze erreicht einen höheren Informationszuwachs als die grüne Entscheidungsgrenze. Die grüne Entscheidungsgrenze trennt die wahren Labels jedoch besser.

Die a posteriori Wahrscheinlichkeit $p(C_n = C^k|\mathbf{x}_n)$ in den Blättern wird ebenfalls beeinflusst. Fällt ein Trainingsbeispiel mit fehlerhaftem Label $\widetilde{C}_n = C^a$ und dem wahrem Label $C_n = C^k$ in das Blatt κ des Baumes b, trägt es zu der Summe $V^a_{b\kappa}$ bei, anstatt zu der Summe $V^k_{b\kappa}$ (Geichung 30), wodurch die resultierenden Wahrscheinlichkeiten $p_{b\kappa}(C_n = C^k)$ und $p_{b\kappa}(C_n = C^a)$ verfälscht werden. Würde in dem Beispiel aus Abb. 7 die grüne Trennebene gewählt und alle dargestellten Trainingsbeispiele für die Zuordnung einer Wahrscheinlichkeitsverteilung zu einem Blatt genutzt werden, erhielte das Blatt, welches die linke Seite repräsentiert, die Wahrscheinlichkeiten $p(C_n = C^{blau}|\mathbf{x}_n) = \frac{8}{12} \approx 0,67$ und $p(C_n = C^{orange}|\mathbf{x}_n) = \frac{4}{12} \approx 0,33$. Unter Verwendung der wahren Labels wären dies $p(C_n = C^{blau}|\mathbf{x}_n) = 1$ und $p(C_n = C^{orange}|\mathbf{x}_n) = 0$.

Aus diesem Grund wird der Random Forest Klassifikator in dieser Arbeit erweitert. Jedem Trainingsbeispiel n werden anstelle einer bestimmten Klasse alle Klassen mit einer bestimmten Wahrscheinlichkeit zugeordnet, der *Wahrscheinlichkeit der Klassenzugehörigkeit* $\mathcal{V}_n = p(C_n|\mathbf{x}_n, \widetilde{C}_n)$. Somit können Unsicherheiten der Trainingslabels berücksichtigt werden. Die Auswahl der Entscheidungsgrenze im Training des CART und die Berechnung der Wahr-

scheinlichkeitsverteilung in den Blättern müssen entsprechend angepasst werden, um dies zu berücksichtigen (Kapitel 4.2.1, Kapitel 4.2.4). Die Bestimmung dieser Wahrscheinlichkeit der Klassenzugehörigkeit $\mathcal{V}_n$ wird in Kapitel 4.2.2 beschrieben. Hierfür wird der Zusammenhang zwischen $\mathcal{V}_n = p(C_n|\mathbf{x}_n, \widetilde{C}_n)$ und den Wahrscheinlichkeiten $p(C_n|\mathbf{x}_n)$ und $p(\widetilde{C}_n|C_n)$ aufgestellt. Während für $p(C_n|\mathbf{x}_n)$ der Belief der vorherigen Iteration genutzt wird, wird $p(\widetilde{C}_n|C_n)$ mit Hilfe der Übergangsmatrix $\overleftarrow{\mathbf{\Gamma}}$ und der Änderungswahrscheinlichkeit g_n angenähert (Kapitel 4.2.3). Durch die Abhängigkeit der individuellen Übergangswahrscheinlichkeiten $p(\widetilde{C}_n|C_n)$ von g_n wird in den Wahrscheinlichkeiten der Klassenzugehörigkeit auch berücksichtigt, dass Änderungen in Segmenten im Bildraum auftreten. Der so erweiterte Random Forest Klassifikator wird in dieser Arbeit *robuster Random Forest* (RRF) genannt.

4.2.1. Auswahl der Entscheidungsgrenze

Eine Möglichkeit, um den Einfluss fehlerhafter Trainingslabels auf die Auswahl der Entscheidungsgrenze in dem Trainingsprozess des RF zu reduzieren, besteht darin, das Auftreten von Fehlern in den Trainingslabels bei der Berechnung des Informationszuwachses $\mathcal{I}$ (Geichung 29) zu berücksichtigen. Dies wird umgesetzt, indem jedes Trainingsbeispiel n nicht einer einzigen Klasse sondern jeder Klasse mit einer bestimmten Wahrscheinlichkeit $\mathcal{V}_n = p(C_n|\mathbf{x}_n, \widetilde{C}_n)$ zugeordnet wird. Diese Wahrscheinlichkeit ist abhängig von den Daten in Form der Merkmalsvektoren $\mathbf{x}_n$ und der beobachteten und ggf. fehlerhaften Labels aus der veralteten Karte $\widetilde{C}_n$.

Die einzigen Werte, die von den fehlerhaften Labels beeinflusst werden, sind die Wahrscheinlichkeiten $p_j(C^k)$ für die beiden Teilmengen $\mathcal{T}_{Mj}$ mit $j \in \{1, 2\}$, die aus der Aufteilung der Trainingsbeispiele $\mathcal{T}_i$ des Knotens i durch die Trennfläche resultieren. In dem ursprünglichen RF erfolgt die Berechnung dieser Wahrscheinlichkeiten, wie in Gleichung 27 beschrieben, mit

$$p_j(C^k) = \frac{1}{N_{Mj}} \sum_{n \in \mathcal{T}_{Mj}} \delta(\widetilde{C}_n = C^k), \tag{38}$$

wobei $\delta(\widetilde{C}_n = C^k)$ dem Wert 1 entspricht, wenn $\widetilde{C}_n = C^k$ und andernfalls den Wert 0 annimmt. Unter der Verwendung der beobachteten Labels zum Training würde diese Gleichung zu $p_j(\widetilde{C}^k)$ anstatt zu $p_j(C^k)$ führen. Die wahren Labels sind jedoch unbekannt. Stattdessen wird jedem Trainingsbeispiel jede Klasse mit einer bestimmten Wahrscheinlichkeit, der Wahrscheinlichkeit der Klassenzugehörigkeit $\mathcal{V}_n^k = p(C_n = C^k|\mathbf{x}_n, \widetilde{C}_n)$ zugeordnet. Diese werden aus den beobachteten Variablen geschätzt (Kapitel 4.2.2). Die Zuordnung eines Trainingsbeispieles zu einer der beiden Äste des Baumes M_1 oder M_2 bei der Evaluierung einer Entscheidungsgrenze, welche der Term $\delta(\widetilde{C}_n = C^k)$ in Gleichung 27 realisiert, wird dementsprechend angepasst. Für die Berechnung der Wahrscheinlichkeiten $p_j(C^k)$ ergibt sich nun:

$$p_j(C^k) = \frac{1}{N_{Mj}} \sum_{n \in \mathcal{T}_{Mj}} p(C_n = C^k|\mathbf{x}_n, \widetilde{C}_n) = \frac{1}{N_{Mj}} \sum_{n \in \mathcal{T}_{Mj}} \mathcal{V}_n^k. \tag{39}$$

Die Auswahl der Entscheidungsgrenze erfolgt dann auf Basis des Informationszuwachses $\mathcal{I}$ (Gleichung 29), wobei die darin verwendeten Entropien $\mathcal{E}_1$ und $\mathcal{E}_2$ (Gleichung 28) auf Ba-

sis der in Gleichung 39 vorgestellten Berechnung der Wahrscheinlichkeiten $p_j(C^k)$ beruhen. Die Berechnung der Wahrscheinlichkeit der Klassenzugehörigkeit $\mathcal{V}_n$ wird in Kapitel 4.2.2 erläutert.

4.2.2. Wahrscheinlichkeit der Klassenzugehörigkeit

Für die Bestimmung der Wahrscheinlichkeit der Klassenzugehörigkeit $\mathcal{V}_n = p(C_n|\mathbf{x}_n, \widetilde{C}_n)$ eines Pixels n wird die statistische Beziehung zwischen den drei beteiligten Variablen C_n, $\widetilde{C}_n$ und $\mathbf{x}_n$ betrachtet. Die gemeinsame Wahrscheinlichkeit $p(C_n, \mathbf{x}_n, \widetilde{C}_n)$ wird mit Hilfe eines Bayes'schen Netzes (Abb. 8) bestimmt. Variablen werden in dem Bayes'schen Netz als Knoten dargestellt. Beobachtete Variablen, hier das beobachtete Label $\widetilde{C}_n$ der Karte und der Merkmalsvektor $\mathbf{x}_n$, werden grau dargestellt und das unbekannte Klassenlabel C_n weiß. Durch Kanten, die zwei Knoten verbinden, wird der statistische Zusammenhang beider Variablen dargestellt. Gerichtete Kanten entsprechen dabei einer bedingten Wahrscheinlichkeit.

Bezogen auf die drei oben genannten Variablen ergibt sich das Bayes'sche Netz in Abb. 8. Hierbei wird angenommen, dass die Merkmalsausprägung in Form des Merkmalsvektors $\mathbf{x}_n$ von der Objektklasse bzw. dem Label C_n abhängt. Das aktuelle Label C_n hängt von den beobachteten Label $\widetilde{C}_n$ aus der Karte ab. Diese Abhängigkeit ist gegeben, da die Karte die Labels zu einem früheren Zeitpunkt darstellt. Einige Änderungen der Objekte und somit auch der Labels in der Zeit zwischen der Kartenerstellung und der Sensoraufnahme sind demnach wahrscheinlicher als andere. Ein direkter Zusammenhang zwischen den veralteten Labels und den aktuellen Merkmalsvektoren wird nicht angenommen.

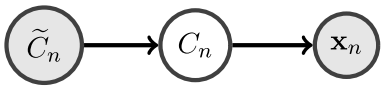


Abbildung 8: Bayes'sches Netz der Variablen: Merkmalsvektor $\mathbf{x}_n$, unbekanntes Klassenlabel C_n und beobachtetes Label $\widetilde{C}_n$. Graue Knoten: beobachtete Variablen. Weiße Knoten: unbeobachtete Variablen. Gerichtete Kanten: bedingte Wahrscheinlichkeit.

Entsprechend dem so entstandenen Graphen wird die gemeinsame Wahrscheinlichkeit $p(C_n, \mathbf{x}_n, \widetilde{C}_n)$ wie folgt faktorisiert [Bishop, 2006]:

$$p(C_n, \mathbf{x}_n, \widetilde{C}_n) = p(\mathbf{x}_n|C_n) \cdot p(C_n|\widetilde{C}_n) \cdot p(\widetilde{C}_n). \tag{40}$$

Der Zusammenhang zweier bedingter Wahrscheinlichkeiten $p(a|b)$ und $p(b|a)$ ist durch das Theorem von Bayes gegeben [Bishop, 2006]:

$$p(a|b) = \frac{p(b|a) \cdot p(a)}{p(b)}. \tag{41}$$

Mit dem Satz von Bayes kann die Gleichung 40 wie folgt umgeformt werden:

$$\begin{aligned} p(C_n, \mathbf{x}_n, \widetilde{C}_n) &= p(\mathbf{x}_n|C_n) \cdot p(C_n|\widetilde{C}_n) \cdot p(\widetilde{C}_n) \\ &= \frac{p(C_n|\mathbf{x}_n) \cdot p(\mathbf{x}_n)}{p(C_n)} \cdot \frac{p(\widetilde{C}_n|C_n) \cdot p(C_n)}{p(\widetilde{C}_n)} \cdot p(\widetilde{C}_n) \\ &= p(C_n|\mathbf{x}_n) \cdot p(\mathbf{x}_n) \cdot p(\widetilde{C}_n|C_n). \end{aligned}$$

Mit Hilfe der Bayesregel wird die benötigte posteriori Wahrscheinlichkeit bestimmt:

$$p(C_n|\mathbf{x}_n, \widetilde{C}_n) = \frac{p(C_n, \mathbf{x}_n, \widetilde{C}_n)}{p(\mathbf{x}_n, \widetilde{C}_n)} = \frac{p(C_n|\mathbf{x}_n) \cdot p(\widetilde{C}_n|C_n) \cdot p(\mathbf{x}_n)}{p(\mathbf{x}_n, \widetilde{C}_n)}. \tag{42}$$

Die Wahrscheinlichkeit $p(\mathbf{x}_n, \widetilde{C}_n)$ im Nenner der Gleichung 42 ist eine Randverteilung von $p(C_n, \mathbf{x}_n, \widetilde{C}_n)$ und kann aus der Summe über alle möglichen Werte von C_n bestimmt werden. Da $p(\mathbf{x}_n)$ unabhängig von C_n ist, kann es aus der Summe herausgehoben werden und kürzt sich mit dem gleichen Term im Zähler der Gleichung 42 heraus:

$$\begin{aligned} p(C_n|\mathbf{x}_n, \widetilde{C}_n) &= \frac{p(C_n|\mathbf{x}_n) \cdot p(\widetilde{C}_n|C_n) \cdot p(\mathbf{x}_n)}{p(\mathbf{x}_n, \widetilde{C}_n)} \\ &= \frac{p(C_n|\mathbf{x}_n) \cdot p(\widetilde{C}_n|C_n) \cdot p(\mathbf{x}_n)}{\sum_a \left(p(C_n = C^a|\mathbf{x}_n) \cdot p(\widetilde{C}_n|C_n = C^a) \cdot p(\mathbf{x}_n)\right)} \\ &= \frac{p(C_n|\mathbf{x}_n) \cdot p(\widetilde{C}_n|C_n) \cdot p(\mathbf{x}_n)}{p(\mathbf{x}_n) \cdot \sum_a \left(p(C_n = C^a|\mathbf{x}_n) \cdot p(\widetilde{C}_n|C_n = C^a)\right)} \\ &= \frac{p(C_n|\mathbf{x}_n) \cdot p(\widetilde{C}_n|C_n)}{\sum_a \left(p(C_n = C^a|\mathbf{x}_n) \cdot p(\widetilde{C}_n|C_n = C^a)\right)}. \end{aligned}$$

Die resultierende Wahrscheinlichkeit $p(C_n|\mathbf{x}_n, \widetilde{C}_n)$ leitet sich also aus der a posteriori Wahrscheinlichkeit $p(C_n|\mathbf{x}_n)$ und der individuellen Übergangswahrscheinlichkeit $p(\widetilde{C}_n|C_n)$ ab:

$$\mathcal{V}_n = p(C_n|\mathbf{x}_n, \widetilde{C}_n) = \frac{p(C_n|\mathbf{x}_n) \cdot p(\widetilde{C}_n|C_n)}{\sum_a \left(p(C_n = C^a|\mathbf{x}_n) \cdot p(\widetilde{C}_n|C_n = C^a)\right)}. \tag{43}$$

Für die a posteriori Wahrscheinlichkeiten $p(C_n|\mathbf{x}_n)$ werden die Beliefs $Bel(C_n|\mathbf{d})$ der vorherigen Iteration verwendet. Die individuellen Übergangswahrscheinlichkeiten $p(\widetilde{C}_n|C_n)$ werden im folgenden Kapitel 4.2.3 beschrieben.

4.2.3. Individuelle Übergangswahrscheinlichkeiten für jedes Trainingsbeispiel

Die Wahrscheinlichkeit der Klassenzugehörigkeit $\mathcal{V}_n$ eines Trainingsbeispiels n beinhaltet die Übergangswahrscheinlichkeit $p(\widetilde{C}_n|C_n)$. Die bereits vorgestellten Übergangswahrscheinlichkeiten $p(\widetilde{C}|C)$ aus der Übergangsmatrix $\overleftarrow{\mathbf{\Gamma}}$ können in diesem Fall nicht für jedes Trainings-

beispiel verwendet werden[2]. Obwohl dies aus statistischer Sicht korrekt ist, sagt die Übergangsmatrix $\overleftarrow{\mathbf{\Gamma}}$ nichts über die Wahrscheinlichkeit einer Änderung eines bestimmten beobachteten Labels $\widetilde{C}_n$ an einem bestimmten Trainingsbeispiel n aus. In vielen Fällen würden sonst die Übergangswahrscheinlichkeiten auch bei Trainingsbeispielen mit fehlerhaftem Label auf keine Änderung hindeuten, da eine Änderung im gesamten Gebiet eher selten vorkommt. Die Verwendung von $p(\widetilde{C}|C)$ ohne Modifikation führt demnach zu keiner Kompensation der fehlerhaften Trainingslabels. Daher wird eine individuelle Übergangsmatrix $\overleftarrow{\mathbf{\Gamma}}_n$ mit den Elementen $p(\widetilde{C}_n|C_n)$ für jedes Trainingsbeispiel n verwendet. Für die Annäherung dieser individuellen Übergangswahrscheinlichkeiten wird die Änderungswahrscheinlichkeit $g_n = p(\widetilde{C}_n \neq C_n)$ (Kapitel 4.5) verwendet. Die iterative Bestimmung der Änderungswahrscheinlichkeit basiert auf Heuristiken. Eine der Heuristiken basiert auf der Annahme, dass Pixel, deren Label sich geändert hat, in Segmenten im Bildraum auftreten. Die Wahrscheinlichkeit, dass sich die Klasse des Pixels n nicht ändert, ist $p(\widetilde{C}_n = C_n) = 1 - g_n$ und dient als Wert der Hauptdiagonalen der Übergangsmatrix $\overleftarrow{\mathbf{\Gamma}}_n$. Außerhalb der Hauptdiagonalen werden die Elemente von $\overleftarrow{\mathbf{\Gamma}}$ zunächst übernommen. Damit die Summe der Wahrscheinlichkeiten über alle Klassen 1 ergibt, erfolgt jedoch eine Skalierung der Werte außerhalb der Diagonalen. Die Berechnung der individuellen Übergangswahrscheinlichkeiten ist demnach wie folgt definiert:

$$p(\widetilde{C}_n = C^k | C_n = C^j) = \begin{cases} 1 - g_n & \text{wenn } k = j \\ g_n \cdot \frac{p(\widetilde{C}=C^k|C=C^j)}{\sum_{m=1}^{K} \delta(m \neq j) \cdot p(\widetilde{C}=C^m|C=C^j)} & \text{wenn } k \neq j, \end{cases} \tag{44}$$

wobei die Wahrscheinlichkeiten $p(\widetilde{C} = C^k | C = C^j)$, wie bereits erwähnt, aus der Übergangsmatrix $\overleftarrow{\mathbf{\Gamma}}$ stammen.

Die so entstandenen individuellen Übergangswahrscheinlichkeiten $p(\widetilde{C}_n|C_n)$ für jedes Trainingsbeispiel können nun für die Berechnung der Wahrscheinlichkeit der Klassenzugehörigkeit $\mathcal{V}_n$ (Kapitel 4.2.2) verwendet werden.

4.2.4. Wahrscheinlichkeit in den Blättern

Zur Berechnung der a posteriori Wahrscheinlichkeit $p(C_n|\mathbf{x}_n)$ eines Primitivs n müssen die a posteriori Wahrscheinlichkeiten $p_{b\kappa}(C_n|\mathbf{x}_n)$ aller Blätter κ jedes Baumes b im RF bestimmt werden. In dem ursprünglichen RF wird hierfür ein Teil der Trainingsbeispiele durch jeden Baum bis zu den Blättern geschleust. In Blatt κ werden dann die Trainingsbeispiele pro Klasse gezählt (vgl. Gleichung 30):

$$V_{b\kappa}^k = \sum_{n \in \kappa} \delta(\widetilde{C}_n = C^k). \tag{45}$$

Wie bei der Auswahl der Entscheidungsgrenze (Kapitel 4.2.1) ist diese Summe fehlerhaft, wenn einige der Trainingslabels aufgrund deren Herkunft aus einer veralteten Karte fehlerhaft sind. Bei dem RRF wird daher auch an dieser Stelle die Zuordnung zu einer Klasse, welche in Gleichung 30 durch den Term $\delta(\widetilde{C}_n = C^k)$ umgesetzt wird, ersetzt durch die Wahrscheinlichkeiten

[2] Mit Hilfe des Indexes n werden auch im weiteren Verlauf dieser Arbeit die Übergangswahrscheinlichkeiten $p(\widetilde{C}|C)$, welche sich auf das gesamte Bild beziehen, und die individuellen Übergangswahrscheinlichkeiten $p(\widetilde{C}_n|C_n)$, welche sich auf ein konkretes Trainingsbeispiel n beziehen, unterschieden.

der Klassenzugehörigkeit $\mathcal{V}_n = p(C_n|\mathbf{x}_n, \widetilde{C}_n)$:

$$V_{b\kappa}^k = \sum_{n \in \kappa} p(C_n = C^k|\mathbf{x}_n, \widetilde{C}_n). \tag{46}$$

Anschließend werden die Werte für $V_{b\kappa}^k$ eines Blattes κ normiert und als Wahrscheinlichkeit $p_{b\kappa}(C^k)$ interpretiert.

Die Klassifikation eines neuen Primitivs erfolgt wie bei dem ursprünglichen RF (Kapitel 3.6).

4.3. Integration der Kartenlabels als Beobachtungen

Die Karte stellt die Objekte bzw. die Klassen zu einem früheren Zeitpunkt dar und enthält somit Informationen über die Position und Form von Objekten, solange diese noch in den Sensordaten vorhanden sind. Diese Information können für die Klassifikation genutzt werden. Besonders hilfreich ist dies bei der Klassifikation von Objekten, bei denen sich der Klassifikator unsicher ist z.B. bei Mischpixeln ode in Schattenbereichen. In dieser Arbeit wird daher das CRF Modell aus Kapitel 3.7 erweitert, sodass die Kartenlabels als weitere Beobachtungen berücksichtigt werden.

Das Label der Karte $\widetilde{C}_n$ jedes Pixels n wird in Form eines beobachteten Knotens mit aufgenommen. Der statistische Zusammenhang zwischen dem Label C_n und dem beobachteten Label $\widetilde{C}_n$ wird durch eine Kante integriert. Eine Kante zwischen dem beobachteten Label $\widetilde{C}_n$ und den Daten $\mathbf{d}$ wird nicht verwendet, da kein direkter Zusammenhang der Beobachtungen besteht. Der resultierende Graph des erweiterten CRF ist in Abb. 9 dargestellt.

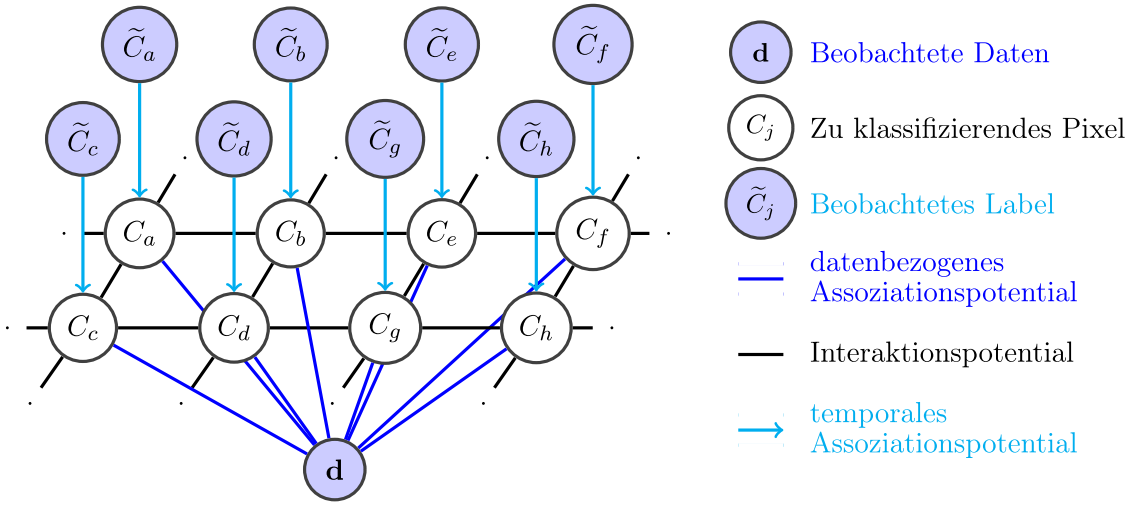


Abbildung 9: Graphenstruktur des CRF für die Klassifikation eines Bildes aus einem Zeitpunkt unter Berücksichtigung der Labels einer veralteten Karte.

Aus dem Graph resultiert die Faktorisierung der a posteriori Wahrscheinlichkeit zur Bestimmung der optimalen Labelkonfiguration $\mathbf{C}$ des gesamten Bildes:

$$p(\mathbf{C}|\mathbf{d}, \widetilde{\mathbf{C}}) = \frac{1}{\mathcal{Z}} exp\Big(\sum_{n \in \mathcal{S}} \big(A_d(C_n, \mathbf{d}) + A_z(C_n, \widetilde{C}_n)\big) + \sum_{n \in \mathcal{S}} \sum_{m \in \mathcal{O}_n} I(C_n, C_m, \mathbf{d}) \Big). \tag{47}$$

$A_z(C_n, \widetilde{C}_n)$ beschreibt hierbei das temporale Assoziationspotential, welches die unbekannten Labels mit den beobachteten Labels verbindet. Da nun zwei verschiedene Arten an Assoziationspotentialen verwendet werden, wird das Assoziationspotential, welches bereits in dem klassischen CRF aus Kapitel 3.7 verwendet wird und die Labels C_n mit den Daten $\mathbf{d}$ verbindet, als datenbezogenes Assoziationspotential $A_d(C_n, \mathbf{d})$ bezeichnet. Dieses datenbezogene Assoziationspotential $A_d(C_n, \mathbf{d})$ wird mit Hilfe der a posteriori Wahrscheinlichkeit des RRF entsprechend Gleichung 35 modelliert und das Interaktionspotential $I(C_n, C_m, \mathbf{d})$ mit Hilfe des kontrastsensitiven Potts-Modells entsprechend Gleichung 36.

Die beobachteten Labels $\widetilde{C}$ beschreiben, wenn sie aus einer veralteten Karte stammen, die Klassen zu einem früheren Zeitpunkt. Besonders wenn der Zeitraum zwischen der Kartenerstellung und der Sensoraufnahme vergleichsweise gering ist, kann davon ausgegangen werden, dass ein Großteil der in der Karte abgebildeten Objekte gleich geblieben ist. Außerdem sind einige Arten von Änderungen wahrscheinlicher als andere. So ist es beispielsweise wahrscheinlicher, dass ein neues Gebäude gebaut wird als dass eines abgerissen wird. Demnach sind die aktuellen unbekannten Labels C abhängig von den beobachteten Labels $\widetilde{C}$. Dieser Zusammenhang kann durch Übergangswahrscheinlichkeiten $\overrightarrow{\gamma} = p(C|\widetilde{C})$ modelliert werden. Kapitel 4.4 zeigt eine Möglichkeit diese Übergangswahrscheinlichkeiten aus den Labels der Karte und den Labels, welche aus der Klassifikation einer vorherigen Iteration resultieren, zu berechnen.

Da die Übergangswahrscheinlichkeiten im Allgemeinen eine sehr kleine Wahrscheinlichkeit für eine Änderung anzeigen, würde die Modellierung mit den Übergangswahrscheinlichkeiten eine Aufrechterhaltung der veralteten Kartenlabels unterstützen. Diese temporale Glättung des Klassifikationsergebnisses an die veraltete Karte wird durch eine Gewichtung θ_n des temporalen Assoziationspotentials jedes Pixels n, mit der der Einfluss des beobachteten Labels auf das Klassifikationsergebnis modelliert wird, verringert.

Das Gewicht θ_n ist vergleichbar mit der Änderungswahrscheinlichkeit g_n aus Kapitel 4.5. Während g_n, welche bereits in Kapitel 4.2.3 für die Berechnung der individuellen Übergangswahrscheinlichkeiten verwendet wurde, in Änderungsgebieten einen großen Wert aufweist, sollte der Wert für θ_n an diesen Stellen klein sein, damit das Label $\widetilde{C}_n$ der veralteten Karte möglichst wenig Einfluss hat. Dementsprechend kann das Gewicht θ_n als Wahrscheinlichkeit, dass sich die Klasse des Pixels n nicht ändert $p(\widetilde{C}_n = C_n)$, interpretiert werden.

Bestimmt wird das Gewicht θ_n eines Pixels n wie die Änderungswahrscheinlichkeit g_n in einem iterativen Prozess (Kapitel 4.5). Ausgehend von einem Startwert θ_n^0 wird auch das Gewicht θ_n^{i+1} für die nächste Iteration $i+1$ an jedem Pixel n nach der Klassifikation mit dem Gewicht θ_n^i und dem Änderungsbild B^i aktualisiert. Die Aktualisierung erfolgt wie in Kapitel 4.5 beschrieben.

Das temporale Assoziationspotential A_z hängt von den Übergangswahrscheinlichkeiten $\overrightarrow{\gamma}_{ka} = p(C = C^k|\widetilde{C} = C^a)$ der Klassen a und k, die in der Übergangsmatrix $\overrightarrow{\Gamma}$ zusammengefasst sind, und der Gewichtung θ_n für jedes Pixel n ab:

$$A_z(C_n, \widetilde{C}_n) = \theta_n \cdot ln\big(p(C = C^k|\widetilde{C} = C^a)\big). \tag{48}$$

Zur Bestimmung der optimalen Labelkonfiguration $\hat{\mathbf{C}}$ wird innerhalb der Inferenz die Wahrscheinlichkeit $p(\mathbf{C}|\mathbf{d}, \widetilde{\mathbf{C}})$ (Gleichung 47) maximiert:

$$\hat{\mathbf{C}} = \arg\max_{\mathbf{C}} \left(p(\mathbf{C}|\mathbf{d}, \widetilde{\mathbf{C}})\right). \tag{49}$$

Die Inferenz erfolgt auch hier approximativ basierend auf der LBP (Kapitel 3.7).

4.4. Bestimmung der Übergangsmatrix

In der in dieser Arbeit vorgestellten Methodik werden zwei Arten an Übergangswahrscheinlichkeiten benötigt, um den Zusammenhang zwischen den unbekannten Labels C und den beobachteten Labels $\widetilde{C}$ zu modellieren: $\overleftarrow{\gamma} = p(\widetilde{C}|C)$ und $\overrightarrow{\gamma} = p(C|\widetilde{C})$. Die Pfeilnotation wird hierbei genutzt, um anzuzeigen, ob die bedingte Wahrscheinlichkeit von C bei gegebenem $\widetilde{C}$ oder von $\widetilde{C}$ bei gegebenem C verwendet wird.

Die Übergangswahrscheinlichkeiten $\overleftarrow{\gamma}$ werden zur Berechnung der individuellen Übergangswahrscheinlichkeiten $\overleftarrow{\gamma}_n$ (Gleichung 44, Kapitel 4.2.3) verwendet, mit denen wiederum die Wahrscheinlichkeiten der Klassenzugehörigkeit $\mathcal{V} = p(C|\mathbf{x}, \widetilde{C})$ bestimmt werden können. Die Übergangswahrscheinlichkeiten $\overrightarrow{\gamma}$ werden für das temporale Assoziationspotential benötigt (Gleichung 48, Kapitel 4.3).

Bei bekannter gemeinsamer Verteilung von C und $\widetilde{C}$ kann die bedingte Wahrscheinlichkeit bestimmt werden [Bishop, 2006]:

$$p(\widetilde{C}|C) = \frac{p(\widetilde{C}, C)}{p(C)}, \tag{50}$$

$$p(C|\widetilde{C}) = \frac{p(\widetilde{C}, C)}{p(\widetilde{C})}. \tag{51}$$

Im Allgemeinen ist aber weder die gemeinsame Wahrscheinlichkeit $p(\widetilde{C}, C)$ noch die Wahrscheinlichkeit $p(C)$ bekannt.

In dieser Arbeit wird daher repräsentativ für die wahren Labels das Klassifikationsergebnis der vorherigen Iteration verwendet, um die Wahrscheinlichkeitsverteilungen zu bestimmen. Zur Vereinfachung der Notation wird der Iterationsindex bei der Klasse C weggelassen.

Die Wahrscheinlichkeit $p(C_n = C^k)$, dass aus N Pixeln ein beliebig gewähltes Pixel n die Klasse C^k aufweist, ist gegeben durch $p(C_n = C^k) = \#C^k/N$, wobei $\#C^k$ die Anzahl der Pixel mit dem Label C^k im Klassifikationsergebnis ist. Entsprechend wird mit Hilfe der Anzahl der Pixel mit dem Label C^a in der Karte die Wahrscheinlichkeit $p(\widetilde{C}_n = C^a) = \#\widetilde{C}^a/N$ bestimmt. Die Anzahl an Pixeln $\#(C^k \wedge \widetilde{C}^a)$, die das Label C^k in dem Klassifikationsergebnis aufweisen und das Label C^a in der Karte, wird für die gemeinsame Wahrscheinlichkeit $p(C_n = C^k, \widetilde{C}_n = C^a) = \#(C^k \wedge \widetilde{C}^a)/N$ verwendet.

Die Übergangswahrscheinlichkeiten $\overleftarrow{\gamma}_{ka}$ und $\overrightarrow{\gamma}_{ka}$ zweier Klassen C^k und C^a ergeben sich demnach entsprechend der Gleichungen 50 und 51 aus:

$$\overleftarrow{\gamma}_{ka} = p(\widetilde{C} = C^a | C = C^k) = \frac{\sum_{n=1}^{N} \delta((C_n = C^k) \wedge (\widetilde{C}_n = C^a))}{\sum_{n=1}^{N} \delta(C_n = C^k)} = \frac{\#(C^k \wedge \widetilde{C}^a)}{\#C^k}, \tag{52}$$

$$\overrightarrow{\gamma}_{ka} = p(C = C^k | \widetilde{C} = C^a) = \frac{\sum_{n=1}^{N} \delta((C_n = C^k) \wedge (\widetilde{C}_n = C^a))}{\sum_{n=1}^{N} \delta(\widetilde{C}_n = C^a)} = \frac{\#(C^k \wedge \widetilde{C}^a)}{\#\widetilde{C}^a}. \tag{53}$$

Die so bestimmten Übergangswahrscheinlichkeiten basieren auf dem Klassifikationsergebnis der vorherigen Iteration bzw. der Initialisierung. Demnach wird angenommen, dass das Klassifikationsergebnis jeder Iteration überwiegend korrekt ist, sodass die Übergangswahrscheinlichkeiten die tatsächlichen Übergänge zwischen den einzelnen Klassen ausreichend gut repräsentieren.

4.5. Änderungswahrscheinlichkeit

Die Änderungswahrscheinlichkeit g_n spiegelt die Wahrscheinlichkeit, dass sich das beobachtete Label $\widetilde{C}_n$ von Pixel n geändert hat, wieder. Wie bereits in dem Überblick in Kapitel 4.1 dargestellt, wird g_n iterativ, basierend auf einem binären Änderungsbild B, welches potenzielle Änderungen im Bild zeigt, aktualisiert. Die Bestimmung des Änderungsbildes unter Berücksichtigung mehrerer Heuristiken und die Aktualisierung der Änderungswahrscheinlichkeit wird in diesem Kapitel genauer beleuchtet.

Wie in Kapitel 2.1 eingeführt, besteht eine Möglichkeit zur Detektion von Änderungen zwischen zwei Zeitpunkten im Vergleich zweier Labelbilder. Im Bereich der Kartenaktualisierung werden hierbei die Labels der veralteten Karte mit den resultierenden Labels der Klassifikation von aktuellen Fernerkundungsdaten verglichen. Aus diesem Vergleich lässt sich ein binäres Änderungsbild B erstellen, welches für jedes Pixel n anzeigt, ob eine Änderung des Labels vorliegt, also das Kartenlabel $\widetilde{C}_n$ nicht dem Klassifikationsergebnis C_n entspricht, oder keine Änderung aufgetreten ist ($C_n = \widetilde{C}_n$):

$$B_n = \begin{cases} 1 & \text{wenn } C_n \neq \widetilde{C}_n \\ 0 & \text{wenn } C_n = \widetilde{C}_n. \end{cases} \tag{54}$$

Ein Nachteil dieses einfachen Vergleiches ist, dass auch Fehlklassifikationen als Änderungen interpretiert werden. Eine Unterscheidung von Fehlklassifikationen und tatsächlichen Änderungen kann mit Hilfe zusätzlicher Information über die Art der Daten und der Änderungen erfolgen.

Bei der Kartenaktualisierung durch die Klassifikation aktueller Fernerkundungsdaten beschreiben sowohl die Karte als auch die Daten Objekte, jedoch zu zwei unterschiedlichen Zeitpunkten. Die fehlerhaften Labels der Karte entstehen demnach z.B. durch Änderungen dieser Objekte zwischen diesen Zeitpunkten. Andere Ursachen für Labelfehler, wie beispielsweise Fehler bei der Georeferenzierung, werden in dieser Arbeit nicht modelliert. Resultieren Labelfehler

aus der Änderung ganzer Objekte, sind die Pixel, deren Klasse sich geändert hat, nicht zufällig über das Bild verteilt, sondern formen Segmente im Bildraum. Bei vereinzelten Pixeln, deren Labels sich nach der Klassifikation von den Kartenlabels unterscheiden, wird von einer Fehlklassifikation ausgegangen.

Ein weiterer Indikator für potenzielle Fehlklassifikationen bei hochaufgelösten Daten ist Schlagschatten. Bei passiven Sensoren ist die Qualität der spektralen Information dort relativ schwach, da nicht genügend Licht zu den Objekten innerhalb des Schattens gelangt. Dies führt zu sich stark ähnelnden Merkmalen der verschiedenen Klassen, sodass eine Unterscheidung der Klassen durch die Merkmale innerhalb der Klassifikation oft nicht möglich ist. Außerdem führt die schwache spektrale Information zu Problemen bei der Bildzuordnung (engl. Matching), wenn aus zwei Bildern des Gebietes ein digitales Oberflächenmodell erzeugt werden soll. Demnach können die Höhenmerkmale in Schlagschattenbereichen ebenfalls fehlerhaft sein.

In digitalen Oberflächenmodellen befinden sich außerdem meist fehlerhafte Höhenwerte in Bereichen, in denen ein vertikaler Höhensprung vorliegt, z.B. an Dachrändern. Neben der Tatsache, dass manche Pixel (Mischpixel), die Objekte unterschiedlicher Klassen und somit unterschiedlicher Höhen abdecken, können Fehler in diesen Bereichen auch bei der Erzeugung des Höhenmodells entstehen. Werden beispielsweise Stereobilder verwendet, sind tiefe Bodenbereiche neben dem Höhensprung oft nur in einem Bild zu sehen, sodass kein Höhenwert bestimmt werden kann.

Diese Probleme sind nicht unabhängig von der Auflösung. Die Objekte, welche in den Bildern detektiert werden können, unterscheiden sich beispielsweise stark, je nachdem ob hoch oder niedrig aufgelöste Daten vorliegen. Dementsprechend ist auch die Größe der Änderungen im Bildraum abhängig von der Auflösung. Auch das Vorkommen von Schlagschatten kann abhängig von der Auflösung betrachtet werden. Obwohl auch in niedrig aufgelösten Bildern Schatten entstehen können, z.B. durch steile Berge, sind Schlagschatten meist in niedrig aufgelösten Bildern nicht relevant. Schlagschatten entstehen bei hochaufgelösten Bildern beispielsweise neben Bäumen oder Gebäuden. Das Problem mit vertikalen Höhensprüngen entsteht ebenfalls bei der Verarbeitung hochaufgelöster Daten, da sie meist an Dachrändern auftreten. Das Problem mit Mischpixeln ist unabhängig von der Auflösung. Jedoch ist anzumerken, dass im Allgemeinen die Auflösung und Klassestruktur aufeinander abgestimmt werden, sodass relativ wenige Pixel Mischpixel sind.

Zusammenfassend lassen sich drei Annahmen zur Unterscheidung von Änderungen, die nur fälschlicherweise aufgrund von Fehlklassifikationen als solche interpretiert werden, und tatsächlichen Änderungen von Objekten treffen:

- An Objektgrenzen entstehen vermehrt Fehlklassifikationen.
- Änderungen erfolgen in größeren Segmenten im Bildraum.
- In Bereichen im Schlagschatten ist eine korrekte Klassifikation oft nicht möglich.

Angelehnt an diese Annahmen wird das Änderungsbild B angepasst. Hierfür wird folgender Ablauf verwendet:

1. Erosion mit einem quadratischen Strukturelement der Seitenlänge u. Hierdurch werden schmale Strukturen entfernt, die aufgrund von Fehlklassifikationen an Objektgrenzen entstehen. Auch kleine isolierte Segmente werden hiermit entfernt. Wie oben beschrieben, können diese Fehlklassifikationen aus der Klassifikation von Mischpixeln resultieren. Ist bei hoch aufgelösten Daten ein Oberflächenmodell vorhanden, können auch Fehler im Oberflächenmodell die Fehlklassifikationen verursachen.

2. Zusammenschluss von Pixeln, deren Label sich geändert hat, zu Segmenten. Hierbei befinden sich zusammenhängende Gebiete von Pixeln mit geändertem Label in einem Segment. Die Nachbarschaft wird durch die 4 direkt angrenzenden Pixel definiert. Alle Pixel jedes Segmentes besitzen demnach den Wert 1 in dem Änderungsbild B, was auf eine Änderung hindeutet. Wird das Segment entfernt, bedeutet dies, dass allen Pixeln dieses Segmentes der Wert 0 in B zugewiesen wird.

3. Entfernung von Segmenten, deren Fläche kleiner ist als ein Schwellwert s. Dies resultiert aus der Annahme, dass Änderungen im Bildraum in größeren Segmenten auftreten. Bei kleineren Segmenten wird von einer vorherigen Fehlklassifikation der Pixel ausgegangen.

4. Entfernung von Segmenten, die vermutlich von Schlagschatten betroffen sind. Hierfür werden der Median und der Mittelwert der Intensität in den Segmenten (Med_{Bal} bzw. Mit_{Bal}) und in dem ganzen Bild (Med_{Bild} bzw. Mit_{Bild}) bestimmt. Wenn $Mit_{Bal} < Mit_{Bild}/2$ und $Med_{Bal} < Med_{Bild}/2$, also wenn die Pixel in dem Segment deutlich dunkler sind als im ganzen Bild, wird angenommen, dass das Segment im Schlagschatten liegt. Bei Segmenten im Schlagschatten wird angenommen, dass die spektralen Informationen aus den Sensoren eine zu schlechte Qualität aufweisen, um Änderungen detektieren zu können. Außerdem wird angenommen, dass eine Fehlklassifikation aufgrund des Schattens wahrscheinlicher ist als eine tatsächliche Änderung in diesem Bereich. Segmente im Schlagschatten werden demnach entfernt.

5. Dilatation mit einem quadratischen Strukturelement der Seitenlänge u. Aufgrund der Erosion zur Entfernung schmaler Strukturen (s. Schritt 1) stellen die Grenzen der Änderungssegmente die Grenzen der geänderten Objekte im Bild nicht korrekt dar. Aus diesem Grund erfolgt hier eine Dilatation.

Für die Entfernung von schmalen langen Segmenten, die vermutlich aus Fehlklassifikationen an Objektgrenzen resultieren, könnte auch morphologisches Opening verwendet werden. In dieser Arbeit wurde dies in eine unabhängige Erosion und Dilatation aufgespalten, um Bereiche, die sich an wenigen Pixeln berühren, vor der Analyse der Größe und Helligkeit zu trennen. Abb. 10 zeigt ein Beispiel zur Verdeutlichung des Unterschiedes zwischen der unabhängigen Erosion und einem morphologischen Opening. In den rot markierten Bereichen in Abb. 10 ist zu erkennen, dass einige Bereiche durch wenige Pixel verbunden sind und demnach als ein

großes Segment interpretiert werden. Diese Verbindungen bleiben auch nach einem morphologischen Opening bestehen. Durch eine Erosion werden die Verbindungen unterbrochen und jeder Bereich wird als ein individuelles Segment analysiert und anschließend ggf. entfernt.

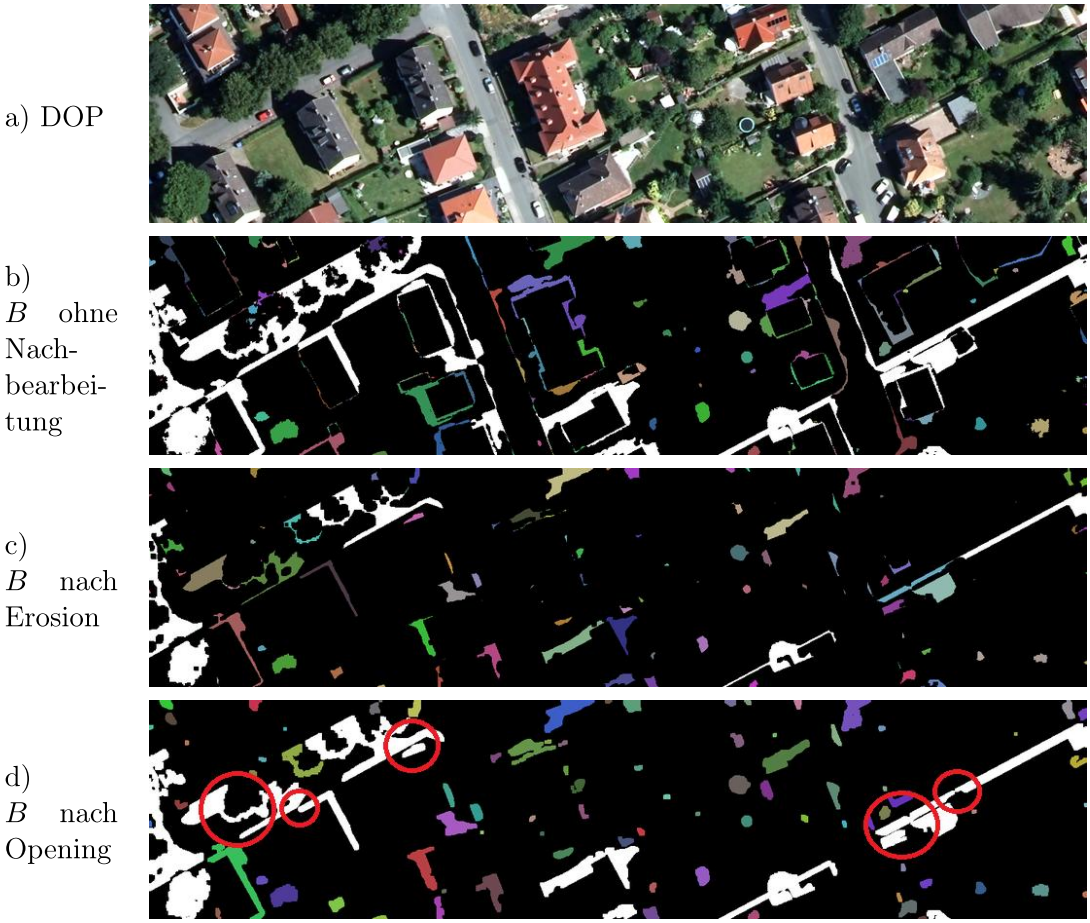
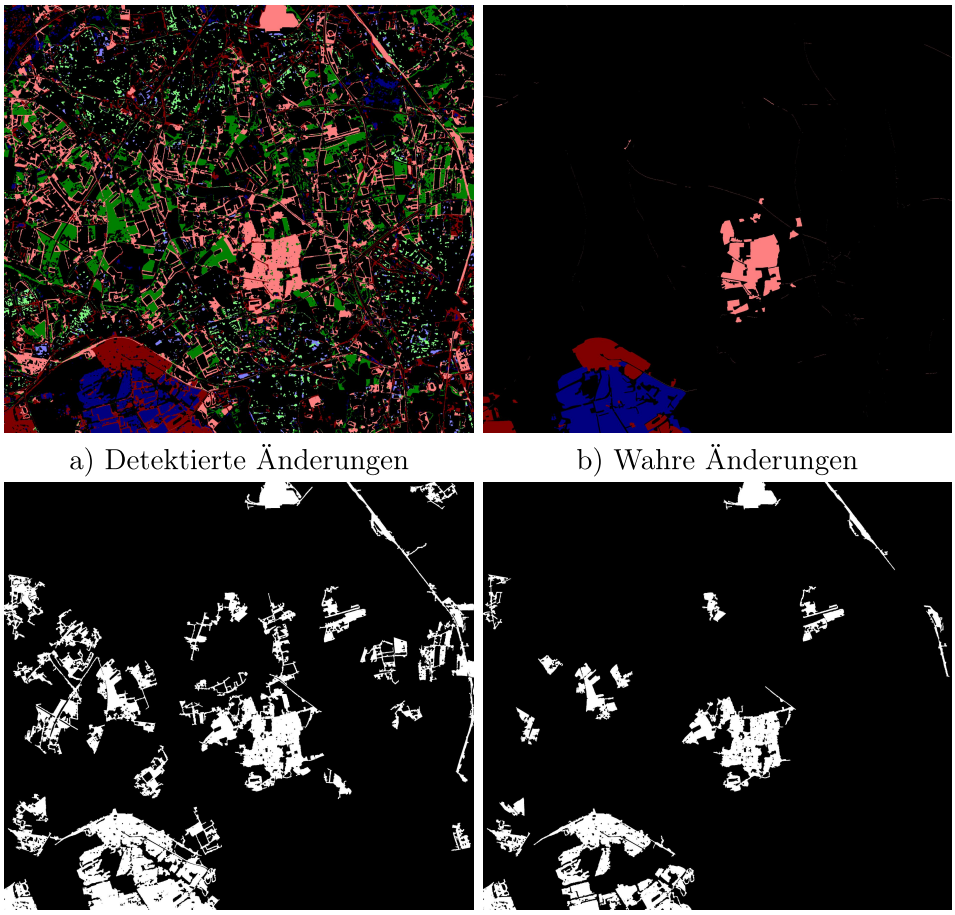


Abbildung 10: Vergleich der eingefärbten Änderungsbilder B nach der Erosion und dem Opening.
a) DOP. b) Änderungsbild B ohne Nachbearbeitung. c) Änderungsbild B nach der Erosion. d) Änderungsbild B nach dem Opening.
Weiß: Änderung in großen Segmenten (bleiben). Farbe: Änderung in kleinen Segmenten (werden entfernt). Rot markiert: Berührungspunkte von Segmenten nach dem Opening.

Weitere Informationen über das Gebiet und die Daten könnten zu zusätzlichen Unterscheidungskriterien führen. Ist beispielsweise bekannt, dass eine Art der Änderung nicht aufgetreten ist, z.B. die Änderung von der Klasse *Gebäude* zu der Klasse *Baum*, kann diese Änderung, wenn sie detektiert wird, als Fehlklassifikation gewertet werden. Die Klassifikationsmethode in dieser Arbeit soll jedoch auf verschiedene Arten von Fernerkundungsdaten anwendbar sein, sodass keine weiteren Annahmen getroffen und die Änderungsbilder B nicht weiter angepasst werden.

Bei der Erstellung des Änderungsbildes B wird die Art der Änderung nicht berücksichtigt. In einigen Fällen kann es jedoch hilfreich sein, verschiedene Arten an Änderungen zunächst unabhängig von den anderen zu betrachten, beispielsweise wenn klassenspezifische Anforderungen an neue Objekte, wie z.B. eine bestimmte Größe, vorliegen. Segmente an Änderungen,

wie sie zuvor extrahiert wurden, können aus mehreren Segmenten verschiedener Arten an Änderungen bestehen. Besonders bei niedrig aufgelösten Daten, wie Satellitendaten, können größere Segmente an Änderungen, die aus mehreren kleinen Segmenten unterschiedlicher Arten an Änderungen zusammengesetzt sind, auf mehrere zusammenhängende Fehlklassifikationen hindeuten. Abb. 11 zeigt an einem Beispiel wie sich die Änderungsbilder B unterscheiden können, wenn die Art der Änderung bei der Unterscheidung von Änderungen, die nur fälschlicherweise aufgrund von Fehlklassifikationen als solche interpretiert werden, und tatsächlichen Änderungen von Objekten berücksichtigt wird. Im Folgenden wird daher eine Methode zur Bestimmung des Änderungsbildes gezeigt, bei der die Art der Änderung berücksichtigt wird.



a) Detektierte Änderungen

b) Wahre Änderungen

c) Art der Änderung nicht berücksichtigt

d) Art der Änderung berücksichtigt

Abbildung 11: Vergleich des Änderungsbildes B abhängig davon, ob die Art der Änderung bei der Bestimmung des Änderungsbildes B berücksichtigt wurde.
Legende a), b) Schwarz: keine Änderung. Farbe: eine bestimmte Art der Änderung. c), d) Schwarz: keine Änderung. Weiß: Änderung.

Während bei der Erstellung des Änderungsbildes B ohne Berücksichtigung der Art der Änderung angenommen wird, dass Pixel mit geändertem Label Segmente mit einer gewissen Größe im Bildraum bilden, wird bei der Bestimmung des Änderungsbildes unter Berücksichtigung der Art der Änderung angenommen, dass ein neues Objekt der Klasse k eine gewisse Größe besitzt. Somit formen Pixel, die ein neues Objekt dieser Klasse zeigen, ebenfalls Segmente im Bildraum, die eine Mindestgröße überschreiten.

Bei der Bestimmung des Änderungsbildes B wird zunächst ein Änderungsbild ${}^kB$ pro Klasse k erzeugt:

$$ {}^kB_n = \begin{cases} 1 & \text{wenn } C^k = C_n \neq \widetilde{C}_n \\ 0 & \text{wenn } C^k = C_n = \widetilde{C}_n. \end{cases} \tag{55} $$

Die so entstandenen Änderungsbilder ${}^kB$ werden mit dem oben beschriebenen, auf Heuristiken beruhenden Ablauf verarbeitet. Aus den einzelnen Änderungsbildern ${}^kB$ wird anschließend wieder ein gesamtes Änderungsbild B erzeugt:

$$ B_n = \begin{cases} 1 & \text{wenn } \sum_{k=1}^{K} {}^kB_n > 0 \\ 0 & \text{wenn } \sum_{k=1}^{K} {}^kB_n = 0. \end{cases} \tag{56} $$

Dabei ist anzumerken, dass hierbei keine Konflikte auftreten. Pixel, die in mindestens einem Änderungsbild ${}^kB$ als Änderung markiert sind, werden auch in dem Änderungsbild B als solches markiert. Welche Art von Änderung aufgetreten ist, wird, wie bereits beschrieben, bei dem Änderungsbild B nicht berücksichtigt.

Neben der Annahme, dass Änderungen größere Segmente im Bildraum bilden, wurde bei der Erstellung von B ohne Berücksichtigung der Art der Änderung noch die Annahmen getroffen, dass an Objektgrenzen vermehrt Fehlklassifikationen entstehen und dass in Bereichen im Schlagschatten eine korrekte Klassifikation oft nicht möglich ist. In der hier vorliegenden Arbeit wird davon ausgegangen, dass diese Annahmen für neue Objekte jeder Klasse zutreffen, sodass die Heuristiken, die auf diesen Annahmen beruhen, auch bei der Erstellung des Änderungsbildes unter Berücksichtigung der Art der Änderung verwendet werden können. Dabei ist anzumerken, dass die Anzahl an Fehlklassifikationen an den Objektgrenzen und auch die Anzahl an Objekten, die sich im Schlagschatten befinden, im Allgemeinen von der Klasse abhängt. Dies wird zur Vereinfachung des Modells in dieser Arbeit jedoch nicht berücksichtigt.

Mit den Einträgen des Änderungsbildes B_n für jedes Pixel n kann nun die Änderungswahrscheinlichkeit g_n aktualisiert werden. Ausgehend von einem Startwert $g^0 = 1/K$ wird die Wahrscheinlichkeit g_n entweder erhöht (wenn $B_n = 1$) oder gesenkt (wenn $B_n = 0$). Die Bestimmung der Änderungswahrscheinlichkeit g_n^{i+1} für die nächste Iteration $i+1$ aus g_n^i und B_n^i der aktuellen Iteration i erfolgt mit Hilfe einer zuvor definierten Aktualisierungsfunktion f_a. Die neue Wahrscheinlichkeit muss dabei zwischen 0 und 1 liegen. Im Folgenden werden zwei alternative Aktualisierungsfunktionen, f_{a1} und f_{a2}, vorgestellt.

Eine Möglichkeit ist die einfache Addition bzw. Subtraktion einer Konstanten c:

$$ g_n^{i+1} = f_{a1}(g_n^i, B_n) = \begin{cases} min[g_n^i + c, 1 - \zeta] & \text{wenn } B_n^i = 1 \\ max[g_n^i - c, \zeta] & \text{wenn } B_n^i = 0. \end{cases} \tag{57} $$

ζ ist hierbei eine Konstante mit einem kleinen Wert, nahe 0, um numerische Probleme zu vermeiden.

Es kann auch eine quadratische Funktion gewählt werden. Dadurch kann beispielsweise er-

reicht werden, dass eine bereits große Änderungswahrscheinlichkeit sich innerhalb einer Aktualisierung stärker verkleinert, wenn das Änderungsbild keine Änderung anzeigt, als sie sich vergrößern würde, wenn das Änderungsbild eine Änderung anzeigt. Um eine quadratische Funktion zu definieren, müssen mindestens drei Punkte festgelegt werden. In der folgenden Funktion werden hierfür die zwei Punkte $(g_n^0 = g_n^1 = 0)$ und $(g_n^0 = g_n^1 = 1)$ verwendet und $g_n^1 = g_n^0 \cdot 2$ falls eine Änderung vorliegt und andernfalls $g_n^1 = g_n^0/2$:

$$g_n^{i+1} = f_{a2}(g_n^i, B_n) = \begin{cases} min[\frac{1}{2\cdot(1/K)-1} \cdot (g_n^i)^2 + (1 - \frac{1}{2\cdot(1/K)-1}) \cdot g_n^i, 1-\zeta] & \text{wenn } B_n^i = 1 \\ max[-\frac{1}{2\cdot(1/K)-1} \cdot (g_n^i)^2 + (1 + \frac{1}{2\cdot(1/K)-1}) \cdot g_n^i, \zeta] & \text{wenn } B_n^i = 0. \end{cases} \tag{58}$$

Durch die kleine Konstante ζ wird vermieden, dass die Wahrscheinlichkeit die extremen Werte 0 oder 1 annimmt. Diese Werte würden exaktes Wissen darüber, ob sich das Label eines Pixels n geändert hat oder nicht, suggestieren, welches jedoch nicht vorliegt. Davon abgesehen ist eine Aktualisierung durch f_{a2} dann nicht mehr möglich, da $g_n^{i-1} = f_{a2}(0, B_n) = 0$. Eine schematische Darstellung der beiden Funktionen, sowie der Begrenzung durch eine Konstante ζ sind in Abb. 12 zu sehen.

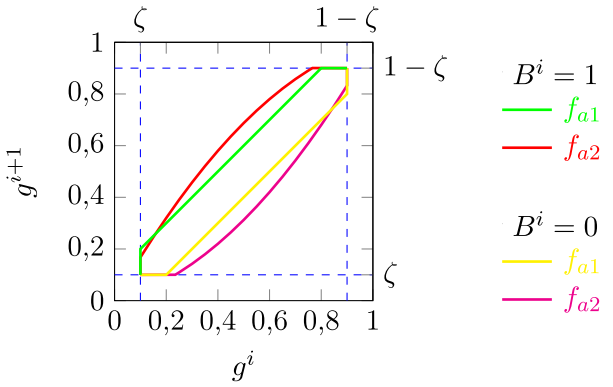


Abbildung 12: Zwei Aktualisierungsfunktionen f_{a1} und f_{a2} der Änderungswahrscheinlichkeit g.

Die Gewichtung θ_n des temporalen Assoziationspotentials (Kapitel 4.3), die den Einfluss des Kartenlabels auf das Klassifikationsergebnis reguliert, wird ähnlich aktualisiert, wie die Änderungswahrscheinlichkeit g_n. In Änderungsgebieten sollte der Wert für θ_n möglichst klein sein, damit das Label $\widetilde{C}_n$ der veralteten Karte möglichst wenig Einfluss hat. Demnach sinkt θ_n, wenn der Eintrag des Änderungsbildes $B_n = 1$ ist, während g_n steigt. Bei der Aktualisierung ändert sich somit nur die Richtung der Anpassung durch B_n. Erfolgt die Aktualisierung von g_n mit der Gleichung 57 wird θ_n wie folgt aktualisiert:

$$\theta_n^{i+1} = f_{a1}(\theta_n^i, B_n) = \begin{cases} min[\theta_n^i + c, 1-\zeta] & \text{wenn } B_n^i = 0 \\ max[\theta_n^i - c, \zeta] & \text{wenn } B_n^i = 1. \end{cases} \tag{59}$$

Wird hingegen Gleichung 58 für die Aktualisierung von g_n genutzt, gilt:

$$\theta_n^{i+1} = f_{a2}(\theta_n^i, B_n) = \begin{cases} min[\frac{1}{2\cdot(1/K)-1} \cdot (\theta_n^i)^2 + (1 - \frac{1}{2\cdot(1/K)-1}) \cdot \theta_n^i, 1-\zeta] & \text{wenn } B_n^i = 0 \\ max[-\frac{1}{2\cdot(1/K)-1} \cdot (\theta_n^i)^2 + (1 + \frac{1}{2\cdot(1/K)-1}) \cdot \theta_n^i, \zeta] & \text{wenn } B_n^i = 1. \end{cases} \tag{60}$$

Hierbei ist anzumerken, dass, aufgrund der unterschiedlichen Startwerte von g_n und θ_n, nicht immer $\theta_n = 1 - g_n$ gilt.

Am Ende dieses Abschnittes ist zu erwähnen, dass es sich bei dem hier vorgestellten Modell für die Änderungswahrscheinlichkeit um eine Heuristik handelt. Weder die Ermittlung der Änderungsgebiete noch der Aktualisierungsschritt nach der Klassifikation sind aus anderen statistischen Größen hergeleitet worden, sondern basieren auf Annahmen, wie sich Änderungen in der Klassifikation ausprägen. Außerdem basiert die iterative Aktualisierung auf der Annahme, dass die Wahrscheinlichkeit für eine Änderung steigt, wenn die Anzahl an Iterationen, in denen sich das klassifizierte Label von dem Kartenlabel unterscheidet, steigt. Diese Annahme ist nicht immer korrekt, da beispielsweise auch ein Pixel, deren Label sich nicht geändert hat, in mehreren Iterationen falsch klassifiziert sein kann. Die Änderungswahrscheinlichkeit dieses Pixels steigt demnach auch mit jeder Iteration.

4.6. Klassifikation basierend auf der robusten logistischen Regression

In Kapitel 4.1 wurde der iterative Klassifikationsalgorithmus basierend auf dem robusten Random Forest vorgestellt. Unabhängig vom verwendeten Basisklassifikator liegt dabei ein Vorteil in der Integration der Kartenlabels als Beobachtung innerhalb eines erweiterten CRF Modells (Kapitel 4.3). Ein Nachteil des RRF ist, dass dieser sowohl das Klassifikationsergebnis als auch die Beliefs der vorherigen Iteration benötigt. Für die Initialisierung musste daher die Standardvariante des Random Forest genutzt werden. Außerdem beeinflussen Heuristiken, die bei der Berechnung der Änderungswahrscheinlichkeit genutzt werden, das Trainingsergebnis des RRF. Die robuste logistische Regression (RLR) von Bootkrajang und Kabán [2012] (Kapitel 3.4) basiert ausschließlich auf den beobachteten Daten und den Labels der alten Karte und ist somit unabhängig von einem solchen iterativen Prozess. Da die Gewichtung θ_n des temporalen Assoziationspotentials dennoch iterativ aktualisiert werden muss, wird in diesem Kapitel der Ablauf des Klassifikationsalgorithmus vorgestellt, wenn die Modellierung des datenbezogenen Assoziationspotentials A_d mit der RLR erfolgt. Dieser Algorithmus wird in der vorliegenden Arbeit mit dem Akronym *RLRK* (**R**obuste **L**ogistische **R**egression mit **K**artenbeobachtung) abgekürzt.

An dieser Stelle ist zu erwähnen, dass im Training der RLR die Fehler in den Trainingslabels nur durch Übergangswahrscheinlichkeiten $\overleftarrow{\gamma} = p(\widetilde{C}|C)$ modelliert werden. Dieses Modell beinhaltet keinen Zusammenhang zwischen dem Auftreten der Fehler und den Daten.

In Abb. 13 ist die Initialisierung zu sehen. Im Vergleich zu der bereits vorgestellten Initialisierung in Abb. 5 besteht der wesentliche Unterschied in der Verwendung der RLR anstelle

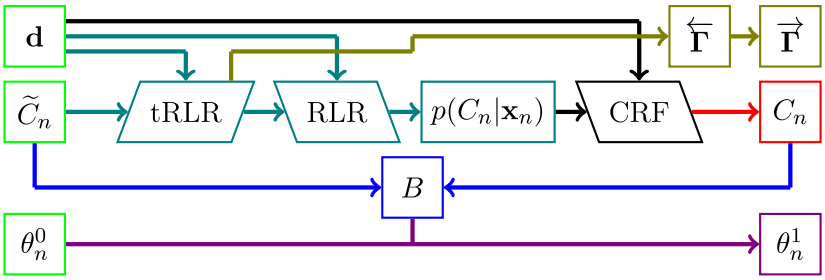


Abbildung 13: Initialisierungsprozess für die **RLRK** basierend auf der RLR. Rechteckige Boxen: Variablen. Parallelogramme: Prozesse. Parallelogramm mit spitzem Winkel oben rechts: Trainingsprozess. Die Farben in der Abbildung werden im Text aufgegriffen.

des RF. Das Training der RLR (s. Kapitel 3.4) erfolgt mit den Merkmalsvektoren $\mathbf{x}_n$ und den Labels $\widetilde{C}_n$ aus der Karte (Abb. 13, cyan, tRLR).

Anders wie beim RF wird hier die Übergangsmatrix $\overleftarrow{\mathbf{\Gamma}}$ bereits mitgeschätzt (Abb. 13, gelbgrün). Diese enthält die Übergangswahrscheinlichkeiten $\overleftarrow{\gamma}_{ka} = p(\widetilde{C} = C^a|C = C^k)$ zwischen den Klassen C^k und C^a. In dem temporalen Assoziationspotential wird jedoch die Übergangswahrscheinlichkeit $\overrightarrow{\gamma}_{ka} = p(C = C^k|\widetilde{C} = C^a)$ verwendet, da angenommen wird, dass die aktuellen Labels von den Labels der veralteten Karte abhängen. Nach dem Satz von Bayes (Gleichung 41) kann $\overleftarrow{\gamma}_{ka}$ mit Hilfe der Wahrscheinlichkeitsverteilungen $p(C_n = C^k)$ und $p(\widetilde{C} = C^a)$ in $\overrightarrow{\gamma}_{ka}$ überführt werden:

$$\overrightarrow{\gamma}_{ka} = p(C = C^k|\widetilde{C} = C^a) = \frac{p(\widetilde{C} = C^a|C = C^k) \cdot p(C = C^k)}{p(\widetilde{C} = C^a)} = \frac{\overleftarrow{\gamma}_{ka} \cdot p(C = C^k)}{p(\widetilde{C} = C^a)}. \quad (61)$$

Die Verteilung der Labels $p(C = C^k)$ ist im Allgemeinen unbekannt. Um dennoch Übergangswahrscheinlichkeiten $\overrightarrow{\gamma}_{ka}$ aus $\overleftarrow{\gamma}_{ka}$ zu berechnen, wird angenommen, dass sie ungefähr der Verteilung der beobachteten Labels gleicht ($p(C = C^k) \approx p(\widetilde{C} = C^k)$). Diese Annahme trifft im Allgemeinen nicht zu, da von Unterschieden der Labels aus der Karte und den zu bestimmenden Labels ausgegangen wird. Dennoch kann $p(\widetilde{C} = C^k)$, besonders bei einer kleinen Anzahl an fehlerhaften Trainingslabels, als Annäherung von $p(C = C^k)$ betrachtet werden.

Mit Hilfe der trainierten RLR wird die a posteriori Wahrscheinlichkeit $p(C_n|\mathbf{x}_n)$ für jedes Pixel mit dem Merkmalsvektor $\mathbf{x}_n$ bestimmt (Abb. 13, cyan, RLR). Diese fließen in das datenbezogene Assoziationspotential des CRF (Kapitel 3.7) ein (Abb. 13, schwarz). Obwohl die Übergangswahrscheinlichkeiten $\overrightarrow{\gamma}$ bereits bekannt sind, wird der statistische Zusammenhang zwischen den Labels C und den beobachteten Labels $\widetilde{C}$ an dieser Stelle noch nicht in Form eines temporalen Assoziationspotentials im CRF berücksichtigt, damit die initiale Änderungsdetektion nicht durch die Beobachtung der Labels der veralteten Karte beeinflusst wird. In das kontrastsensitive Potts-Modell des Interaktionspotentials fließen wie auch beim klassischen CRF die Daten $\mathbf{d}$ ein.

Als Ergebnis der Klassifikation mit Hilfe des CRF entstehen die Beliefs $Bel(C_n|\mathbf{d})$ abhängig von den Daten $\mathbf{d}$, welche jedoch nicht weiter verwendet werden und daher in Abb. 13 nicht abgebildet sind, und die Labels C_n (Abb. 13, rot). Mit diesen Labels und denen der veralteten

Karte wird das Änderungsbild B (Kapitel 4.5) bestimmt (Abb. 13, blau). Die Änderungsgebiete werden daraufhin für die Aktualisierung der Gewichte θ nach Gleichung 59 verwendet (Abb. 13, lila). Als Startwert für die Aktualisierung wird $\theta_n^0 = 0$ angenommen, sodass in der Initialisierung das Kartenlabel, wie zuvor beschrieben, noch nicht als weitere Beobachtung in den Klassifikationsprozess einfließt. Die Änderungswahrscheinlichkeit g_n, welche für das Training des RRF benötigt wurde, muss bei der Verwendung der RLR nicht ermittelt werden, da der RLR nur auf Basis der Merkmalsvektoren $\mathbf{x}_n$ und der Labels $\widetilde{C}_n$ der alten Karte trainiert wird.

Nach der Initialisierung folgt der iterative Prozess (Abb. 14), bestehend aus der Klassifikation mit dem erweiterten CRF (Abb. 14, schwarz), der Bestimmung des Änderungsbildes B (Abb. 14, blau) und der Aktualisierung der Gewichte θ_n (Abb. 14, lila). Die a posteriori Wahrscheinlichkeiten $p(C_n|\mathbf{x}_n)$ und die Übergangsmatrix $\overrightarrow{\mathbf{\Gamma}}$ müssen nicht erneut bestimmt werden, da die RLR (Kapitel 3.4) zum Training lediglich die Merkmalsvektoren $\mathbf{x}_n$ und die beobachteten Labels $\widetilde{C}_n$ benötigt, welche in dem iterativen Prozess jedoch nicht geändert werden. Lediglich die Klassifikation mit dem CRF muss aufgrund der sich ändernden Gewichte θ_n des temporalen Assoziationspotentials der beobachteten Kartenlabels in jeder Iteration wiederholt werden. In Abb. 14 ist der iterative Prozess mit dunkelgrauen Pfeilen hinterlegt.

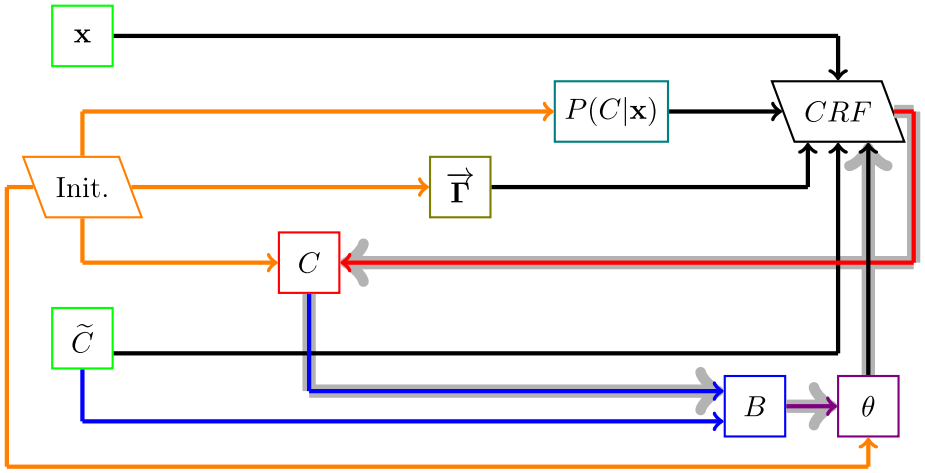


Abbildung 14: Iterativer Prozessablauf des **RLRK**. Rechteckige Boxen: Variablen. Parallelogramme: Prozesse. Init: Initialisierungsprozess. Die Farben in der Abbildung werden im Text aufgegriffen.

Als Abbruchkriterium für den iterativen Prozess kann auch hier eine zuvor definierte Anzahl an maximalen Iterationen N_{it} verwendet werden oder eine maximale Anzahl an Pixeln, deren Gewicht θ sich in einer Iteration ändert.

4.7. Multitemporale Klassifikation

Im multitemporalen Fall sind Fernerkundungsdaten zu Z unterschiedlichen Zeitpunkten von einem bestimmten Gebiet aufgenommen worden. Dementsprechend ist für jedes Primitiv n in jedem Zeitpunkt $z \in \{1, ..., Z\}$ ein Merkmalsvektor $\mathbf{x}_n^z$ bekannt. Zu bestimmen sind die Klassen C_n^z. Wie im monotemporalen Fall ist zusätzlich eine veraltete Karte gegeben. Diese

liefert für jedes Pixel n des ersten Zeitpunktes z^1 ein beobachtetes Label $\widetilde{C}_n$. Während die Karte in dieser Arbeit die gleiche Auflösung aufweist wie der erste Zeitpunkt, können die Daten der unterschiedlichen Zeitpunkte verschieden aufgelöst sein, solange die Klassenstruktur identisch bleibt. Die Klassenstruktur hängt wiederum von der verwendeten Karte ab, da die Trainingslabels aus der Karte extrahiert werden.

Das multitemporale CRF für die gemeinsame Klassifikation aller Daten beruht auf dem Modell von Hoberg et al. [2015], welches jedoch an den konkreten Fall, dass eine veraltete Karte existiert, angepasst wurde.

In Kapitel 4.7.1 wird zunächst auf das multitemporale CRF eingegangen. Anschließend wird in Kapitel 4.7.2 eine sequenzielle Trainingsprozedur eingeführt. In Kapitel 4.7.3 wird der komplette Ablauf der multitemporalen Klassifikation dargestellt.

4.7.1. Multitemporales CRF Modell

Das multitemporale CRF besteht, wie das monotemporale CRF in Kapitel 3.7, aus beobachteten und unbeobachteten Knoten, welche Zufallsvariablen darstellen, und Kanten, welche die statistischen Abhängigkeiten zwischen den Knoten repräsentieren. Die Knoten des in dieser Arbeit vorgestellten multitemporalen CRF bilden sowohl die zu bestimmenden Labels C_n^z und die Daten $\mathbf{d}_z$ zu jedem Zeitpunkt z, als auch die Kartenlabels $\widetilde{C}_n$, wobei die Daten $\mathbf{d}_z$ und die Kartenlabels $\widetilde{C}_n$ beobachtet sind. Die Daten $\mathbf{d}_z$ und die Labels C_n^z des selben Zeitpunktes werden über ein Assoziationspotential $A_d(C_n^z, \mathbf{d}^z)$ verknüpft. Ein Interaktionspotential $I_r(C_n^z, C_m^z, \mathbf{d}^z)$ beschreibt den statistischen Zusammenhang zweier räumlich benachbarter Labels C_n^z und C_m^z. Da angenommen wird, dass der erste Zeitpunkt $z = 1$ zeitlich am nächsten an dem Zeitpunkt der Kartenerstellung liegt, werden die Labels C_n^1 des ersten Zeitpunktes mit den Kartenlabels $\widetilde{C}_n$ verbunden. Dies erfolgt in Form eines temporalen Assoziationspotentials $A_z(C_n^1, \widetilde{C}_n)$, wie in Kapitel 4.3 beschrieben. Ein direkter statistischer Zusammenhang der Labels der alten Karte zu späteren Zeitpunkten wird nicht angenommen, sodass auch keine Kante die Kartenlabels und die Labels späterer Zeitpunkte verbindet.

Im Gegensatz zu dem monotemporalen CRF werden in dem multitemporalen CRF auch temporale Nachbarn miteinander verbunden. Der statistische Zusammenhang zweier temporal benachbarter Knoten C_n^z und C_l^o wird dabei in einem temporalen Interaktionspotential $I_z(C_n^z, C_l^o)$ beschrieben. Wie bei der Integration der Kartenlabels (Kapitel 4.3) wird angenommen, dass die Labels zum Zeitpunkt z von den Labels aus einem vorherigen Zeitpunkt $z-1$ abhängen. Ein direkter statistischer Zusammenhang zu Labels früherer Zeitpunkte wird nicht angenommen, sodass lediglich direkte temporale Nachbarn durch Kanten verbunden werden.

Abb. 15 zeigt beispielhaft den Graph eines multitemporalen CRF für zwei Zeitpunkte, wobei die Bodenpixelgröße der Fernerkundungsdaten zum frühen Zeitpunkt $z = 1$ halb so groß ist wie die zum Zeitpunkt $z = 2$. Die Daten $\mathbf{d}^z$ zu jedem Zeitpunkt z mit den Kanten zu den Labels des entsprechenden Zeitpunktes sind für eine bessere Übersicht in Abb. 15 nicht dargestellt.

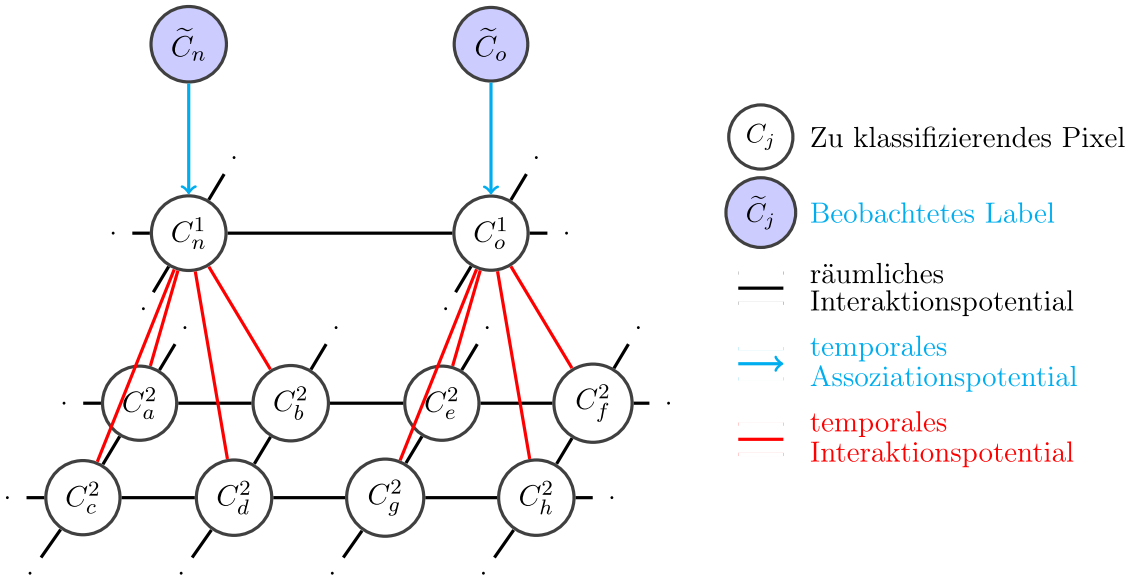


Abbildung 15: Graphenstruktur des CRF für die Klassifikation von zwei Bildern aus unterschiedlichen Zeitpunkten unter Berücksichtigung der Labels einer veralteten Karte. Die Auflösung des Bildes zum Zeitpunkt $z = 2$ ist doppelt so hoch wie die des Bildes aus $z = 1$. Die Daten $\mathbf{d}^1$ und $\mathbf{d}^2$ sind nicht abgebildet.

Basierend auf dem Graphen in Abb. 15 kann die a posteriori Wahrscheinlichkeit wie folgt modelliert werden:

$$p(\mathbf{C}|\mathbf{d},\widetilde{\mathbf{C}}) = \frac{1}{\mathcal{Z}} \cdot exp\Bigg(\sum_z \Big(\sum_{n\in\mathcal{S}^z} A_d(C_n^z,\mathbf{d}^z) + \sum_{n\in\mathcal{S}^z}\sum_{m\in\mathcal{O}_n^z} I_r(C_n^z,C_m^z,\mathbf{d}^z) + \sum_{n\in\mathcal{S}^z}\sum_{l\in\mathcal{L}_n^o} I_z(C_n^z,C_l^o)\Big) + \sum_{n\in\mathcal{S}^1} A_z(C_n^1,\widetilde{C}_n)\Bigg). \tag{62}$$

Die Menge aller Daten $\mathbf{d}$ besteht hier aus den Daten aller Zeitpunkte $\{\mathbf{d}^1,...,\mathbf{d}^Z\}$. Die Partitionsfunktion $\mathcal{Z}$ ist, wie in Gleichung 34, eine Normalisierungskonstante um die Potentiale in Wahrscheinlichkeiten umzuwandeln. $\mathcal{S}^z$ ist die Menge aller Bildprimitive zum Zeitpunkt z und $\mathcal{O}_n^z$ die Menge der räumlich benachbarten Knoten von Primitiv n zum Zeitpunkt z. $\mathcal{L}_n^o$ stellt die Menge der zeitlich benachbarten Knoten von Primitiv n dar, o entspricht der an z angrenzenden Epoche und kann somit $z-1$ oder $z+1$ repräsentieren. $\mathcal{L}_n^o$ enthält somit, falls sowohl ein vorheriger als auch ein nachfolgender Zeitpunkt existiert, Knoten zu unterschiedlichen Zeitpunkten. Das temporale Interaktionspotential $I_z(C_n^z, C_l^o)$ verknüpft die Labels C_n^z und C_l^o der räumlich überlappenden Pixel n und l zu den benachbarten Zeitpunkten z und o. Das räumliche Interaktionspotential, welches zwei benachbarte Labels C_n^z und C_m^z eines Zeitpunktes z miteinander verknüpft, wird nun, zur besseren Unterscheidung zum temporalen Interaktionspotential I_z, mit I_r abgekürzt.

Wie in dem monotemporalen CRF aus Kapitel 4.3 kann das Assoziationspotential $A_d(C_n^z,\mathbf{d}^z)$ mit der a posteriori Wahrscheinlichkeit $p(C_n^z|\mathbf{x}_n^z(\mathbf{d}^z))$ eines beliebigen probabilistischen Klassifikators, in dieser Arbeit des RRF, modelliert werden (Gleichung 35). Als räumliches Interaktionspotential $I_r(C_n^z, C_m^z, \mathbf{d}^z)$ wird das kontrastsensitive Potts-Modell genutzt (Gleichung

36). Das temporale Assoziationspotential $A_z(C_n^1, \widetilde{C}_n)$ basiert, wie in Kapitel 4.3 eingeführt, auf der Übergangswahrscheinlichkeit $\overrightarrow{\gamma} = p(C|\widetilde{C})$ und der Gewichtung θ (Gleichung 48).

Das temporale Interaktionspotential $I_z(C_n^z, C_l^o)$ hängt von der Übergangswahrscheinlichkeit $\overrightarrow{\gamma}_{ka} = p(C_n^z = C^k|C_l^o = C^a)$ ab, also der Wahrscheinlichkeit von Label $C_n^z = C^k$ zum Zeitpunkt z gegeben dem Label $C_l^o = C^a$ zum Zeitpunkt o:

$$I_z(C_n^z, C_l^o) = \theta_n^z \cdot ln\big(p(C^z|C^o)\big). \tag{63}$$

Die Gewichte θ_n^z des temporalen Interaktionspotentials für jedes Pixel n zum Zeitpunkt z sollen auch hier, wie bei dem temporalen Assoziationspotential $A_z(C_n^1, \widetilde{C}_n)$, eine zeitliche Überglättung vermeiden. Das temporale Interaktionspotential in [Hoberg et al., 2015] basiert ebenfalls auf einer Übergangswahrscheinlichkeit, besitzt jedoch keine individuelle Gewichtung θ_n^z, sondern nur einen Gewichtsfaktor für alle temporalen Interaktionspotentiale.

Die Übergangswahrscheinlichkeit $\overrightarrow{\gamma}_{ka} = p(C^z = C^k|C^o = C^a)$ und das Gewicht θ_n^z werden zusammen mit den Parametern des temporalen und räumlichen Assoziationspotentials in einem sequenziellen Training, welches in Kapitel 4.7.2 erläutert wird, bestimmt.

4.7.2. Sequenzielles Training

Abgesehen von dem räumlichen Interaktionspotential I_r (Gleichung 36) müssen die Parameter der Potentiale aus den Trainingsdaten bestimmt werden. Trainiert werden müssen:

- Klassifikatoren zur Bestimmung der lokalen a posteriori Wahrscheinlichkeiten $p(C_n^z|\mathbf{x}^z)$ des Assoziationspotentials A_d für jeden Zeitpunkt z,
- die Übergangswahrscheinlichkeiten $p(C^1|\widetilde{C})$ des temporalen Assoziationspotentials A_z,
- die Übergangswahrscheinlichkeiten $p(C^z|C^o)$ und $p(C^o|C^z)$ des temporalen Interaktionspotentials I_z und
- die Gewichte θ_n^z für jedes Pixel für das temporale Assoziations- bzw. das Interaktionspotential zu jedem Zeitpunkt.

Die Trainingsdaten bestehen aus den Merkmalsvektoren $\mathbf{x}_n^z$ jedes Pixels n zu jedem Zeitpunkt z und den Labels der veralteten Karte $\widetilde{C}_n$ für jedes Pixel n des ersten Zeitpunktes.

Ein möglicher Ansatz läge darin, die Daten aus jedem Zeitpunkt unabhängig von den anderen mit Hilfe der veralteten Karte zu trainieren und anschließend zu klassifizieren. Anschließend müssten aus den resultierenden Labelbildern die Parameter des temporalen Interaktionspotentials I_z ermittelt werden. Im Extremfall könnten sich jedoch die Objekte in der veralteten Karte und die Objekte in den Sensordaten, etwa aufgrund einer sehr großen zeitlichen Differenz, zu stark voneinander unterscheiden. Die Anzahl an fehlerhaften Trainingslabels könnte zu groß werden für eine korrekte Klassifikation, da selbst bei der Verwendung eines robusten Trainingsprozesses von mehr korrekten als fehlerhaften Trainingsbeispielen pro Ballung einer Klasse im Merkmalsraum ausgegangen wird.

Aus diesem Grund wird in dieser Arbeit sequenziell trainiert und klassifiziert. Es wird also zunächst Zeitpunkt $z = 1$ mit Hilfe der Kartenlabels und den Merkmalsvektoren aus den Daten $\mathbf{d}^1$ trainiert und anschließend klassifiziert, sodass die Labels C_n^1 bekannt sind. Hierfür wird der zuvor vorgestellte Algorithmus aus Kapitel 4.1.2 basierend auf dem RRF verwendet. Die Labels aus der Klassifikation des Zeitpunktes $z = 1$ werden anschließend mit den Merkmalsvektoren der Daten $\mathbf{d}^2$ als Trainingsdaten für den nächsten Zeitpunkt $z = 2$ verwendet. Das Klassifikationsergebnis fungiert demnach als veraltete Karte des Zeitpunktes $z = 2$. Aufgrund der unterschiedlichen Aufnahmezeitpunkte der Daten $\mathbf{d}^1$ und $\mathbf{d}^2$ und aufgrund von Fehlern im Klassifikationsergebnis sind auch die Trainingslabels für den Zeitpunkt $z = 2$ fehlerbehaftet. Auch hier wird daher für das Training und die Klassifikation der in dieser Arbeit vorgestellte Algorithmus verwendet. Nach der Klassifikation sind nun die Labels C_n^2 bekannt und dienen zum Training für den Zeitpunkt $z = 3$, sofern dieser vorhanden ist. Dies wird fortgesetzt bis die Fernerkundungsdaten zu allen Zeitpunkten klassifiziert wurden, wobei immer das Klassifikationsergebnis C_n^i für das Training des nächsten Zeitpunktes $z = i + 1$ verwendet wird.

Zum Zeitpunkt z bestehen demnach die Trainingsdaten aus den Merkmalsvektoren $\mathbf{x}_n^z$ und den Labels $\widetilde{C}_n^z = C_n^{z-1}$. Training und Klassifikation erfolgt wie im monotemporalen Szenario (vgl. Kapitel 4.1). Daraus resultiert die a posteriori Wahrscheinlichkeit $p(C_n^z|\mathbf{x}_n^z)$ und, für $z = 1$, die Übergangswahrscheinlichkeiten $p(C^1|\widetilde{C})$ sowie die Gewichte θ_n^1. Für die anderen Zeitpunkte entstehen ebenfalls die Übergangswahrscheinlichkeiten $p(C^z|\widetilde{C}^z)$ und $p(\widetilde{C}^z|C^z)$ und die Gewichte θ_n. Wegen $\widetilde{C}_n^z = C_n^{z-1}$ können die Gewichte als θ_n^z zum Zeitpunkt z und die Übergangswahrscheinlichkeiten als $p(C^z|C^{z-1})$ und $p(C^{z-1}|C^z)$ interpretiert werden. Die Übergangswahrscheinlichkeiten $p(C^z|C^{z+1})$ und $p(C^{z+1}|C^z)$ ergeben sich aus dem Training des Zeitpunktes $z+1$, sofern Daten zum Zeitpunkt $z+1$ vorhanden sind. Die Übergangswahrscheinlichkeiten werden, wie in Kapitel 4.4 beschrieben, aus den Klassifikationsergebnissen der zwei Zeitpunkte bestimmt. Für die Berechnung von $p(C^1|\widetilde{C})$ werden jedoch die Kartenlabels und das Klassifikationsergebniss vom ersten Zeitpunkt verwendet.

Diese Vorgehensweise weist jedoch auch einige Nachteile auf. Der erste Nachteil ist die Möglichkeit der Fehlerakkumulation aufgrund des sequenziellen Trainings. Im Allgemeinen können die Klassen im Merkmalsraum nicht optimal durch den Klassifikator getrennt werden. Werden Pixel zum Zeitpunkt $z = i$ fehlerhaft klassifiziert, erfolgt das Training des Klassifikators zum Zeitpunkt $z = i + 1$ mit diesen fehlerhaften Labels. Der resultierende Klassifikator für den Zeitpunkt $z = i+1$ weist die Pixel nun ggf. ebenfalls der falschen Klasse zu. Diese Art an Fehlern in den Trainingslabels, die auftreten, wenn Klassifikationsergebnisse als Trainingslabels verwendet werden, tritt meist an Trainingsbeispielen auf, deren Merkmale nicht repräsentativ für die eigene Klasse sind. Ein Beispiel für Objekte mit nicht typischen Merkmalen sind Gebäude mit einer abweichenden Dachfarbe. Sind beispielsweise die Dächer der Gebäude in der Karte rot und die Dächer der neuen Gebäude grau, wird das neue Gebäude vermutlich fehlklassifiziert. Der Klassifikator des nächsten Zeitpunktes wird nun mit den fehlerhaften Labels antrainiert. Dementsprechend wird dieser das Haus mit dem grauen Dach ebenfalls fehlerhaft klassifizieren und ggf. auch weitere neue Häuser mit grauen Dächern. In der Trainingsprozedur werden diese Art an Fehlern im Allgemeinen nicht mit modelliert. Auch einige robuste Klassi-

fikatoren, wie der RRF, fangen diese Fehler nicht ab, sodass die Fehlklassifikationen bestehen bleiben. Zusätzlich können bei jeder Klassifikation der Daten eines neuen Zeitpunktes weitere Fehler hinzukommen.

Ein weiterer Nachteil der sequenziellen Prozedur ist die unabhängige Verwendung der Daten unterschiedlicher Zeitpunkte. Abgesehen von dem Klassifikationsergebnis der vorherigen Epoche wird jeder Zeitpunkt komplett unabhängig von den anderen Zeitpunkten betrachtet. Abhängigkeiten zwischen den Daten unterschiedlicher Zeitpunkte werden nicht mit berücksichtigt. Aus diesem Grund wird nach dem sequenziellen Training eine gemeinsame Klassifikation aller Daten mit Hilfe des multitemporalen CRF durchgeführt.

4.7.3. Ablauf der multitemporalen Klassifikation

In diesem Kapitel wird auf den gesamten Prozessablauf bei der Klassifikation eines multitemporalen Datensatzes eingegangen.

Zunächst werden mit dem sequenziellen Training (Kapitel 4.7.2) für den ersten Zeitpunkt die Übergangswahrscheinlichkeiten $p(C^1|\widetilde{C})$ und die Gewichte θ_n^1 und für jeden weiteren Zeitpunkt die Übergangswahrscheinlichkeiten $p(C^z|C^o)$ und $p(C^o|C^z)$, sowie die Gewichte θ_n^z bestimmt. Diese bilden das temporale Assoziations- und Interaktionspotential (A_z und I_z). Ebenfalls wird im sequenziellen Training für jeden Zeitpunkt der RRF Klassifikator trainiert, um die a posteriori Wahrscheinlichkeit für das datenbezogene Assoziationspotential A_d zu bestimmen. Die Parameter des kontrastsensitiven Potts-Modells, mit dem das räumliche Interaktionspotential I_r modelliert wird, werden nicht trainiert, sondern zuvor empirisch festgelegt.

Anschließend kann die optimale Labelkonfiguration $\hat{\mathbf{C}}$ durch die Maximierung der Wahrscheinlichkeit $p(\mathbf{C}|\mathbf{d}, \widetilde{\mathbf{C}})$ aus Gleichung 62 bestimmt werden (Gleichung 49). Da, wie bei dem monotemporalen CRF, die exakte Bestimmung im Allgemeinen nicht möglich ist, erfolgt auch hier die Inferenz mit Hilfe der LBP (Kapitel 3.7). Bei der LBP werden nun iterativ die Beliefs $Bel(C_n)$ an jedem Knoten n jedes Zeitpunktes z aktualisiert, indem die Knoten an ihre angrenzenden Kanten und die Kanten an ihre angrenzenden Knoten Nachrichten übermitteln. Da zeitlich benachbarte Knoten mit dem temporalen Interaktionspotential verbunden sind, erfolgt dieser Nachrichtenaustausch auch zwischen zwei aufeinander folgenden Zeitpunkten z und $z+1$ in beide Richtungen. Nachdem ein Abbruchkriterium erreicht ist und somit das LBP endet, wird für jeden Knoten das Label C des maximalen Beliefs gewählt. Das Ergebnis der multitemporalen Klassifikation besteht demnach aus den Labels für jedes Bildprimitiv n zu jedem Zeitpunkt z.

5. Aufbau der Experimente

In diesem Kapitel wird der Aufbau der Experimente zur Untersuchung des verwendeten Verfahrens zur Merkmalsselektion, der gewählten Parameter und zur Evaluierung des in dieser Arbeit vorgestellten Klassifikationsalgorithmus beschrieben. Die Ergebnisse der Experimente folgen in Kapitel 6. Zunächst werden die fünf zur Verfügung stehenden Datensätze in Kapitel 5.1 vorgestellt. Da nicht alle Daten zum Training genutzt werden können, wird in Kapitel 5.2 eine Möglichkeit zur Auswahl der Trainingsbeispiele unter Berücksichtigung der fehlerhaften Trainingslabels beschrieben. In Kapitel 5.3 werden die Merkmale, die aus den Daten abgeleitet wurden, gezeigt. Die verwendete Methode zur Merkmalsselektion wird in Kapitel 5.4 bezüglich der Verwendung von Labels aus einer veralteten Karte analysiert. Anschließend werden in Kapitel 5.5 verschiedene Genauigkeitsmaße zur Evaluierung der Experimente vorgestellt, die auf dem Vergleich zwischen dem Klassifikationsergebnis und der Referenz beruhen. In Kapitel 5.6 folgt die Definition der Fragen, welcher in den experimentellen Untersuchungen nachgegangen wird. Aufgrund des engen Zusammenhanges zwischen den Fragen und den Experimenten enthält dieses Kapitel auch das Vorgehen bei den experimentellen Untersuchungen.

5.1. Datensätze

Für die experimentellen Untersuchungen der vorgestellten Ansätze werden fünf Datensätze genutzt: Hameln, Las Vegas, Vaihingen, Herne und Husum. Ein Überblick über die Datensätze gibt Tabelle 2. Die Reihenfolge der Datensätze in der Tabelle ergibt sich aus der Aufteilung in Datensätze mit realen Veränderungen über die Zeit (Hameln, Las Vegas) und simulierten Veränderungen (Vaihingen, Herne und Husum), sowie aus der Bodenpixelgröße. Spezifische Angaben zu jedem Datensatz, wie z.B. die Herkunft, werden in den folgenden Unterkapiteln erläutert.

5.1.1. Hameln

Der Datensatz *Hameln* beinhaltet Fernerkundungsdaten aus den Jahren 2010 und 2016, sowie Kartenmaterial aus den Jahren 1955 und 2010, welche den Innenstadtbereich der Stadt Hameln zeigen. Hameln ist eine Stadt in Niedersachsen (Deutschland) und hat sich baulich zwischen 2010 und 2016 kaum verändert; lediglich am Stadtrand gibt es einige Neubauten. Abb. 16 zeigt unten einen Überblick der Stadt. In dieser Arbeit erfolgt aufgrund der wenigen Änderungen eine Konzentration auf zwei Bereiche (genannt N1 und N2), in denen Neubauten entstanden sind (Abb. 16, unten, rote Kennzeichnung). Die Gebiete sind in Abb. 16 zusätzlich vergrößert für die Jahre 2010 und 2016 zu sehen. Das Neubaugebiet in N1 ist deutlich auf der rechten Seite des Bildes zu erkennen.

1955 existierten kaum Gebäude in den beiden Gebieten N1 und N2. Für Analysen mit der veralteten Karte von 1955 wird daher ein größeres Gebiet (NA) betrachtet (Abb. 16, unten, blaue Kennzeichnung). Es existiert jedoch lediglich eine Referenz für die Gebiete N1 und N2, sodass die Evaluierung des Gebietes NA nur in dem Überlappungsbereich von N1 und NA

	Hameln	Las Vegas, US	Vaihingen	Herne	Husum
Sensordaten	Luftbild	Satellitendaten: Landsat	Luftbild	Satellitendaten: RapidEye	Satellitendaten: RapidEye
spektrale Kanäle	R, G, B, NIR	R, G, B, NIR, IR	R, G NIR	R, G, NIR	R, G, NIR
Höhendaten	DGM, DOM	-	DOM	-	-
Anzahl Gebiete	3	1	10	1	1
Größe $[m^2]$	N1: 703 x 270 N2: 257 x 184 NA: 1913 x 573	75000 x 66000	ca. je 180 x 230	13535 x 12345	17735 x 9990
Größe [pix]	N1: 3515 x 1351 N2: 1286 x 921 NA: 9565 x 2864	2500 x 2200	ca. je 2000 x 2500	2707 x 2469	3547 x 1998
Bodenpixelgröße [m]	$0,20$ $0,50$ $5,00$ DOP DOM DGM	30	0,09	5	5
Karten	real 1955, 2010	real 1986, 1991, 2000	3 simulierte	2 simulierte	2 simulierte
Multitemporal	Ja	Ja	Nein	Nein	Nein
Aufnahmezeit	2010, 2016	1991, 2000, 2016	2008	2009	2010
Klassen	Gebäude Versiegelte Fläche Vegetation	Bebauung Wasser Boden	Gebäude Versiegelte Fläche niedrige Vegetation Bäume	Bebauung Wald Boden	Bebauung Wald Boden
Hauptänderung	niedrige Veg. → Gebäude, Vers. Fläche	Boden → Bebauung	Vers. Fläche, niedrige Veg., Bäume → Gebäude	Wald → Boden; Wald, Boden → Bebauung	Wald → Boden; Wald, Boden → Bebauung

Tabelle 2: Überblick der Datensätze.

erfolgen kann. Das DOP und die veraltete Karte des Gebietes NA befinden sich in Anhang A (Abb. A.1).



Abbildung 16: DOPs von *Hameln*. Oben - Gebiet N1 (rechts) und N2 (links) im Jahr 2010. Mitte - Gebiet N1 (rechts) und N2 (links) im Jahr 2016. Unten - Überblick mit den Gebieten N1 und N2 in rot und NA in blau markiert.

Die Sensordaten des Datensatzes *Hameln* bestehen aus hoch aufgelösten Orthophotos (vier Kanäle, 20 cm Bodenpixelgröße), die im Frühjahr 2010 bzw. im Sommer 2016 entstanden sind, sowie digitalen Oberflächen- und Geländemodellen. In dem Orthophoto von 2010 sind noch Gebäudewände zu sehen; Bei der Orthophotoerstellung diente vermutlich ein DGM anstatt eines DOMs als Grundlage. Das DOM wurde für den Zeitpunkt 2010 durch ein Bildzuordnungsverfahren aus überlappenden, orientierten Luftbildern während eines Vorverarbeitungsschrittes extrahiert. Für den Zeitpunkt 2016 basiert das DOM auf Laserscandaten. Das DGM entstand aus verschiedenen Erfassungsmethoden, wodurch sich auch dessen Qualität räumlich stark unterscheidet. Die zwei veralteten Karten wurden aus unterschiedlichen topographischen Datenbeständen abgeleitet. Die erste Karte zeigt die Stadt um 1955 im Maßstab 1:2500 und stammt aus dem Archiv des Katasteramtes der Stadt Hameln [Stadt Hameln, Abteilung Stadtentwicklung und Planung, 2016]. Es handelt sich hierbei um eine analoge Karte der Stadtmitte, vermutlich ein Teil der *Deutschen Grundkarte* (DGK), die bereits eingescannt und georeferenziert zur Verfügung stand. Ein visueller Vergleich zu den Objekten der Karte von 2010, deren Georeferenzierung als korrekt angenommen wird, zeigte jedoch, dass die

Georeferenzierung der alten Karte noch fehlerbehaftet war. Da das Straßenbild besonders im Stadtzentrum größtenteils gleich geblieben ist, konnten die Mittelpunkte von Straßenkreuzungen als Passpunkte fungieren, mit denen die Parameter einer Affintransformation für die Aktualisierung der Georeferenzierung geschätzt werden konnten. Die anschießende Vektorisierung der eingescannten Karte erfolgte manuell.

Die zweite Karte zeigt den Zeitpunkt um etwa 2010. Sie basiert hauptsächlich auf den Daten des *Amtlichen Topographisch-Kartographischen Informationssystems* (ATKIS®) und des *Amtlichen Liegenschaftskatasterinformationssystems* (ALKIS®)[Arbeitsgemeinschaft der Vermessungsverwaltungen der Länder der Bundesrepublik Deutschland (AdV), 2008]. Da nicht alle Klassen aus ALKIS und ATKIS auch in der DGK existieren, wurde die Klassenstruktur vereinfacht, indem alle Objekte einer der folgenden Klassen zugeordnet wurden: *Gebäude*, *versiegelte Fläche*, *Vegetation*, *Eisenbahnlinie* und *Wasser*. Die Objekte der Klasse *Gebäude* stammen hierbei aus ALKIS, während die anderen Objekte ATKIS entnommen wurden.

Beide Karten basieren auf Datenbeständen, die über einen längeren Zeitraum und mit unterschiedlichen Methoden erfasst wurden. Dementsprechend sind die Begrenzungen einiger Objekte fehlerhaft. Abb. 17 zeigt dies beispielhaft an einem Ausschnitt von N1 für die Klasse *Gebäude*. Mögliche Gründe hierfür sind Fehler bei der Georeferenzierung oder aber auch unterschiedliche Darstellungen der Gebäuderänder aufgrund von Dachüberhängen.



Abbildung 17: Ausschnitt aus Gebiet N1 des Datensatzes *Hameln* mit Fehlern in der Karte. Rot: Umrisse der Gebäude in der Karte aus 2010. Hintergrund: DOP aus 2016.

Ausschließlich für die beiden Gebiete N1 und N2 wurde eine Referenz erstellt, indem die Bereiche in den DOPs aus den Jahren 2010 und 2016 manuell gelabelt wurden. Die Klassenstruktur orientiert sich hierbei an dem Kartenmaterial.

Tabelle 3 zeigt den prozentualen Anteil an Änderungen in den Gebieten N1 und N2. Die Berechnungen erfolgten durch den Vergleich der Karte des älteren Zeitpunktes und der Referenz des Aktuelleren. Dies ist von Bedeutung, da die Referenz von 2010, welche manuell mit Hilfe des DOP erstellt wurde, sich von der Karte aus 2010, die vom LGLN bereit gestellt wurde, besonders an Gebäudegrenzen unterscheidet. Da ein Großteil der Änderungen durch das Bauen neuer Gebäude verursacht wurde, ist auch prozentual die Anzahl an Pixeln der Klasse *Gebäude* bezogen auf die Anzahl aller Pixel in dem Gebiet in Tabelle 3 gelistet. Der Anteil der Gebäudepixel des Jahres 2010 bezieht sich hierbei auf die veraltete Karte.

	Änderungen			Anteil *Gebäude*		
	1955-2010	2010-2016	1955-2016	1955	2010	2016
N1	20	15	23	3	9	12
N2	35	17	39	13	23	27

Tabelle 3: Anteil an Änderungen und der Klasse *Gebäude* in den Gebieten des Datensatzes *Hameln*. Links - Die Anzahl an Pixeln, deren Label aus der Karte sich von dem aus der Referenz unterscheiden, bezogen auf die Anzahl aller Pixel in %. Rechts - Die Anzahl an Pixeln mit dem Label *Gebäude* bezogen auf die Anzahl aller Pixel in %.

5.1.2. Las Vegas

Die Sensordaten des Datensatzes *Las Vegas* sind georeferenzierte *Landsat* Bilder der Jahre 1986, 1991, 2000 und 2016 und zeigen die Stadt Las Vegas in Nevada (USA). Tabelle 4 zeigt, welche Daten mit welchem Sensor und zu welchem Zeitpunkt aufgenommen wurde.

Name	Sensor(en)	verw. Bänder	GSD	Zeitpunkte
Landsat 4,5	Multispectral Scanner (MSS)	1, 2, 4	60 m	1986
Landsat 4,5	Thematic Mapper (TM)	1 - 5	30 m	1991
Landsat 7	Enhanced Thematic Mapper Plus (ETM+)	1 - 5	30 m	2000
Landsat 8	Operational Land Imager (OLI) Thermal Infrared Sensor (TIRS)	2 - 6	30 m	2016

Tabelle 4: Landsat Daten. GSD = Bodenpixelgröße.

Las Vegas ist eine stark expandierende Großstadt. Abb. 18 zeigt den Zuwachs der Stadt von 1986 über 1991 und 2000 bis zum Jahr 2016. Westlich der Stadt liegt der Colorado River, der durch den Hoover-Damm zum Lake Mead aufgestaut wird. Abb. 18 zeigt ein Gebiet, welches sowohl die Stadt als auch Teile des Flusses und den See beinhaltet. In den Experimenten werden ausschließlich die Daten dieses Gebietes verwendet.

Bei dem *Las Vegas* Datensatz entstand die veraltete Karte durch manuelles Labeln der Landsat Daten aus dem Jahr 1986. Als Klassen wurden *Bebauung*, *Wasser* und *Boden* unterschieden, da diese bei der niedrigen Auflösung von 60 m noch relativ gut erkennbar sind. Für die

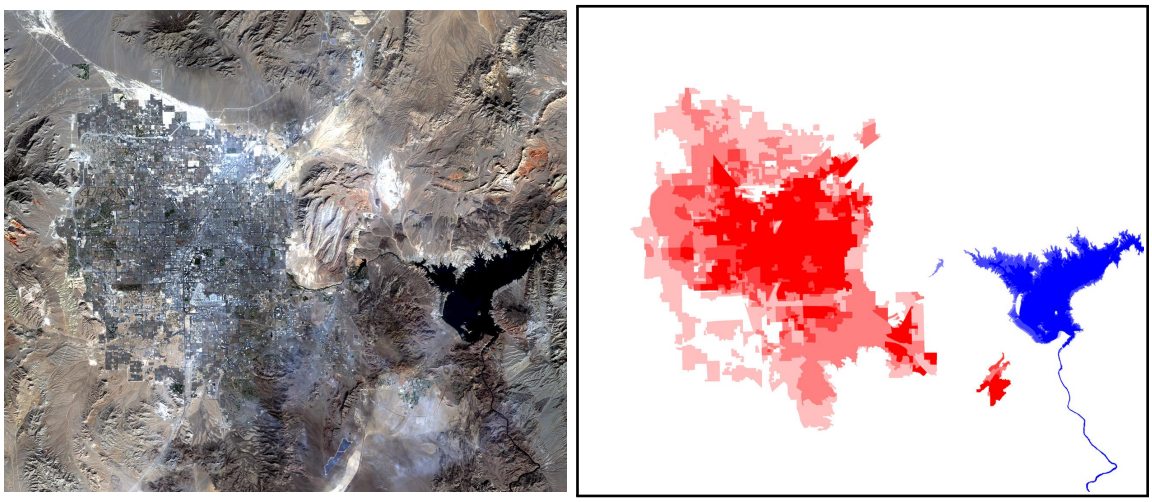


Abbildung 18: Links - Satellitenbild (RGB) von *Las Vegas* aus dem Jahr 2016. Rechts - Visualisierung des Wachstums der Stadt Las Vegas von 1986 bis 2016.
Rot: *Bebauung* [Je dunkler der Farbton, desto älter ist die Bebauung]. Blau: *Wasser*. Weiß: *Boden*.

Klassifikation wurde dieses Landsat Bild nicht weiter verwendet. Die Referenz für den *Las Vegas* Datensatz wurde ebenfalls manuell erstellt und die Klassenstruktur entspricht der aus der Karte. In Anhang A, Abb. A.3, befinden sich die Karte von 1986 und die RGB Bilder aller folgenden Zeitpunkte.

Tabelle 5 zeigt neben der Anzahl der Pixel mit unterschiedlichen Labeln zwischen den einzelnen Jahren auch den prozentualen Anteil an Pixeln der Klasse *Bebauung* im Bild. Hierbei ist zu berücksichtigen, dass sich das Label weniger Pixel in mehreren Zeiträumen ändert, sodass der prozentuale Anteil an Pixeln mit verändertem Label nicht addiert werden kann.

Zeitraum	Änderungen	Jahr	Anteil *Bebauung*
1986 - 1991	2	1986	5
1991 - 2000	7	1991	7
2000 - 2016	7	2000	13
1991 - 2016	13	2016	19
1986 - 2016	15		

Tabelle 5: Anteil an Änderungen und der Klasse *Bebauung* in dem Datensatz *Las Vegas*. Links - Die Anzahl an Pixeln, deren Labels der Referenz zweier Zeitpunkte verschieden sind, bezogen auf die Anzahl aller Pixel in %. Rechts - Die Anzahl an Pixeln mit dem Label *Bebauung* bezogen auf die Anzahl aller Pixel in %.

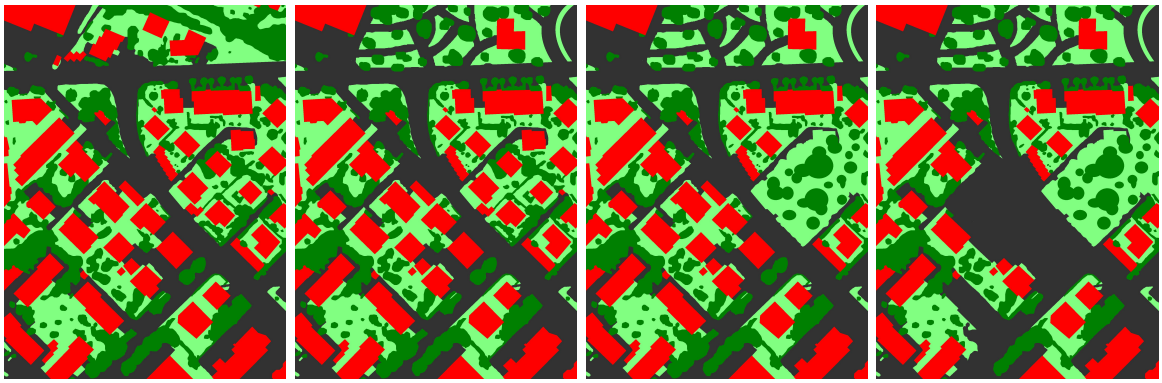
5.1.3. Vaihingen

Der *Vaihingen* Datensatz ist Teil der ISPRS 2D Labelling Challenge [Wegner et al., 2015] und besteht aus 10 Gebieten, zu denen Orthophotos, DOMs und Referenzen bereit gestellt sind. Das DOM wurde mit Hilfe von dichten Bildzuordnungsverfahren (engl. dense image matching) aus überlappenden Luftbildern bestimmt. In dieser Arbeit werden die Gebiete mit den Nummern 1, 3, 5, 13, 17, 21, 23, 30, 32 und 37 verwendet, da für diese alle drei oben

genannten Datentypen zur Verfügung stehen. Obwohl auch für Gebiet Nr. 26 die Datentypen bereit gestellt werden, findet es keine Verwendung in den Experimenten, da es als einziges Gebiet Wasser in Form eines Flusses enthält. Um die Klassifikationsergebnisse der Gebiete aus Vaihingen gut vergleichen zu können, wurde daher entschieden, das Gebiet Nr. 26 nicht zu verwenden. Die verwendeten Gebiete zeigen unterschiedliche Bereiche der Stadt Vaihingen an der Enz in Baden-Württemberg (Deutschland). In Anhang A (Abb. A.4) befindet sich eine Übersicht der Gebiete. Zur Vereinfachung der Notation und zur Unterscheidung der Gebiete des Datensatzes *Vaihingen* von den Gebieten N1 und N2 des Datensatzes *Hameln* werden die Gebiete von *Vaihingen* in dieser Arbeit mit dem Buchstaben A und ihrer Nummer gekennzeichnet: A1, A3, A5, A13, A17, A21, A23, A30, A32 und A37.

Die Referenz besteht aus 6 Klassen: *versiegelte Fläche*, *Gebäude*, *niedrige Vegetation*, *Baum*, *Auto* und *Sonstiges*. Da in Karten im Allgemeinen nicht die Klassen *Auto* und *Sonstiges* vorkommen, wird die Klasse *Auto* der Klasse *versiegelte Fläche* zugeordnet und *Sonstiges* der Klasse *niedrige Vegetation*.

Für den *Vaihingen* Datensatz liegen keine veralteten Daten vor. Die Labels der veraltenen Karte wurden daher durch manuelle Änderungen einiger Labels in den vorhandenen Referenzdaten simuliert. Um das Wachsen an bebauten Gebieten zu simulieren, wurden größtenteils Objekte der Klasse *Gebäude* aus der Referenz durch *versiegelte Fläche*, *niedrige Vegetation* und *Baum* ersetzt. Die Simulation anderer Arten von Änderungen erfolgte ebenfalls, z.B. durch Hinzufügen von Bäumen, jedoch bei einer geringeren Anzahl an Primitiven. Die daraus resultierenden simulierten Karten sehen alle realistisch aus. Für jedes Gebiet wurden drei Karten mit jeweils einem anderen Grad an Veränderungen erstellt. Abb. 19 zeigt beispielhaft die Referenz und die drei Karten für Gebiet A30. Tabelle 6 zeigt die Änderungen in % und den prozentualen Anteil der Klasse *Gebäude*.



(a) Referenz (b) Karte 1 (c) Karte 2 (d) Karte 3

Abbildung 19: Referenz und drei Karten des Gebietes A30 des Datensatzes *Vaihingen*. Rot: *Gebäude*. Grau: *versiegelte Fläche*. Hellgrün: *niedrige Veg.*. Dunkelgrün: *Baum*.

5.1.4. Herne

Der Datensatz *Herne* zeigt die Landschaft nördlich der Stadt Herne in Nordrhein-Westfalen (Deutschland). Es sind beispielsweise Teile von Recklinghausen, Herten und Marl zu sehen.

Name des Gebietes	Änderungen			Anteil *Gebäude*			
	Karte 1	Karte 2	Karte 3	Referenz	Karte 1	Karte 2	Karte 3
A1	7	14	20	38	34	29	24
A3	7	21	28	27	24	18	15
A5	11	23	31	50	43	33	25
A13	17	24	29	18	20	18	17
A17	20	27	36	17	20	17	16
A21	11	27	29	20	18	18	17
A23	14	23	29	18	15	12	10
A30	10	14	22	26	25	22	19
A32	12	16	19	25	26	20	20
A37	9	11	15	23	23	29	25
Mittelwert	12	20	26	26	25	22	19

Tabelle 6: Anteil an Änderungen und der Klasse *Gebäude* in den Gebieten des Datensatzes *Vaihingen*. Links - Die Anzahl an Pixeln, deren Label aus der Karte sich von dem aus der Referenz unterscheiden, bezogen auf die Anzahl aller Pixel in %. Rechts - Die Anzahl an Pixeln mit dem Label *Gebäude* bezogen auf die Anzahl aller Pixel in %.

Zwischen den Städten befinden sich Ackerland und einige Wälder. Das Falschfarbenbild (NIR-R-G) befindet sich in Abb. A.5 in Anhang A.

Die Daten von *Herne* sind Teil des Datensatzes aus [Hoberg et al., 2015] und bestehen aus dem roten, grünen und nahen infraroten Band eines RapidEye Bildes. Die Referenz ist ebenfalls verfügbar und besteht aus den Klassen *Siedlung*, *Industrie*, *Acker*, *Grünfläche*, *Wald*, *Straßen* und *Besondere*. Wie in [Hoberg et al., 2015] wurden die Klassen *Siedlung* und *Industrie* zu der Klasse *Bebauung* zusammengeführt, da die beiden Klassen in den Satellitendaten nicht klar unterschieden werden können. *Acker* und *Grünfläche* bilden die Klasse *Boden*. Die Klasse *Straßen* enthält lediglich Überlandstraßen außerhalb des Ortsgebietes und ist deutlich unterrepräsentiert in den Daten, sodass sie der Klasse *Bebauung* zugeordnet wird. Ebenfalls aufgrund der verhältnismäßig sehr geringen Anzahl an Pixeln mit der Klasse *Besondere* wird diese in die Klasse *Boden* integriert.

Der Datensatz *Herne* ist monotemporal und enthält keine veraltete Karte. Obwohl in [Hoberg et al., 2015] der Datensatz im multitemporalen Kontext Verwendung fand, erfolgte dort eine Manipulation der Sensordaten, um Satellitenbilder eines zweiten Zeitpunktes zu simulieren. Da jedoch in dieser Arbeit eine veraltete Karte benötigt wird, wurde diese durch manuelle Änderungen in der vorhandene Referenz simuliert. Die Entstehung neu bebauter Gebiete wurde durch die Änderung des Labels einiger Pixel der Klasse *Bebauung* aus der Referenz zu *Boden* und *Wald* in der Karte simuliert. Zusätzlich wurde für die Simulation von Entwaldung Pixel der Klasse *Boden* und *Bebauung* in das Kartenlabel *Wald* geändert. Wie bei dem Datensatz *Vaihingen* wurden größere Gebiete geändert, um die Änderungen realistisch zu halten. Für eine erweiterte Analyse wurden zwei Karten mit unterschiedlichen großen Änderungsgebieten erstellt. In der ersten Karte besitzen 5% der Pixel ein anderes Label verglichen zur Referenz, während in der zweiten Karte 17% der Pixel von einer simulierten Änderung betroffen sind. Der Anteil an bebauten Gebieten in der Referenz beträgt 33%, in den Karten sind es 30%

(erste Karte) bzw. 24% (zweite Karte). Die Referenz und die beiden Karten sind ebenfalls in Anhang A (Abb. A.5) ersichtlich.

5.1.5. Husum

Der Datensatz *Husum* beinhaltet die Stadt Husum in Schleswig-Holstein (Deutschland) und deren Umgebung. Neben kleineren Dörfern wie Wittbeck, Schwesing oder Wester-Ohrstedt ist besonders viel Ackerland vorhanden. Die Klasse *Boden* dominiert daher stark diesen Datensatz. In Abb. A.6 in Anhang A befindet sich das Falschfarbenbild (NIR-R-G).

Wie bei dem Datensatz *Herne* wurden auch die Daten von *Husum* bereits in [Hoberg et al., 2015] verwendet. Sie bestehen ebenfalls aus dem roten, grünen und nahen infraroten Band von RapidEye Bildern, und die gegebene Referenz beinhaltet die selben Klassen *Siedlung*, *Industrie*, *Acker*, *Grünfläche*, *Wald*, *Straßen* und *Besondere*, die hier zu den drei Klassen *Boden*, *Bebauung* und *Wald* zusammengeführt wurden. Durch die manuelle Änderung der Labels einiger Pixel in der Referenz wurden auch bei *Husum* zwei veraltete Karten simuliert. In der ersten Karte besitzen 10% der Pixel ein geändertes Label, in der zweiten Karte 20%. Die Änderungsarten entsprechen denen des Datensatzes *Herne*. Die daraus resultierenden Karten sehen ebenfalls realistisch aus. Die Pixel der Klasse *Bebauung* decken hier 11% der Referenz ab, in der ersten bzw. zweiten Karte sind es 10% bzw. 9%. Die Referenz und die beiden Karten sind in Abb. A.6 in Anhang A abgebildet.

5.2. Auswahl der Trainingsbeispiele

Ein Trainingsbeispiel n besteht aus dem Merkmalsvektor $\mathbf{x}_n(\mathbf{d})$, der aus den Daten $\mathbf{d}$ bestimmt wurde, und dem Label aus der veralteten Karte $\widetilde{C}_n$. Die Anzahl an zur Verfügung stehenden Trainingsbeispielen ist demnach sehr groß, da jedes Pixel der Fernerkundungsdaten, welches auch in der Karte abgebildet ist, für das Training genutzt werden kann. Die mögliche Anzahl an Trainingsbeispielen ist jedoch aufgrund des zur Verfügung stehenden Speichers und zur Begrenzung der Rechenzeit eingeschränkt. Deshalb werden N_T Trainingsbeispiele aus der gesamten Anzahl an möglichen Trainingsbeispielen N_{Tg} ausgewählt.

In den Experimenten werden zufällig N_T Trainingsbeispiele ausgewählt. Die Auswahl erfolgt in jeder Iteration vor dem Training, um den Prozess durch einen weiteren Zufallsfaktor zu stabilisieren.

Besonders bei Trainingsdatensätzen, bei denen die Anzahl an Trainingsbeispielen einer oder mehrerer Klassen deutlich kleiner ist als die Anzahl der anderen Klassen, wird im Allgemeinen eine Auswahl bevorzugt, bei der die Anzahl der Trainingsbeispiele pro Klasse in etwa konstant bleibt. Zunächst wird hierfür die tatsächliche Anzahl der Trainingsbeispiele $\#C^i$ jeder Klasse C^i bestimmt. Anschließend wird ein Trainingsbeispiel mit einer Wahrscheinlichkeit von $min(1, \frac{N_{TK}}{\#C^i})$ für das Training verwendet, wobei N_{TK} die gewünschte Anzahl an Trainingsbeispielen pro Klasse ist. Es wird nur ein Prozentsatz $s_i = \frac{N_{TK}}{\#C^i}$ aller zur Verfügung stehenden Trainingsbeispiele der Klasse C^i für das Training genutzt.

Durch die Verwendung der Labels einer veralteten Karte, ist die tatsächliche Anzahl der Trainingsbeispiele $\#C^i$ pro Klasse C^i nicht bekannt. Stattdessen wird die Anzahl $\#\widetilde{C^k}$ der Trainingsbeispiele mit dem beobachteten Label $\widetilde{C^k}$ bestimmt.

Ist die Übergangswahrscheinlichkeit $\overrightarrow{\gamma} = p(C^i|\widetilde{C}^k)$ bekannt, kann die tatsächliche Anzahl $\#C^i$ der Trainingsbeispiele jeder Klasse C^i aus $\widetilde{C}^k$ abgeschätzt werden:

$$
\begin{aligned}
p(\widetilde{C^k}) &= \frac{\#\widetilde{C^k}}{N_T}, \\
p(C^i) &= \frac{1}{J} \cdot \sum_{k}^{K} p(C^i|\widetilde{C}^k) \cdot p(\widetilde{C^k}), \\
\#C^i &= p(C^i) \cdot N_T,
\end{aligned}
$$

mit $J = \sum_a^K \sum_b^K p(C^i|\widetilde{C}^k)$ für die Normierung der Wahrscheinlichkeiten.

Bei dem iterativen Algorithmus auf Basis des RRF basiert, wie in Kapitel 4.1.2 beschrieben, die Auswahl bei der Initialisierung mit dem RF zunächst direkt auf den beobachteten Labels, da die Übergangswahrscheinlichkeiten noch unbekannt sind. Nachdem diese aus dem Ergebnis der Initialisierung und der Karte, wie in Gleichung 53 beschrieben, bestimmt wurden, erfolgt die Schätzung der tatsächlichen Anzahl der Trainingsbeispiele jeder Klasse.

Bei der Klassifikationsmethode auf Basis der RLR (Kapitel 4.6) kann keine ausgewogene Auswahl wie zuvor beschrieben durchgeführt werden, da ansonsten die Bestimmung der Übergangswahrscheinlichkeiten $\overleftarrow{\gamma}$ im Training verfälscht wird. Aus diesem Grund werden hier N_T Trainingsbeispiele zufällig und unabhängig von dem beobachteten Klassenlabel ausgewählt. Im Folgenden wird kurz gezeigt, warum die Übergangswahrscheinlichkeit $\overleftarrow{\gamma}$ durch eine ausgewogene Auswahl beeinflusst wird.

Da die Übergangswahrscheinlichkeit $\overrightarrow{\gamma}$ für die oben genannte Umrechnung bei dem Training der RLR noch nicht bekannt ist, muss die Anzahl $\#\widetilde{C^i}$ der Trainingsbeispiele mit dem beobachteten Label $\widetilde{C^i}$ für eine ausgewogene Auswahl verwendet werden. Der Zusammenhang des Prozentsatzes $s_i \leq 1$ der Klasse C^i mit der gewünschten Anzahl an Trainingsbeispielen pro Klasse N_{TK} ist demnach:

$$
N_{TK} = s_i \cdot \sum_{n=1}^{N} \left(\delta(\widetilde{C}_n = C^i)\right) = s_i \cdot \sum_{k=1}^{K} \Big(\sum_{n=1}^{N} \left(\delta((C_n = C^k) \wedge (\widetilde{C}_n = C^i))\right)\Big),
$$

wobei δ die Kronecker-Delta-Funktion ist. Um den Einfluss dieses klassenspezifischen Prozentsatzes auf die Übergangswahrscheinlichkeiten $\overleftarrow{\gamma}$ zu verdeutlichen, wird im Folgenden ein Prozentsatz s_i in die Gleichung 52 zur Bestimmung der Übergangswahrscheinlichkeiten $\overleftarrow{\gamma}$ integriert.

$$
\overleftarrow{\gamma}_{ka}(s_i) = \frac{\sum_{n=1}^{N} \left(\delta_{si} \cdot \delta((C_n = C^k) \wedge (\widetilde{C}_n = C^a))\right)}{\sum_{b=1}^{K} \Big(\sum_{n=1}^{N} \left(\delta_{si} \cdot \delta((C_n = C^k) \wedge (\widetilde{C}_n = C^b))\right)\Big)}, \tag{64}
$$

mit $\delta_{si} = \delta(\widetilde{C_n} = C^i) \cdot (s_i - 1) + 1$. δ_{si} ist eine Funktion, die dem Prozentsatz s_i entspricht, wenn $(\widetilde{C_n} = C^i)$ und 1 wenn $(\widetilde{C_n} \neq C^i)$. Bei einem Vergleich mit der ursprünglichen Gleichung 52

ist ein Einfluss des Prozentsatz s_i durch δ_{si} zu erkennen:

$$\begin{aligned}
\overleftarrow{\gamma}_{ka} &= \frac{\sum_{n=1}^{N}\left(\delta((C_n = C^k) \wedge (\widetilde{C}_n = C^a))\right)}{\sum_{n=1}^{N}\left(\delta(C_n = C^k)\right)} \\
&= \frac{\sum_{n=1}^{N}\left(\delta((C_n = C^k) \wedge (\widetilde{C}_n = C^a))\right)}{\sum_{b=1}^{K}\left(\sum_{n=1}^{N}\left(\delta((C_n = C^k) \wedge (\widetilde{C}_n = C^b))\right)\right)} \\
&\neq \\
\overleftarrow{\gamma}_{ka}(s_i) &= \frac{\sum_{n=1}^{N}\left(\delta_{si} \cdot \delta((C_n = C^k) \wedge (\widetilde{C}_n = C^a))\right)}{\sum_{b=1}^{K}\left(\sum_{n=1}^{N}\left(\delta_{si} \cdot \delta((C_n = C^k) \wedge (\widetilde{C}_n = C^b))\right)\right)}.
\end{aligned}$$

In den Experimenten wird demnach bei allen Klassifikatoren, die auf dem Random Forest basieren, eine ausgewogene Auswahl der Trainingsbeispiele durchgeführt. Wenn die Übergangswahrscheinlichkeiten $\overrightarrow{\gamma}$ bereits in der Initialisierung bestimmt wurden, wird die tatsächliche Anzahl $\#C^i$ der Trainingsbeispiele jeder Klasse C^i zunächst geschätzt. Ansonsten wird die Anzahl der Trainingsbeispiele $\#\widetilde{C}^i$ aus der veralteten Karte verwendet.

Bei den Klassifikatoren basierend auf der logistischen Regression erfolgt eine zufällige Auswahl der Trainingsbeispiele ohne Berücksichtigung des Labels. Die Auswahl ist demnach nicht ausgewogen, um die Bestimmung der Übergangswahrscheinlichkeiten $\overleftarrow{\gamma}$ im Training nicht zu verfälschen.

Bei allen Experimenten werden etwa 1% der verfügbaren Daten zum Training verwendet. Die einzige Ausnahme bildet die Klassifikation des Gebietes NA aus dem Datensatz *Hameln*, aufgrund der sehr hohen Anzahl zur Verfügung stehender Trainingsbeispiele. Hier wurden nur 0,01% der 9565×2864 Pixel verwendet.

5.3. Merkmale

Aus den Eingangsdaten jedes Datensatzes wird zunächst ein Satz an Merkmalen extrahiert. Jeder Satz enthält dabei zunächst die spektralen Merkmale bestehend aus den ursprünglichen Bändern, Helligkeit, Farbton, Sättigung und NDVI. Falls Höhendaten vorhanden sind, wird die Höhe über Grund ebenfalls als Merkmal aufgenommen. Da bei dem *Vaihingen* Datensatz nur ein DOM vorhanden ist, wurde das DGM, wie in Kapitel 3.2.4 beschrieben, aus dem DOM angenähert. Die Größe der beiden Strukturelemente basiert auf Annahmen über die minimale und maximale Größe abgebildeter Objekte in dem DOM. In *Vaihingen* sind die Objekte über dem Gelände im Allgemeinen Häuser und Bäume. Der Schwellwert S^{Diff} zur Kombination der beiden Höhenmodellen DHM^{gross} und DHM^{klein} wird auf $1m$ gesetzt, da angenommen wird, dass jedes Gebäude die Höhe von $1m$ überschreitet. Die minimale Breite eines Objektes, welches bei der Annäherung des DGM berücksichtigt werden soll, wird mit $1m$ angenommen. Daraus resultiert die Größe des kleineren Strukturelementes $S^{klein} = 2 \cdot \lfloor \frac{0{,}5}{GSD} \rfloor + 1$. Die größten Objekte über der Geländeoberfläche sind in Vaihingen Industriegebäude. Resultierend aus der Annahme, dass die Gebäude nicht breiter als $40m$ sind, ist die Größe des größeren Strukturelementes $S^{gross} = 2 \cdot \lfloor \frac{20}{GSD} \rfloor + 1$. Durch die Glättungsoperationen mit einem Medianfilter und

einem Gaußfilter sollen aus dem Opening resultierende Artefakte reduziert werden. Die Breite der Artefakte wird mit etwa $4m$ bis $5m$ angenommen. Die Glättung erfolgt daher mit einem Medianfilter der Größe $S^{Median} = 2 \cdot \lfloor \frac{2}{GSD} \rfloor + 1$ und einem Gaußfilter der Standardabweichung $\sigma^{GaussDGM} = 2 \cdot \lfloor \frac{2,5}{GSD} \rfloor + 1$.

Durch die Glättung der spektralen Merkmale *Rot*, *Grün*, *Blau*, *NIR*, *IR*, *NDVI* und *nDOM* (soweit vorhanden) mit Hilfe eines Gaußfilters mit einer Standardabweichung der Größe $\sigma = 2$ entstehen weitere, weniger durch Rauschen beeinträchtigte, Merkmale. Bereiche der Vegetation weisen meist einen hohen Wert bei dem Merkmal NDVI auf. Vereinzelte Pixel können jedoch einen deutlich geringeren Wert aufzeigen. Dies geschieht beispielsweise bei Bäumen, aufgrund von kleinen lokalen Schatten. Aus diesem Grund wird zusätzlich das Maximum der NDVI-Werte in einer 5×5 Umgebung des Pixels als zusätzliches Merkmal berechnet.

Für die Berechnung der Texturmerkmale Energie, Kontrast, Homogenität und Korrelation wird die 5×5 Umgebung eines Pixels genutzt. Die Haralick-Merkmale basieren auf einer Co-Occurance Matrix mit einer Distanz von $d = 1$ und einer Richtung von $\alpha = 0°$. Die Grauwerte wurden aus den RGB Werten bestimmt:

$$g = 0,299 \cdot g_R + 0,587 \cdot g_G + 0,114 \cdot g_B.$$

Ist, wie bei dem Vaihingen Datensatz, kein blauer Kanal vorhanden sondern der nahe Infrarote, wird der nahe infrarote Kanal verwendet:

$$g = 0,299 \cdot g_{NIR} + 0,587 \cdot g_R + 0,114 \cdot g_G.$$

Ein weiteres verwendetes Merkmal ist angelehnt an die HOG Merkmale. Wie in 3.2.3 beschrieben, werden der Gradientenbetrag $|h|$ und die Richtung ϕ bestimmt und die gewichteten Histogramme berechnet. Es werden 9 Richtungsklassen zwischen 0° und 180° betrachtet, wobei die Richtung $\phi - 180°$ für die Einteilung verwendet wird, wenn ϕ größer als 180° ist. Innerhalb jeder Richtungsklasse werden nun die Histogrammeinträge über alle Pixel normiert. Für jedes Primitiv werden anschließend die Histogrammeinträge aller Richtungsklassen summiert und durch die Anzahl an Klassen dividiert. Das so entstandene Merkmal ist vergleichbar mit dem Mittelwert der Gradienten innerhalb der Region. Der Unterschied besteht in der Normierung aller Pixel innerhalb einer Richtungsklasse, wodurch Primitive hervorgehoben werden, die mindestens einen Histogrammeintrag, der innerhalb seiner Richtungsklasse relativ groß ist, besitzen. Dieses Merkmal wird in dieser Arbeit als MHOG bezeichnet, wobei das M für *Mittel* steht.

Nach dem Extrahieren erfolgt eine Normalisierung jedes Merkmals über alle Pixel auf den Bereich zwischen 0 und 1. Die komplette Liste der extrahierten Merkmale findet sich in Tabelle 7. Die theoretischen Grundlagen sind in Kapitel 3.2 zu finden. Aus diesem Pool an Merkmalen werden anschließend die Merkmale selektiert, die relevant für die Klassifikation sind. Hierfür wird des Verfahren aus Kapitel 3.2.5 verwendet. Im folgenden Kapitel wird diese Methode in Hinblick auf die Verwendung von Trainingslabels, die aus einer veralteten Karte stammen, analysiert.

Gruppe	Untergruppe	Bezeichnung	Hameln	Las V.	Vaih.	Herne	Husum
Spektral	Originalbänder	Blau	x	x			
		Grün	x	x	x	x	x
		Rot	x	x	x	x	x
		NIR	x	x	x	x	x
		IR		x			
	HSV-Farbraum	Helligkeit	x	x	x	x	x
		Farbton	x	x	x	x	x
		Sättigung	x	x	x	x	x
	NDVI		x	x	x	x	x
Höhe	nDOM		x		x		
Kontext	Max(NDVI)		x	x	x	x	x
	Gaußfilter	Gauß(Blau)	x	x			
	$\sigma = 2$	Gauß(Grün)	x	x	x	x	x
		Gauß(Rot)	x	x	x	x	x
		Gauß(NIR)	x	x	x	x	x
		Gauß(IR)		x			
		Gauß(NDVI)	x	x	x	x	x
		Gauß(nDOM)	x		x		
Textur	Haralick	Energie	x	x	x	x	x
		Kontrast	x	x	x	x	x
		Homogenität	x	x	x	x	x
		Korrelation	x	x	x	x	x
Struktur	MHOG		x	x	x	x	x

Tabelle 7: Extrahierte Merkmale.

5.4. Analyse der Merkmalsselektionsmethode

In dieser Arbeit wird der Selektionsalgorithmus aus [Frénay et al., 2014] verwendet, basierend auf der Transinformation [Shannon, 1948], wie in Kapitel 3.2.5 beschrieben. Frénay et al. [2014] analysierten die Methode in Hinblick auf die Verwendung fehlerhafter Trainingslabels und erweitern sie entsprechend. In dieser Arbeit treten die Fehler jedoch nicht zufällig auf, wie in Frénay et al. [2014], sondern sie resultieren aus der Verwendung einer veralteten Karte. In diesem Kapitel wird daher untersucht, ob die Erweiterung auch in diesem Fall sinnvoll ist.

Frénay et al. [2014] zeigen an einem Beispiel, in dem das Label zufällig ausgewählter Trainingsbeispiele geändert wurde, dass diese Art der Merkmalsselektion durch fehlerhafte Trainingslabels beeinflusst wird und zu schlechteren Ergebnissen in einer darauf folgenden kNN (eng. k-Nearest-Neighbor) Klassifikation führt.

Die Ursache für die beobachtete Verschlechterung der Ergebnisse in den Experimenten läge bei der Schätzung der Transinformation (Gleichung 15) mit fehlerbehafteten Trainingslabels. Ein fehlerhaftes Trainingsbeispiel mit der korrekten Klasse C^k und der beobachteten Klasse C^a geht bei der Schätzung der bedingten Wahrscheinlichkeit (Gleichung 14) in die Berechnung der Entropie $\hat{\mathcal{E}}(\mathbf{x}|C = C^a)$ ein, anstatt in die Berechnung von $\hat{\mathcal{E}}(\mathbf{x}|C = C^k)$. Folglich ist sowohl die Schätzung der Entropie $\hat{\mathcal{E}}(\mathbf{x}|C = C^a)$ und $\hat{\mathcal{E}}(\mathbf{x}|C = C^k)$ fehlerhaft. Durch die Herausnahme oder Hinzunahme eines Trainingsbeispieles kann die Distanz $\frac{\epsilon_\xi(n)}{2}$ zu dem ξ-ten Nachbarn von einigen Trainingsbeispielen vergrößert oder verkleinert werden. Viel wichtiger

sei jedoch, dass $\epsilon_\xi(n)$ für ein Trainingsbeispiel mit falschem Label sehr groß werden könne. Ist das Trainingsbeispiel n mit fehlerhaftem Label $\widetilde{C}_n = C^a$ umgeben von Trainingsbeispielen der Klasse C^k, da das Trainingsbeispiel n korrekterweise der Klasse C^k angehört, liefert die Schätzung der Entropie einen zu kleinen Wert.

Frénay et al. [2014] stellen zur Behebung dieser Beeinflussung eine Selektionsmethode vor, die mit fehlerhaften Trainingslabels umgehen kann. Das Auftreten fehlerhafter Labels wird dabei wie folgt modelliert: Ein Trainingsbeispiel mit der wahren Klasse C^a besitzt mit einer Wahrscheinlichkeit $p_e(a)$ ein fehlerhaftes Label, welches zufällig aus den verbleibenden Klassen gewählt ist. Dies entspricht einer spezifischen Form des NAR Modells (Kapitel 2.3), bei der lediglich die K Wahrscheinlichkeiten p_e anstatt $K \times K$ Übergangswahrscheinlichkeiten für die Modellierung der Fehler in den Labels genutzt werden. In einem EM-Algorithmus werden dann die wahren Labels und die Wahrscheinlichkeiten p_e prädiziert. Die Schätzung des wahren Labels beruht dabei auf der Annahme, dass fehlerhaft gelabelte Trainingsbeispiele von korrekten Trainingsbeispielen umgeben sind.

Die Analyse und die Erweiterung des Selektionsalgorithmus von Frénay et al. [2014] basiert auf diesem Fehlermodell, welches jedoch für die in dieser Arbeit behandelte Problemstellung im Allgemeinen nicht ausreichend ist. Da die Labels der Trainingsbeispiele aus einer veralteten Karte stammen, resultieren die Fehler aus Änderungen der Landbedeckung zwischen zwei Zeitpunkten und sind nicht durch reinen Zufall entstanden. Entsteht ein neues Objekt, sind die Merkmale dieses Objektes einander recht ähnlich. Besonders die Merkmale benachbarter Pixel, die das gleiche Objekt zeigen, unterscheiden sich kaum. So liegen solche Pixel mit fehlerhaftem Label nicht nur im Objektraum, sondern auch im Merkmalsraum nahe beieinander und bilden somit Ballungen im Merkmalsraum. Im Gegensatz zu Ballungen aus Trainingsbeispielen mit korrekten Labels ist der Abtsand zwischen den Trainingsbeispielen in Ballungen aus fehlerhaften Trainingsbeispielen im Mittel etwas größer, da sie aus deutlich weniger Trainingsbeispielen bestehen.

Abb. 20 zeigt beispielhaft an zwei Klassen in einem 2D Merkmalsraum den Unterschied bei der Fehlermodellierung. Auf der rechten Seite wurde vereinzelten Trainingsbeispielen zufällig anstatt des korrekten Labels *Blau* das Label *Grün* zugewiesen (Abb. 20, rechts, rot). Deren nächste Nachbarn des gleichen Klassenlabels sind demnach relativ weit entfernt (Abb. 20, rechts, orangefarbende Linien). Anders sieht es auf der linken Seite von Abb. 20 aus, wo ein komplexeres Fehlermodell, basierend auf der Annahme, dass fehlerhafte Trainingsbeispiele Ballungen im Merkmalsraum bilden, angenommen wird. Hier ist die Distanz der fehlerhaften Trainingsbeispiele zu den nächsten Nachbarn der gleichen beobachteten Klasse nicht deutlich größer als jene der korrekten Labels. Obwohl die Merkmale eines neuen Objektes sich nicht unbedingt alle stark ähneln, kann dennoch davon ausgegangen werden, dass jedes Pixel dieses Objektes ähnliche Merkmale aufweist, wie mindestens ein zweites Pixel dieses Objektes. Meist sind es zwei im Bildraum benachbarte Pixel eines Objektes, die ähnliche Merkmale aufweisen und daher auch im Merkmalsraum nah beieinander liegen.

Nachdem gezeigt wurde, wodurch sich die beiden Fehlermodelle unterscheiden, wird im Folgenden analysiert, ob fehlerhafte Trainingslabels, die aus der Verwendung veralteter Karten

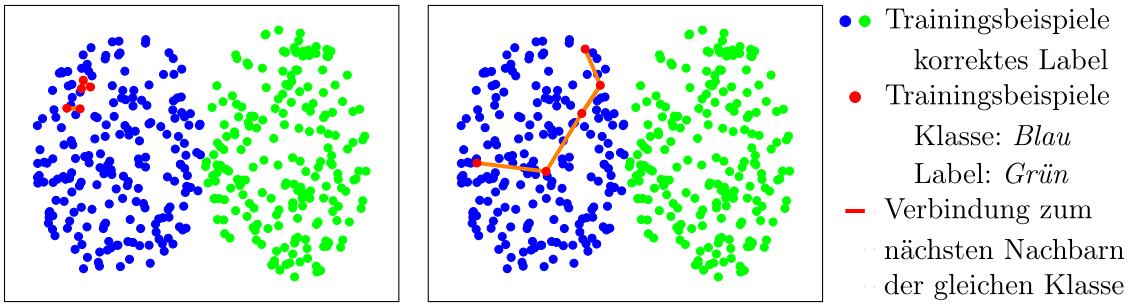


Abbildung 20: Trainingsbeispiele mit fehlerhaftem Label, die im Merkmalsraum Ballungen bilden (links) und zufälliges Auftreten fehlerhafter Labels (rechts). Die orangefarbenen Linien zeigen für jedes fehlerhaft gelabelte Trainingsbeispiel den nächsten Nachbarn der gleichen Klasse. Bilden die fehlerhaften Trainingsbeispiele Ballungen, ist auch bei diesen Trainingsbeispielen diese Distanz relativ klein und somit die orangefarbenen Linie sehr kurz.

stammen, einen ähnlichen Einfluss auf das Selektionsergebnis ausüben, wie die Fehler, deren Auftreten ausschließlich von dem Klassenlabel abhängen und somit dem NAR Modell entsprechen.

Frénay et al. [2014] zeigen auf, dass die klassenbezogene Entropie sich durch die Hinzunahme bzw. Herausnahme eines Trainingsbeispieles ändert. Dieser Einwand bleibt unabhängig von der Verteilung der fehlerhaften Trainingsbeispiele im Merkmalsraum bestehen. In der Gleichung 11, mit deren Hilfe die Entropie $\hat{\mathcal{E}}(\mathbf{x}|C = C^k)$ geschätzt wird, ändert sich sowohl die Anzahl N und somit $\Psi(N)$ und die Normierung mit $\frac{1}{N}$ der Summe $\sum_{n=1}^{N} ln(\epsilon_\xi(n))$. Auch die zweifache Distanz $\epsilon_\xi(n)$ einiger Trainingsbeispiele ändern sich. In [Frénay et al., 2014] wird jedoch bereits darauf hingewiesen, dass dies vermutlich nicht zu den starken Unterschieden der MI zwischen der Verwendung fehlerfreier und fehlerbehafteter Trainingslabels führt.

Das größte Problem sei laut Frénay et al. [2014] die veränderte Distanz bei den fehlerhaften Trainingslabels. Bilden die Trainingsbeispiele einer Klasse Ballungen im Merkmalsraum und liegen Fehler in den Trainingslabels vor, die lediglich vom Objekttyp abhängen, so wird, wie in Abb. 20 zu sehen ist, die Distanz $\epsilon_\xi(n)$ verhältnismäßig groß. Da die Trainingsbeispiele mit fehlerhaften Labels in dieser Arbeit Ballungen im Merkmalsraum bilden, sind die Distanzen zum nächsten Nachbarn kaum größer als bei den korrekten Beispielen. Eine leichte Vergrößerung der Distanz ist jedoch aufgrund der geringeren Dichte der Ballungen zu erwarten. Dies führt zu einer Verringerung der Entropie $\hat{\mathcal{E}}(\mathbf{x}|C = C^k)$ und somit auch zu einer Verringerung der Transinformation. Die Tatsache, dass bei der Merkmalsselektion die MI-Werte relativ zueinander betrachtet werden, könnte ebenfalls den negativen Einfluss der Fehler in den Labels reduzieren, da die Labelfehler bei allen Merkmalskonstellationen vorhanden sind. Dies müsste jedoch noch genauer untersucht werden.

Die von Frénay et al. [2014] vorgeschlagene Erweiterung beruht ebenfalls auf der Annahme, dass Trainingsbeispiele mit falschem Label zu einer starken Verringerung der MI führen und dadurch als Ausreißer erkannt werden. Außerdem wird davon ausgegangen, dass das Auftreten der fehlerhaften Labels auf dem Zufall beruht und ausschließlich von dem Klassenlabel abhängt. Da diese Annahme bei der Verwendung von Labels aus veralteten Karten zum Trai-

ning nicht zutrifft, wird in der vorliegenden Arbeit die Erweiterung nicht verwendet, sondern das ursprüngliche Verfahren, wie in Kapitel 3.2.5 beschrieben.

5.5. Genauigkeitsmaße

Für die Bewertung des Klassifikationsergebnisses wird eine pixelbasierte *Konfusionsmatrix* $\mathcal{KM}$ aus den Labelbildern der Klassifikation und der Referenz erstellt. Ein Element $\mathcal{KM}_{ij}$ in der Zeile i und der Spalte j setzt sich dabei aus der Anzahl aller Pixel mit dem Label i aus der Klassifikation und Label j aus der Referenz zusammen. Auf der Hauptdiagonalen der Konfusionsmatrix erscheint demnach die Anzahl korrekt klassifizierter Pixel pro Klasse.

Das Verhältnis der Anzahl korrekt klassifizierter Pixel zu deren Gesamtanzahl ergibt die *Gesamtgenauigkeit GG*:

$$\text{Gesamtgenauigkeit} \quad GG = \frac{\sum_i \mathcal{KM}_{ii}}{\sum_i \sum_j \mathcal{KM}_{ij}}. \tag{65}$$

Aussagen bezüglich einzelner Klassen können jedoch nicht auf Grundlage der Gesamtgenauigkeit getroffen werden. Besonders wenn die Anzahl an Pixeln pro Klassen stark variiert und demnach wenige Klassen dominieren, ist die Gesamtgenauigkeit für die alleinige Bewertung der Ergebnisse nicht gut geeignet.

Für eine detailliertere Analyse werden daher oft die *Vollständigkeit VG* [Heipke et al., 1997], auch *User's Accuracy*, *Recall* oder *Completeness* genannt, und *Korrektheit KG* [Heipke et al., 1997], auch *Producer's Accuracy*, *Precision* oder *Correctness* genannt, bevorzugt. Die Vollständigkeit ist hierbei das Verhältnis der richtig erkannten Pixel einer Klasse zur Gesamtanzahl der Pixel dieser Klasse in der Referenz und beschreibt somit, wie viele Pixel einer bestimmten Klasse korrekt detektiert wurden:

$$\text{Vollständigkeit} \quad VG_i = \frac{\mathcal{KM}_{ii}}{\sum_j \mathcal{KM}_{ij}}. \tag{66}$$

Die Korrektheit hingegen gibt das Verhältnis zwischen den korrekt klassifizierten Pixeln einer Klasse und der Gesamtanzahl der Pixel dieser Klasse im Klassifikationsergbnis an. Sie beschreibt demnach wie viele Pixel, die in der Klassifikation einer bestimmten Klasse zugeordnet wurden, auch in der Referenz dieser Klasse entsprechen:

$$\text{Korrektheit} \quad KG_i = \frac{\mathcal{KM}_{ii}}{\sum_j \mathcal{KM}_{ji}}. \tag{67}$$

Für eine Aussage über das Klassifikationsergebnis bezogen auf eine Klassen werden beide Genauigkeitsmaße kombiniert. Eine Möglichkeit bildet das harmonische Mittel aus Vollständigkeit und Korrektheit; der *F1 Wert* [Wegner et al., 2015]:

$$\text{F1 Wert} \quad F1_i = 2 \cdot \frac{KG_i \cdot VG_i}{KG_i + VG_i} = \frac{2}{\frac{1}{KG_i} + \frac{1}{VG_i}}. \tag{68}$$

Eine Aussage für alle Klassen kann beispielsweise mit dem Mittelwert des F1 Wertes $MF1 = \frac{1}{K}\sum_i F1$ über alle Klassen K erfolgen.

5.6. Aufbau und Zielsetzung der Experimente

5.6.1. Merkmalsselektion

Bevor die Klassifikation der Daten erfolgen kann, müssen die Merkmale, auf deren Basis die Klassifikation erfolgt, ausgewählt werden. Obwohl diese Arbeit ihren Fokus auf das Training und die Klassifikation legt, soll dennoch gezeigt werden, dass auch für die Merkmalsselektion keine manuell erstellten Labels benötigt werden.

Die Merkmalsselektion erfolgt daher mit der Rückwärtssuche basierend auf der Transinformation (Kapitel 3.2.5). Der Einfluss von Fehlern in den Labels, die entstehen, wenn die Labels aus einer veralteten Karte extrahiert werden, sollen entsprechend Kapitel 5.4 bei dieser Methode gering sein. In den Experimenten wird diese Annahme überprüft, indem die Selektionsergebnisse, die unter Verwendung von Karten mit einer unterschiedlichen Anzahl an Fehlern in den Labels entstehen, verglichen werden. In Kapitel 6.1 wird daher folgende Fragen untersucht:

- Hat die Tatsache, dass die beobachteten Labels aus einer veralteten Karte stammen und somit fehlerbehaftet sind, Einfluss auf die Merkmalsselektion?

Die Anzahl der zu selektierenden Merkmale wird nicht festgelegt. Stattdessen wird mit Hilfe der Transinformation, die mit bzw. ohne ein Merkmal bestimmt wurde, entschieden, ob dieses Merkmal selektiert wird oder nicht. Wie bei der Klassifikation wird davon ausgegangen, dass die beobachteten Labels aus einer veralteten Karte stammen, sodass auch die Merkmalsselektion auf diese Labels zurückgreift. Für die Evaluierung der Merkmalsselektion wurde zusätzlich die Merkmalsselektion mit der Referenz durchgeführt; Diese Auswahl findet aber keine Verwendung in den weiteren Experimenten (Kapitel 6.1). Um den Einfluss der fehlerhaften Trainingslabels gering zu halten, wird dabei nur der nächste Nachbar ($k = 1$) betrachtet. Außerdem wird, aus Speicher- und Rechenzeitgründen nur etwa 1% der Daten zufällig und unabhängig von dem Klassenlabel gewählt und für die Merkmalsselektion verwendet.

Die Merkmalsselektion wird für jeden Datensatz und jede veraltete Karte unabhängig durchgeführt. Somit kann bei dem Training und der Klassifikation mit einer bestimmten Karte auf die entsprechenden Merkmale zurückgegriffen werden.

Stehen mehrere Gebiete zur Verfügung, wie bei den Datensätzen *Hameln* und *Vaihingen*, erfolgt die Merkmalsselektion zunächst unabhängig für jedes Gebiet. Im Anschluss werden die Listen der ausgewählten Merkmale kombiniert, sodass sich die selektierten Merkmale aller Gebiete eines Datensatzes in dem gemeinsamen Merkmalsvektor befinden.

Das Gebiet NA des *Hameln* Datensatzes wurde bei der Merkmalsselektion nicht verwendet, da keine Referenz für das gesamte Gebiet vorhanden ist und somit keine Evaluierung des Selektionsergebnisses erfolgen kann. In den multitemporalen Experimenten (Kapitel 6.4) wird daher die Merkmalsauswahl verwendet, die auf der Karte von 2010 und den Daten von 2016

der Gebiete N1 und N2 basiert. Damit ist es das einzige Experiment in dieser Arbeit, bei dem nicht die gleiche veraltete Karte für die Merkmalsselektion und die Klassifikation verwendet wird. Außerdem ist zu erwähnen, dass die Bestimmung der Merkmalsvektoren aus Daten unabhängig für jeden Zeitpunkt erfolgt.

Bei dem multitemporalen Datensatz *Las Vegas* werden neben der Karte aus 1986 auch die Referenzen der Jahre 1991 und 2000 als veraltete Karte interpretiert, um, wie bei den simulierten Datensätzen, durch mehrere Karten mit einer unterschiedlich großen Anzahl an Änderungen den Selektionsalgorithmus auch in dieser Hinsicht zu untersuchen. Die verwendeten Daten zur Merkmalsselektion stammen aus dem Jahr 2016. Die Daten der Jahre 1991 und 2000 wurden hier nicht verwendet.

Die Experimente für die Merkmalsselektion und die selektierten Merkmale aller Datensätze sind in Kapitel 6.1 zu finden.

5.6.2. Parametertests

Die verwendeten Klassifikatoren enthalten einige Parameter, auch Hyperparameter genannt, die vor dem Training festgesetzt werden müssen und die im Training nicht verändert werden. Tabelle 8 zeigt die untersuchten Hyperparameter und die Werte, die für die anschließende Klassifikation genutzt wurden.

Die Wahl der Parameter der logistischen Regression (σ, Φ) und des Random Forest (T_{max}, N_G, N_B, N_F) erfolgte aufgrund von Erfahrungswerten für diese Basisklassifikatoren bei deren Anwendung auf Fernerkundungsdaten. Auch die Glättungsparameter des Potts Modells (β_0, β_1) wurden aus Erfahrung festgelegt. Die anderen Parameter resultieren aus Annahmen bezogen auf die gegebenen Daten. Die Diagonale der Übergangsmatrix γ_0 wurde aufgrund der Annahme, dass sich ein Großteil der Labels der veralteten Karte nicht geändert hat und somit nur ein Teil der Labels fehlerhaft ist, mit 80 % bzw. 0,8 belegt. Die Wahl der Parameter zur Aktualisierung der Änderungswahrscheinlichkeit erfolgte abhängig von der Auflösung der Fernerkundungsdaten. Nur die Aktualisierungsfunktion wurde unabhängig von der Auflösung mit f_{a1} anstatt f_{a2} festgelegt, damit auch kleine Werte in wenigen Iterationen stark anwachsen können. Die Anzahl der Iterationen wurde auf maximal 20 begrenzt, da sich das Ergebnis nach weiteren Iterationen meist kaum verändert. Die Anzahl an verwendeten Trainingsdaten ist durch den Speicher und der zur Verfügung stehenden Rechenzeit begrenzt.

Diese Werte könnten beispielsweise mit Hilfe eines Greedy Ansatzes angepasst werden, sodass sich das Klassifikationsergebnis verbessert. Doch auch hier würden korrekt gelabelte Daten für die Bewertung der Parameterwahl benötigt. Da die Trainingslabels in dieser Arbeit jedoch fehlerbehaftet sind, wird von einer solchen Anpassung abgesehen. Dennoch soll der Einfluss der Parameter auf das Klassifikationsergebnis geprüft werden. Hierfür werden in den Parameterstudien verschiedene Werte für die Parameter gewählt und die resultierenden Klassifikationsergebnisse mit Hilfe der manuell erstellten Referenz analysiert.

In Kapitel 6.2 wird die folgende Fragen experimentell untersucht:

- Wie sensitiv ist das Klassifikationsergebnis bezogen auf die zuvor zu definierenden Parameter?

Für die Analyse werden jeweils alle Parameter, bis auf einen, als konstant angesehen. Für diesen zu testenden Parameter werden verschiedene Werte angenommen und für jeden gewählten Wert wird ein Training mit anschließender Klassifikation durchgeführt.

Obwohl der iterative Algorithmus basierend auf dem RRF (**RRFK**), wie in Kapitel 4.1 beschrieben, die Neuerungen dieser Arbeit beinhaltet, wird dieser aufgrund der deutlich längeren Prozessierzeit nicht für die Analyse aller Parameter verwendet. Die Parameter der RLR, des RF und des Potts Modells haben keinen direkten Bezug zu der Aktualisierung der Gebiete mit potenziellen Änderungen. Obwohl indirekt, durch den Einfluss der Parameter auf das Klassifikationsergebnis, auch eine Beziehung vorhanden ist, wird davon ausgegangen, dass Parameterwerte, die bei der Verwendung des RF zu guten Ergebnissen führen, auch für den **RRFK** gut geeignet sind. Aus diesem Grund wird für die Parameter des RF auch dieser für die Analyse verwendet (**RF**); für die Parameter der RLR ist dies die RLR (**RLR**). Die Klassifikation erfolgt beim **RF** und der **RLR** mit dem klassischen CRF. Die Klassifikatoren sind noch einmal in Tabelle 9 zusammengefasst. Die Wahl des Klassifikators für die jeweilige Untersuchung ist in Tabelle 8 ersichtlich. Das Klassifikationsergebnis wird anschließend mit der Referenz verglichen und der MF1 (Kapitel 5.5) bestimmt. Für die Analyse der Parameterwahl werden für eine bessere Übersicht nur 4 monotemporale Szenarien betrachtet.

Die Analyse des Einflusses auf die Klassifikation hoch aufgelöster Fernerkundungsdaten erfolgt mit dem Datensatz *Hameln*. Hierbei werden die aktuellen Sensordaten aus 2016 und die Karte von 2000 verwendet. Um unterschiedliche Grade von Änderungen zu berücksichtigen, werden beide Gebiete N1 und N2 betrachtet.

Die Untersuchung der niedrig aufgelösten Bilder erfolgt mit dem *Las Vegas* Datensatz. Die aktuellen Sensordaten stammen aus dem Jahr 2016 und als Kartenmaterial wird die Karte aus 1986 und die Referenz aus 2000 verwendet, um auch hier unterschiedliche Mengen an Änderungen bei der Analyse zu betrachten.

5.6.3. Monotemporale Klassifikation

Das Hauptziel der experimentellen Untersuchungen dieser Arbeit ist es, zu testen, ob die Neuentwicklungen im Rahmen dieser Arbeit zu einer Verbesserung der Klassifikationsergebnisses führen. Im Bereich der monotemporalen Klassifikation sind diese hauptsächlich der robuste Random Forest, welcher Fehler in den Trainingslabels berücksichtigt, und die Integration der Kartenlabels als Beobachtung in das CRF. Die Fragen, die in den Experimenten bezogen auf die monotemporale Klassifikation untersucht werden, lauten:

a) Verbessert sich die Klassifikation durch die Berücksichtigung der Fehler im Training im Vergleich zu der Annahme, die Labels seien fehlerfrei?

Symbol	Kontext	Kapitel	Beschreibung	Methodik	Testbereich [Schritt]	verwendeter Wert
σ	rob. Log. Regr.	3.3	Regularisierungsparameter	**RLR**	0,5 - 3 [0,5], 4, 5, 10 - 50 [10], 100	10
Φ		3.3	Merkmalsraumtransformation		linear, quadratisch	quadr.
γ_0		3.4	Diagonale der initialen Übergangsmatrix		0,1 - 0,9 [0,1]	0,8
T_{max}	Random Forest	3.5	Maximale Tiefe eines Baumes	**RF**	2 - 9 [1]	7
N_G		3.5	Anzahl der getesteten Trennflächen		50 - 300 [50]	100
N_B		3.6	Anzahl der Bäume		10 - 100 [10]	40
N_F		3.6	Anzahl der Merkmale		1, 3, 5, 7	$\sqrt{F} = 3$
β_0	k.s. Potts-Modell	3.7	Glättungsf./ Einfluss von I_r		0,5, 1, 1,5, 2	1
β_1		3.7	Einfluss vom datenabhängigen Term		0,1, 0,25, 0,5, 0,75, 0,9	0,5
u	Aktualisierung der Änderungs-wahrscheinlichkeit	4.5	minimale Breite der Segmente	**RRFK**	*hag*: 0, 0,2, 1, 2 m *nag*: 0, 30 m	*hag*: 1 m *nag*: 30 m
s			minimale Größe der Segmente		*hag*: 0, 2^2, 8^2, 250^2 m^2 *nag*: 0, 250^2, 360^2, 750^2 m^2	*hag*: 8^2 m^2 *nag*: 250^2 m^2
Schatten		4.5	Verwendung von Schattendetektion		ja/nein	*hag*: ja *nag*: nein
f_{a1}, f_{a2}		4.5	Aktualisierungsfunktion		f_{a1}, f_{a2}	f_{a1}
$N_{iteration}$	Iteration		Maximale Anzahl an Iterationen des Prozesses aus Training und Klassifikation		5 - 100 [1]	20

Tabelle 8: Zusammenfassung der Parameter. nag = niedrig aufgelöst, hag = hoch aufgelöst.

b) Wie wirkt sich die Integration der Kartenlabels als Beobachtung auf das Klassifikationsergebnis aus?

c) Führt das Ersetzen der Zuordnung eines Trainingsbeispieles zu einer Klasse durch die Zuordnung eines Trainingsbeispieles zu allen Klassen mit einer bestimmten Wahrscheinlichkeit zu einem besseren Klassifikationsergebnis?

d) Ist das Klassifikationsergebnis der vorgestellten Methode, basierend auf dem neu entwickelten RRF, besser als das Ergebnis der von Bootkrajang und Kabán [2012] eingeführten RLR?

e) In welchem Rahmen beeinflusst die Anzahl der Fehler in den Trainingslabels das Klassifikationsergebnis?

f) Führt der vorgestellte Klassifikationsalgorithmus sowohl bei hoch, als auch bei niedrig aufgelösten Datensätzen zu einem verbesserten Ergebnis?

Frage a beschäftigt sich mit dem Vergleich der Ergebnisse unter Verwendung der klassischen Klassifikatoren ohne die Berücksichtigung von Fehlern in den Labels mit denen von Klassifikatoren, die Fehler in den Trainingslabels im Training modellieren. Die Klassifikation ohne die Berücksichtigung von Fehlern in den Labels der Trainingsdaten erfolgt mit Hilfe des RF (Kapitel 3.6) und der LR (Kapitel 3.3), die innerhalb eines klassischen CRFs (Kapitel 3.7) verwendet werden. Die Labels resultieren aus der Inferenz des CRFs. Diese Methoden werden in den Experimenten mit **RF** bzw. **LR** bezeichnet und entsprechen den Standardklassifikationsverfahren der Literatur. Die Klassifikation unter Berücksichtigung der Fehler im Training basiert auf dem RRF bzw. der RLR. Der erste Klassifikator, in den Experimenten mit **RRF** bezeichnet, entspricht dem iterativen Algorithmus wie in Kapitel 4.1 beschrieben, mit einer Ausnahme. Anstatt des erweiterten CRF-Modells, welches die Kartenlabels als weitere Beobachtung in den Klassifikationsprozess integriert, wird auch hier das klassische CRF verwendet, um die Ergebnisse mit dem **RF** vergleichen zu können. Der Klassifikator auf Basis der RLR (**RLR**) entspricht der Initialisierung des in Kapitel 4.6 vorgestellten Algorithmus (s. Abb. 13). Auch hierbei wird die RLR für die Modellierung des datenbezogenen Assoziationspotenzials in einem klassischen CRF verwendet.

Frage b bezieht sich auf die Klassifikation mit dem erweitertem CRF unter Berücksichtigung der Kartenlabels (Kapitel 4.3). Hierbei werden die Klassifikatoren verwendet, die die fehlerbehafteten Trainingslabels sowohl im Training als auch als Beobachtungen in der Klassifikation berücksichtigen. Damit entspricht der Klassifikaor unter Verwendung des RRF dem in dieser Arbeit neu entwickelte Iterative Algorithmus aus Kapitel 4.1, der in den Experimenten **RRFK** genannt wird. Der Klassifikator unter Verwendung der RLR entspricht dem iterativen Algorithmus aus Kapitel 4.6 und wird in den Experimenten mit **RLRK** abgekürzt. Das K in dem Akronym steht dabei für **K**artenbeobachtung. Für die Beantwortung der Frage b werden die Klassifikationsergebnisse mit denen der Methoden ohne Beobachtung der Kartenlabels, **RRF** und **RLR**, verglichen.

Frage c beleuchtet den Umstand, dass im RRF jedem Trainingsbeispiel anstelle einer bestimmten Klasse alle Klassen mit einer bestimmten Wahrscheinlichkeit zugeordnet werden.

Um zu analysieren, ob die Verwendung dieser Wahrscheinlichkeiten der Klassenzugehörigkeit zu einer Verbesserung des Klassifikationsergebnisses führt, wird jedem Trainingsbeispiel in jeder Iteration des **RRFK** nur eine Klasse zugeordnet, was der Ersetzung des RRF durch den klassischen RF in dem iterativen Algorithmus entspricht. Demnach muss hier nach jeder Iteration ein Label gewählt werden. Eine Möglichkeit besteht darin, die Labels aus dem Klassifikationsergebnis der vorherigen Iteration zu verwenden. Lediglich bei der Initialisierung stammen die Labels aus der veralteten Karte. Diese Klassifikationsmethode wird in der Beschreibung der Experimente als **ERFK** (Akronym für **E**rgebnisbasierter **R**andom **F**orest mit **K**artenbeobachtung) bezeichnet. Bei der zweiten Möglichkeit werden zum Training jeder Iteration die Labels der Karte verwendet. Der iterative Prozess mit der Aktualisierung des temporalen Assoziationspotenzials bleibt ansonsten unverändert. Bezeichnet wird diese Methode in den Experimenten mit **KRFK** (Akronym für **K**artenbasierter **R**andom **F**orest mit **K**artenbeobachtung).

Frage d bezieht sich auf den Vergleich der Ergebnisse des **RRF** und den Ergebnissen der **RLR**. Neben dem Basisklassifikator, LR bzw. RF, unterscheidet sich hierbei auch die Modellierung der Fehler in den Trainingslabels. Während die **RLR** ausschließlich einen Zusammenhang zwischen dem Auftreten der Fehler und dem Klassenlabel berücksichtigt (Abb. 1, NAR Modell), modelliert der **RRF** auch einen Zusammenhang zwischen dem Auftreten eines Fehlers und den Daten (Abb. 1, NNAR Modell). Bei dem Vergleich der Klassifikationsergebnisse lässt sich jedoch nicht eindeutig unterscheiden, ob die Unterschiede aufgrund der Fehlermodelle oder der Klassifikatoren entstanden sind.

Eine Zusammenfassung der verwendeten Methoden findet sich in Tabelle 9.

<table>
<tr><th>Basiskl.</th><th>Abkürzung</th><th>robustes Training</th><th>Trainingslabels</th><th>Karte als Beob.</th><th>Iterativ</th></tr>
<tr><td rowspan="3">Log. Regr.</td><td>LR</td><td>nein</td><td rowspan="3">Labels der Karte</td><td rowspan="2">nein</td><td rowspan="2">nein</td></tr>
<tr><td>RLR</td><td rowspan="2">ja</td></tr>
<tr><td>RLRK</td><td>ja</td><td>ja</td></tr>
<tr><td rowspan="5">Random Forest</td><td>RF</td><td rowspan="3">nein</td><td rowspan="2">Labels der Karte</td><td>nein</td><td>nein</td></tr>
<tr><td>KRFK</td><td rowspan="3">ja</td><td rowspan="4">ja</td></tr>
<tr><td>ERFK</td><td>Klassifikationsergebnis (vorherige It.)</td></tr>
<tr><td>RRF</td><td rowspan="2">ja</td><td rowspan="2">Wahrsch. der Klassenzugehörigkeit</td><td>nein</td></tr>
<tr><td>RRFK</td><td>ja</td></tr>
</table>

Tabelle 9: Übersicht der verschiedenen Klassifikatoren, die in den Experimenten verglichen werden.

Während sich die Fragen a bis d mehr auf die Methodik beziehen, werden für die Beantwortung der Fragen e und f Anforderungen an die Datensätze, die in den Experimenten verwendet werden, untersucht.

In Frage e geht es um den Umfang an Fehlern in den Trainingslabels, der von den Verfahren kompensiert werden kann. Um zu untersuchen wie sich das Klassifikationsergebnis bei steigender Anzahl fehlerhafter Trainingslabels verändert, wurden bei drei Datensätzen zwei bzw. drei veraltete Karten simuliert. Die Anzahl an Pixeln mit fehlerhaftem Label unterscheidet sich

bei den simulierten Karten eines Datensatzes. Bei dem realen Datensatz *Las Vegas* wurden bei den monotemporalen Experimenten immer die Satellitendaten von 2016 verwendet. Die Referenzen der Daten aus 1991 und 2000 können demnach ebenfalls als veraltete Karten interpretiert werden. Mit der Karte aus 1986 stehen somit ebenfalls drei Karten mit unterschiedlich großen Änderungsgebieten zur Verfügung.

Für die Beantwortung der letzten Frage (f) werden Fernerkundungsdaten unterschiedlicher Auflösungen benötigt. In den Experimenten werden sowohl hoch aufgelöste Fernerkundungsdaten mit einer Bodenpixelgröße von $0,08m$ für *Vaihingen* und $0,2m$ für *Hameln*, als auch mittel bis niedrig aufgelöste Satellitendaten mit einer Bodenpixelgröße von $5m$ für *Herne* und *Husum* und $30m$ für *Las Vegas* mit dem vorgestellten Algorithmus klassifiziert. Somit werden die Experimente an Daten mit vier unterschiedlichen Auflösungen, die von einer Bodenpixelgröße von $0,08m$ bis zu einer Bodenpixelgröße von $30m$ reichen, getestet.

Die Evaluierung der Ergebnisbilder aus der Klassifikation erfolgt durch den Vergleich mit einer manuell erstellten Referenz und den daraus abgeleiteten Bewertungskriterien Gesamtgenauigkeit, Korrektheit, Vollständigkeit und mittlerem F1 Wert (Kapitel 5.5).

Für den Vergleich werden drei Konfigurationen unterschieden:

1. **Gesamtes Bild**

   Bei der ersten Konfiguration werden alle Pixel des Bildes für die Berechnung der Bewertungskriterien verwendet. Mit dieser Konfiguration soll gezeigt werden, wie gut die Methode insgesamt funktioniert, denn das resultierende Labelbild stellt die aktualisierte Karte dar. Damit sind die resultierenden Gesamtgenauigkeitsangaben ein Maß für die Genauigkeit der gesamten aktualisierten Karte.

2. **Nur Änderungsgebiete**

   Bei der zweiten Konfiguration werden nur jene Pixel für die Evaluierung verwendet, deren Kartenlabels sich von dem Label in der Referenz unterscheiden und die demnach von einer Änderung betroffen sind. Diese Konfiguration zeigt also, wie gut die neuen Objekte erkannt wurden.

3. **Änderungsdetetektion**

   Die dritte Konfiguration konzentriert sich auf die Änderungsdetetektion für das gesamte Bild. Sie ist demnach eng mit der ersten Konfiguration verbunden. Änderungen sind Pixel, deren Label aus der Klassifikation sich von dem Label aus der Karte unterscheidet. Das Label selbst wird daher nicht mitberücksichtigt, sodass das Klassifikationsergebnis auf zwei Klassen, *Änderung* und *keine Änderung*, reduziert wird. Diese Unterscheidung ist von Bedeutung, wenn nur das Auftreten einer Änderung, nicht aber die Art der Änderung benötigt wird. Dies ist beispielsweise bei einer halbautomatischen Kartenaktualisierung der Fall, wenn die Anpassung der veränderten Objekte in einem Nachbearbeitungsschritt manuell erfolgt.

5.6.4. Multitemporale Klassifikation

Für den multitemporalen Fall stehen lediglich zwei Datensätze zur Verfügung: *Hameln* und *Las Vegas*. An Ihnen sollen die folgenden Fragen untersucht werden:

a) Wird das Klassifikationsergebnis durch eine sequenzielle Klassifikation besser, verglichen zu einer direkten Klassifikation ohne Verwendung von veralteten Fernerkundungsdaten?

b) Wie ändert sich das Ergebnis durch die Integration von temporalem Kontext?

Für die Evaluierung werden die Ergebnisse von drei verschiedenen Methoden untersucht, die in Tabelle 10 zusammengefasst sind.

Methode	Abkürzung	Kapitel	Beschreibung
Direkte Kl.	**KD**	4.1	Monotemporale Klassifikation (**RRFK**). Die Trainingslabels stammen aus der veralteten Karte. Die Daten der Zwischenzeitpunkte werden nicht genutzt.
Sequenzielle Kl.	**KS**	4.7.2	Monotemporale Klassifikation (**RRFK**). Die Trainingslabels stammen von dem Klassifikationsergebnis des vorherigen Zeitpunktes. Nur bei dem ersten Zeitpunkt stammen sie aus der veralteten Karte.
Multitemporale Kl.	**KM**	4.7.3	Multitemporale Klassifikation mit dem Multitemporalen CRF aus Kap. 4.7.1. Die Parameter aus dem Training der sequenziellen Klassifikation **KS** werden übernommen.

Tabelle 10: Übersicht der verschiedenen Klassifikatoren, die in den Experimenten mit den multitemporalen Datensätzen verglichen werden.

Frage a bezieht sich auf den Vergleich der Ergebnisse einer direkten Klassifikation und der sequenziellen Klassifikation aller Zeitpunkte. Bei der direkten Klassifikation **KD** erfolgt die Klassifikation der Sensordaten eines Zeitpunktes unabhängig von den Daten der anderen Zeitpunkte. Für die Klassifikation wird die neu entwickelten Methode **RRFK** verwendet (Kapitel 4.1). Der **RRFK** wird auch bei der sequenzielle Klassifikation **KS** verwendet. Anders als bei der direkten Klassifikation stammen die Trainingslabels nicht aus der veralteten Karte, sondern von dem Klassifikationsergebnis des vorherigen Zeitpunktes. Nur das Training für den ersten Zeitpunkt erfolgt mit den Labels aus der veralteten Karte.

Frage b beleuchtet das Ergebnis der multitemporalen Klassifikation **KM** mit Hilfe des multitemporalen CRFs aus Kapitel 4.7.1. Für die Modellierung der Potenziale werden die Trainingsergebnisse der sequenziellen Klassifikation genutzt (Kapitel 4.7.2). Für die Evaluierung wird das Ergebnis mit denen der anderen Methoden **KD** und **KS** verglichen. Auch hier erfolgt der Vergleich anhand der Bewertungskriterien Gesamtgenauigkeit, Korrektheit, Vollständigkeit und mittlerer F1 Wert (Kapitel 5.5). Die drei Konfigurationen, die bereits in Kapitel 5.6.3 bei der monotemporalen Klassifikation dargelegt wurden, werden bei dem Vergleich ebenfalls unterschieden.

6. Experimente

6.1. Merkmalsselektion

Den Prozess der Merkmalsselektion wird zunächst an dem Datensatz *Hameln* gezeigt. Die Selektion bei den anderen Datensätzen folgt im Anschluss.

Hameln

Die Selektion erfolgt mit Hilfe der Rückwärtssuche basierend auf der Transinformation (MI) (s. Kapitel 3.2.5). Hierbei werden iterativ Merkmale aus dem vorhandenen Satz an Merkmalen entfernt. Die Entscheidung, welches Merklmal entfernt wird, erfolgt anhand der MI unter der Herausnahme jedes einzelnen Merkmals. Abb. 21 zeigt die Transinformation MI für die erste Iteration, in der noch alle Merkmale im Satz enthalten sind. Auf der X-Achse ist aufgezeigt, welches Merkmal nicht verwendet wurde. Die Punkte links in der Grafik, mit „Gemeinsam“ gekennzeichnet, zeigen die MI mit dem kompletten Satz an berechneten Merkmalen. Für die rot gekennzeichneten Punkte wurde die MI unter Verwendung der Labels aus der veralteten Karte bestimmt, für die blauen die Labels aus der Referenz.

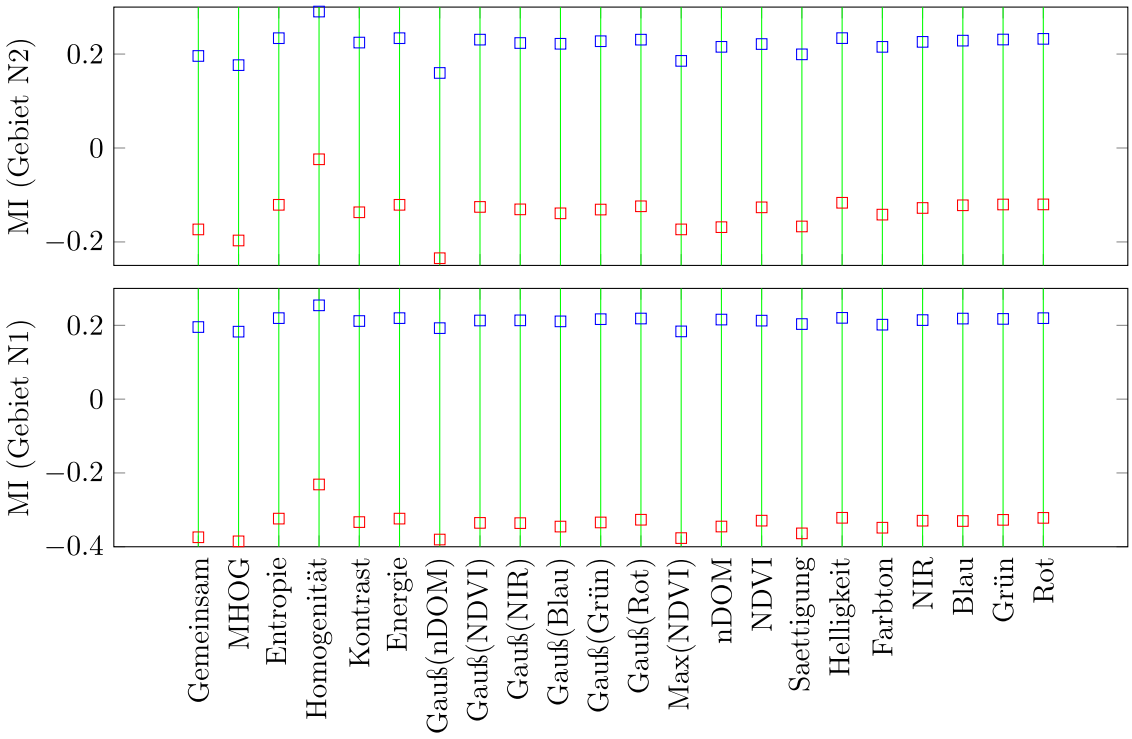


Abbildung 21: MI für die erste Iteration. Oben - MI für Gebiet N2 des Datensatzes *Hameln*. Unten - MI für Gebiet N1 des Datensatzes *Hameln*.
X-Achse: nicht verwendetes Merkmal. Rot: Verwendung der veralteten Karte. Blau: Verwendung der Referenz.

Es ist deutlich zu erkennen, dass der absolute Wert der MI durch die Verwendung fehlerhafter Labels sinkt. Die relativen Positionen zueinander sind jedoch sehr ähnlich, sodass der Einfluss

auf die Merkmalsauswahl gering ausfällt. In diesem Beispiel wird in beiden Fällen das Haralick-Merkmal Homogenität aus dem Satz an Merkmalen entfernt, da die MI ohne dieses Merkmal besonders groß ist.

Für jede Iteration kann somit nach der Elimination eines Merkmales ein MI-Wert bestimmt werden. Abb. 22 zeigt diesen Wert abhängig von der Anzahl der verbleibenden Merkmale.

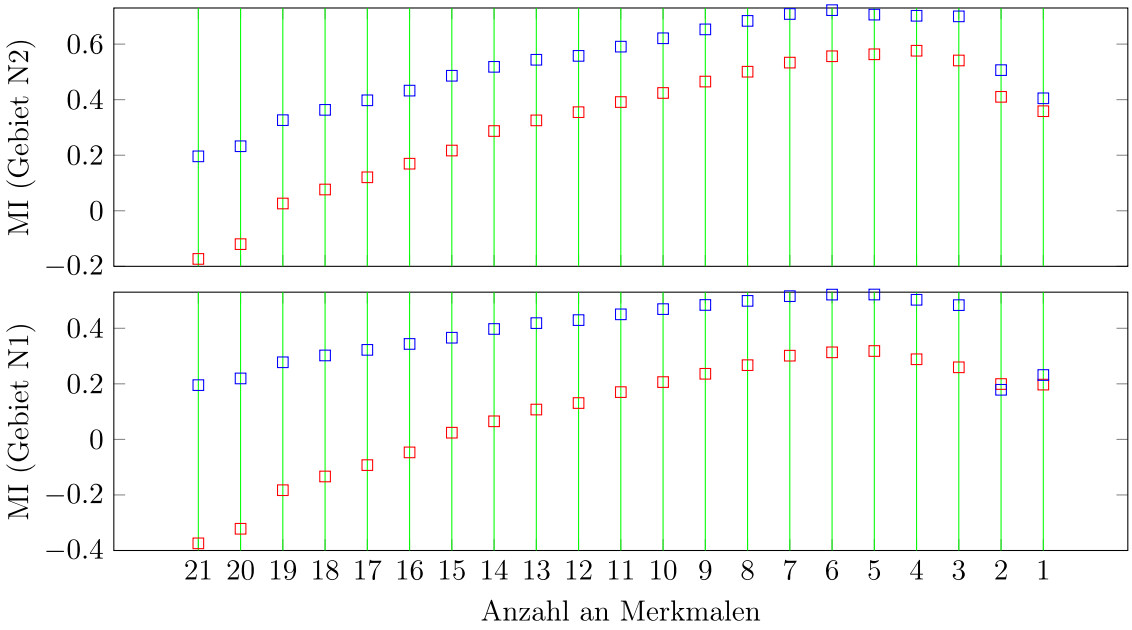


Abbildung 22: MI abhängig von der Anzahl der verbleibenden Merkmale. Oben - MI für Gebiet N2 des Datensatzes *Hameln*. Unten - MI für Gebiet N1 des Datensatzes *Hameln*.
Rot: Verwendung der veralteten Karte. Blau: Verwendung der Referenz.

Auch hier zeigt sich beim Vergleich der MI aus Referenz und veralteter Karte, dass die MI unter Verwendung der fehlerfreien Labels zwar fast immer größer als für Labels aus der veralteten Karte ist, dass der Verlauf der beiden Kurven jedoch sehr ähnlich ist. In beiden Fällen steigt die MI bis zu einem Maximum an und fällt dann wieder leicht ab. Bei Gebiet N1 ist das Maximum erreicht, wenn noch 5 Merkmale genutzt werden, sowohl unter Verwendung der Referenz als auch unter Verwendung der veralteten Karte. Bei dem Gebiet N2 sinkt die MI unter Verwendung der Referenz bereits bei weniger als 6 verbleibenden Merkmalen, während das Maximum bei Verwendung der veralteten Karte mit 4 verbleibenden Merkmalen erreicht wird. Die Merkmale, die bei dem Erreichen des Maximums noch in dem Satz an Merkmalen vorhanden sind, werden in den Merkmalsvektor $\mathbf{x}$ aufgenommen.

Tabelle 11 zeigt für jede Iteration, welches Merkmal herausgenommen wurde. Die Merkmale, die bei der maximalen MI verwendet wurden, sind grau markiert. Diese Auflistung kann auch als Ranking entsprechend der Relevanz des Merkmales für die Trennung der Klassen interpretiert werden, wobei die wichtigsten am spätesten eliminiert werden und somit in der Tabelle oben stehen.

Wie zu erwarten, sind die Listen beider Gebiete nicht identisch, aber es sind ähnliche Tendenzen zu sehen. Die Homogenität und Helligkeit fallen in den ersten zwei Iterationen heraus

Iteration	N1		N2	
	Karte	Referenz	Karte	Referenz
21	Gauß(nDOM)	Gauß(nDOM)	Gauß(nDOM)	Gauß(nDOM)
20	Gauß(Blau)	Gauß(Blau)	Gauß(Blau)	NDVI
19	Max(NDVI)	Gauß(NIR)	Gauß(NIR)	Gauß(Blau)
18	Sättigung	Max(NDVI)	Max(NDVI)	Sättigung
17	MHOG	MHOG	Grün	MHOG
16	NIR	Farbton	Sättigung	Max(NDVI)
15	Farbton	Sättigung	MHOG	Blau
14	Gauß(NIR)	Grün	Farbton	Farbton
13	Kontrast	Kontrast	nDOM	Gauß(NIR)
12	Gauß(NDVI)	Gauß(NDVI)	Gauß(NDVI)	Kontrast
11	NDVI	NDVI	Kontrast	Gauß(Grün)
10	nDOM	nDOM	NDVI	Gauß(NDVI)
9	Gauß(Grün)	Gauß(Grün)	Gauß(Grün)	nDOM
8	Grün	NIR	NIR	NIR
7	Gauß(Rot)	Gauß(Rot)	Blau	Grün
6	Blau	Energie	Energie	Rot
5	Entropie	Entropie	Entropie	Entropie
4	Energie	Blau	Gauß(Rot)	Energie
3	Rot	Rot	Rot	Gauß(Rot)
2	Helligkeit	Helligkeit	Helligkeit	Helligkeit
1	Homogenität	Homogenität	Homogenität	Homogenität

Tabelle 11: Liste der Merkmale für die Gebiete N1 und N2 aus dem Datensatz *Hameln*. Die Position des Merkmales gibt an, bei welcher Iteration (links) es aus dem Satz an Merkmalen herausgenommen wurde.
Grau: Von dem Algorithmus selektierte Merkmale.

und auch die Haralick-Merkmale Entropie und Energie folgen innerhalb der nächsten vier Iterationen. Von den Texturmerkmalen scheint daher der Kontrast am ehesten für die Unterscheidung der Klassen geeignet zu sein, obwohl auch dieser bereits in den Iteration 11-13 herausgenommen wird. Das Strukturmerkmal MHOG hingegen ist für Gebiet N1 unter den ersten fünf Merkmalen und auch bei Gebiet N2 bleibt es unter Verwendung der Kartenlabels bis zur 15. Iteration in dem Satz enthalten. Die Höhe des nDOM, das Maximum des NDVI und der blaue Kanal scheinen für die Unterscheidung der Klassen essenziell zu sein, da bei deren vollständiger Entfernung eine starke Verkleinerung der MI zu beobachten ist.

Ein Vergleich der selektierten Merkmale unter Verwendung der veralteten Karte mit jenen unter Verwendung der Referenz zeigt die gleichen Tendenzen. Dennoch scheinen die Fehler in den Labels bei der Verwendung der veralteten Karte Einfluss auf die Selektion auszuüben. Das Strukturmerkmal MHOG wird im Gebiet N2 nur ausgewählt, wenn die Labels aus der Referenz stammen, nicht aber, wenn die Labels der Karte verwendet werden.

Durch die Glättung bei der Bestimmung einiger Merkmale nimmt das Rauschen der Merkmale ab, sodass die Größe der Ballungen von Primitiven einer Klasse im Merkmalsraum abnimmt. Objektgrenzen verschwimmen jedoch in solchen Merkmalen. Da nur wenige Pixel in der Nähe von Objekträndern liegen, werden geglättete Merkmale von dem Selektionsalgorithmus besser eingestuft als ungeglättete. Damit der Klassifikationsalgorithmus die Möglichkeit hat, auch

diese Ränder korrekt zu detektieren, werden die besten drei Merkmale, falls sie geglättet sind, zusätzlich ungeglättet in den Merkmalsvektor $\mathbf{x}$ aufgenommen.

Da die beiden Gebiete N1 und N2 zwei unterschiedliche Ausschnitte der Stadt Hameln zeigen, wird deren Merkmalsauswahl für eine einheitliche Auswahl der ganzen Stadt kombiniert. Hierbei werden nur die Selektionsergebnisse unter Verwendung der veralteten Karte genutzt. Der gemeinsame Merkmalsvektor ist in der ersten Spalte von Tabelle 12 zu sehen.

Für die Merkmalsselektion unter Verwendung des Datensatzes *Hameln* kann festgehalten werden, dass die Verwendung von fehlerbehafteten Labels aus einer veralteten Karte das Selektionsergebnis zwar beeinflusst, die selektierten Merkmale unter Verwendung der Kartenlabels jedoch gleiche Tendenzen aufweisen wie jene unter Verwendung der Referenz. Besonders wichtige Merkmale für eine spätere Klassifikation sind u.A. der NDVI, das nDOM und der blaue Kanal.

Las Vegas

Bei dem *Las Vegas* Datensatz wird kein deutliches Maximum der MI abhängig von der Anzahl der verbliebenden Merkmale festgestellt. Stattdessen steigt die MI erst an und fängt bei 4-5 Merkmalen an etwas zu springen, wie in Abb. 23 zu sehen ist. Dabei unterscheiden sich die Werte der MI jedoch kaum. Es werden demnach die ersten 5 Merkmale gewählt, anstatt den Maximalwert als Grenze zu nutzen. Auch hier werden die besten drei Merkmale, falls sie geglättet sind, zusätzlich in ungeglätteter Form selektiert. Die ausgewählten Merkmale sind ebenfalls in Tabelle 12 gezeigt.

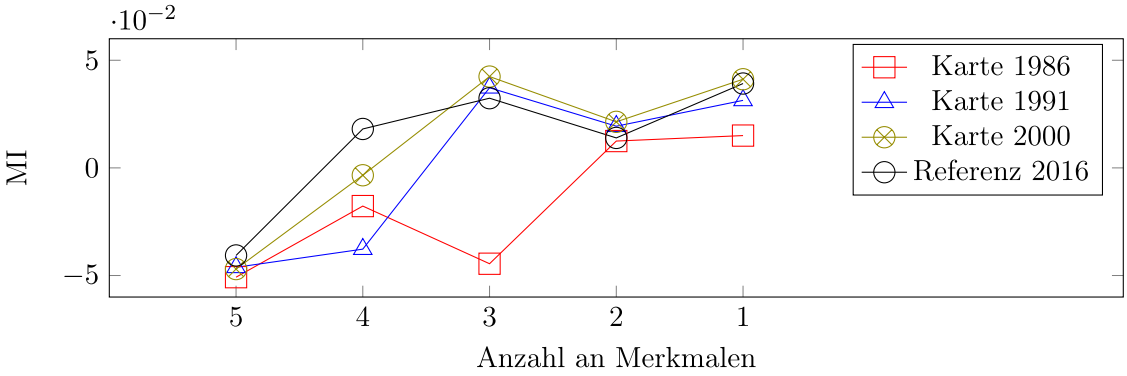


Abbildung 23: MI von *Las Vegas* abhängig von der Anzahl an verbleibenden Merkmalen.

Die Relevanz der Merkmale ist bei dem *Las Vegas* Datensatz abhängig von der verwendeten Karte. Abb. 24 zeigt für jedes Merkmal, in der wievielten Iteration des Selektionsprozesseses es eliminiert wurde. Je höher der Iterationszähler ist, desto wichtiger ist auch das Merkmal. Es ist zu erkennen, dass die Relevanz für den Großteil der Merkmale unabhängig von der Karte ist. Besonders die Ergebnisse der Selektion unter Verwendung der Karten aus 1991 und 2000 ähneln sich stark. Verglichen zur Referenz von 2016 wird das Merkmal des Farbtons unter Verwendung der veralteten Karten deutlich wichtiger eingestuft, wobei die RGB Kanäle und das Maximum des NDVI bei der Referenz als wichtiger eingestuft werden. Unter Verwendung der Karte aus 1986 zeigen besonders der gaußgefilterte rote Kanal und die Helligkeit einen

Hameln	Las Vegas			
2010	1986	1991	2000	2016
Blau	Rot	Blau	Blau	Blau
NIR	IR	IR	IR	IR
Sättigung	Helligkeit	Farbton		
nDOM				
NDVI	NDVI	NDVI	NDVI	NDVI
Max(NDVI)				Max(NDVI)
Gauß(Blau)	Gauß(Rot)	Gauß(Blau)	Gauß(Blau)	Gauß(Blau)
Gauß(NIR)		Gauß(IR)	Gauß(IR)	Gauß(IR)
Gauß(nDOM)				
	Gauß(NDVI)	Gauß(NDVI)	Gauß(NDVI)	Gauß(NDVI)
MHOG	MHOG		MHOG	

Tabelle 12: Selektierte Merkmale der Datensätze *Hameln* und *Las Vegas*. Die Jahreszahl zeigt das Jahr aus dem die Karte stammt. Die Sensordaten sind in beiden Fällen aus dem Jahr 2016.

hohen Wert auf, während die Relevanz der geglätteten infraroten und blauen Bänder deutlich niedriger ist als unter Verwendung der anderen Karten. Eine mögliche Begründung könnten

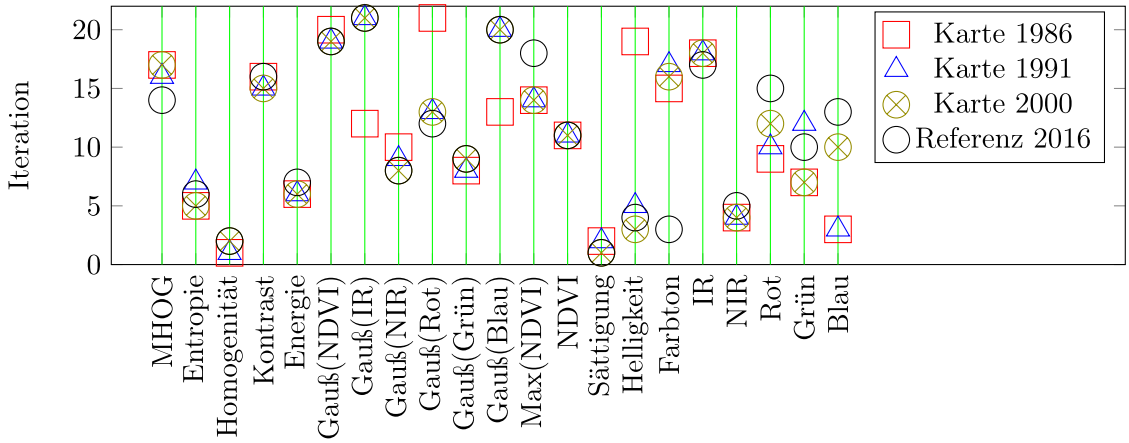


Abbildung 24: Iteration, in dem Merkmal aus der Liste entfernt wurde. Ein hoher Wert deutet auf ein wichtiges Merkmal hin. Datensatz: *Las Vegas*.

die unterschiedliche Merkmalsausprägung von alten und neuen bebauten Gebieten (vgl. RGB Bild in Abb. 18) und der starke Zuwachs der Klasse *Bebauung* (vgl. Tabelle 5) sein.

Bei der Merkmalsselektion mit dem *Las Vegas* Datensatz kann festgehalten werden, dass das Selektionsergebnis kaum von der Anzahl der Fehler in den Labels abhängt. Obwohl die selektierten Merkmale unter Verwendung von Karten aus unterschiedlichen Zeitpunkten nicht identisch sind, sind doch gleiche Tendenzen erkennbar. Dies ist ebenfalls bei dem Vergleich mit dem Ergebnis aus der Selektion unter Verwendung der Labels aus der Referenz zu beobachten. Unabhängig von dem Jahr aus dem die Karte stammt, enthält die Merkmalsauswahl den NDVI, den infraroten Kanal und einen Kanal aus dem sichbaren Bereich in geglätteter Form.

Vaihingen

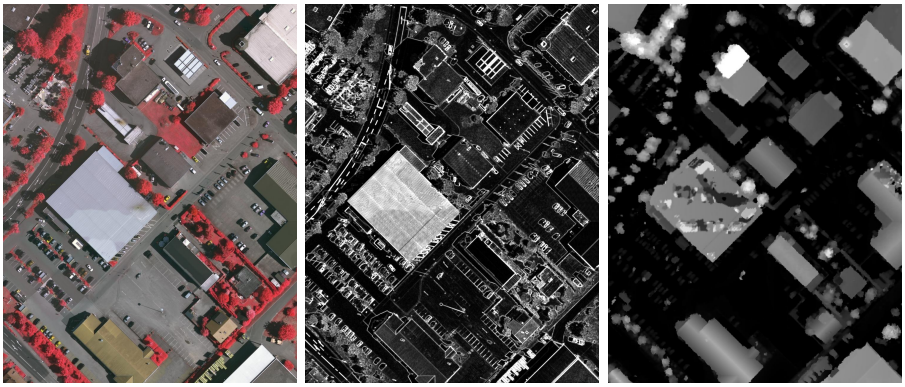
Beim *Vaihingen* Datensatz wird für jede simulierte Version der veralteten Karte (Karte 1, Karte 2 und Karte 3) eine Liste an selektierten Merkmalen bestimmt (Tabelle 13). Hierfür wurden zunächst für jedes Gebiet individuell die wichtigsten Merkmale selektiert und anschließend zusammengefasst. Dementsprechend werden für die Klassifikation alle Merkmale verwendet, die für mindestens ein Gebiet selektiert wurden. Wie beim *Hameln* Datensatz wurden dabei jene Merkmale selektiert, die bei maximaler MI im Satz waren. Das nDOM, der NDVI und die originalen Bänder werden meist in geglätteter Form von dem Algorithmus selektiert. Die ungefilterte Variante der besten drei Merkmale wird daher, wie auch bei den vorherigen Datensätzen, zusätzlich ausgewählt.

Die Merkmalsauswahl aus der veralteten Karte mit den geringsten Änderungen (Karte 1) und den meisten Änderungen (Karte 3) sind identisch und auch die Merkmalsauswahl der Karte 2 unterscheidet sich lediglich durch das Fehlen der Merkmale Rot und Gauß(Rot). Die Merkmalsauswahl unter Verwendung der Referenz als Karte beinhaltet den roten Kanal ebenfalls nicht. Bei allen Kombinationen aus Gebiet und Karte liegt die maximale MI bei 3 - 6 verbliebenen Merkmalen. Unter den drei wichtigsten Merkmalen befinden sich bei allen Kombinationen der NDVI und das nDOM. Eine Außnahme bildet Gebiet 13 unter Verwendung der Karte 1, bei dem das NDVI Merkmal erst an 5. Stelle steht und, da unter Verwendung von 4 Merkmalen die maximale MI erreicht wurde, nicht ausgewählt wurde. Die drei wichtigsten Merkmale bilden stattdessen der grüne Kanal, der nah infrarote Kanal und das nDOM. Eine weitere Ausnahme bildet Gebiet 17 unter Verwendung der Referenz, bei dem der NDVI erst an 4. Stelle steht. Der grüne Kanal des Orthophotos befindet sich bei einem Großteil der Kombinationen neben dem NDVI und nDOM unter den drei wichtigsten Merkmalen. In den anderen Fällen wird an dessen Stelle einer der beiden anderen Kanäle ausgewählt. Das MHOG Merkmal ist bei Gebiet 13 und 37 unabhängig von der genutzten Karte selektiert worden. Diese Gebiete zeigen eine ähnliche Bebauungsstruktur: In beiden Fällen handelt es sich um ein Wohngebiet mit Einfamilienhäusern und Vegetation in Form von Bäumen und Gras. Wird die Referenz als Karte verwendet, befindet sich das MHOG-Merkmal auch bei anderen Gebieten (23, 30 und 32) unter den selektierten Merkmalen. Auch diese haben, bis auf Gebiet 32, eine ähnliche Bebauungsstruktur wie die Gebiete 13 und 37. Gebiet 32, ein Gewerbegebiet, bildet hierbei eine Ausnahme. In Abb. 25b ist zu erkennen, dass das MHOG Merkmal in diesem Gebiet bei einem relativ großen Gebäude in der Mitte des Bildes besonders prägnant ist und somit ein gutes Merkmal für die Klassifikation dieses Gebäudes darstellt. Die Höhe in dem nDOM (Abb. 25c), welches im Allgemeinen ein gutes Merkmal für die Gebäudedetektion darstellt, ist bei vielen Pixeln dieses Gebäudes jedoch fehlerhaft, vermutlich verursacht durch Fehler bei der Bildzuordnung.

Die Experimente zur Merkmalsselektion bezogen auf den *Vaihingen* Datensatz zeigen, dass sich die Selektionsergebnisse unter Verwendung der Karten mit einer unterschiedlichen Anzahl an Fehlern in den Labels sehr ähneln.

Vaihingen			
Referenz	Karte 1	Karte 2	Karte 3
NIR	NIR	NIR	NIR
	Rot		Rot
Grün	Grün	Grün	Grün
Sättigung	Sättigung	Sättigung	Sättigung
nDOM	nDOM	nDOM	nDOM
NDVI	NDVI	NDVI	NDVI
Max(NDVI)	Max(NDVI)	Max(NDVI)	Max(NDVI)
Gauß(NIR)	Gauß(NIR)	Gauß(NIR)	Gauß(NIR)
	Gauß(Rot)		Gauß(Rot)
Gauß(Grün)	Gauß(Grün)	Gauß(Grün)	Gauß(Grün)
Gauß(nDOM)	Gauß(nDOM)	Gauß(nDOM)	Gauß(nDOM)
Gauß(NDVI)	Gauß(NDVI)	Gauß(NDVI)	Gauß(NDVI)
MHOG	MHOG	MHOG	MHOG

Tabelle 13: Selektierte Merkmale des Datensatzes *Vaihingen*.



(a) DOP (b) MHOG (c) nDOM

Abbildung 25: DOP, MHOG Merkmal und nDOM von Gebiet 32 aus *Vaihingen*.

Herne und Husum

Die Merkmalsauswahl in den Datensätzen *Herne* und *Husum* wird wie bei dem Datensatz *Vaihingen* für jede Variante der veralteten Karte individuell durchgeführt. Die resultierenden Merkmale sind in Tab. 14 zu sehen. Für den Datensatz *Husum* werden unter Verwendung beider veralteter Karten die gleichen Merkmale selektiert. Auch bei dem Datensatz *Herne* unterscheidet sich die Auswahl lediglich durch das Fehlen des Merkmales Max(NDVI) unter Verwendung der Karte mit den meisten Änderungen zur Referenz (Karte 2).

Bei beiden Datensätzen basiert die Auswahl auf dem MHOG-Merkmal, dem grünen und nahen infraroten Kanal, sowie dem NDVI. Unterschiedlich ist, dass bei *Husum* der NDVI, der zuvor mit einem Maximumfilter bearbeitet wurde, selektiert wurde. Bei *Herne* wurde der NDVI auch in der originalen Auflösung sowie mit einem Gaußfilter geglättet ausgewählt. Bei *Husum* liegt dafür der grüne Kanal zusätzlich zu der ungefilterten Form auch mit einem Gaußfilter gefiltert vor, während bei *Herne* nur der ursprüngliche Kanal selektiert wurde.

Die Merkmalsselektion für den Datensatz *Herne* zeigt ebenfalls, dass sich die Auswahl der

Herne			Husum		
Referenz	Karte 1	Karte 2	Referenz	Karte 1	Karte 2
Grün	Grün	Grün	Grün	Grün	Grün
NIR	NIR	NIR	NIR	NIR	NIR
NDVI	NDVI	NDVI			
Max(NDVI)	Max(NDVI)		Max(NDVI)	Max(NDVI)	Max(NDVI)
Gauß(NDVI)	Gauß(NDVI)	Gauß(NDVI)			
			Gauß(Grün)	Gauß(Grün)	Gauß(Grün)
Gauß(NIR)	Gauß(NIR)	Gauß(NIR)	Gauß(NIR)	Gauß(NIR)	Gauß(NIR)
MHOG	MHOG	MHOG	MHOG	MHOG	MHOG

Tabelle 14: Selektierte Merkmale der Datensätze *Herne* und *Husum*.

Merkmale unter der Verwendung von Karten mit unterschiedlich vielen Fehlern in den Labels sehr ähneln. Bei dem Datensatz *Husum* wurden sogar unter Verwendung der Referenz und der beiden Karten die gleichen Merkmale gewählt.

Diskussion

In Kapitel 5.6.2 wurde die Frage gestellt, ob die Verwendung der Labels aus einer veralteten Karte einen Einfluss auf das Selektionsergebnis hat. Da sich die Auswahl der Merkmale unter Verwendung der Referenz anstatt der veralteten Karte unterscheidet, ist ein gewisser Einfluss vorhanden. Dennoch sind viele Merkmale sowohl unter Verwendung der Karte als auch der Referenz selektiert worden. Bei dem Datensatz *Husum* ist die Auswahl sogar komplett identisch. Außerdem lassen sich Ähnlichkeiten zwischen den Merkmalen erkennen, die auf Basis einer der beiden Varianten selektiert wurden. So wird beispielsweise unter Verwendung der Karte aus 1986 bei dem Datensatz *Las Vegas* der rote Kanal, unter Verwendung der Referenz jedoch der blaue Kanal selektiert.

Obwohl ein Einfluss der Fehler in den Trainingslabels auf die Merkmalsselektion beobachtet wurde, werden in den Experimenten die Merkmale, die unter Verwendung der veralteten Karte selektiert wurden, verwendet. Da bei den Selektionsergebnissen unter Verwendung der Karten mit unterschiedlich vielen Fehlern in den Labels die gleichen Tendenzen zu beobachten waren, wird vermutet, dass der Einfluss auf das Klassifikationsergebnis in den weiteren Experimenten eher gering ist.

6.2. Einfluss der Parameter

Parameter der robusten logistischen Regression

Abb. 26 zeigt den MF1 abhängig von jeweils einem Parameter der RLR. Der Parameter σ wurde eingeführt, um eine sehr große Steigung der Sigmoid Funktion und somit eine Überanpassung der LR an die Trainingsdaten zu verhindern. Bei den Datensätzen *Hameln* und *Las Vegas* unter Verwendung der Karte von 2000 ändert sich der MF1 um maximal 3,2%. Hierbei liegt der kleinste Wert bei $\sigma = 0,5$, gefolgt von dem MF1 bei $\sigma = 1$. Ist σ über 5, ändert sich

der F1 Wert weniger als 0,5%. Unter Verwendung der stark veralteten *Las Vegas* Karte von 1986 wird eine Schwankung des MF1 von bis zu 16,6% festgestellt. Hierbei liegt das Minimum des MF1 (73,3%) bei $\sigma = 10$ und das Maximum (89,8%) bei $\sigma = 2,5$. Besonders die Detektion der Klasse *Bebauung* verursacht diesen starken Unterschied. Während die Vollständigkeit bei $\sigma = 10$ den sehr geringen Wert von 22,9% aufweist, liegt sie unter Verwendung von $\sigma = 2,5$ bei 75,2%. Die Gewichtskoeffizienten $\mathbf{w}$ sind bei $\sigma = 10$ deutlich größer als bei $\sigma = 2,5$, was auf eine größere Steigung der Sigmoidfunktion und eine stärkere Überanpassung hindeutet.

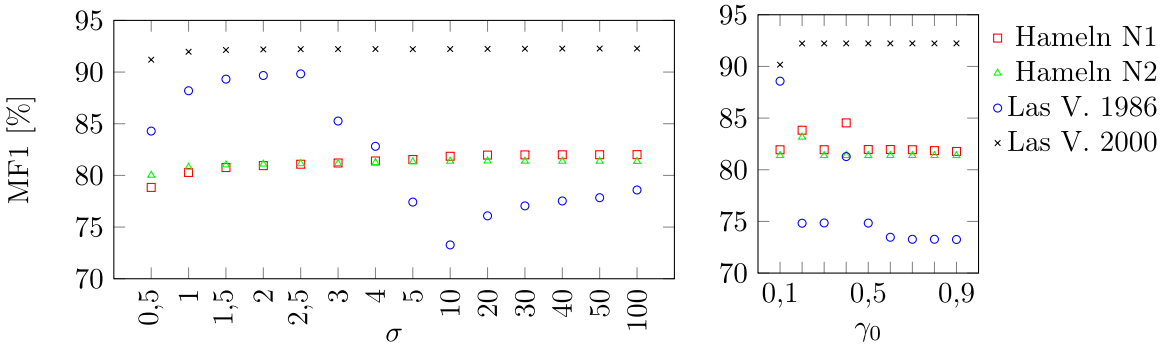


Abbildung 26: MF1 [%] abhängig von je einem Parameter (σ, γ_0) der robusten logistischen Regression.

Die Näherungswerte der Übergangswahrscheinlichkeiten $\overleftarrow{\gamma}_0$ und der Parameter $\mathbf{w}$ werden aufgrund der Nichtlinearität der Funktion $E(\mathbf{w}^0)$ (Gleichung 18) für die iterative Bestimmung des Minimums von $E(\mathbf{w})$ benötigt. Da die Grenzen der Klassen im Merkmalsraum unbekannt sind, werden die Parameter $\mathbf{w}$ mit 0 initialisiert. Danach wird die LR durchgeführt. Die resultierenden Parameter $\mathbf{w}$ dienen als Initialwerte für die RLR. Für die Initialisierung der Übergangswahrscheinlichkeiten wird die Wahrscheinlichkeit, dass keine Änderungen auftreten, auf γ_0 festgesetzt:

$$\overleftarrow{\gamma}_{ij} = \begin{cases} \gamma_0 & \text{wenn } i = j \\ \frac{1-\gamma_0}{K-1} & \text{wenn } i \neq j \end{cases} \tag{69}$$

Der Einfluss von γ_0 auf das Klassifikationsergebnis ist in Abb. 26 rechts zu sehen. Abb. 27 vergleicht zusätzlich die im Training geschätzten Übergangswahrscheinlichkeiten (Abb. 27, Kreise) mit den Übergangswahrscheinlichkeiten, die mit Hilfe der Referenz berechnet wurden (Abb. 27, Linie). Der Übersicht halber werden nur die Übergangswahrscheinlichkeiten, die stark von 0 abweichen und nicht auf der Hauptdiagonale liegen, gezeigt.

In Abb. 27 ist zu sehen, dass die geschätzten Werte (Kreise) meist unter den realen (Linie) liegen und somit die Wahrscheinlichkeit für das Auftreten von Änderungen unterschätzt wird. Besonders bei Werten von über 0,5 für γ_0 ist dies zu beobachten. Außerdem ist der geschätzte Wert relativ unabhängig von dem Startwert, wenn γ_0 größer als 0,5 ist. Dies könnte daran liegen, dass nach der Initialisierung zunächst die Übergangswahrscheinlichkeiten aktualisiert werden. Die Übergangswahrscheinlichkeiten γ_0 werden daher in weiteren Iterationen nicht direkt für die Aktualisierung der Parameter $\mathbf{w}$ verwendet.

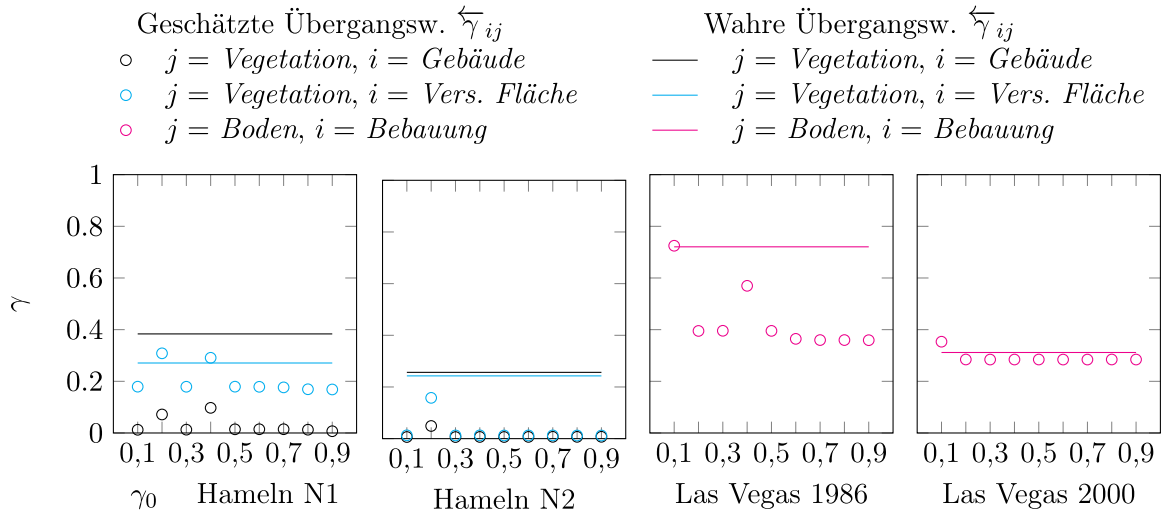


Abbildung 27: Übergangswahrscheinlichkeit $\overleftarrow{\gamma}_{ij} = P(\widetilde{C} = C^j | C = C^i)$ abhängig von dem Startwert γ_0.

Während die RLR die Klassen im Merkmalsraum nur mit linearen Entscheidungsgrenzen trennt, werden durch eine Merkmalsraumtranformation mit der Funktion Φ auch nichtlineare Grenzen ermöglicht. Getestet wurde hier eine quadratische Erweiterung. Diese führt bei dem *Hameln* Datensatz zu einer leichten Verbesserung des MF1 von 79,9 % auf 81,9 % (N1) bzw. von 80,7 % auf 81,4 % (N2). Während das Ergebnis des *Las Vegas* Datensatzes unter Verwendung der Karte von 2000 mit einem MF1 von 92,1% für die lineare Variante und 92,2% für die quadratische Erweiterung ziemlich gleich bleibt, sinkt der MF1 von 90,0 % (lineare Variante) auf 73,3 % (quadratische Erweiterung) unter Verwendung der Karte aus 1986. Eine mögliche Begründung ist, dass die Bebauung, die zwischen 1986 und 2016 entstanden ist, andere Merkmale aufweist als die, die vor 1986 errichtet wurde. Demnach bilden Pixel der Klasse *Bebauung* im Jahr 2016, die in der Karte als *Boden* gekennzeichnet sind, eine eigene Ballung im Merkmalsraum. Durch die quadratische Erweiterung, durch die auch komplexere Entscheidungsgrenzen ermöglicht werden, kann diese Ballung ggf. von der Ballung der korrekt gelabelten Trainingsbeispiele der Klasse *Bebauung* getrennt werden. Da die Neubauten fälschlicherweise das Label *Boden* im Training aufweisen, werden diese auch entsprechend falsch klassifiziert.

Zusammenfassend ist ein Einfluss durch die Wahl der Parameter σ und γ_0 auf das Klassifikationsergebnis zu erkennen. Besonders stark ist hier der Einfluss bei dem *Las Vegas* Datensatz unter Verwendung der Karte von 1986. Die Ergebnisse mit dem Datensatz *Hameln* und mit dem Datensatz *Las Vegas* unter Verwendung der Karte aus 2000 sind bei Parameter σ jedoch um den gewählten Wert von $\sigma = 10$ relativ konstant. Der MF1 Wert in Abhängigkeit von γ_0 ist für Werte über 0,5% bei allen Experimenten relativ stabil. Der gewählte Wert von $\gamma_0 = 0,8$ liegt in diesem Bereich.

Parameter des Random Forests

Die in diesem Abschnitt vorgestellten Experimente mit dem RF Klassifikator wurden fünf mal wiederholt, um die Stabilität des Ergebnisses ebenfalls zu analysieren. Abb. 28 zeigt den Mittelwert des MF1 abhängig von der Wahl der Parameter des RF. Die Fehlerbalken geben die Standardabweichung bezogen auf die fünf Durchläufe an.

Die gewählte Anzahl an Bäumen N_B des RF scheint einen gewissen Einfluss auf das Klassifikationsergebnis auszuüben. Der MF1 ändert sich jedoch weniger als 2% und es ist kein eindeutiges Maximum oder ein Bereich, bei dem der MF1 Wert sehr groß ist, zu erkennen. Bei einer sehr kleinen Anzahl an Bäumen ist das Ergebnis des RF nicht stabil, was in einer größeren Standardabweichung resultiert.

Auch die Anzahl an zufällig gewählten Trennflächen N_G in jedem Knoten beeinflusst das Klassifikationsergebnis. Die Änderung des MF1 Wertes liegt bei maximal 1,3%. Ein Wert oder Bereich, bei dem der MF1 bei allen Experimenten sehr groß ist, ist nicht zu erkennen.

Bei der Änderung der maximalen Tiefe T_{max} der Bäume ist eine deutliche Änderung des Klassifikationsergebnissis unter Verwendung des *Las Vegas* Datensatzes zu beobachten. Das Maximum des MF1 Wertes liegt mit 81,3% (Karte aus 1986) bzw. 86,7% (Karte aus 2000) bei einer maximalen Tiefe von 7. Der minimale MF1 ist hingegen 75,3% (Karte aus 1986) bei einer Tiefe von 2 bzw. 80,9% (Karte von 2000) bei einer Tiefe von 3. Die Ergebnisse bei so einer geringen Tiefe sind außerdem instabil, was eine große Standardabweichung besonders bei Verwendung der Karte aus 2000 zeigt. Die Ergebnisse des *Hameln* Datensatzes sind mit einer maximalen Differenz von 1,9% des M1 Wertes relativ ähnlich. Mit der gewählten Tiefe von 7 können demnach bei allen vier getesteten Szenarien akzeptable Ergebnisse erzielt werden.

Die Anzahl an Merkmalen N_F, die bei jedem Knoten gewählt werden, ist standardmäßig auf $\sqrt{F} = 3$ festgelegt. Die Experimente zeigten, dass eine Vergrößerung das Ergebnis nicht verbessert. Wird nur ein Merkmal ausgewählt, verbessert sich der MF1 bei den Experimenten mit dem *Hameln* Datensatz. Unter Verwendung der *Las Vegas* Daten verschlechtert sich jedoch der MF1, wenn ein einzelnes Merkmal genutzt wird. Dies liegt besonders an der schlechten Unterscheidung der Klassen *Boden* und *Bebauung*. Die Korrektheit der Klasse *Bebauung* liegt im Mittel bei nur 47,1%, da viele Pixel der Klasse *Boden* als *Bebauung* klassifiziert wurden.

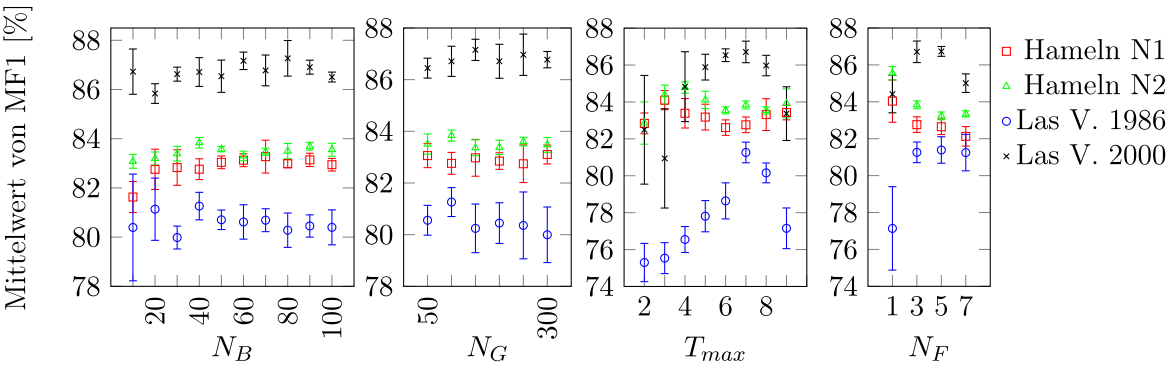


Abbildung 28: Mittelwert von MF1 [%] abhängig von je einem Parameter (N_B, N_G, T_{max}, N_F) des Random Forest.

Zusammenfassend kann festgestellt werden, dass die Parameter des RF einen gewissen Einfluss auf das Klassifikationsergebnis ausüben. Bei der Anzahl an Bäumen N_B und der Anzahl der zufällig gewählten Trennflächen N_G ändert sich der F1 Wert weniger als 2% und es ist keine eindeutige Tendenz zu erkennen. Bei der maximalen Tiefe T_{max} sind zwar stärkere Änderungen des MF1 Wertes zu beobachten, aber mit dem gewählten Wert von $T_{max} = 7$ sind die Ergebnisse bei allen Experimenten relativ gut. Die Anzahl an Merkmalen N_F wurde auf den Standardwert von $N_F = \sqrt{F} = 3$ festgelegt. Mit diesem Wert ist das Ergebnis bei allen getesteten Szenarien akzeptabel.

Parameter des kontrastsensitiven Potts-Modells

Die Sensitivität des Klassifikationsergebnisses auf die Wahl der Parameter β_0 und β_1 im kontrastsensitiven Potts-Modell zeigt Abb. 29 anhand der Werte für MF1. Werden die Ergebnisse des *Hameln* Datensatzes betrachtet, so scheinen beide Parameter keinen großen Einfluss auf das Ergebnis auszuüben, da sich der MF1 nur um maximal 0,8 % ändert. Die folgende Analyse fokussiert sich daher auf den *Las Vegas* Datensatz.

Der Parameter β_0 modelliert den Einfluss des Interaktionspotentials. Bei den Ergebnissen des *Las Vegas* Datensatzes unter Verwendung der Karte aus 1986 unterscheiden sich die Mittelwerte des MF1 der Parameter $\beta_0 = 1$, $\beta_0 = 1,5$ und $\beta_0 = 2$ weniger als 0,4%. Der Mittelwert des MF1 Wert für $\beta_0 = 0,5$ ist deutlich kleiner. Ein kleiner Wert für β_0, wie $\beta_0 = 0,5$, deutet dabei auf einen kleinen Einfluss des Interaktionspotentials hin. Die Berücksichtigung von räumlichem Kontext scheint demnach zu einer Verbesserung des Klassifikationsergebnisses zu führen.

Wird die Karte aus 1986 verwendet, ist der Mittelwerte des MF1 bei $\beta_0 = 0,5$ ebenfalls mit 85,6% am geringsten. Bei $\beta_0 = 1$ ist der Mittelwerte des MF1 mit 86,7% etwa 1% größer und bei $\beta_0 = 1,5$ ist der Mittelwerte des MF1 sogar bei 87,6%. Eine weitere Erhöhung des Wertes auf $\beta_0 = 2$ führt jedoch zu keinem größeren Mittelwerte des MF1 (87,5%).

Der gewählte Wert von 1 zeigt zusätzlich eine relativ kleine Standardabweichung von 0,6% für die veralteten Karten aus 1986 und 2000.

Der Parameter β_1 modelliert den Einfluss der Merkmale. Der Vergleich der MF1 Werte abhängig von dem Parameter β_1 zeigt keinen Wert, der bei allen Experimenten zu einem besonders guten oder schlechten Ergebnis führen. Die Mittelwerte unterscheiden sich um maximal 1,4% bei einer Änderung des β_1 Wertes. Der MF1 des gewählten Wertes von $\beta_1 = 0,5$ ist dabei relativ hoch und weicht bei dem *Las Vegas* Datensatz weniger als 0,3% von dem maximalen MF1 Wert ab, der mit $\beta_1 = 0,75$ (Karte aus 2000) bzw. $\beta_1 = 1$ (Karte aus 1986) erreicht wurde. Die Standardabweichung ist mit weniger als 0,8%, unabhängig von dem Wert für β_1, relativ klein, was auf eine gewisse Stabilität hinweist.

Zusammenfassend kann gesagt werden, dass die Parameter des kontrastsensitiven Potts-Modells bei der Betrachtung des *Hameln* Datensatzes keinen großen Einfluss auf das Klassifikationsergebnis ausüben. Wird der *Las Vegas* Datensatz betrachtet ist zwar ein etwas größerer Einfluss

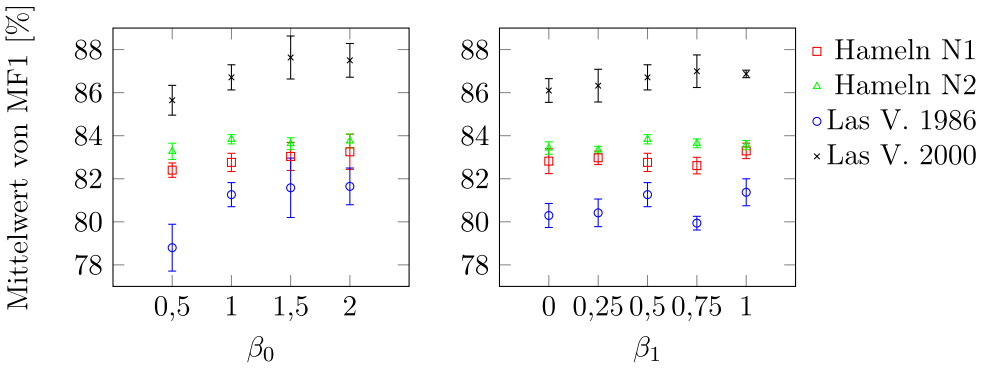


Abbildung 29: Mittelwert von MF1 [%] abhängig von je einem Parameter (β_0, β_1) des kontrastsensitiven Potts-Modells.

als bei dem *Hameln* Datensatz zu erkennen, jedoch sind die Werte $\beta_0 = 1$ und $\beta_1 = 0,5$ so gewählt, dass der MF1 Wert weniger als 1% von dem maximalem MF1 Wert aller getesteter Werte für β_0 und β_1 abweicht.

Parameter des iterativen Aktualisierungsprozesses

Abb. 30 zeigt den MF1 abhängig von der Anzahl der Iterationen innerhalb des **RRFK**. Dies ist nicht zu verwechseln mit der maximalen Anzahl an Iterationen innerhalb der Inferenz des CRFs, die konstant einen Wert von 16 aufweist.

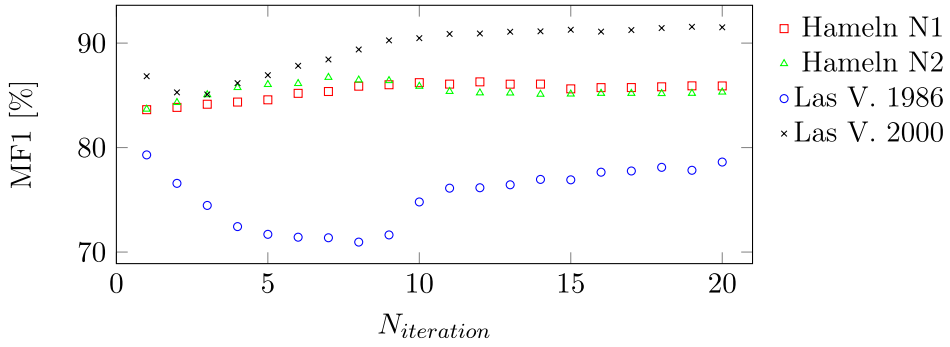
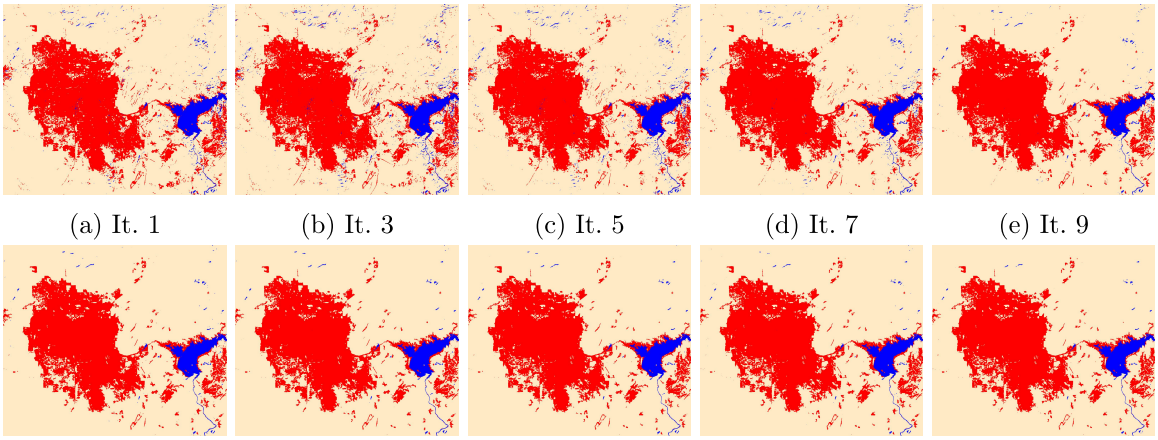


Abbildung 30: MF1 [%] abhängig von der Anzahl der Iterationen des iterativen Klassifikationsprozesses.

Bei dem *Hameln* Datensatz ist zunächst ein leichter Anstieg des MF1 Wertes mit steigender Anzahl der Iterationen zu erkennen. Ab etwa 15 Iterationen bleibt das Ergebnis relativ stabil. Es ist jedoch auch zu erkennen, dass das Ergebnis sich dennoch mit jeder Iteration leicht ändert. Dies resultiert hauptsächlich aus den vielen zufälligen Elementen des RRF. Nach 20 Iterationen wird daher der Prozess beendet, auch wenn sich das Klassifikationsergebnis im Vergleich zur vorherigen Iteration geändert hat.

Bei dem *Las Vegas* Datensatz ändert sich besonders in den ersten 10 Iterationen der MF1 Wert deutlich. Wird die Karte aus dem Jahr 2000 verwendet, sinkt der MF1 Wert in den ersten zwei Iterationen um mehr als 1,5% und steigt daraufhin wieder an. Ab etwa der 10. Iteration ist auch hier das Ergebnis relativ stabil. Der MF1 Wert ist nach der 20. Iteration mit

91,5% um 4,7% größer als in der ersten Iteration. Abb. 31 zeigt die Klassifikationsergebnisse zu einigen Iterationen. Es ist zu erkennen, dass besonders die Anzahl an Fehlklassifikationen in der Wüste um Las Vegas herum, deren Label in der Referenz *Boden* ist, bei den späteren Iterationen deutlich geringer ist, verglichen mit den ersten Iterationen.



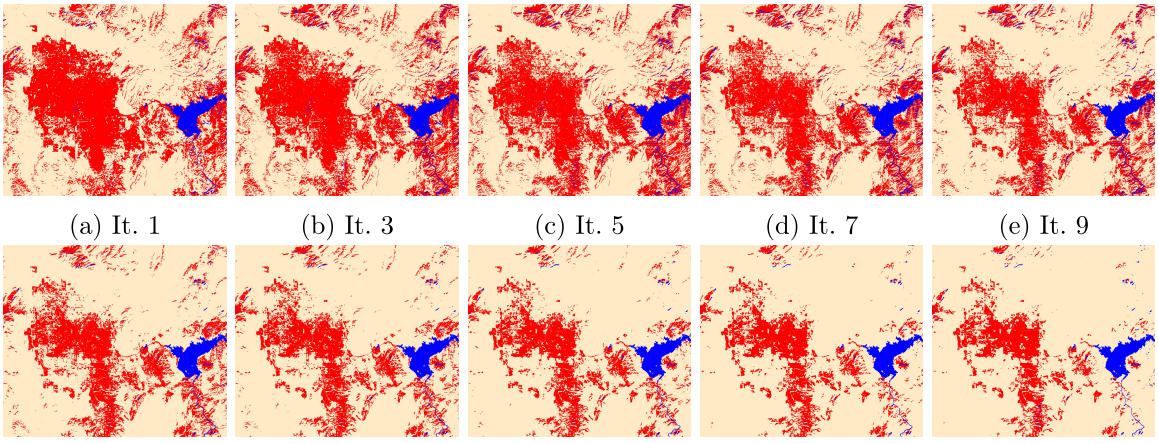
(a) It. 1 (b) It. 3 (c) It. 5 (d) It. 7 (e) It. 9

(f) It. 11 (g) It. 13 (h) It. 15 (i) It. 17 (j) It. 19

Abbildung 31: Klassifikationsergebnis des Datensatzes *Las Vegas* abhängig von der Anzahl der Iterationen. Die Karte stammt aus 2000.
Rot: *Bebauung* Blau: *Wasser*. Beige: *Boden*.

Unter Verwendung der Karte von 1986 sinkt der MF1 Wert von 79,3% bis etwa zur 8. Iteration auf 71,0%. Besonders bei der 10. Iteration erfolgt dann ein starker Anstieg des MF1 Wertes bis auf 74,8%. Von der 10. bis 20. Iteration steigt der MF1 Wert weiter, wenn auch nicht kontinuierlich, bis auf 78,6%. Insgesamt sinkt der MF1 Wert also um etwa 0,7% von der ersten bis zur 20. Iteration. Abb. 32 zeigt das Ergebnis zu den ungeraden Iterationen. Bei der Initialisierung wurden viele Pixel der Klasse *Boden* fälschlicherweise als *Bebauung* klassifiziert. Die Vollständigkeit der Klasse *Boden* liegt bei der Initialisierung mit dem RF bei 78,3% und die Korrektheit der Klasse *Bebauung* bei 52,0%. Die Pixel der Klassen *Bebauung* und *Wasser* wurden größtenteils korrekt erkannt; die Vollständigkeit der Klasse *Bebauung* ist 91,8% und die der Klasse *Wasser* 98,7%. Bis zur 20. Iteration nehmen die fehlerhaft klassifizierten Pixel der wahren Klasse *Boden* deutlich ab. Die Vollständigkeit der Klasse *Boden* steigt demnach und liegt bei der 20. Iteration bei 95,2%. Dafür wird die Klasse *Bebauung* schlechter klassifiziert, sodass die Vollständigkeit auf 44,5% sinkt. Der F1 Wert der Klasse *Boden* steigt demnach von 86,9% (1. Iteration) auf 91,1% (20. Iteration) und jener der Klasse *Bebauung* sinkt von 66,4% auf 55,0%.

Tabelle 15 (links) zeigt den MF1 Wert unter Verwendung der zwei verschiedenen Aktualisierungsfunktionen für die Änderungswahrscheinlichkeit g und das Gewicht θ. Der Unterschied ist bei dem *Las Vegas* Datensatz unter Verwendung der Karte von 2000 mit 1,8% am größten. Hier sind, unter Verwendung der linearen Aktualisierungsfunktion f_{a1}, die Vollständigkeit der Klasse *Boden* (92.8%) und die Korrektheit der Klasse *Bebauung* (77.6%) etwas höher als bei der quadratischen Funktion f_{a2} mit 90.8% bzw. 73.1%. Verursacht wird dies durch die Fehlklassifikation vereinzelter Pixeln der Klasse *Boden* als *Bebauung*. Bei den weiteren durchgeführten Experimenten ist der MF1 unter Verwendung der Funktion f_{a2} etwas größer



(a) It. 1 (b) It. 3 (c) It. 5 (d) It. 7 (e) It. 9

(f) It. 11 (g) It. 13 (h) It. 15 (i) It. 17 (j) It. 19

Abbildung 32: Klassifikationsergebnis des Datensatzes *Las Vegas* abhängig von der Anzahl der Iterationen (Abkürzung für Iteration: It.). Die Karte stammt aus 1986. Rot: *Bebauung*. Blau: *Wasser*. Beige: *Boden*.

als unter Verwendung der Funktion f_{a1}. Die Differenz beträgt jedoch maximal 0,9%. In den folgenden Experimenten wurde f_{a1} verwendet.

In Tabelle 15 sind auch die MF1 Werte zu finden, abhängig davon, ob die Detektion möglicher Schlagschatten bei der Erstellung des Änderungsbildes B (Kapitel 4.5) berücksichtigt wurde oder nicht. Mit einer Differenz von 6,2% ist der MF1 Wert unter Verwendung der Schattendetektion bei dem Las Vegas Datensatz mit der Karte von 1986 deutlich größer als ohne die Schattendetektion. Dieser Unterschied entsteht, da sich in der Umgebung der Stadt Las Vegas viele Dünen befinden, die einen starken Schatten werfen. Der Bereich im Schatten gehört der Klasse *Boden* an, wird aber ohne Berücksichtigung der Schatten teilweise als *Bebauung* klassifiziert. Dies zeigt auch die deutlich höhere Korrektheit der Klasse *Bebauung* von 94.5% unter Berücksichtigung der Schatten verglichen zur Korrektheit von 55.4% ohne deren Berücksichtigung. Auch die Vollständigkeit der Klasse *Boden* ist mit 99.0% höher als 91.1% ohne die Schattendetektion. Dennoch wird in den Experimenten auf die Detektion von Schatten bei den niedrig aufgelösten Datensätzen (*Las Vegas*, *Herne*, *Husum*) verzichtet. Dies hat zwei Gründe. Zum einen sind die Schatten der Gebäude in den hoch aufgelösten Datensätzen deutlich kleiner als die Schatten der Dünen. Wird in der Klassifikation nun angenommen, dass alle erkannten Änderungen in Schattengebieten Fehldetektionen sind, bekommt das Kartenlabel einen größeren Einfluss in diesen Gebieten. Wahre Änderungen könnten übersehen werden. Da bei großen Schatten aufgrund des Geländes ganze Bebauungsgebiete nicht erkannt werden könnten, wird bei niedrig aufgelösten Bildern keine Schattendetektion durchgeführt. Ein weiterer Grund ist, dass die Detektion der Schatten auf der Intensität basiert. Diese ist nicht nur bei Schatten, sondern auch bei Wasser sehr gering. Die Klasse *Wasser* hat, wenn nur die Pixel mit fehlerhaftem Label in der Karte ($C \neq \widetilde{C}$) betrachtet werden, mit 69,8% eine um 1,8% schlechtere Vollständigkeit unter Berücksichtigung von Schatten. Erfolgt die Evaluierung mit allen Pixeln ist die Vollständigkeit immer noch um 0,3% kleiner.

Der Einfluss der minimalen Breite der Segmente u und der minimalen Größe der Segmente s auf den MF1 Wert ist in Abb. 33 dargestellt. Bei den Parametern u und s wurden nur Werte

Datensatz	Aktualisierungsfunktion		Schatten detektiert	
	f_{a1}, linear	f_{a2}, quadr.	ja	nein
Hameln, N1	86,0	86,2	86,0	85,8
Hameln, N2	84,8	85,2	84,8	85,1
Las Vegas, 1986	76,0	76,9	82,2	76,0
Las Vegas, 2000	91,7	89,9	93,2	91,7

Tabelle 15: MF1 [%] abhängig von der Aktualisierungsfunktion und der Schattendetektion.

getestet, die für die Auflösung des Datensatzes auch Sinn ergeben. Beispielsweise wird die minimale Breite von 0,2 m nicht bei dem *Las Vegas* Datensatz mit einer Bodenpixelgröße von 30 m getestet.

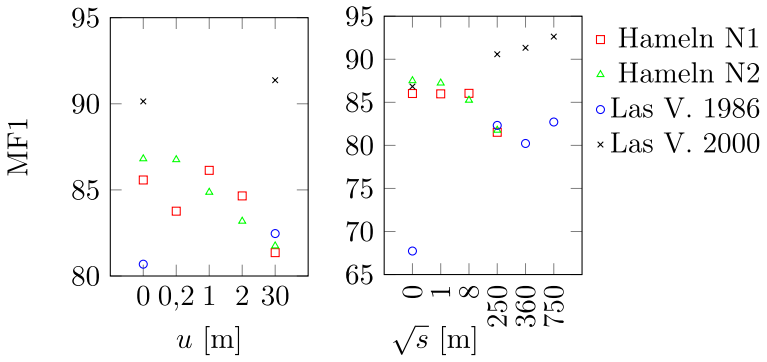


Abbildung 33: MF1 [%] abhängig von den Parametern u (minimale Breite der Segmente) und s (minimale Größe der Segmente).

Die Parameter u und s haben einen starken Einfluss auf das Klassifikationsergebnis. Die maximale Differenz des MF1 Wertes innerhalb einer Testreihe beträgt 5,1% für den Parameter u und 15,0% für den Parameter s. Besonders niedrige MF1 Werte resultieren, wenn die verwendeten Werte nicht zur Auflösung der Fernerkundungsdaten passen. Dies sind besonders große Werte ($u = 30$, $\sqrt{s} \geq 250$) für den Datensatz *Hameln* und besonders kleine Werte ($u = 0$, $\sqrt{s} \leq 8$) für den Datensatz *Las Vegas*. Werden beispielsweise bei der Analyse der minimalen Größe s nur die zur Auflösung passenden Werte betrachtet ist die maximale Differenz 2,5% anstatt 15,0%. Dies war zu erwarten, da das Aussehen der Änderungen stark von den erkennbaren Objekten und somit auch von der Auflösung abhängt. Die verwendeten Werte passen mit $u = 30$ und $\sqrt{s} = 250$ für den Datensatz *Las Vegas* und $u = 1$ und $\sqrt{s} = 8$ für den Datensatz *Hameln* gut zu der Bodenpixelauflösung.

Zusammenfassend kann gesagt werden, dass die Parameter bezogen auf den Aktualisierungsprozess (Anzahl an Iterationen $N_{iteration}$, minimale Breite der Segmente u, minimale Größe der Segmente s, Verwendung von Schattendetektion, Wahl der Aktualisierungsfunktion) einen großen Einfluss auf das Klassifikationsergebnis ausüben. Die Wahl dieser Parameter u und s und die Wahl, ob Schatten detektiert werden oder nicht, erfolgt jedoch nicht zufällig, sondern in Abhängigkeit von den zu klassifizierenden Daten. Dadurch können relativ gute Klassifikationsergebnisse erzielt werden. Die Änderung der Aktualisierungsfunktion hat bei einem Großteil der getesteten Szenarien kaum Einfluss auf das Ergebnis. Bezogen auf die Anzahl an Iterationen $N_{iteration}$ ist bei den meisten getesteten Szenarien ab etwa 15 Iterationen ein rela-

tiv stabiler MF1 Wert zu beobachten. Der gewählte Wert von $N_{iteration} = 20$ liegt innerhalb dieses Bereichs.

Diskussion

In Kapitel 5.6.2 wurde die Frage gestellt, wie sensitiv das Klassifikationsergebnis bezogen auf die zuvor zu definierenden Parameter ist. Die Experimente zeigen, dass die Wahl der Parameter einen Einfluss auf das Ergebnis ausübt. Bei vielen Parametern gibt es jedoch einen Bereich in dem das Ergebnis relativ stabil und der MF1 Wert verhältnismäßig groß ist. Die verwendeten Werte liegen größtenteils in diesem Bereich. Es gibt zwar bei vielen Parametern weitere Werte, bei denen das Ergebnis einen größeren MF1 zeigt, jedoch ist der Unterschied des MF1 Wertes meist gering.

Da die Referenzen lediglich für die Evaluierung zur Verfügung stehen, werden die Werte der Parameter nicht angepasst. In den folgenden Experimente werden daher Werte genutzt, die auf Erfahrungswerten mit den Basisklassifikatoren beruhen bzw. abhängig von der Auflösung der Daten gewählt werden. Dies hat auch den Vorteil, dass die Parameter nicht an die Daten überangepasst werden.

Insgesamt zeigen die Experimente, dass die Werte der Parameter, die in den folgenden Experimenten verwendet werden, zu relativ guten Ergebnissen führen. Auch existiert meist ein relativ großer Bereich, bei dem das Ergebnis stabil ist. Somit wird vermutet, dass die Werte der Parameter keinen großen Einfluss auf das Klassifikationsergebnis nehmen.

6.3. Monotemporale Klassifikation

In diesem Abschnitt werden Ergebnisse der monotemporalen Klassifikation präsentiert. Die Analyse orientiert sich an den Fragen aus Kapitel 5.6.3. Die Ergebnisse der unterschiedlichen Methoden aus Tabelle 9 werden für jeden Datensatz verglichen. In Kapitel 6.3.5 folgt ein kurzes Fazit zu den monotemporalen Experimenten, in dem noch einmal auf alle Fragen aus Kapitel 5.6.3 eingegangen wird. Die Frage f, bei der die Ergebnisse in Hinblick auf die unterschiedlichen Auflösungen der Daten analysiert werden soll, kann erst in diesem Kapitel beantwortet werden.

6.3.1. Hameln

Abb. 34 und Abb. 35 zeigen die GG der unterschiedlichen Methoden bezogen auf das ganze Bild, die geänderten Gebiete und die Änderungsdetektion für die Gebiete N1 und N2. In Tabelle 16 ist neben der GG auch der MF1 dargestellt. Die weiteren Genauigkeitsmaße Vollständigkeit, Korrektheit und F1 Wert finden sich in Tabelle B.1, B.2 und B.3 in Anhang B. Im Folgenden erfolgt die Analyse, die sich an den Fragen aus Kapitel 5.6.3 orientiert. Frage e wird hier nicht berücksichtigt, da sie sich auf die Beeinflussung des Klassifikationsergebnisses durch die Anzahl der Fehler in den Trainingslabels bezieht, die hier konstant ist.

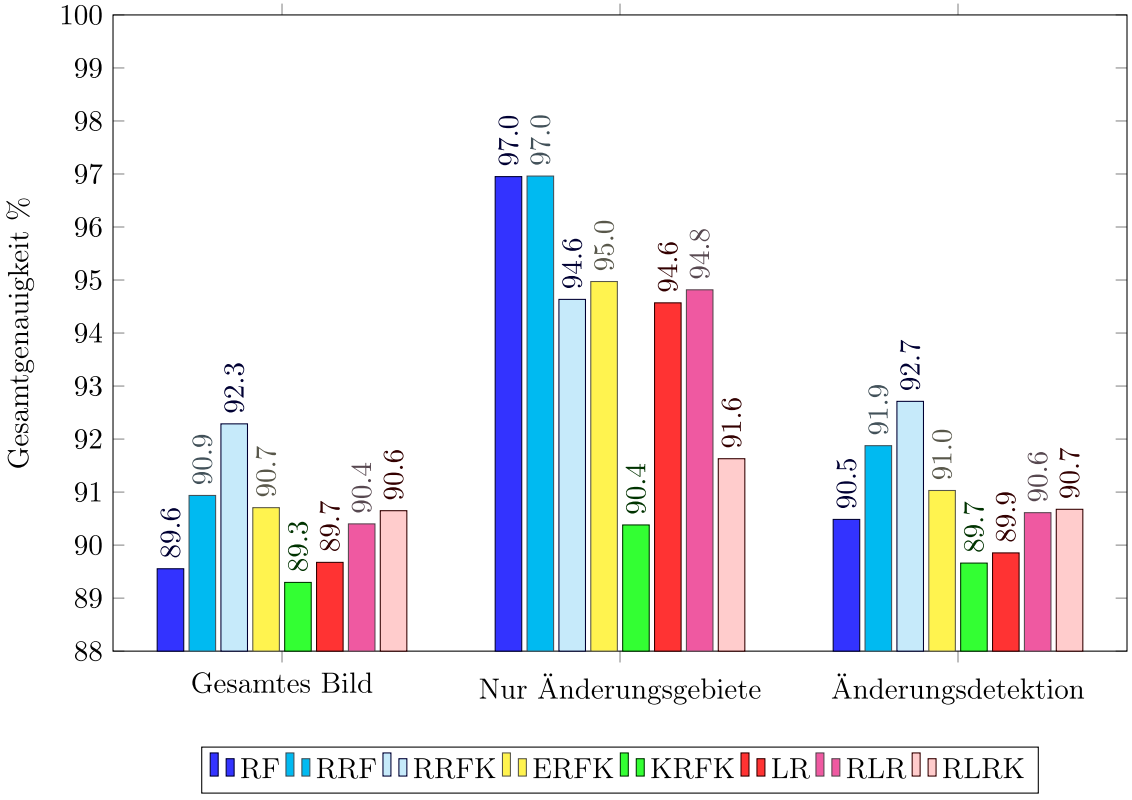


Abbildung 34: Gesamtgenauigkeit [%]. Datensatz: *Hameln* (N1).

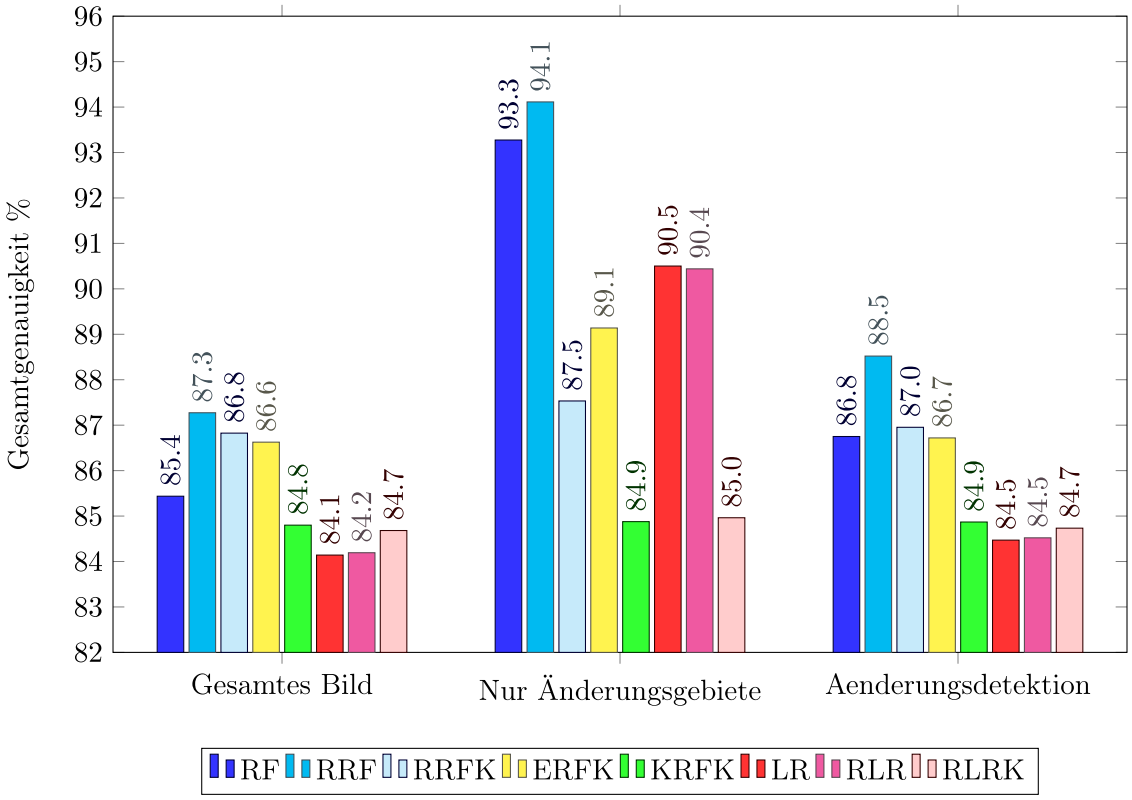


Abbildung 35: Gesamtgenauigkeit [%]. Datensatz: *Hameln* (N2).

Gebiet		Gesamtes Bild		Nur Änderungen		Änderungsdetektion	
		GG	MF1	GG	MF1	GG	MF1
N1	RF	89,6	83,5	**97,0**	85,2	90,5	83,4
	RRF	90,9	84,6	**97,0**	**85,3**	91,9	**85,4**
	RRFK	**92,3**	**85,9**	94,6	74,5	**92,7**	84,3
	ERFK	90,7	81,7	95,0	75,0	91,0	81,9
	KRFK	89,3	82,3	90,4	54,6	89,7	72,8
	LR	89,7	79,7	94,6	73,8	89,9	79,6
	RLR	90,4	81,9	94,8	76,5	90,6	81,0
	RLRK	90,6	82,4	91,6	59,2	90,7	76,3
N2	RF	85,4	83,7	93,3	71,3	86,8	77,5
	RRF	**87,3**	**85,4**	**94,1**	**76,2**	**88,5**	**80,4**
	RRFK	86,8	85,3	87,5	47,7	87,0	66,5
	ERFK	86,6	84,2	89,1	48,4	86,7	69,9
	KRFK	84,8	83,2	84,9	36,0	84,9	55,0
	LR	84,1	81,3	90,5	59,6	84,5	70,1
	RLR	84,2	81,4	90,4	59,5	84,5	70,1
	RLRK	84,7	82,9	85,0	36,4	84,7	55,1

Tabelle 16: Genauigkeitsmaße [%]. Datensatz: *Hameln*. Der größte Wert pro Gebiet ist fett gedruckt.

Frage a) Berücksichtigung der Fehler in den Trainingslabels

Zunächst wird untersucht, ob sich das Klassifikationsergebnis verbessert, wenn die Fehler in den Trainingslabels im Training berücksichtigt werden (Frage a). Bei der Betrachtung der GG und des MF1 Wertes ist eine deutliche Verbesserung des Ergebnisses zu beobachten, wenn die Methode **RRF** anstatt der Methode **RF** verwendet wird. Besonders unter Berücksichtigung des gesamten Bildes steigt die GG relativ deutlich von 89,6% auf 90,9% für N1 und von 85,4% auf 87,3% für N2 an. Auch bei der Änderungsdetektion steigt die GG um 1,4% (N1) bzw. 1,7% (N2). In den geänderten Gebieten ist die GG nur bei Gebiet N2 etwas größer unter Verwendung des **RRF**; Bei Gebiet N1 ist die GG identisch.

Bei dem Vergleich der **LR** mit der **RLR** ist nur eine geringe Verbesserung der GG zu beobachten. Bei Gebiet N1 ist die GG beispielsweise um 0,7% höher unter Berücksichtigung des gesamten Bildes.

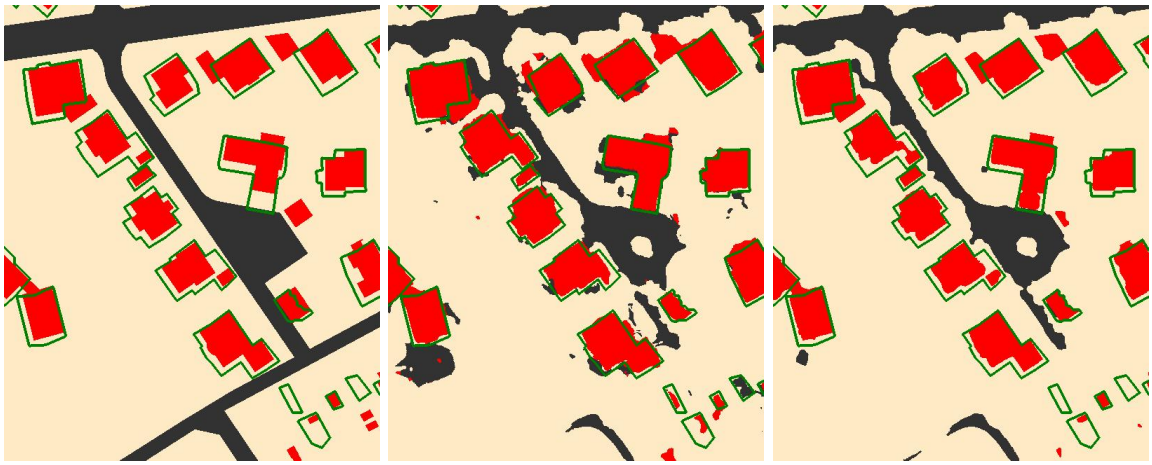
Demnach zeigen diese Ergebnisse, dass sich das Klassifikationsergebnis verbessert, wenn die Fehler in den Trainingslabels im Training berücksichtigt werden.

Frage b) Berücksichtigung der Kartenlabels als weitere Beobachtung

Als nächstes wird die Integration der Karte als Beobachtung betrachtet (Frage b). Bei dem Gebiet N1 ist eine Steigerung der GG von 1,4% zu beobachten, wenn die Methode **RRFK** anstatt der Methode **RRF** verwendet wird. Damit ist die GG der Methode **RRFK** um 2,7% größer als die des **RF** Klassifikators. Ein Vergleich der Ergebnisse der **RLRK** und der **RLR**

zeigt ebenfalls eine leichte Erhöhung der GG von 0,2%. Ein ähnliches Bild ist bei der Änderungsdetektion zu beobachten. Werden jedoch nur die Pixel mit fehlerhaftem Label in der Karte betrachtet, ist die GG mit 94,6% unter Verwendung der **RRFK** um 2,4% kleiner als unter Verwendung des **RRF**, wenn auch immer noch relativ hoch. Eine Verschlechterung des Klassifikationsergebnis in den Änderungsgebieten war bei der Berücksichtigung des Klassenlabels in dem Klassifikationsprozess jedoch zu erwarten. Da die Klassenlabels in diesen Änderungsgebieten nicht korrekt sind, hat die zusätzliche Berücksichtigung dieser Labels einen schlechten Einfluss auf das Klassifikationsergebnis in diesen Gebieten. Auch bei dem Vergleich der Ergebnisse der Methoden **RLR** und **RLRK** ist eine Verschlechterung der GG von 3,2% in den Änderungsgebieten zu beobachten. Die GG liegt bei der **RLRK** nur noch bei 91,6%, während sie bei der **RLR** noch bei 94,8% lag.

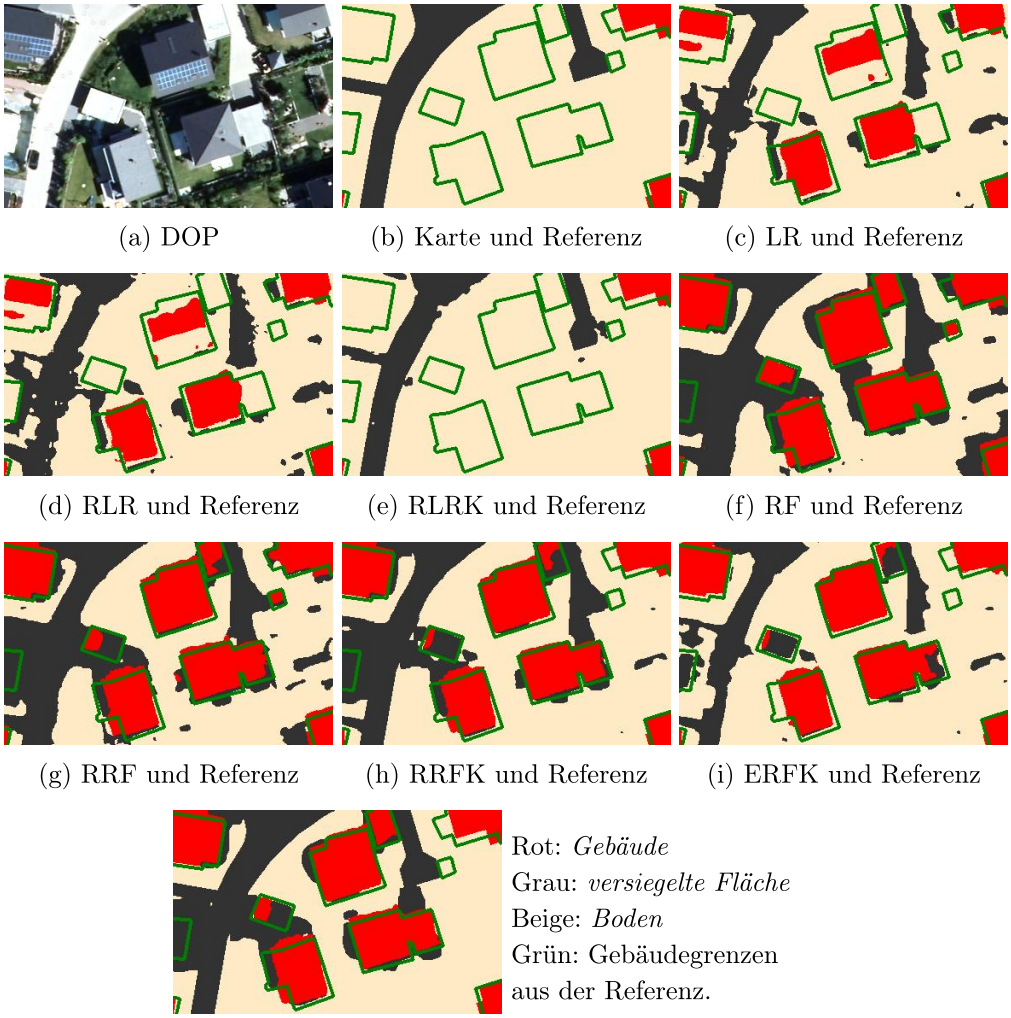
Ein visueller Vergleich zeigt den Einfluss der zusätzlichen Beobachtung der Kartenlabels. In Abb. 36 sind die Karte und die Klassifikationsergebnisse eines Ausschnitte des Gebiets N1 zu sehen. Es ist zu erkennen, dass die Gebäudegrenzen unter Verwendung des **RRFK** durch den Einfluss der Kartenlabels mehr den Grenzen in der veralteten Karte gleichen, als die Gebäudegrenzen im Ergebnis des **RRF**. Bei dem Datensatz *Hameln* stimmen die Gebäudegrenzen aus der veralteten Karte, wie in Kapitel 5.1.1 erläutert, nicht mit denen aus dem DOP und somit auch nicht denen der Referenz überein. Die Ränder werden demnach in diesem Fall besser mit der Methode **RRF** erkannt. Bei neuen Gebäuden, wie in Abb. 37, ist kein großer Unterschied der Gebäudegrenzen zu beobachten.



(a) Karte und Referenz (b) RRF und Referenz (c) RRFK und Referenz

Abbildung 36: Ausschnitt aus Gebiet N1 mit bereits existierenden Gebäuden.
Rot: *Gebäude*. Grau: *versiegelte Fläche*. Beige: *Boden*. Grün: Gebäudegrenzen aus der Referenz.

Am Rand des Neubaugebietes in N1 liegt ein Bereich, bei dem die Gebäude noch nicht gebaut wurden, die Straße aber schon existiert (Abb. 38). In der Karte ist die Straße ebenfalls noch nicht zu erkennen, sodass die Straßenpixel, wie die Pixel in deren Umgebung, das Kartenlabel *Boden* aufweisen. In dem Bereich, in dem Gebäude gebaut werden sollen, ist zu erkennen, dass die Bodenpixel neben der Straße aufgrund der Abnutzung durch die Bauarbeiten ein anderes Erscheinungsbild aufweisen als andere Pixel die ebenfalls der Klasse *Boden* angehören. Während sowohl der **RF** als auch der **RRF** einige dieser Pixel der Klasse *versiegelte Fläche*



(a) DOP (b) Karte und Referenz (c) LR und Referenz

(d) RLR und Referenz (e) RLRK und Referenz (f) RF und Referenz

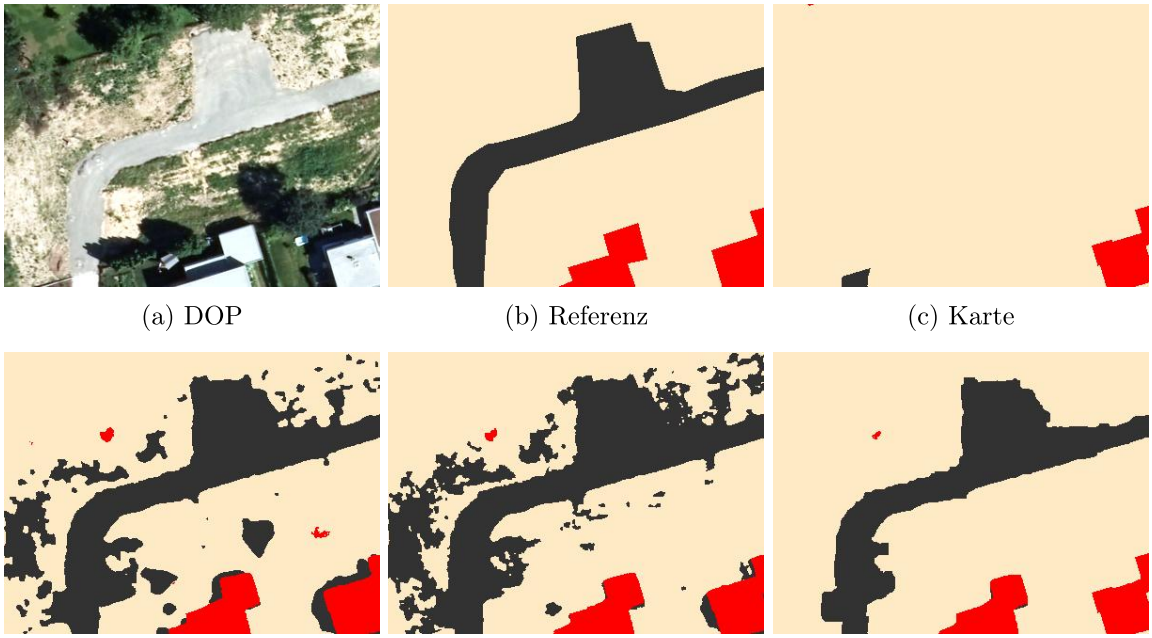
(g) RRF und Referenz (h) RRFK und Referenz (i) ERFK und Referenz

(j) KRFK und Referenz

Abbildung 37: Ausschnitt aus Gebiet N1 des Datensatzes *Hameln* mit neuen Gebäuden.

fehlerhaft zuordnen, werden sie durch den Einfluss des Kartenlabels mit dem **RRFK** korrekt der Klasse *Boden* zugeordnet. Die neu gebaute Straße wurde von beiden Methoden korrekt klassifiziert.

Bei Gebiet N2 zeigt der Vergleich der GG des gesamten Bildes der Methoden **RRF** und **RRFK**, dass die zusätzliche Integration des Kartenlabels in die Klassifikation zu einer Verschlechterung der GG führt. Die GG sinkt von 87,3% auf 86,8%. Dies liegt vermutlich an der schlechteren Detektion der Klasse *Gebäude*, was an der geringeren Vollständigkeit dieser Klasse von 77,3% unter Verwendung des **RRFK** verglichen zu der Vollständigkeit von 86,3% unter Verwendung des **RRF** zu erkennen ist. Ursächlich hierfür könnten die in der Karte fehlerhaft gelabelten Pixel der Klasse *Gebäude* an Gebäuderändern sein. Wie zuvor beschrieben weisen einige Pixel an Gebäuderändern in der Karte ein falsches Label auf, beispielsweise aufgrund von Dachüberhängen, die in der Karte nicht erfasst wurden, oder aufgrund von Fehlern bei der Georeferenzierung. In dem Gebiet N2 gibt es, im Gegensatz zu dem Gebiet N1, jedoch kaum Änderungen zwischen dem Zeitpunkt der Kartenerstellung und der Aufnahme der Bilder in Form von Neubauten. Über 58% der Pixel mit unterschiedlichen Karten- und Referenzlabels werden durch den Algorithmus bei der Bestimmung der Änderungsbilder (Kapitel 4.5) als



(a) DOP (b) Referenz (c) Karte

(d) RF (e) RRF (f) RRFK

Abbildung 38: Ausschnitt aus Gebiet N1 des Datensatzes *Hameln* mit einer neuen Straße. Rot: *Gebäude*. Grau: *versiegelte Fläche*. Beige: *Boden*.

unverändert angesehen, da sie Segmente im Bildraum bilden, die zu klein oder zu schmal sind, um als tatsächliche Änderung verstanden zu werden. Werden nur die Änderungen mit dem Label *Gebäude* in der Referenz betrachtet, werden sogar über 72% der Pixel irrtümlich als *unverändert* gekennzeichnet. Diese Pixel an Gebäuderändern werden in der Klassifikation eher dem Label der Karte zugeordnet, wenn diese als Beobachtung mit einfließt. Die zwei neuen Gebäude wurden jedoch, wie in Abb. 39 zu sehen ist, von allen Methoden erkannt.

Zusammenfassend kann gesagt werden, dass die Integration der Kartenlabels in die Klassifikation zu einer Annäherung des Klassifikationsergebnisses an die veraltete Karte führt. Demnach verbessert sich größtenteils das Ergebnis der Pixel, deren Labels sich nicht geändert haben. Änderungen, die in dem iterativen Klassifikationsalgorithmus bei der Bestimmung der Änderungsbilder als *Änderung* erkannt wurden, werden auch unter Berücksichtigung der Kartenlabels korrekt klassifiziert. Segmente an Pixeln mit geänderten Labels, die zu klein oder zu schmal sind um als *Änderung* erkannt zu werden, wie beispielsweise an Gebäuderändern, werden meist in der Klassifikation dem Kartenlabel zugeordnet. Demnach ähneln die Ränder der bereits in der Karte existierenden Gebäude in dem Klassifikationsergebnis mehr den Rändern in der Karte, wenn die Karte in der Klassifikation berücksichtigt wird. Neue Gebäude werden dennoch auch mit der Klassifikation mit beobachteten Kartenlabels korrekt erkannt.

Frage c) Verwendung der Wahrscheinlichkeiten der Klassenzugehörigkeit

Die nächste zu untersuchende Frage (c) ist, ob das Ersetzen der Zuordnung eines Trainingsbeispieles zu einer Klasse durch die Zuordnung dieses Trainingsbeispieles zu allen Klassen mit der Wahrscheinlichkeit der Klassenzugehörigkeit das Klassifikationsergebnis verbessert. Hierfür wird das Ergebniss des **RRFK** mit den Ergebnissen der Methoden **ERFK** und **KRFK**

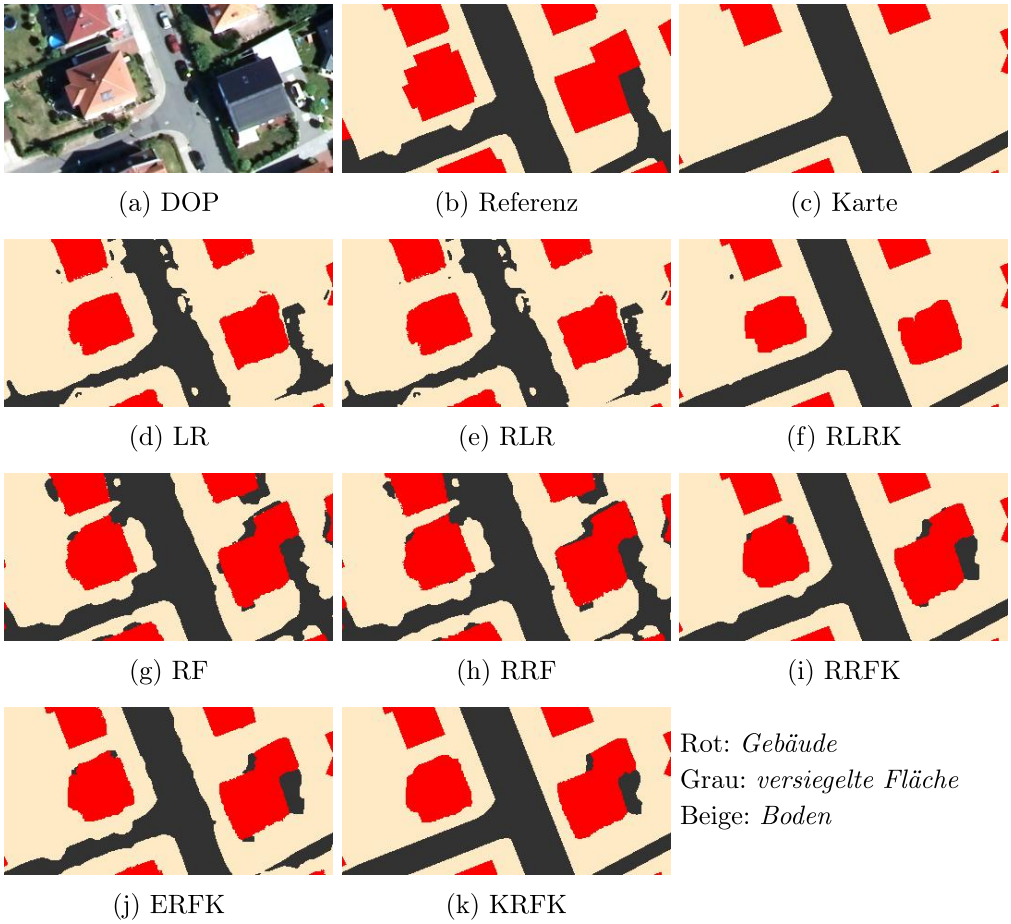


Abbildung 39: DOP und Ergebnisse eines Ausschnittes des Gebietes N2 des Datensatzes *Hameln*, in dem die zwei neuen Gebäude in Gebiet N2 enthalten sind.

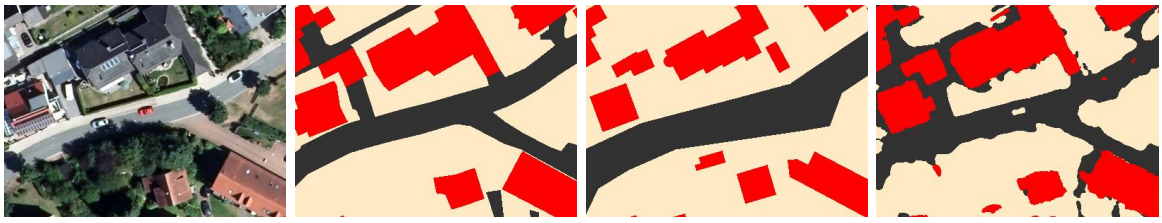
verglichen. Bei dem **ERFK** werden die resultierenden Labels der vorherigen Iteration für das Training verwendet; ausschließlich bei der ersten Iteration werden die Labels der veralteten Karte genutzt. Bei dem **KRFK** stammen die Trainingslabels immer aus der veralteten Karte.

Die GG der Ergebnisse des **ERFK** unter Berücksichtigung des gesamten Bildes für die Gebiete N1 und N2 sind um 1,6% bzw. 0,2% geringer als die GG der Ergebnisse des **RRFK**. Auch bei der Änderungsdetektion ist die GG, mit einer Differenz von 1,7% (N1) bzw. 0,3% (N2), höher bei der Klassifikation mit dem **RRFK**. Werden nur die Pixel, mit geändertem Label in der Karte betrachtet, ist die GG unter Verwendung der Methode **ERFK** um 0,4% (N1) bzw. 1,6% (N2) größer. Dies kann durch den geringeren Einfluss der Kartenlabels in dem Training des **ERFK** erklärt werden. Während in jeder Iteration des **RRFK** das Kartenlabel bei der Berechnung der Wahrscheinlichkeit der Klassenzugehörigkeit berücksichtigt wird, erfolgt das Training jeder Iteration, abgesehen von der ersten, bei dem **ERFK** ohne die direkte Verwendung der Kartenlabels.

Bei der Betrachtung der weiteren Genauigkeitsmaße unter Berücksichtigung des gesamten Bildes ist zu erkennen, dass der F1 Wert der Klasse *versiegelte Fläche* bei dem Ergebnis des **ERFK** mit 68,0% (N1) bzw. 77,2% (N2) deutlich kleiner ist als jener bei dem Ergebnis des **RRFK**, welcher 79,3% für N1 bzw. 83,0% für N2 beträgt (Tabelle B.1). Dies resultiert vor

allem aus der relativ kleinen Vollständigkeit der Klasse *versiegelte Fläche* bei dem Ergebnis des **ERFK** mit 55,3% (N1) bzw. 63,9% (N2). Damit liegt die Vollständigkeit unter jener des **RRFK** mit 74,6% (N1) bzw. 74,0% (N2). Anders als bei der GG ist die Vollständigkeit der Klasse *versiegelte Fläche* auch bei der alleinigen Betrachtung der Änderungsgebiete bei dem Ergebnis des **ERFK** geringer (Tabelle B.2). Eine mögliche Ursache der Fehlklassifikationen von Pixeln der Klasse *versiegelte Fläche*, welche diese niedrige Vollständigkeit im ganzen Bild verursachen, könnte in einer iterativen Verstärkung der Fehler liegen. Dies wird im Folgenden anhand eines Beispiels erläutert.

Die Abbildungen 40 und 41 zeigen Ausschnitte des Gebietes N1, in dem eine Straße mit Autos zu erkennen ist. Autos werden in einer Karte im Allgemeinen nicht erfasst, sodass die entsprechenden Pixel in der Karte und der Referenz das Label *versiegelte Fläche* aufweisen. Die Merkmale dieser Pixel, besonders das Höhenmerkmal *nDOM*, unterscheiden sich jedoch von denen anderer Straßenpixel. In der Initialisierung konnten die Autos daher vom RF nicht sicher einer Klasse zugeordnet werden, sodass einige Autopixel fehlerhaft klassifiziert werden. Hierbei erhalten die Pixel roter Autos meist das Label *Boden* (Abb. 40), während die Pixel anderer Autos teilweise korrekt oder als *Gebäude* klassifiziert werden (Abb. 41). Da in dem **ERFK** die Labels des Klassifikationsergebnises der vorherigen Iteration zum Training verwendet werden, verstärkte sich anscheinend die falsche Annahme des Klassifikators, dass Pixel, die Autos abbilden, der Klasse *Boden* angehören. In Abb. 42 ist beispielhaft für ein nicht rotes Auto gezeigt, wie das Label von *Gebäude* zu *Boden* über mehrere Iterationen wechselt. Im Ergebnisbild von **ERFK** erhalten demnach mehr Autopixel das fehlerhafte Label *Boden* als in dem Ergebnisbild des **RF**. Das selbe Phänomen ist auch in Abb. 40 bei den Straßenpixeln im Schatten der angrenzenden Bäume zu erkennen. In dem Ergebnis des **RF** sind nur wenige dieser Pixel im Schatten, die in der Referenz zu der Klasse *versiegelte Fläche* gehören, fehlerhaft dem Label *Boden* zugeordnet worden. Nach mehreren Iterationen innerhalb des **ERFK** besitzt jedoch ein Großteil der Pixel das falsche Label *Boden*.

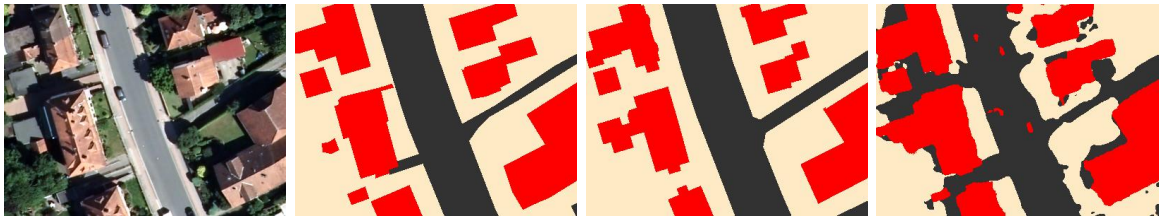


(a) DOP (b) Referenz (c) Karte (d) RF

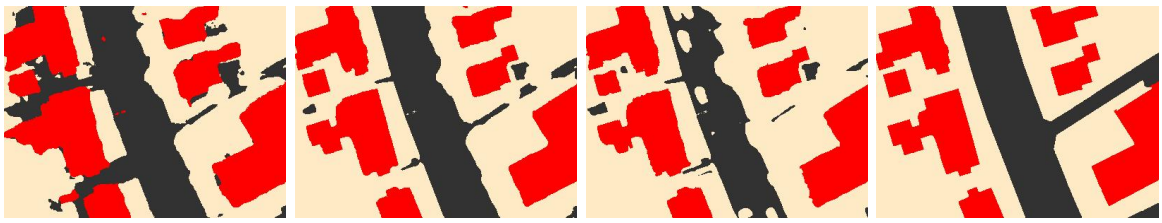


(e) RRF (f) RRFK (g) ERFK (h) KRFK

Abbildung 40: Straße 1 in Gebiet N1 des *Hameln* Datensatzes.
Rot: *Gebäude*. Grau: *versiegelte Fläche*. Beige: *Boden*.



(a) DOP (b) Referenz (c) Karte (d) RF



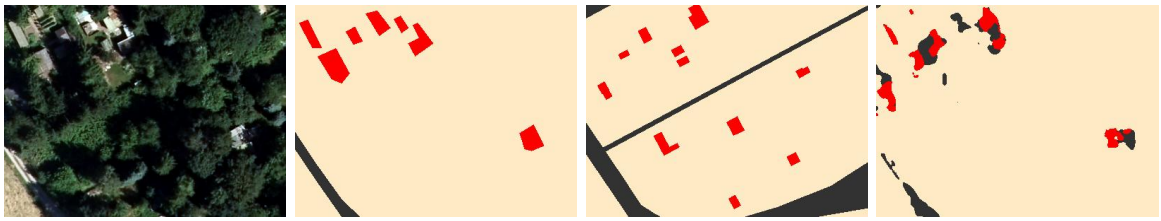
(e) RRF (f) RRFK (g) ERFK (h) KRFK

Abbildung 41: Straße 2 in Gebiet N1 des *Hameln* Datensatzes.
Rot: *Gebäude*. Grau: *versiegelte Fläche*. Beige: *Boden*.

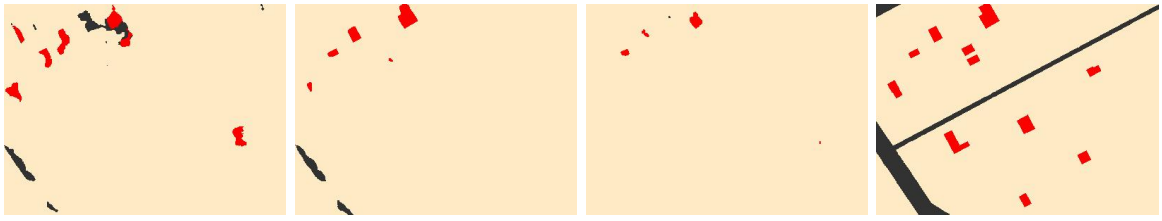


DOP It. 1 It. 3 It. 5 It. 7 It. 9 It. 11 It. 13 It. 15 It. 17 It. 19

Abbildung 42: Ein Auto auf Straße 2 (Abb. 41) in Gebiet N1 des *Hameln* Datensatzes zu unterschiedlichen Iterationsen des **ERFK**.
Rot: *Gebäude*. Grau: *versiegelte Fläche*. Beige: *Boden*.



(a) DOP (b) Referenz (c) Karte (d) RF



(e) RRF (f) RRFK (g) ERFK (h) KRFK

Abbildung 43: Straße 3 in Gebiet N1 des *Hameln* Datensatzes.
Rot: *Gebäude*. Grau: *versiegelte Fläche*. Beige: *Boden*.

Die GG des Ergebnisses der Methode **KRFK** ist sehr gering verglichen zu den Ergebnissen der anderen Methoden, die auf dem RF basieren. In Gebiet N1 ist die GG sogar kleiner als die aller anderen Methoden. Besonders in Änderungsgebieten ist die GG mit 90,4% vergleichsweise sehr gering. Die Vollständigkeit und Korrektheit in den Änderungsgebieten liegt bei den Ergebnissen des **KRFK** unter 45% für die Klassen *Boden* und *versiegelte Fläche*. Bei Gebiet

N2 sind die Vollständigkeit und Korrektheit dieser Klassen in Änderungsgebieten sogar unter 15%. Bei der Methode **KRFK** werden anstatt der Labels aus dem Klassifikationsergebnis der vorherigen Iteration die Labels der veralteten Karte zum Training genommen. Zusätzlich werden die Kartenlabels als Beobachtung in den Klassifikationsprozess integriert. Der Karteneinfluss ist demnach ziemlich hoch, sodass Pixel, deren Merkmalsvektor nicht repräsentativ für die Klasse ist, meist der Klasse aus der Karte zugeordnet werden. Somit werden auch die Autopixel und Schattenpixel der Beispiele aus Abb. 40 und Abb. 41 korrekt als *versiegelte Fläche* klassifiziert. Hat sich das Label eines Pixels jedoch geändert, wird das Pixel in der Klassifikation teilweise dennoch dem Kartenlabel zugeordnet. Dies geschieht vor allem, wenn die Änderung des Labels bei der Bestimmung des Änderungsbildes in Kapitel 4.5 nicht erkannt wurde. Dies erklärt die geringe Vollständigkeit und Korrektheit der Klassen *Boden* und *versiegelte Fläche*. Gut zu erkennen ist dies am Beispiel von Abb. 43. Obwohl der schmale Feldweg und die kleineren Gebäude in diesem Beispiel aufgrund der vielen Bäume in den Fernerkundungsdaten nicht mehr zu erkennen sind, werden sie dennoch von dem **KRFK** detektiert. Größere neue Objekte werden jedoch meist auch mit **KRFK** korrekt erkannt, wie an den Neubauten in Abb. 37 zu sehen ist.

Zusammenfassend kann gesagt werden, dass die Klassifikation mit dem **ERFK** zu einer schlechteren GG und zu einer deutlich niedrigeren Vollständigkeit der Klasse *versiegelte Fläche* führt als der *RRFK*, wenn das Gesamtergebnis betrachtet wird. In den Gebieten, in denen die Labels der veralteten Karte fehlerhaft sind, ist die GG aufgrund des geringeren Karteneinflusses etwas größer unter Verwendung des *ERFK*. Die Ergebnisse des *KRFK* werden stark von den Labels der veralteten Karte beeinflusst. Dies führt zu deutlich schlechteren GG, vor allem, wenn nur die Pixel in den Änderungsgebieten betrachtet werden.

Frage d) Vergleich mit der robusten logistischen Regression

Als letztes werden noch die Klassifikationsergebnisse basierend auf dem RRF mit jenen der RLR verglichen (Frage d). Die GG des Ergebnisses der **RLR** liegt in Gebiet N1 und N2 unter der GG des Ergebnisses der Methode **RRF**. Auch bei dem Vergleich der Ergebnisse der Methoden **RLRK** und **RRFK** ist die GG größer, wenn der **RRFK** für die Klassifikation genutzt wird. Diese Beobachtung bezieht sich sowohl auf die GG des ganzen Bildes, als auch auf die unter der Verwendung der Pixel in Änderungsgebieten und die der Änderungsdetektion. Bei den MF1 Werten (Tabelle 16) ist die Differenz zwischen den Ergebnissen der Klassifikatoren sogar noch deutlicher zu erkennen.

Ein Vergleich der Ergebnisse der Basisklassifikatoren **LR** und **RF** zeigt ebenfalls eine höhere GG bei dem **RF**, wenn nur die Pixel mit fehlerhaftem Label in der Karte betrachtet werden. Dies ist auch bei der Änderungsdetektion zu beobachten. Obwohl in Gebiet N1 die GG des Ergebnis von der **LR** bezogen auf das gesamte Bild um 0,1% größer ist, ist der MF1 Wert bezogen auf das gesamte Bild 3,8% niedriger. Dies zeigt, dass vermutlich der F1 Wert einer oder mehrerer Klassen bei dem Ergebnis der **LR** deutlich kleiner ist. In Tabelle B.1 ist zu erkennen, dass der F1 Wert der Klasse *Gebäude* bei dem Ergebnis des **RF** mit 85,5% deutlich größer ist als der F1 Wert von 74,8% bei dem Ergebnis der **LR**. Besonders die Vollständigkeit

der Klasse *Gebäude* ist mit 61,4% bei dem Ergebnis der **LR** sehr gering, verglichen zu 86,7%, wenn der **RF** verwendet wird. Bei Gebiet N2 sind sowohl die GG als auch der MF1 Wert größer für das Ergebnis des **RF**.

Bei allen Vergleichen (**LR** vs. **RF**, **RLR** vs. **RRF**, **RLRK** vs. **RRFK**) sind die Vollständigkeit der Klassen *Gebäude* und *versiegelte Fläche* und die Korrektheit der Klasse *Boden* deutlich geringer für die Methoden, die auf der logistischen Regression basieren. Dies deutet darauf hin, dass mit den Methoden basierend auf der LR viele Pixel fäschlicherweise das Label *Boden* zugewiesen bekommen. Dieser Unterschied ist bei der Vollständigkeit und Korrektheit des gesamten Gebietes (B.1) und der Pixel aus Änderungsgebieten (B.2) zu erkennen. Der F1 Wert der beiden Klassen *versiegelte Fläche* und *Boden* ist demnach auch meist bei den auf dem RF basierenden Methoden größer.

Abb. 37 zeigt die Ergebnisse der **RRF**, **RRFK**, **RLR** und **RLRK** für einen Teil des Neubaugebietes in Gebiet N1. Werden zunächst die Ergebnisse der **RLR** und der **RLRK** betrachtet, ist zu erkennen, dass bereits mit der **RLR** die Detektion der neuen Gebäude problematisch ist. Besonders die Gebäudeseiten, auf denen Solarpanele angebracht wurden und die somit andere Merkmalscharakteristiken aufweisen, werden fälschlicherweise als *Boden* klassifiziert. Durch die Integration der Kartenlabels als zusätzliche Beobachtungen (**RLRK**) werden auch die restlichen Neubauten nicht mehr erkannt, obwohl die beiden Gebäude im unteren Bereich in der Initialisierung mit der **RLR** fast vollständig erkannt wurden. Der **RRF** hat deutlich weniger Probleme mit den neuen Gebäuden als die **RLR**. Bis auf die Gebäude mit komplett weißem Dach auf der rechten Seite des Bildes und einige Pixel in den Randbereichen werden die Gebäude korrekt klassifiziert. Auch bei der zusätzlichen Beobachtung der Kartenlabels im **RRFK** werden die Gebäude weiterhin als *Gebäude* erkannt.

Es kann zusammengefasst werden, dass die Klassen *versiegelte Fläche* und *Boden* meist von den Methoden, die auf dem RF basieren, besser erkannt werden als von jenen, die auf der LR basieren. Besonders die Vollständigkeit der beiden Klassen ist bei den Ergebnissen der Methoden, die auf der LR basieren, deutlich geringer.

6.3.2. Las Vegas

Abb. 44 zeigt die Gesamtgenauigkeit der Klassifikationsergebnisse unter Berücksichtigung aller Pixel für die Sensordaten von 2016 und den Karten aus den Jahren 1986, 1991 und 2000. Die GG unter Berücksichtigung der geänderten Pixel ist in Abb. 45 dargestellt. Die GG der Änderungsdetektion zeigt ein ähnliches Muster wie die GG des Klassifikationsergebnisses unter Berücksichtigung aller Pixel und wird daher nicht grafisch dargestellt. Die GG und der MF1 der drei Evaluierungskonfigurationen finden sich in Tabelle 17. Die weiteren Genauigkeitsmaße sind in den Tabellen B.4, B.5 und B.6 in Anhang B aufgeführt. In Abb. 46 sind die verschiedenen Karten und die Referenz aus 2016 abgebildet.

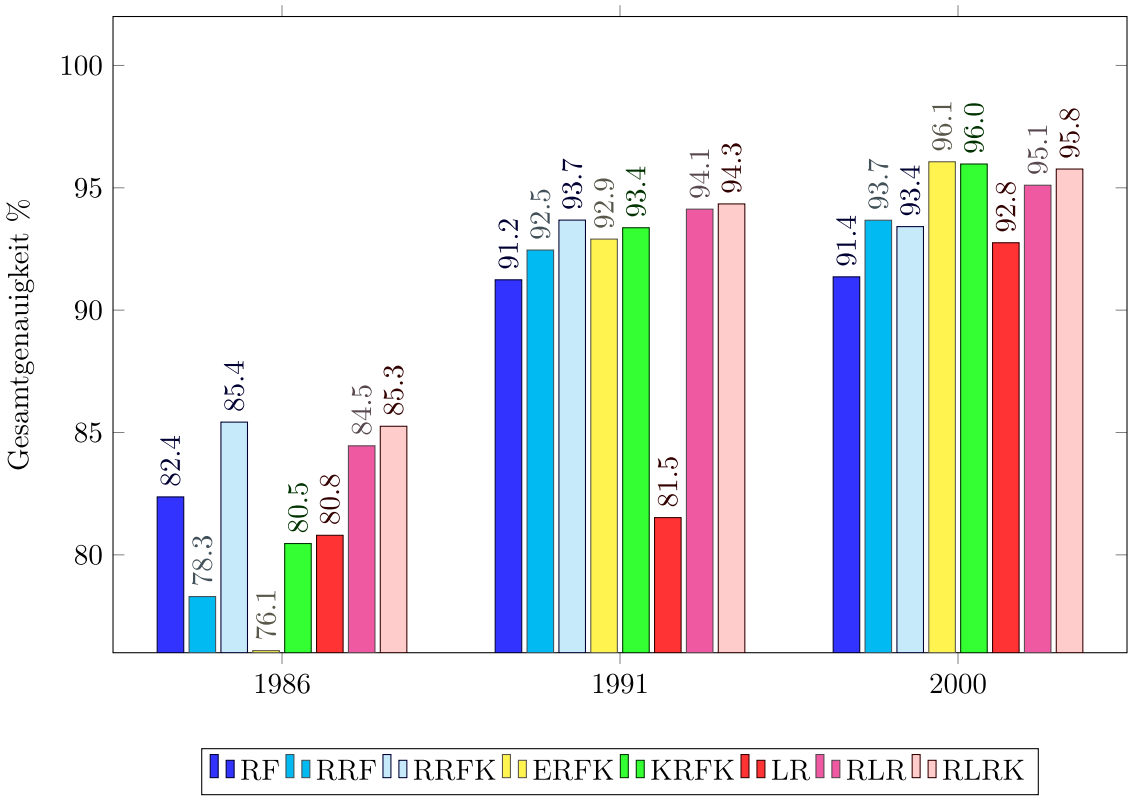


Abbildung 44: Gesamtgenauigkeit [%]. Datensatz: *Las Vegas*. Grundlage: Gesamtes Bild.

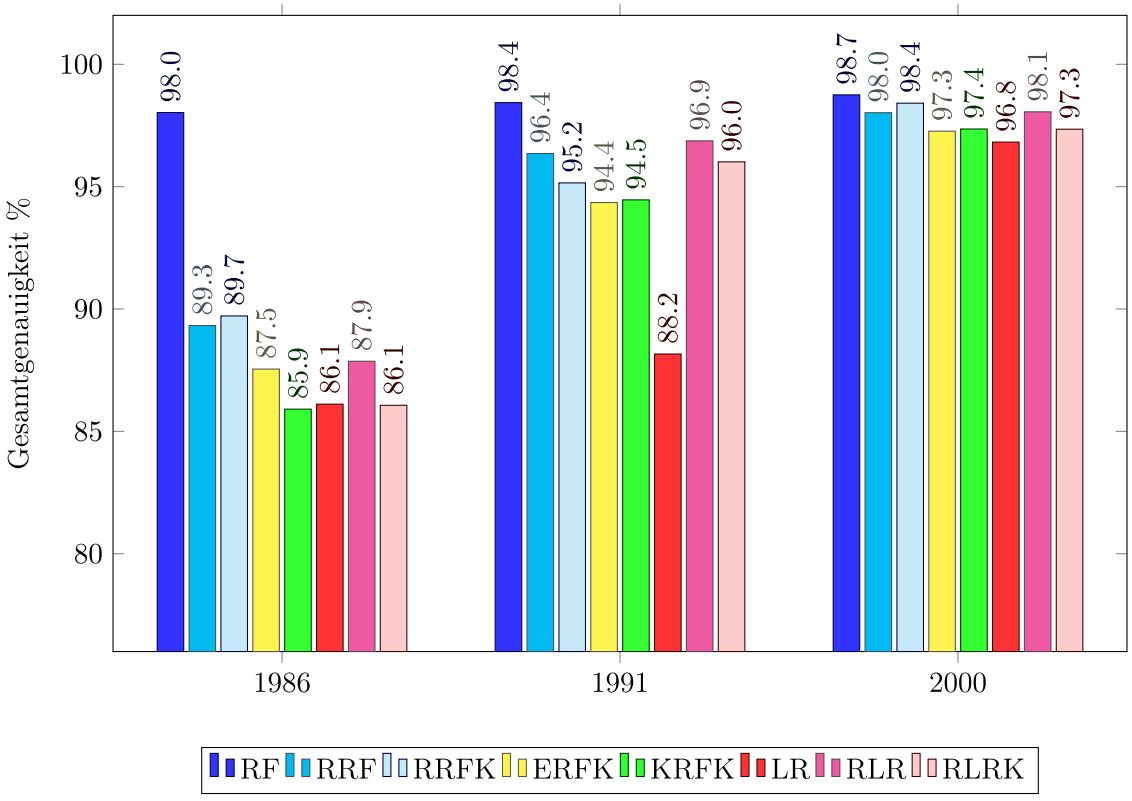
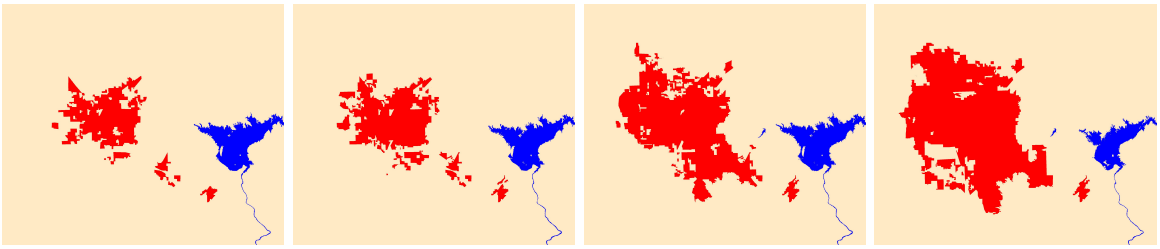


Abbildung 45: Gesamtgenauigkeit [%]. Datensatz: *Las Vegas*. Grundlage: Nur Änderungsgebiete.

Gebiet		Gesamtes Bild		Nur Änderungen		Änderungsdetektion	
		GG	MF1	GG	MF1	GG	MF1
1986	RF	82,4	**80,3**	**98,0**	**77,1**	82,8	**75,1**
	RRF	78,3	65,6	89,3	56,5	78,3	58,0
	RRFK	**85,4**	78,6	89,7	53,2	**85,5**	65,7
	ERFK	76,1	61,0	87,5	49,9	76,1	52,0
	KRFK	80,5	61,7	85,9	38,1	80,5	48,9
	LR	80,8	62,2	86,1	40,9	80,8	49,9
	RLR	84,5	73,3	87,9	48,1	84,5	59,4
	RLRK	85,3	75,1	86,1	35,7	85,3	52,6
1991	RF	91,2	86,6	**98,4**	**76,8**	91,5	84,2
	RRF	92,5	85,9	96,4	74,4	92,5	83,7
	RRFK	93,7	90,7	95,2	66,2	93,7	84,3
	ERFK	92,9	89,3	94,4	67,6	92,9	81,8
	KRFK	93,4	89,4	94,5	60,2	93,4	82,9
	LR	81,5	64,9	88,2	44,4	81,5	50,4
	RLR	94,1	**91,3**	96,9	72,0	94,2	**87,2**
	RLRK	**94,3**	91,1	96,0	62,6	**94,4**	86,7
2000	RF	91,4	87,2	**98,7**	**69,5**	91,9	78,3
	RRF	93,7	87,3	98,0	67,1	93,9	80,4
	RRFK	93,4	91,5	98,4	63,8	94,1	82,4
	ERFK	**96,1**	**93,7**	97,3	59,7	**96,2**	**84,0**
	KRFK	96,0	90,7	97,4	60,3	96,0	83,9
	LR	92,8	88,9	96,8	59,1	92,8	74,6
	RLR	95,1	92,2	98,1	64,7	95,3	83,5
	RLRK	95,8	92,6	97,3	58,9	95,9	83,6

Tabelle 17: Genauigkeitsmaße [%]. Datensatz: *Las Vegas*. Der größte Wert pro Gebiet ist fett gedruckt.



(a) Karte 1986 (b) Karte 1991 (c) Karte 2000 (d) Referenz 2016

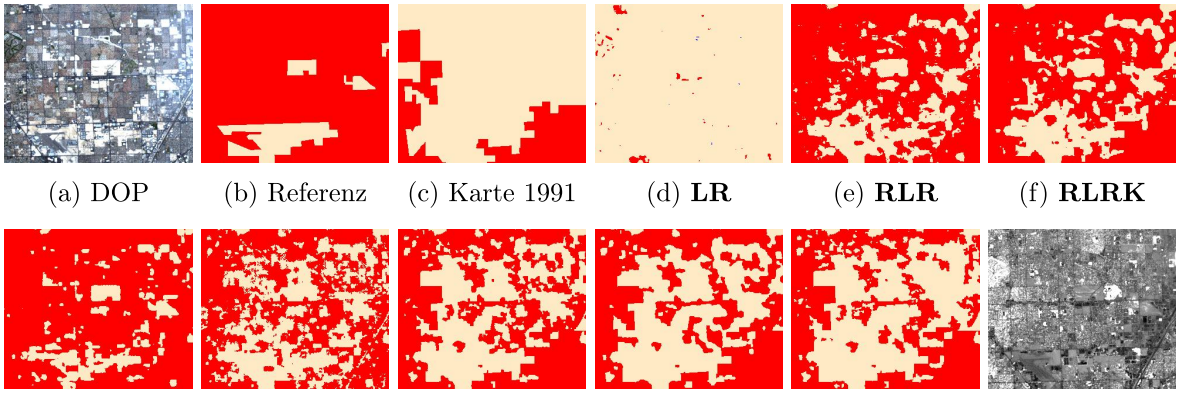
Abbildung 46: Karten und Referenz des Datensatzes *Las Vegas*. Rot: *Bebauung*. Blau: *Wasser*. Beige: *Boden*.

Frage a) Berücksichtigung der Fehler in den Trainingslabels

Zunächst wird wieder analysiert, wie sich die Berücksichtigung der Fehler im Training auf das Klassifikationsergebnis auswirkt (Frage a). Hierfür wird zunächst die GG unter Berücksichtigung des gesamten Bildes betrachtet. Unter Verwendung der Karte aus 2000 (etwa 7% Änderungen) steigt die GG um 2,3% wenn die Klassifikation nicht mit dem **RF** sondern dem **RRF** erfolgt. Auch wenn die **RLR** anstelle der **LR** genutzt wird, steigt die GG um 2,3%. Auch unter Verwendung der Karte aus 1991 (etwa 13% Änderungen) ist die GG des Klassifika-

tionsergebnisses des **RRF** um 1,3% größer als beim **RF**. Die GG der **LR** ist mit 81,5% sogar um 12,6% geringer als die GG unter Verwendung der **RLR**. Die sehr geringe Vollständigkeit des Ergebnisses der **LR** für die Klasse *Bebauung* von 6,6% zeigt, dass kaum Pixel dieser Klasse korrekt erkannt wurden (Tabelle B.4). Auch bei der ausschließlichen Evaluierung für Pixel in Änderungsgebieten ist die GG daher deutlich größer (8,7%) wenn die **RLR** verwendet wird (Abb. 45). Unter Verwendung der aktuelleren Karte von 2000 ist die Differenz der GG der **LR** und **RLR** noch 1,3%. Bei dem Vergleich der Ergebnisse des **RF** und dem **RRF** bezogen auf die Änderungsgebiete ist ein leichter Abfall der GG von 2% (Karte von 1991) bzw. 0,7% (Karte von 2000) zu beobachten, wenn der **RRF** verwendet wird. Dies liegt vor allem daran, dass bei dem **RRF** weniger Pixel der Klasse *Bebauung* zugeordnet werden. Die Vollständigkeit der Klasse *Bebauung* ist jedoch nicht nur in den Änderungsgebieten geringer bei dem **RRF** (Differenz: 19,2% für die Karte von 1991 bzw. 16,8% für die Karte von 2000), sondern auch bezogen auf das ganze Gebiet (Differenz: 15,8% für die Karte von 1991 bzw. 10,2% für die Karte von 2000). Der **RF** klassifiziert stattdessen mehr Pixel der Klasse *Boden* als *Bebauung*, was an der geringeren Korrektheit der Klasse *Bebauung* (Differenz: 14,1% für die Karte von 1991 bzw. 14,6% für die Karte von 2000) und der geringeren Vollständigkeit der Klasse *Boden* (Differenz: 5,5% für die Karte von 1991 bzw. 5,4% für die Karte von 2000) zu erkennen ist.

Abb. 47 und 48 zeigen die Klassifikationsergebnisse eines Neubaugebietes unter Verwendung der Karte aus 1991 und 2000. Hier ist sowohl zu erkennen, dass, wie bereits festgestellt wurde, die Pixel der Klasse *Bebauung* deutlich besser mit der **RLR** als mit der **LR** klassifiziert werden, als auch, dass mit dem **RRF** weniger Pixel korrekt als *Bebauung* klassifiziert werden als mit dem **RF**.

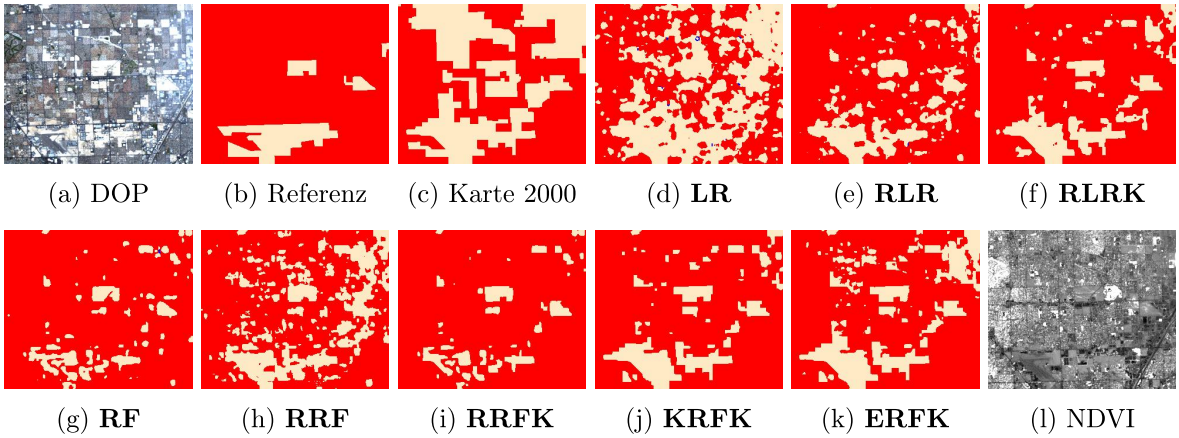


(a) DOP (b) Referenz (c) Karte 1991 (d) **LR** (e) **RLR** (f) **RLRK**

(g) **RF** (h) **RRF** (i) **RRFK** (j) **KRFK** (k) **ERFK** (l) NDVI

Abbildung 47: Ausschnitt aus dem Datensatz *Las Vegas* mit einem Neubaugebiet. Karte von 1991. Rot: *Bebauung*. Blau: *Wasser*. Beige: *Boden*.

Noch nicht betrachtet wurde zuvor die Klassifikation unter Verwendung der ältesten Karte von 1986 (etwa 15% Änderungen). Hier ist die GG des **RRF** bezogen auf das gesamte Bild um 4,1% geringer als bei dem **RF**. In den Änderungsgebieten sind es sogar 8,7%. Die schlechte GG ist besonders auf die schlechte Vollständigkeit der Klasse *Bedauung* von 27,6% zurückzuführen. Auch die Ergebnisse der Methoden **LR** und die **RLR** haben mit 2,8% und 22,9% eine sehr geringe Vollständigkeit der Klasse *Bedauung*. Bei der Methode **RF** ist die Korrektheit der Klasse *Bedauung* mit 53,6% gering. In Abb. 49 sind die Klassifikationsergebnisse unter



(a) DOP (b) Referenz (c) Karte 2000 (d) **LR** (e) **RLR** (f) **RLRK**

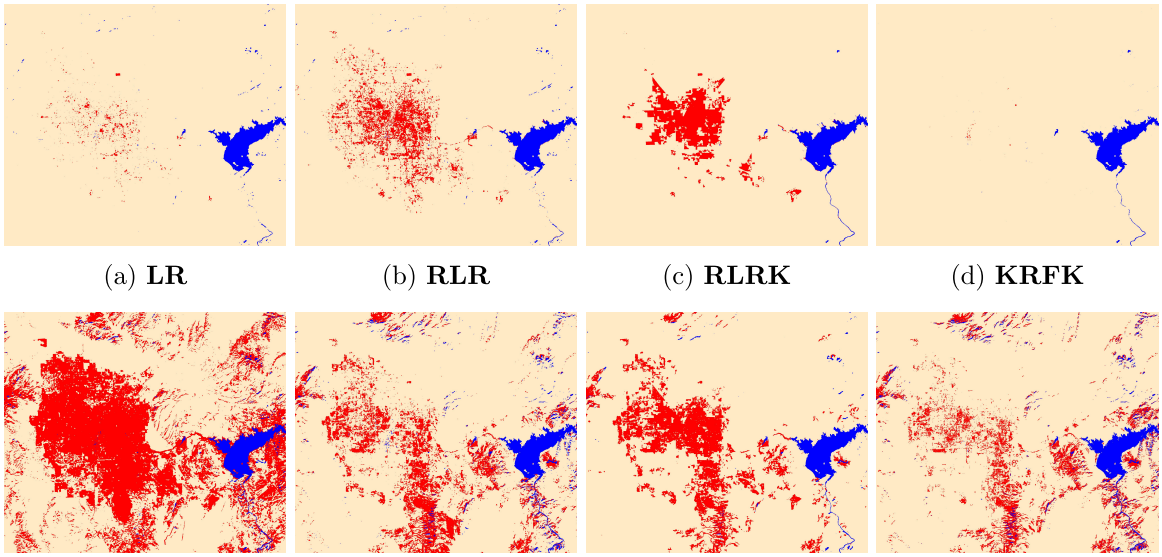
(g) **RF** (h) **RRF** (i) **RRFK** (j) **KRFK** (k) **ERFK** (l) NDVI

Abbildung 48: Ausschnitt aus dem Datensatz *Las Vegas* mit einem Neubaugebiet. Karte von 2000. Rot: *Bebauung*. Blau: *Wasser*. Beige: *Boden*.

Verwendung der Karte von 1986 abgebildet. Es ist zu erkennen, dass die **LR** fast keine Pixel als *Bebauung* klassifiziert und auch mit der **RLR** wurden größtenteils die bereits in der Karte vorhandenen Pixel aus der Innenstadt als *Bebauung* erkannt. Der **RF** detektiert den Großteil der Stadt korrekt als *Bebauung*, was auch die hohe Vollständigkeit von 97,2% dieser Klasse zeigt. Viele der Pixel mit der Klasse *Boden* werden bei dem **RF** jedoch fehlerhaft als *Bebauung* klassifiziert. Der **RRF** detektiert einige neue Pixel der Klasse *Bebauung* korrekt und klassifizieren einige Pixel der Klasse *Boden* als *Bebauung*.

Die Klassifikation der Fernerkundungsdaten von 2016 mit der Karte von 1986 stellt eine besondere Herausforderung dar. Einerseits besitzen 93,0% der Pixel, deren Klasse sich zwischen 1986 und 2016 geändert haben, im Jahr 1986 das Label *Boden* und im Jahr 2016 das Label *Bebauung*. Die Änderungen bestehen also zum größten Teil aus Neubaugebieten. Zum Anderen gibt es mehr Pixel mit der neuen Klasse *Bebauung* und der alten Klasse *Boden*, als Pixel, die sowohl in der Karte als auch in der Referenz das Label *Bebauung* aufweisen. Während nur 5,4% der Pixel des Bildes bereits 1986 der Klasse *Bebauung* zugeordnet waren, ändern sich 14,0% der Pixel des ganzen Bildes von *Boden* zu *Bebauung*. Dementsprechend gibt es mehr Trainingslabels der Klasse *Bebauung*, die fehlerhaft das Label *Boden* aufweisen. Bereits in Kapitel 1.2 wurde erwähnt, dass jede Art von Objekten zum Großteil korrekt durch die Trainingsdaten repäsentiert werden sollte, um fehlerhafte Labels als solche detektieren zu können. Dies ist hier nicht der Fall.

Zusammenfassend kann gesagt werden, dass sich das Klassifikationsergebnis unter Verwendung der Karten aus 1991 und 2000 verbessert, wenn die Fehler in den Trainingsdaten berücksichtigt werden. Nur in den Änderungsgebieten sinkt die GG leicht ab, wenn der **RRF** anstatt des **RF** verwendet wird. Wird die Karte aus 1986 verwendet, ist die Klassifikation unabhängig von dem getesteten Klassifikator problematisch. Dies liegt daran, dass die Trainingslabels der Klasse *Bebauung* zum Großteil das fehlerhafte Label *Boden* aufweisen und nicht das korrekte Label *Bebauung*.



(a) **LR** (b) **RLR** (c) **RLRK** (d) **KRFK**

(e) **RF** (f) **RRF** (g) **RRFK** (h) **ERFK**

Abbildung 49: Ergebnisse für den Datensatz *Las Vegas* unter Verwendung der Karte aus 1986. Rot: *Bebauung*. Blau: *Wasser*. Beige: *Boden*.

Frage b) Berücksichtigung der Kartenlabels als weitere Beobachtung

Bei der Integration der Kartenlabels als Beobachtung (Frage b) ist unter Verwendung der Karte von 1986 eine Verbesserung der Gesamtklassifikation zu beobachten. Die GG des Ergebnisses des **RRFK** ist 7,1% höher als die des **RRF**. Auch die GG des **RLRK** ist 0,8% höher als die der **RLR**. Die Verbesserung ist, wie in Abb. 49 zu sehen, besonders auf die Klasse *Bebauung* bezogen.

Bei dem Ergebnis der **RLRK** ist der starke Einfluss der Karte von 1986 zu erkennen, da ein Großteil der Änderungen nicht erkannt wurde. Dies ist auch bei der Evaluierung der Änderungsdetektion an der geringen Vollständigkeit der Klasse *Änderungen* von 7,5% zu erkennen (Tabelle B.6 in Anhang B). Da bei der **RLR** kaum Pixel der Klasse *Bebauung* zugeordnet wurden, erhöht sich durch den Einfluss der Karte die GG unter Berücksichtigung aller Pixel. Die GG bei ausschließlicher Beachtung der Änderungsgebiete sinkt hingegen.

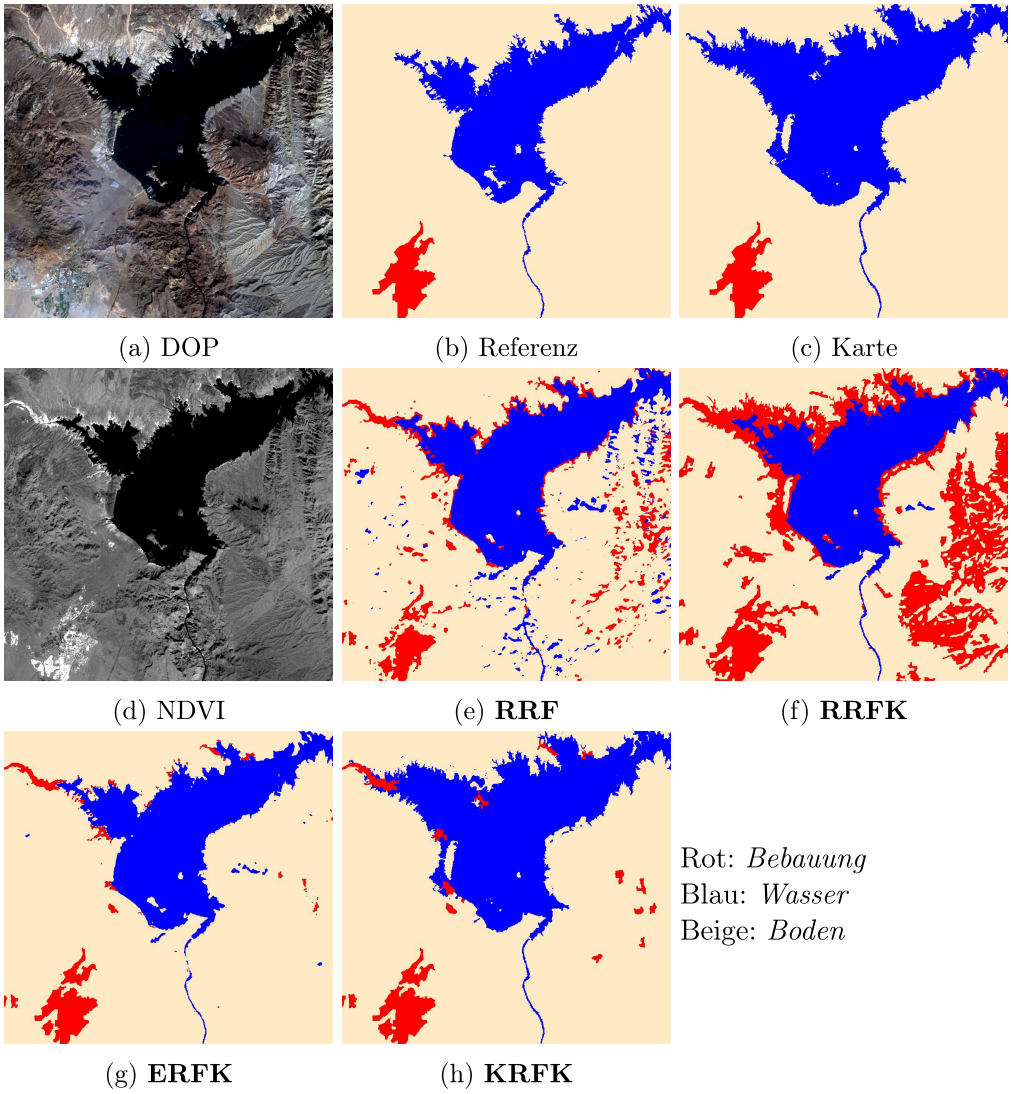
Bei der **RRFK** wirkt sich der Einfluss der Kartenlabels in der Klassifikation nicht nur in den unveränderten Gebieten, sondern auch in den Änderungsgebieten positiv aus. Die GG steigt für das gesamte Bild um 7,1%, und für die Änderungsgebiete um 0,4%, verglichen zur GG des **RRF**. Damit ist bei dem Ergebnis des **RRFK** die GG für das gesamte Bild um 3% höher als bei dem des **RF**. Dieser positive Einfluss auf das Ergebnis des **RRFK** in den Änderungsgebieten kann dadurch erklärt werden, dass sich durch die Integration der Kartenlabels das Klassifikationsergebnis des gesamten Bildes innerhalb der Iteration verbessert und somit auch die Wahrscheinlichkeiten der Klassenzugehörigkeit, mit denen das Training in der nächsten Iteration erfolgt, verbessert werden. Anders als beim **ERFK** basiert das Training nicht ausschließlich auf den Labels der vorherigen Iteration. Beim **RRFK** fließen auch die Beliefs des CRFs der vorherigen Iteration in die Wahrscheinlichkeiten der Klassenzugehörigkeit mit ein.

Das Ergebnis des **RRFK** unter Verwendung der Karte von 1991 ist, bezogen auf das gesamte Gebiet, mit einer GG von 93,7% etwas besser als das des **RRF** mit einer GG von 92,5%. In den Änderungsgebieten ist die GG des **RRFK** jedoch mit einer Differenz von 1,2% kleiner als die GG mit dem **RRF**. Auch bei dem Vergleich der Ergebnisse der **RLR** mit der **RLRK** ist die Verbesserung der GG bezogen auf das gesamte Gebiet und eine leichte Verschlechterung in den Änderungsgebieten zu beobachten, wenn das Kartenlabel in den Klassifikationsprozess integriert wird. Diese Anpassung an das Kartenlabel bei den Methoden **RLRK** und **RRFK** ist auch in dem Neubaugebiet in Abb. 47 zu erkennen.

Werden zum Training die Labels der Karte aus 2000 verwendet, erhöht sich ebenfalls die GG bezogen auf das gesamte Bild, wenn die **RLRK** anstelle der **RLR** verwendet wird. In den Änderungsgebieten sinkt auch hier die GG etwas. Der Vergleich der Ergebnisse des **RRF** und des **RRFK** zeigt hingegen ein umgekehrtes Bild. So verbessert sich die GG in den Änderungsgebieten um 0,4% bei der Hinzunahme der Kartenlabels in den Klassifikationsprozess. Die GG bezogen auf das gesamte Bild sinkt um 0,3%. Bei Betrachtung weiterer Genauingkeitsmaße ist zu erkennen, dass sich die Vollständigkeit der Klasse *Bebauung* im Ergebnis des **RRFK**, verglichen zu der im Ergebnis des **RRF**, um 11% verbessert hat, die Korrektheit jedoch um 10% niedriger geworden ist. Durch den größeren Einfluss der Karte werden demnach mehr Pixel der Klasse *Bebauung* zugeordnet. In den Änderungsgebieten, die, wie bereits angemerkt, zu einem Großteil der Klasse *Bebauung* angehören, führt dies zu einer Verbesserung der GG. Wird das gesamte Gebiet betrachtet, verschlechtert sich die GG minimal (0,3%).

Abb. 50 zeigt einige Klassifikationsergebnisse für einen Ausschnitt des *Las Vegas* Datensatzes. Hier ist deutlich zu erkennen, dass um den See herum und auf der rechten Seite unterhalb des Sees viele Pixel der Klasse *Boden* von dem **RRFK** als *Bebauung* klassifiziert wurden. Diese Pixel scheinen demnach ähnliche Merkmale aufzuweisen wie die Pixel der Klasse *Bebauung*. Es ist auch zu erkennen, dass die vielen kleineren Segmente an Pixeln der Klasse *Boden*, die in dem Ergebnis des **RRF** das Label *Wasser* aufweisen, in dem Ergebnis des **RRFK** korrekt der Klasse *Boden* zugeordnet wurden. Diese Segmente liegen meist in Schatten, die durch starke Höhenunterschiede im Gelände, z.B. aufgrund von Dünen, entstehen. Da Wasser ebenfalls durch sehr niedrige Grauwerte charakterisiert wird, wird Schatten oft fehlerhaft der Klasse *Wasser* zugewiesen. Daher ist im Ergebnis des **RRFK** die Korrektheit der Klasse *Wasser* mit 88,5% bezogen auf das gesamte Bild deutlich größer als die des Ergebnisses des **RRF** mit 66,5%.

Zusammenfassend kann gesagt werden, dass die zusätzliche Beobachtung der Kartenlabels in der Klassifikation bei der Methode basierend auf der robusten logistischen Regression zu einer Annäherung der Klassifikationsergebnisse an die Labels der veralteten Karte führt. Demnach steigt die GG, wenn alle Pixel berücksichtigt werden, und sinkt, wenn nur die Pixel in den Änderungsgebieten für die Evaluierung betrachtet werden. Bei der Methode basierend auf dem RRF ist dies ebenfalls zu beobachten, wenn zum Training die Labels der Karte aus dem Jahr 1991 verwendet werden. Bei der Verwendung der stark veralteten Karte von 1986 verbessert sich zusätzlich die GG in den Änderungsgebieten, wenn die Labels der veralteten Karte in der Klassifikation verwendet werden.



(a) DOP (b) Referenz (c) Karte

(d) NDVI (e) **RRF** (f) **RRFK**

(g) **ERFK** (h) **KRFK**

Abbildung 50: Ausschnitt aus *Las Vegas* mit einem Seegebiet. Karte von 2000.

Frage c) Verwendung der Wahrscheinlichkeiten der Klassenzugehörigkeit

Als nächstes wird das Klassifikationsergebnis des **RRFK** mit denen des **ERFK** und **KRFK** verglichen, um zu analysieren, ob die Erweiterung des RF, dass jedem Trainingsbeispiel alle Klassen mit einer bestimmten Wahrscheinlichkeit zugeordnet werden, das Klassifikationsergebnis verbessert (Frage c). Wird die Karte von 1986 zum Training verwendet, ist die GG des gesamten Bildes von **RRFK** deutlich größer als die GG des Ergebnisses der Methoden **ERFK** (Differenz: 9,3%) und **KRFK** (Differenz: 4,9%). Auch die GG in den Änderungsgebieten ist bei dem **RRFK** um 2,2% größer als die beim **ERFK** und um 3,8% größer als beim **KRFK**. Besonders bei der Klassifikation mit dem **KRFK** wurde kaum ein Pixel der Klasse *Bebauung* als solches klassifiziert, wie in Abb. 49 zu erkennen ist. Auch das Ergebnis des **ERFK** (Abb. 49) beinhaltet viele fehlklassifizierte Pixel, die entweder in der Referenz oder in der Karte der Klasse *Bebauung* zugeordnet sind. Dies ist auch an der geringen Vollständigkeit von 14,9% und Korrektheit von 31,8% der Klasse *Bebauung* für das gesamte Ergebnis des **ERFK** zu erkennen.

Unter Verwendung der aktuelleren Karten von 1991 und 2000 zeigt das Ergebnis des **RRFK**, wenn nur die Änderungen betrachtet werden, ebenfalls eine höhere GG als die Ergebnisse der beiden Methoden **ERFK** und **KRFK**. Der Unterschied liegt hierbei zwischen 0,7% und 1,1%, abhängig davon welche Karte genutzt wird (aus 1991 oder 2000) und mit welcher Methode verglichen wird (**ERFK** oder **KRFK**). Auch in den Neubaugebieten in Abb. 47 und Abb. 48 ist zu erkennen, dass die Methode **RRFK** im Vergleich zu den Methoden **ERFK** und **KRFK** bessere Ergebnisse liefert.

Bezogen auf das gesamte Gebiet ist die GG des **RRFK** nur unter Verwendung der Karte von 1991 leicht größer als die des **ERFK** und **KRFK**. Wird die Karte von 2000 verwendet, ist die GG des **RRFK** mit 93,4% kleiner als die der **ERFK** mit 96,1% und die der **KRFK** mit 96,0%. Wie bereits beschrieben, wurden viele Pixel der Klasse *Boden* durch die Methode **RRFK** fehlerhaft der Klasse *Bebauung* zugeordnet. Bei den Methoden **ERFK** und **KRFK** wurden deutlich weniger Pixel der Klasse *Boden* als *Bebauung* klassifiziert. Dies ist auch in dem Ausschnitt in Abb. 50 zu erkennen. Die Korrektheit der Klasse *Bebauung* ist bei dem Ergebnis des **ERFK** mit 93,9% und bei dem Ergebnis des **KRFK** mit 92,7% deutlich größer als die jene bei dem Ergebnis des **RRFK** mit 76,8%.

Bei dem Vergleich der Ergebnisse des **RRFK** und des **KRFK** fällt auf, dass die Korrektheit der Klasse *Wasser* um 16,9% kleiner ist, wenn der **KRFK** anstelle des **RRFK** verwendet wird. Dies kann an dem starken Einfluss der Kartenlabels bei dem **KRFK** liegen. Am Rand des Sees gibt es einige Pixel der Klasse *Boden*, die in der Karte das Label *Wasser* besitzen (Abb. 50). Der See ist demnach in den Aufnahmen 2016 etwas kleiner als in denen aus 2000. Auch die Merkmale der Pixel in diesem Bereich unterscheiden sich etwas von den Merkmalen der anderen Pixel der Klasse *Boden*, beispielsweise durch einen höheren NDVI (Abb. 50). Durch den starken Einfluss der Karte und vermutlich auch durch eine gewisse Unsicherheit bei der Zuordnung basierend auf den untypischen Merkmalen dieser Pixel hat der **KRFK** den Großteil der geänderten Pixel weiterhin als *Wasser* klassifiziert. Der **RRFK** hingegen hat die Pixel am Rand des Sees fäschlicherweise als *Bebauung* klassifiziert. Diese Art der Fehlklassifikation am Rand des Sees tritt auch bei dem Ergebnis anderer Klassifikatoren auf. Besonders an der Mündung werden einige Pixel mit der Klasse *Boden* als *Bebauung* klassifiziert. Erklärt kann dies durch den hohen NDVI Wert. Da Vegetation in der Wüste, die die Stadt Las Vegas umgibt und der Klasse *Boden* zugeordnet ist, im Allgemeinen nicht vorkommt, ist der NDVI bei den meisten Pixeln der Klasse *Boden* relativ gering. In den Bereichen der Klasse *Bebauung* ist jedoch oft auch Vegetation abgebildet (s. Abb. 47), sodass diese Fehlklassifikation am Rand des Sees zu erwarten war.

Zusammenfassend kann gesagt werden, dass die GG der Ergebnisse des **RRFK** größer ist als jene bei den Methoden **ERFK** und **KRFK**, wenn die älteren Karten aus 1986 und 1991 verwendet werden. In den Änderungsgebieten ist die GG bei allen Experimenten etwas größer, wenn der **RRFK** genutzt wird und nicht der **ERFK** oder **KRFK**.

Frage d) Vergleich mit der robusten logistischen Regression

Zur Beantwortung der Frage d werden die Ergebnisse der Methoden basierend auf dem RRF (**RRF** bzw. **RRFK**) mit denen der Methoden basierend auf der RLR (**RLR** bzw. **RLRK**) verglichen. Anders als bei dem Datensatz *Hameln* sind die Ergebnisse der **RLR** und der **RLRK** hier sehr gut. Unter Verwendung der Karten von 1991 und 2000 ist die GG des gesamten Bildes etwas größer als die der Ergebnisse des **RRF** und **RRFK**. Werden nur die Pixel aus den Änderungsgebieten betrachtet, ist die GG des **RRF** um 0,5% niedriger als die der **RLR** unter Verwendung der Karte von 1991 und um 0,1% niedriger für die Karte von 2000. Während unter Verwendung der Karte aus 1991 auch die GG der Pixel in den Änderungsgebieten bei dem **RRFK** um 0,8% kleiner ist als bei der **RLRK**, ist die GG bei der aktuelleren Karte von 2000 um 1,1% größer. Bei der Verwendung der Karte von 1986 ist die Differenz der GG des gesamten Bildes mit 0,1% relativ gering, wenn das Kartenlabel bei der Klassifikation berücksichtigt wird (**RRFK**, **RLRK**). Mit der **RLRK** wird jedoch, wie bereits erwähnt, kaum eine Änderung erkannt (Abb. 49). Obwohl auch mit der **RRFK** nur wenige Änderungen erkannt werden, ist die Vollständikeit der *Änderung* von 31,8% größer verglichen zu der Vollständigkeit von 7,5% der **RLRK**.

Zusammenfassend ergeben sich in *Las Vegas* aus den Methoden basierend auf der robusten logistischen Regression, verglichen zu den Methoden basierend auf dem robusten Random Forest, sehr gute Ergebnisse.

Frage e) Einfluss der Anzahl an fehlerhaften Labels

Als letztes wird untersucht, in welchem Rahmen die Anzahl der Fehler in den Trainingslabels das Klassifikationsergebnis beeinflusst (Frage e). Unabhängig von dem Klassifikator kann beobachtet werden, dass die GG des gesamten Bildes mit sinkender Anzahl von Änderungen steigt. Einzige Ausnahme bildet der **RRFK**. Hier sinkt die GG um 0,3% wenn die aktuellere Karte aus 2000 anstatt der aus 1991 verwendet wird. Dies liegt, wie bereits beschrieben, an der fehlerhaften Klassifikation einiger Pixel mit der wahren Klasse *Boden* als *Bebauung* (s. Abb. 50). Der MF1 Wert des gesamten Bildes vergrößert sich ebenfalls mit sinkender Anzahl von Änderungen bei allen Klassifikationsmethoden. Auch bei dem **RRFK** steigt der MF1 Wert von 90,7% für die Karte aus 1991 auf 91,5% für die Karte aus 2000. Auch wenn ausschließlich die Pixel in den Änderungsgebieten betrachtet werden, steigt die GG mit sinkender Anzahl von Fehler in den Trainingslabels unabhängig von dem Klassifikator.

Treten zu viele Fehler auf, wie bei der Klassifikation mit der Karte aus 1986, ist das Ergebnis aller Klassifikatoren relativ ungenau. Die Detektion der falschen Labels ist besonders fehleranfällig, wenn, wie in diesem Fall, einige Trainingsbeispiele eine Ballung im Merkmalsraum bilden, in der mehr Trainingsbeispiele mit falschem Label, als mit korrektem Label vorkommen. Dennoch ist zu anzumerken, dass die GG der neu entwickelten Methode **RRFK** bezogen auf das gesamte Gebiet hier höher ist als die GG aller anderen Methoden.

Zusammenfassend ist bei allen Methoden eine Verringerung des MF1 Wertes bei steigender Anzahl an Fehlern in den Labels zu beobachten.

6.3.3. Vaihingen

Der Datensatz *Vaihingen* besteht aus 10 verschiedenen Gebieten. Da eine individuelle Evaluierung aller Gebiete den Rahmen dieser Arbeit überschreiten würde, wird für die Beantwortung der Fragen der Median der GG über alle Gebiete betrachtet. Im Anschluss werden drei Gebiete beispielhaft genauer beleuchtet.

Abb. 51 zeigt den Median der GG über alle Gebiete des *Vaihingen* Datensatzes. Der Median der GG, wenn ausschließlich die Pixel in Änderungsgebieten betrachtet werden, ist in Abb. 52 dargestellt. Der Median der GG der Änderungsdetektion ist durchschnittlich etwas (1,7%) größer als der Median der GG des Klassifikationsergebnisses unter Berücksichtigung aller Pixel. Beide zeigen, wie bereits bei den Datensätzen *Hameln* und *Las Vegas*, ein ähnliches Muster, sodass der Median der GG der Änderungsdetektion nicht grafisch dargestellt wird. Den Median der GG und des MF1 Wertes sowohl für die Klassifikation unter Berücksichtigung aller Pixel und unter ausschließlicher Berücksichtigung der Pixel in Änderungsgebieten, als auch für die Änderungsdetektion sind in Tabelle 16 zu sehen. Die weiteren Genauigkeitsmaße finden sich in den Tabellen B.7, B.8, B.9, B.10, B.11 und B.12 in Anhang B. Für eine bessere Lesbarkeit wird in der folgenden Analyse nicht mehr erwähnt, dass es sich bei den Genauigkeitsmaßen um den Median über alle Gebiete handelt.

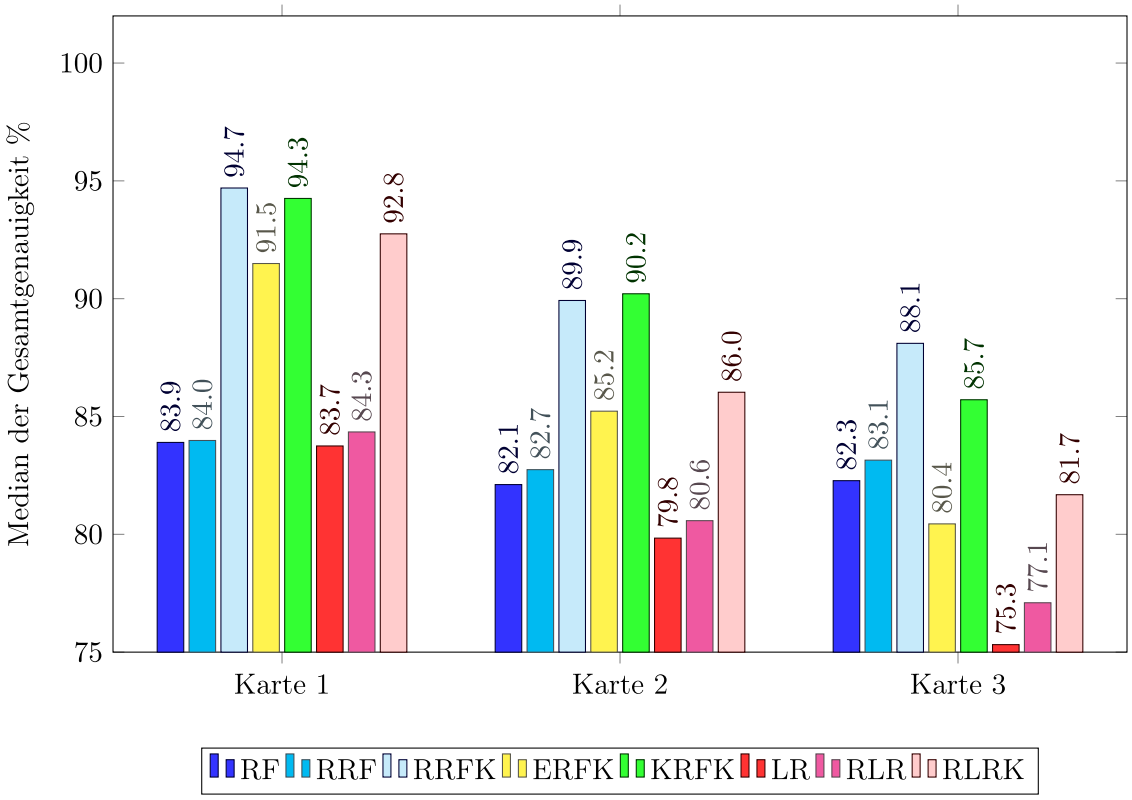


Abbildung 51: Median der Gesamtgenauigkeit über alle Gebiete [%]. Datensatz: *Vaihingen*. Grundlage: Gesamtes Bild.

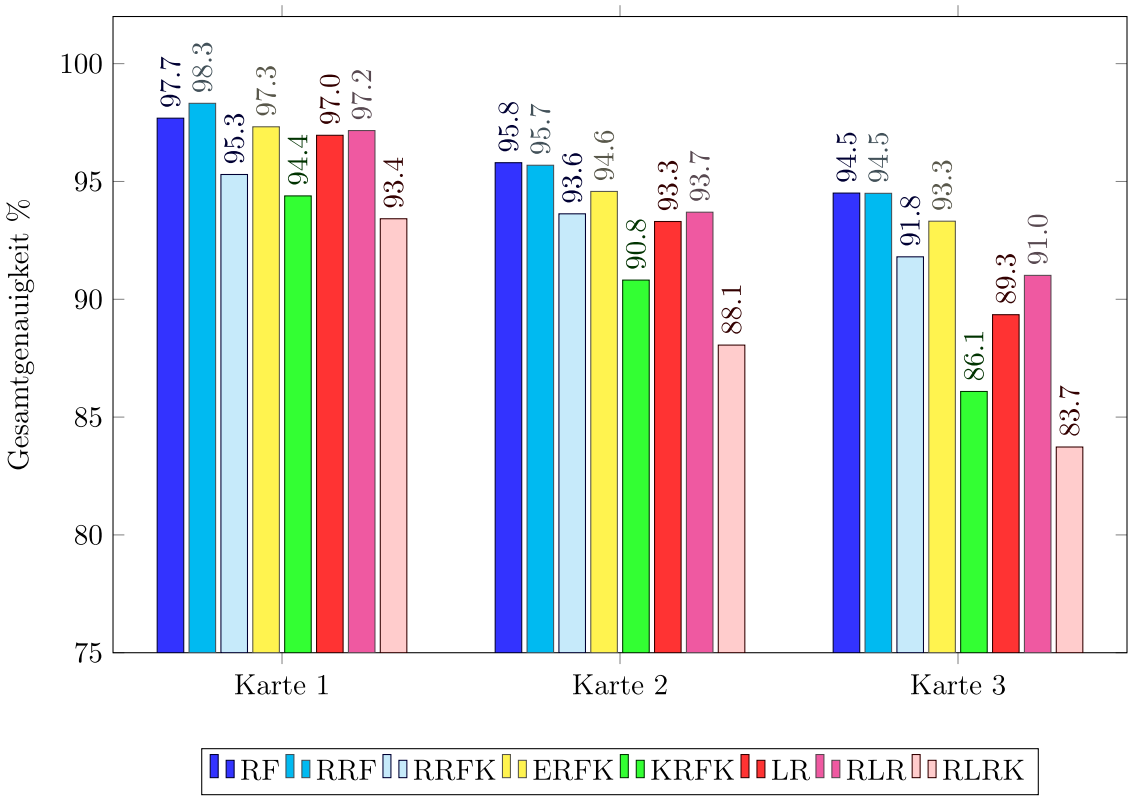


Abbildung 52: Gesamtgenauigkeit [%]. Datensatz: *Vaihingen*. Grundlage: Nur Änderungsgebiete.

Frage a) Berücksichtigung der Fehler in den Trainingslabels

Zunächst wird wieder die Auswirkung der Berücksichtigung der Fehler im Training auf das Klassifikationsergebnis untersucht (Frage a). Die GG bezogen auf das gesamte Bild ist, unabhängig von der verwendeten Karte, etwas höher, wenn die Fehler im Training berücksichtigt werden. So steigt die GG um 0,1% (Karte 1), 0,6% (Karte 2) und 0,8% (Karte 3), wenn die Klassifikation mit dem **RRF** anstatt des **RF** erfolgt. Bei der Verwendung der **RLR** anstelle der **LR** steigt die GG um 0,6% (Karte 1), 0,8% (Karte 2) und 1,8% (Karte 3). In den Änderungsgebieten ist die GG bezogen auf die **RLR** ebenfalls etwas höher als die GG bezogen auf die **LR**. Bei den RF basierten Methoden lässt sich in den Änderungsgebieten nur unter Verwendung der Karte 1 (etwa 12% Änderungen) eine leichte Verbesserung von 0,6% unter Berücksichtigung der Labelfehler beobachten. Wird mit den Labels der Karten 2 (etwa 20% Änderungen) oder 3 (etwa 26% Änderungen) trainiert, ist die GG etwa gleich (Differenz: maximal 0,1%).

Zusammenfassend ist eine Verbesserung der GG bezogen auf das gesamte Bild durch die Berücksichtigung der Fehler im Training zu beobachten.

Frage b) Berücksichtigung der Kartenlabels als weitere Beobachtung

Bei der zusätzlichen Berücksichtigung der Kartenlabels in der Klassifikation (Frage b) steigt die GG des gesamten Bildes stark an. Die Differenzen der GG werden in Tabelle 19 präsentiert.

Gebiet		Gesamtes Bild		Nur Änderungen		Änderungsdetektion	
		GG	MF1	GG	MF1	GG	MF1
Karte 1	RF	83,9	81,3	97,7	74,2	84,6	73,0
	RRF	84,0	81,1	**98,3**	**74,9**	84,9	73,4
	RRFK	**94,7**	**94,0**	95,3	58,5	**95,1**	**85,7**
	ERFK	91,5	89,9	97,3	65,7	92,2	78,6
	KRFK	94,3	93,5	94,4	46,4	94,7	82,9
	LR	83,7	82,3	97,0	72,3	85,1	72,9
	RLR	84,3	83,2	97,2	72,3	85,3	73,2
	RLRK	92,8	92,5	93,4	52,5	93,3	81,3
Karte 2	RF	82,1	79,7	**95,8**	**77,0**	83,9	77,6
	RRF	82,7	79,9	95,7	76,0	84,7	78,2
	RRFK	89,9	89,9	93,6	65,9	**92,9**	**88,5**
	ERFK	85,2	83,7	94,6	71,5	87,8	79,9
	KRFK	**90,2**	**90,6**	90,8	56,8	91,3	85,7
	LR	79,8	78,2	93,3	67,9	81,9	75,7
	RLR	80,6	79,3	93,7	68,9	82,5	75,6
	RLRK	86,0	86,7	88,1	45,6	87,3	76,7
Karte 3	RF	82,3	79,0	**94,5**	73,1	84,2	80,1
	RRF	83,1	80,1	**94,5**	**75,7**	84,6	81,0
	RRFK	**88,1**	**87,1**	91,8	63,5	**90,6**	**86,4**
	ERFK	80,4	79,8	93,3	70,5	83,9	80,2
	KRFK	85,7	84,6	86,1	47,4	86,8	76,2
	LR	75,3	74,2	89,3	67,1	78,9	76,9
	RLR	77,1	77,4	91,0	66,7	80,2	77,3
	RLRK	81,7	81,9	83,7	43,4	83,8	75,4

Tabelle 18: Genauigkeitsmaße [%]. Datensatz: *Vaihingen*. Der größte Wert pro Gebiet ist fett gedruckt.

Ein Vergleich der GG der Ergebnisse der Klassifikatoren **RRF** und **RRFK** zeigen Differenzen von 10,7% (Karte 1), 7,2% (Karte 2) und 5,0% (Karte 3). Die Ergebnisse der **RLR** und **RLRK** unterscheiden sich mit einer Differenz der GG von 8,5% (Karte 1), 5,4% (Karte 2), 4,6% (Karte 3) stark, wobei auch hier die **RLRK** die höhere GG aufweist. Der MF1 Wert bezogen auf das gesamte Bild steigt ebenfalls, wenn die Kartenlabels in der Klassifikation berücksichtigt werden. Werden nur die Änderungsgebiete betrachtet, sinkt der MF1 Wert, wie auch die GG. Die Unterschiede der MF1 Werte (Tabelle. 20) sind jedoch meist deutlich größer als die Unterschieden der GG. Besonders die Differenz des MF1 Wert unter alleiniger Verwendung der Pixel in den Änderungsgebieten liegt bei den Methoden der logistischen Regression über 20%, wenn die Karten mit einem hohen Anteil an Änderungen (Karte 2, Karte 3) genutzt werden.

Werden nur die Änderungsgebiete bei der Berechnung der GG berücksichtigt, sinkt die GG bei der Hinzunahme des Kartenlabels in die Klassifikation. Bei der **RLRK** ist die GG um 3,8% (Karte 1), 5,6% (Karte 2) und 7,3% (Karte 3) kleiner als bei der **RLR**. Wird der **RRFK** verwendet ist die GG nur um 3,0% (Karte 1), 2,1% (Karte 2) und 2,7% (Karte 3) geringer als bei dem **RRF** (s. Tabelle 19). Diese Differenz ist verglichen zu der prozentualen Anzahl an Pixeln, deren Label aus der Karte sich von dem aus der Referenz unterscheidet,

Differenz	Karte	Gesamtes Bild	Nur Änderungsgebiete
GG(**RRFK**) - GG(**RRF**)	Karte 1	10,7	-3,0
	Karte 2	7,2	-2,1
	Karte 3	5,0	-2,7
GG(**RLRK**) - GG(**RLR**)	Karte 1	8,5	-3,8
	Karte 2	5,4	-5,6
	Karte 3	4,6	-7,3

Tabelle 19: Differenz in % zwischen der GG des Ergebnisses mit zusätzlicher Beobachtung des Kartenlabels in der Klassifikation (GG(**RRFK**), GG(**RLRK**)) und der GG des Ergebnisses ohne Beobachtung des Kartenlabels in der Klassifikation (GG(**RRF**), GG(**RLR**)). Datensatz *Vaihingen*.

Differenz	Karte	Gesamtes Bild	Nur Änderungsgebiete
MF1(**RRFK**) - MF1(**RRF**)	Karte 1	12,9	-16,4
	Karte 2	10,0	-10,1
	Karte 3	7,0	-12,2
MF1(**RLRK**) - MF1(**RLR**)	Karte 1	9,3	-19,8
	Karte 2	7,4	-23,3
	Karte 3	4,5	-23,3

Tabelle 20: Differenz in % zwischen dem MF1 des Ergebnisses mit zusätzlicher Beobachtung des Kartenlabels in der Klassifikation (MF1(**RRFK**), MF1(**RLRK**)) und der MF1 des Ergebnisses ohne Beobachtung des Kartenlabels in der Klassifikation (MF1(**RRF**), MF1(**RLR**)). Datensatz *Vaihingen*.

mit 12% für Karte 1, 20% für Karte 2 und 26% für Karte 3, sehr klein. Eine leichte Reduktion der Genauigkeit in den Änderungsgebieten, in denen die Kartenlabels nicht korrekt sind, kann außerdem erwartet werden, wenn der Einfluss des Kartenlabels auf das Klassifikationsergebnis hinzukommt.

Es kann zusammengefasst werden, dass die GG bezogen auf das gesamte Bild durch die Berücksichtigung der Labels aus der Karte in der Klassifikation stark ansteigt. Werden nur die Pixel in den Änderungsgebieten betrachtet, ist zwar eine Reduktion der GG zu beobachten, jedoch ist diese im Vergleich zu dem Anteil an Pixeln mit fehlerhaftem Kartenlabel sehr klein.

Frage c) Verwendung der Wahrscheinlichkeiten der Klassenzugehörigkeit

Für die Beantwortung von Frage c, welche sich mit den Auswirkungen der Wahrscheinlichkeiten der Klassenzugehörigkeit beschäftigt, werden die Ergebnisse des **RRFK** mit denen des **ERFK** und des **KRFK** verglichen.

Bei dem Vergleich der GG bezogen auf das gesamte Gebiet unter Verwendung des **RRFK** und des **ERFK** ist zu erkennen, dass die GG von **ERFK** mit einer Differenz von 3,2% (Karte 1), 4,7% (Karte 2) und 7,7% (Karte 3) deutlich niedriger ist als als jene von **RRFK**. Werden nur die Pixel in Änderungsgebieten betrachtet, ist die GG mit einer Differenz von 2,0% (Karte 1), 1,0% (Karte 2) und 1,5% (Karte 3) etwas größer unter Verwendung des **ERFK**.

Dies lässt auf einen geringeren Einfluss der Kartenlabels innerhalb des **ERFK** schließen. Da die Kartenlabels für die Berechnung der Wahrscheinlichkeiten der Klassenzugehörigkeit innerhalb des **RRFK** benötigt werden, fließen sie indirekt auch in das Training jeder Iteration mit ein. Bei dem **ERFK** basiert das Training hingegen auf dem Klassifikationsergebnis der vorherigen Iteration. Die Kartenlabels werden nur in der ersten Iteration verwendet. Der geringere Einfluss des Kartenlabels ist daher nachvollziehbar.

Bei dem **KRFK** werden lediglich die Labels der Karte zum Training jeder Iteration verwendet. In die Ermittlung der Wahrscheinlichkeiten der Klassenzugehörigkeit für das Training des **RRFK** fließen auch andere Variablen, wie z.B. der Belief der vorherigen Iteration, mit ein. Der Einfluss der Kartenlabels ist demnach bei dem **KRFK** größer als bei dem **RRFK**. In den Experimenten ist dies an der geringeren GG in den Änderungsgebieten unter Verwendung der **KRFK** zu erkennen. Besonders bei Karte 3 mit etwa 26% Änderungen ist die Differenz mit 5,7% relativ hoch. Werden alle Pixel bei der Berechnung der GG berücksichtigt, ist das Ergebnis der **RRFK** unter Verwendung der Karte 1 (etwa 12% Änderungen) mit einer Differenz von 0,4% der GG etwas besser und unter Verwendung der Karte 2 (etwa 20% Änderungen) mit einer Differenz von 0,3% der GG etwas schlechter als das des **KRFK**. Wird die Karte 3 (etwa 26% Änderungen) genutzt, ist die GG des **RRFK** mit einer Differenz von 2,4% höher als bei dem **KRFK**.

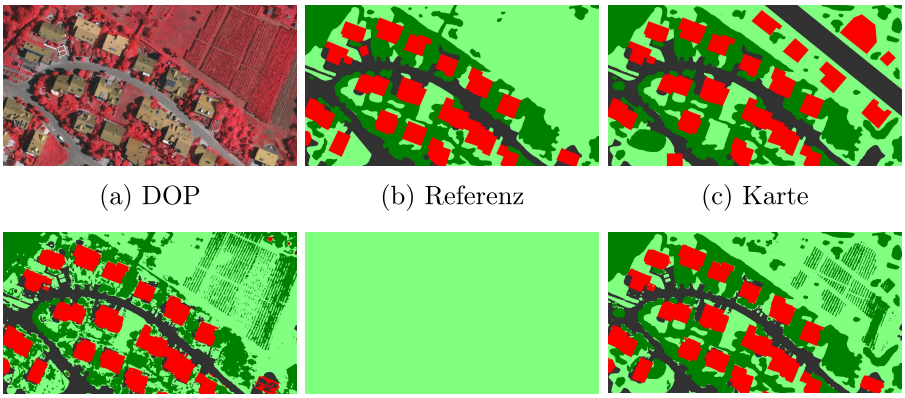
Bei den Methoden **ERFK** und **KRFK** konnte zusätzlich die Problematik beobachtet werden, dass in einigen Experimenten eine Klasse stark unterdrückt wird, sodass deren Korrektheit oder Vollständigkeit und somit der F1 Wert sehr klein ist. Tabelle 21 zeigt den minimalen F1 Wert des gesamten Gebietes über alle Experimente von *Vaihingen*. Bei dem **RRFK** ist der minimale F1 Wert immer über 70%.

	Gebäude	*Baum*	*niedr. Veg.*	*vers. Fläche*
RRFK	81,2	70,5	73,6	83,4
ERFK	0,0	0,0	0,0	0,0
KRFK	26,8	0,0	61,4	61,6
RF	78,4	65,1	56,4	72,2
RRF	80,6	63,1	55,3	74,0
RLRK	61,7	66,1	68,9	69,9
RLR	52,5	58,7	51,1	65,9
LR	48,9	54,4	50,2	64,9

Tabelle 21: Minimaler F1 Wert von allen 10 Gebieten mit allen drei Karten (demnach aller 30 Experimente) des Datensatzes *Vaihingen* [%]. Grundlage: Gesamtes Bild.

Ein F1 Wert von 0 entstand bei der Klassifikation der Gebiete A13 und A17. Bei dem Gebiet A13, unter Verwendung der Karten mit einer großen Anzahl an Fehlern (Karte 2 und Karte 3), wird mit dem **KRFK** die Klasse *Baum* größtenteils als *niedrige Vegetation* klassifiziert. Mit dem **ERFK** wird die Klasse *niedrige Vegetation* hingegen größtenteils als *Baum* klassifiziert. Das Gebiet wird am Ende dieses Kapitels nochmal genauer erläutert (s. Abb. 56). Bei Gebiet A17 wird unter Verwendung der Karte 1 mit wenig Fehlern das gesamte Gebiet vollständig der Klassen *niedrige Vegetation* zugeordnet, sodass der MF1 Wert der anderen Klassen 0 entspricht. Da bei dem **ERFK** die Labels der vorherigen Iteration für das Training verwendet

werden, kann sich die Grenze im Merkmalsraum mit jeder Iteration weiter von der optimalen Grenze entfernen. Dies passiert auch bei Gebiet A17 unter Verwendung der Karte 1. Der MF1 Wert ist hier zunächst bei 80.6% und sinkt mit jeder Iteration. Bei der 15. Iteration wird die Klasse *Baum* bereits nicht mehr erkannt und bei der 20. Iteration werden auch die Pixel der Klassen *Gebäude* und *versiegelte Fläche* nicht mehr korrekt zugeordnet. Wie in Abb. 53 zu erkennen ist, ist die Klassifikation von Gebiet 17 besonders durch den Weinberg auf der rechten Seite des Bildes problematisch. Dieser hat Merkmale der Klassen *Baum* und *niedrige Vegetation* und einige der Pixel besitzen zusätzlich ein fehlerhaftes Label der Klassen *Gebäude*, *Baum* oder *versiegelte Fläche*.



(a) DOP (b) Referenz (c) Karte

(d) **RF** (e) **ERFK** (f) **RRFK**

Abbildung 53: Gebiet A17 aus *Vaihingen*. Karte 1 (19.5% Änderungen). Rot: *Gebäude*. Grau: *versiegelte Fläche*. Hellgrün: *niedrige Veg.*. Dunkelgrün: *Baum*.

Zusammenfassend lässt sich festhalten, dass größtenteils das Ergebnis des **RRFK** verglichen mit dem des **ERFK** und dem des **KRFK**, sowohl unter Betrachtung aller Pixel, als auch unter ausschließlicher Betrachtung der Pixel in den Änderungsgebieten, eine relativ große GG aufweist. Obwohl die GG bei dem Ergebnis des **ERFK** in den Änderungsgebieten aufgrund des geringeren Einflusses der Kartenlabels etwas größer ist als bei dem Ergebnis des **RRFK**, ist die GG doch deutlich kleiner, wenn alle Pixel im Bild bei der Evaluierung betrachtet werden. Bei dem Ergebnis des **KRFK** ist der Einfluss des Kartenlabels wiederum stärker zu beobachten, sodass die GG in den Änderungsgebieten kleiner ist als bei dem Ergebnis des **RRFK**.

Frage d) Vergleich mit der robusten logistischen Regression

Ein Vergleich der Ergebnisse in Hinblick auf die unterschiedlichen Basisklassifikatoren (Frage d) zeigt, dass die Klassifikation basierend auf dem RRF zu höheren Genauigkeiten, besonders in den Änderungsgebieten, führt. Wird zunächst die GG des gesamten Gebietes betrachtet, ist die GG des **RRF** verglichen zur **RLR** noch um 0,3% schlechter, wenn die Karte 1 (etwa 12% Änderungen) verwendet wird. Besitzen mehr Trainingsbeispiele ein fehlerhaftes Label (Karte 2, Karte 3) ist die GG des **RRF** mit einer Differenz von 2,1% (Karte 2) bzw. 6,0% (Karte 3) deutlich größer als die der **RLR**. Bei dem Vergleich der Ergebnisse des **RRFK** und der **RLRK** ist der Unterschied noch ausgeprägter. Hier ist die Differenz der GG 1,9% (Karte

1), 3,9% (Karte 2) und 6,4% (Karte 3), wobei der **RRFK** die größere GG besitzt. In den Änderungsgebieten ist die GG unabhängig von der Karte größer, wenn der **RRF** anstatt der **RLR** bzw. der **RRFK** anstelle der **RLRK** verwendet wird.

Außerdem ist zu beobachten, dass die Integration des Kartenlabels in den Klassifikationsprozess einen größeren positive Einfluss auf die GG des gesamten Bildes ausübt, wenn der RRF als Basisklassifikator verwendet wird. Auch der negativen Einfluss durch die Beobachtung er Kartenlabels auf die GG in den Änderungsgebieten ist bei der Verwendung der Methoden mit dem RRF geringer. Tabelle 19 zeigt die Differenzen der GG der Methoden **RRFK** und **RRF** bzw. **RLRK** und **RLR**, abhängig von der verwendeten Karte. Der Zuwachs der GG des gesamten Gebietes, durch die Berücksichtigung des Kartenlabels in der Klassifikation, ist deutlich größer bei den Methoden basierend auf dem RRF. Wie bereits festgestellt wurde, verringert sich die GG, die nur aus den Pixeln mit fehlerhaftem Label in der Karte berechnet wird, wenn das Kartenlabel in der Klassifikation mit berücksichtigt wird. Dieser negative Effekt ist jedoch bei den RRF Methoden kleiner als bei den RLR Methoden. Dies ist besonders bei dem Vergleich der Differenzen der MF1 Werte zu sehen (Tabelle 20). Hier ist beispielsweise bei der Verwendung von Karte 3 (etwa 26% Änderungen) die Differenz des MF1 Wertes zwischen der **RLRK** und der **RLR** 23,3% und jene zwischen dem **RRFK** und dem **RRF** lediglich 12,2%.

Zusammenfassend zeigt sich, dass die Methoden der logistischen Regression besonders bei einer größeren Anzahl an Änderungen größtenteils zu einem schlechteren Ergebnis führen, als die entsprechenden Methoden basierend auf dem Random Forest.

Frage e) Einfluss der Anzahl an fehlerhaften Labels

Als letztes wird untersucht, wie das Klassifikationsergebnis durch die Anzahl der Fehler in den Trainingslabels beeinflusst wird (Frage e). Auch hier lässt sich wie bei dem Datensatz *Las Vegas* beobachten, dass die GG mit zunehmender Anzahl an fehlerhaften Labels sinkt. Dies ist bei der GG des gesamten Bildes zu beobachten und auch bei der GG, wenn nur die Pixel in Änderungsgebieten betrachtet werden. Einzige Ausnahme bildet der **RF** und der **RRF**. Wird das gesamte Gebiet betrachtet, liegt die GG zunächst bei 81,3% (**RF**) bzw. 81,1% (**RRF**) für Karte 1, sinkt auf 79,7% (**RF**) bzw. 79,9% (**RRF**) für Karte 2 und steigt dann wieder leicht auf 79,0% (**RF**) bzw. 80,1% (**RRF**) für Karte 3. Der Unterschied ist jedoch ziemlich gering. In den Änderungsgebieten sinkt die GG auch bei diesen beiden Methoden.

Wird die GG des gesamten Bildes betrachtet, fällt auf, dass die Verbesserung durch die Berücksichtigung der fehlerhaften Labels im Training (GG(**RRF**)-GG(**RF**)) mit steigender Anzahl der fehlerhaften Labels zunimmt. Die Verbesserung durch die Integration der Kartenlabels in der Klassifikation (GG(**RRFK**)-GG(**RRF**)) nimmt hingegen ab (Tabelle 19). Dies ist verständlich, da die Berücksichtigung des Kartenlabels vor allem die Klassifikation in den Bereichen verbessert, die sich nicht geändert haben. Die Anzahl der korrekt gelabelten Pixel in der Karte nimmt aber bei steigender Anzahl an fehlerhaften Labels ab, sodass auch die Verbesserung aufgrund der zusätzlichen Beobachtung der Kartenlabels abnimmt.

An dieser Stelle soll noch darauf hingewiesen werden, dass die neu entwickelte Methode **RRFK** bei der Verwendung der Karte 3, welche mit etwa 26% die meisten Änderungen beinhaltet, zu der besten GG bezogen auf das gesamte Gebiet von 88,1% führt. In den Änderungsgebieten ist die GG der **RRFK** mit 91,8% ebenfalls sehr hoch, besonders wenn berücksichtigt wird, dass in den veralteten Karten alle Labels in diesen Bereichen fehlerhaft sind.

Zusammenfassend lässt sich auch bei dem Datensatz *Vaihingen* erkennen, dass die GG mit steigender Anzahl an Änderungen im Allgemeinen abnimmt.

Im Folgenden werden drei Gebiete (A1, A13 und A32) beispielhaft genauer beleuchtet. Bei Gebiet A1 ist die Klassifikation mit dem **RRFK** sehr gut verlaufen, während bei dem Gebiet A32 viele Fehlklassifikationen zu beobachten waren. Als drittes Beispiel wird Gebiet A13 gezeigt, da hier eine Klasse mit den Klassifikatoren **KRFK** und **ERFK** nicht detektiert wurde. In allen drei vorgestellten Experimenten wurde die Karte mit der größten Anzahl an fehlerhaften Trainingslabels verwendet.

Gebiet A1

Abb. 54 zeigt das DOP von Gebiet A1 und die Klassifikationsergebnisse sowie die Referenz. Im DOP sind durch Kreise fünf Bereiche gekennzeichnet, bei denen sich die Klassifikationsergebnisse stark unterscheiden.

Besonders das Gebäude in dem dunkelblauen Kreis wurde ausschließlich von der Methode **RRFK** vollständig detektiert. Eine mögliche Ursache für die fehlerhafte Detektion durch die meisten Methoden könnte die geringe Gebäudehöhe sein.

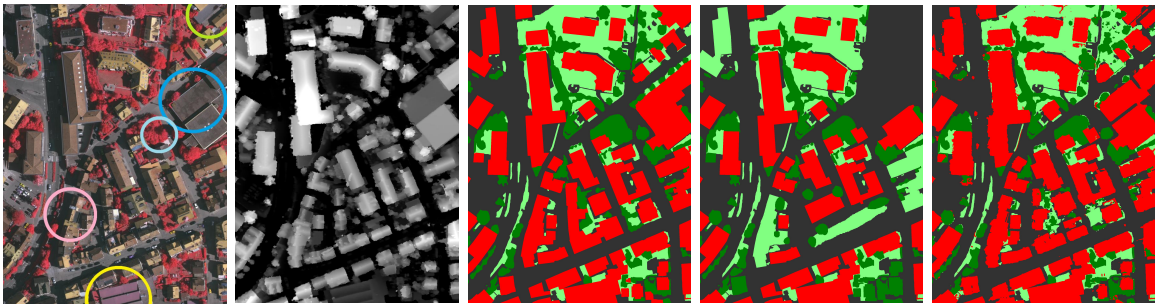
In der rechten oberen Ecke des Gebietes A1, in hellgrün gekennzeichnet, befinden sich drei Gebäude. Das mittlere Gebäude ist ebenfalls sehr niedrig und wird daher nicht von allen Methoden korrekt erkannt. Lediglich der **RRFK** und der **RF** detektieren es teilweise. Der **KRFK** klassifiziert auch das rechte der drei Häuser als *versiegelte Fläche*. Auch das durch einen rosa Kreis gekennzeichnete Gebäude wurde in den Randbereichen nur mit dem **RRFK** vollständig korrekt klassifiziert. Dies könnte ebenso an der geringen Gebäudehöhe am Rand liegen oder an dem Schatten in den Randbereichen des Gebäudes. Die Vollständigkeit der Klasse *Gebäude* ist mit 95,2%, bezogen auf das ganze Bild, deutlich größer unter Verwendung der Methode **RRFK**, als bei der Verwendung des **ERFK** (89,9%), **KRFK** (76,7%), **RF** (87,0%), **RLRK** (75,2%), **RLR** (75,5%) oder **LR** (73,6%). Insgesamt sind etwa 0,2% der Pixel nur im Ergebnis des **RRFK** korrekt klassifiziert. Fast alle (98,7%) dieser Pixel gehören zu der Klasse *Gebäude*.

In gelb gekennzeichnet ist ein Gebäude, bei dem aufgrund der Dachkonstruktion einige Bereiche im Schatten liegen. Diese werden von Klassifikatoren, bei denen das Kartenlabel als Beobachtung in die Klassifikation mit einfließt, der Klasse *Gebäude* zugeordnet. Lediglich bei

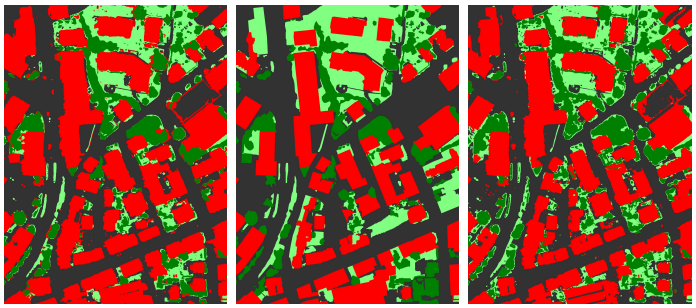
dem **ERFK** wird ein kleiner Teil als *versiegelte Fläche* klassifiziert, obwohl die Karte berücksichtigt wird. Dies könnte, wie zuvor angesprochen, an dem Training ohne den direkten Einfluss der Kartenlabels liegen.

Als letztes soll auf das Gebiet in hellblau hingewiesen werden, in dem sich in der Karte ein Gebäude befindet, in der Referenz jedoch Bäume. Dies wurde aufgrund der geringen Größe nicht als Änderung detektiert, sodass der Einfluss des Kartenlabels relativ hoch ist. Der **RRFK**, wie auch der **KRFK**, hat diesen Bereich daher weiterhin als *Gebäude*, anstelle von *Baum*, klassifiziert.

Insgesamt lässt sich feststellen, dass bei dem Gebiet A1 der **RRFK** ein sehr gutes Ergebnis erzeugt. Etwa 0,2% der Pixel wurden außerdem nur mit dieser neu entwickelten Methode korrekt klassifiziert.



(a) DOP (b) DOM (c) Referenz (d) Karte (e) **RRFK**



Rot: *Gebäude*
Grau: *versiegelte Fläche*
Hellgrün: *niedrige Veg.*
Dunkelgrün: *Baum*

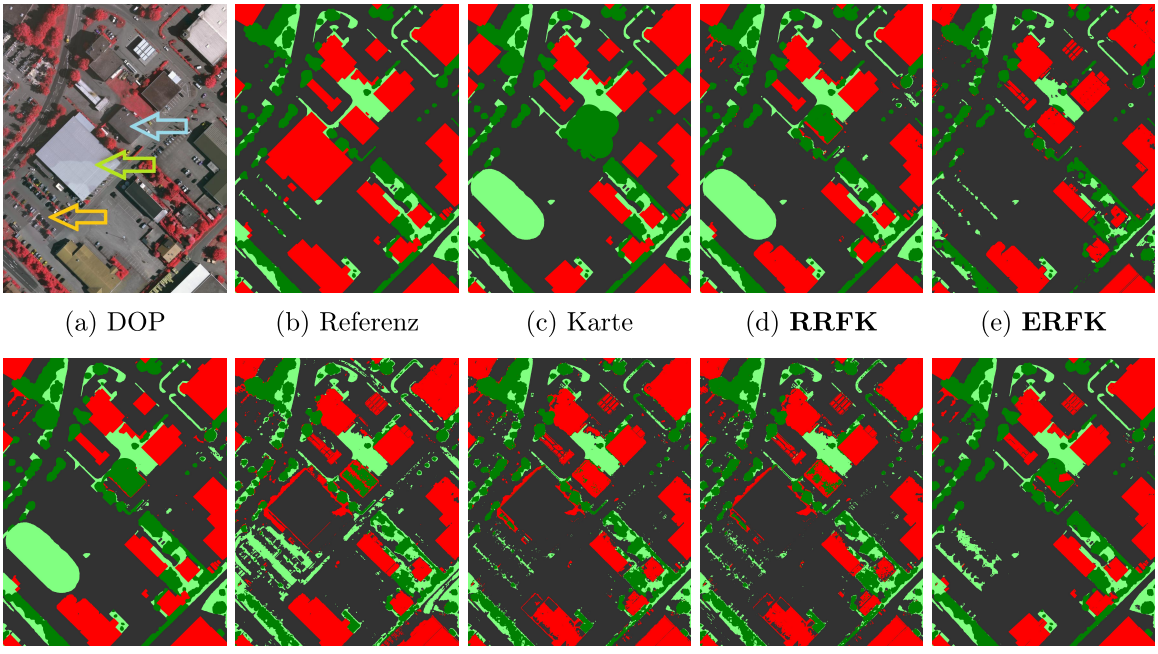
(f) **ERFK** (g) **KRFK** (h) **RF**

Abbildung 54: Gebiet A1 aus *Vaihingen*. Karte 3 (20,1% Änderungen). Auf die Gebiete, die mit Kreisen hervorgehoben werden, wird im Text eingegangen.

Gebiet A32

Das DOP, die Referenz und die Klassifikationsergebnisse von Gebiet A32 sind in Abb. 55 dargestellt. Hier ist die GG des **RRFK** mit 86,7%, bezogen auf das ganze Bild, etwa so groß wie bei dem **ERFK** und **KRFK** mit jeweils 86,8% und kleiner als bei der **RLRK** (88,7%).

Bereits in Kapitel 6.1 wurde das große Industriegebäude in der Mitte des Gebietes (Abb. 55, grüner Pfeil) als problematisch angesehen. Zum einen fallen die Pixel dieses Gebäudes aufgrund der einzigartigen Dachstruktur (s. MHOG in Abb. 25) im Merkmalsraum nicht in die Ballungen anderer Pixel der Klasse *Gebäude* und zum anderen ist das Höhenmerkmal



(a) DOP (b) Referenz (c) Karte (d) **RRFK** (e) **ERFK**

(f) **KRFK** (g) **RF** (h) **LR** (i) **RLR** (j) **RLRK**

Abbildung 55: Gebiet A32 aus *Vaihingen*. Karte 3 (19,4% Änderungen). Auf die Gebiete, die mit Pfeilen hervorgehoben werden, wird im Text eingegangen.
Rot: *Gebäude*. Grau: *versiegelte Fläche*. Hellgrün: *niedrige Veg.*. Dunkelgrün: *Baum*.

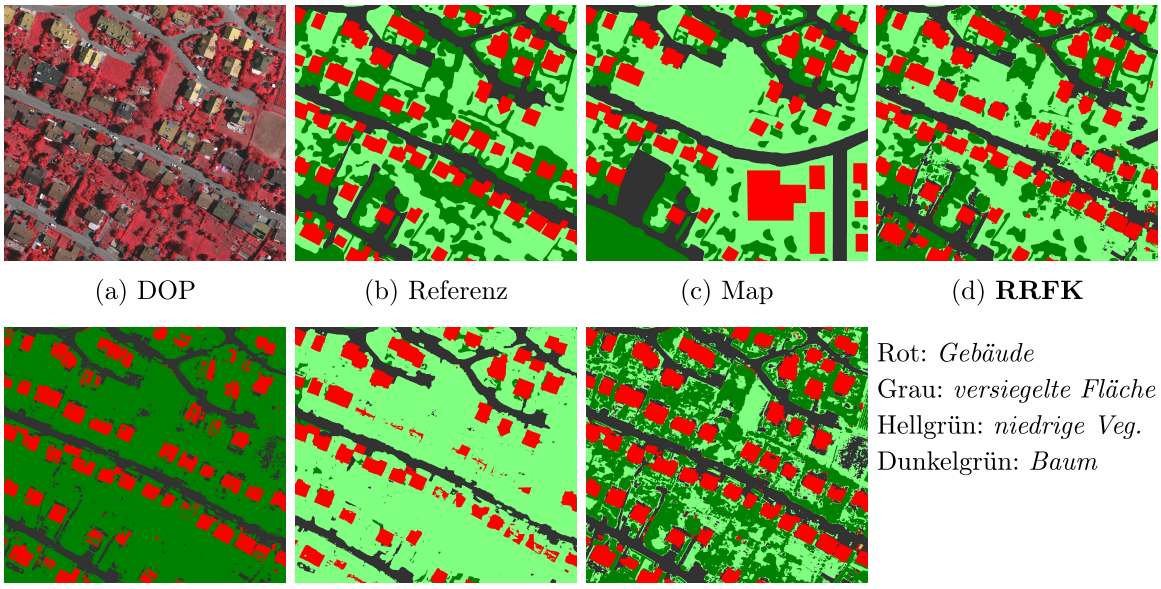
(s. nDOM in Abb. 25) fehlerhaft. Außerdem sind die Pixel des Gebäudes in der Karte als *versiegelte Fläche* gelabelt. Keine Methode hat dieses Gebäude korrekt klassifiziert.

Neben dem Gebäude befindet sich ein Parkplatz, welcher in der Karte größtenteils der Klasse *niedrige Vegetation* angehört (Abb. 55, gelber Pfeil). Schon bei der Klassifikation mit dem **RF** werden einige Pixel der vermeintlichen Grünfläche als *niedrige Vegetation* und andere als *versiegelte Fläche* klassifiziert. Durch den starken Einfluss der Karte bleibt die Grünfläche auch in dem Klassifikationsergebnis unter Verwendung der Methoden **RRFK** und **KRFK** erhalten. Mit der **RLRK** wird nur ein Teil der Grünfläche fälschlicherweise der *niedrige Vegetation* zugeordnet. Bereits bei dem Klassifikationsergebnis der **LR** und der **RLR** ist ein Großteil des Parkplatzes korrekt klassifiziert worden. Die LR bzw. RLR scheint sich demnach bei der Klassifikation dieses Bereiches deutlich sicherer zu sein als der RF, sodass auch mit der Berücksichtigung der Kartenlabels der Parkplatz größtenteils als *versiegelte Fläche* klassifiziert wird. Ähnliches ist bei einem Haus neben dem großen Industriegebäude (Abb. 55, blauer Pfeil) zu beobachten, welches in der Karte als *Baum* abgebildet ist und von den Klassifikatoren **RRFK**, **KRFK**, **RLRK** und **RF** zu einem Großteil ebenfalls dieser Klasse zugeordnet wird.

Zusammenfassend ist in Gebiet A32 zu erkennen, dass die Klassifikation der neu entwickelten Methode **RRFK** ebenfalls fehlerhaft sein kann, wenn ein Großteil der Trainingsbeispiele in einer Ballung im Merkmalsraum fehlerhafte Labels aufweist. Auch Fehler in den Fernerkundungsdaten werden bei dem **RRFK** nicht berücksichtigt und können somit zu fehlerhaften Zuordnungen in der Klassifikation führen.

Gebiet A13

Das DOP, die Referenz und die Klassifikationsergebnisse für Gebiet A13 sind in Abb. 56 dargestellt. An dem Klassifikationsergebnis des **RF** ist zu erkennen, dass die Klassen *niedrige Vegetation* und *Baum* schwer in dem Bild zu trennen sind. Ein Grund hierfür ist die Tatsache, dass sich die Ballungen der Trainingsbeispiele beider Klassen im Merkmalsraum überlappen. In diesem Gebiet kommt hinzu, dass einige Trainingsbeispiele der einen Klasse fälschlicherweise das Label der anderen Klasse in der veralteten Karte aufweisen. Bei den Methoden **ERFK** und **KRFK** führt dies zu einer Fehlklassifikation von immer mehr Pixeln einer Klasse, bis alle Pixel der Klasse *niedrige Vegetation* (**ERFK**) bzw. *Baum* (**KRFK**) falsch erkannt werden. In Tabelle 22 wird die Vollständigkeit und Korrektheit bezogen auf das gesamte Gebiet präsentiert. Da die beiden Klassen *Baum* und *niedrige Vegetation* von jeweils einem der beiden Klassifikatoren nicht erkannt wurden, ist die Vollständigkeit hier 0. Mit dem **RRFK** wurden ein Großteil der Pixel dieser Klassen korrekt klassifiziert, was auch der hohe F1 Wert der beiden Klassen zeigt: 80,3% für *niedrige Vegetation* und 78,7% für *Baum*, jeweils bezogen auf das gesamte Bild.



(a) DOP (b) Referenz (c) Map (d) **RRFK**

(e) **ERFK** (f) **KRFK** (g) **RF**

Abbildung 56: Gebiet A13 aus *Vaihingen*. Karte 3 (28,8% Änderungen).

	Vollständigkeit				Korrektheit			
	Geb.	*Baum*	*niedr. Veg.*	*vers. Fl.*	*Geb.*	*Baum*	*niedr. Veg.*	*vers. Fl.*
RRFK	97,3	69,6	85,6	89,5	91,0	90,5	75,6	84,7
ERFK	76,0	99,3	0,0	73,4	96,6	42,3	-	77,2
KRFK	73,8	0,0	93,3	71,6	98,7	0,2	45,8	95,8

Tabelle 22: Vollständigkeit und Korrektheit aller Klassen für das Gebietes A13 des Datensatzes *Vaihingen* [%]. Grundlage: Gesamtes Bild.

Insgesamt kann für das Gebiet A13 festgehalten werden, dass die Klassen *niedrige Vegetation* und *Baum* mit den Methoden **KRFK** und **ERFK** nicht ausreichend unterschieden werden

konnten, sodass eine der beiden Klassen kaum oder gar nicht in der Klassifikation erkannt wurde.

6.3.4. Herne und Husum

Aufgrund der vielen Ähnlichkeiten der beiden Datensätze *Herne* und *Husum* werden sie in diesem Kapitel gemeinsam evaluiert. Die Gesamtgenauigkeiten der Experimente bezogen auf den *Herne* Datensatz sind in Abb. 57 abgebildet; jene der Experimente bezogen auf den *Husum* Datensatz sind in Abb. 58 dargestellt. Die GG der Änderungsdetektion bezogen auf beide Datensätze wurden nicht grafisch dargestellt, da sie ebenfalls ein ähnliches Muster wie die GG des Klassifikationsergebnisses unter Berücksichtigung aller Pixel zeigen. Die Werte der GG und des MF1 finden sich in den Tabellen 23 und 24. In Anhang B werden zusätzlich weitere Genauigkeitsmaße präsentiert (Datensatz *Herne*: Tabellen B.13, B.14, B.15; Datensatz *Husum*: Tabellen B.16, B.17 und B.18).

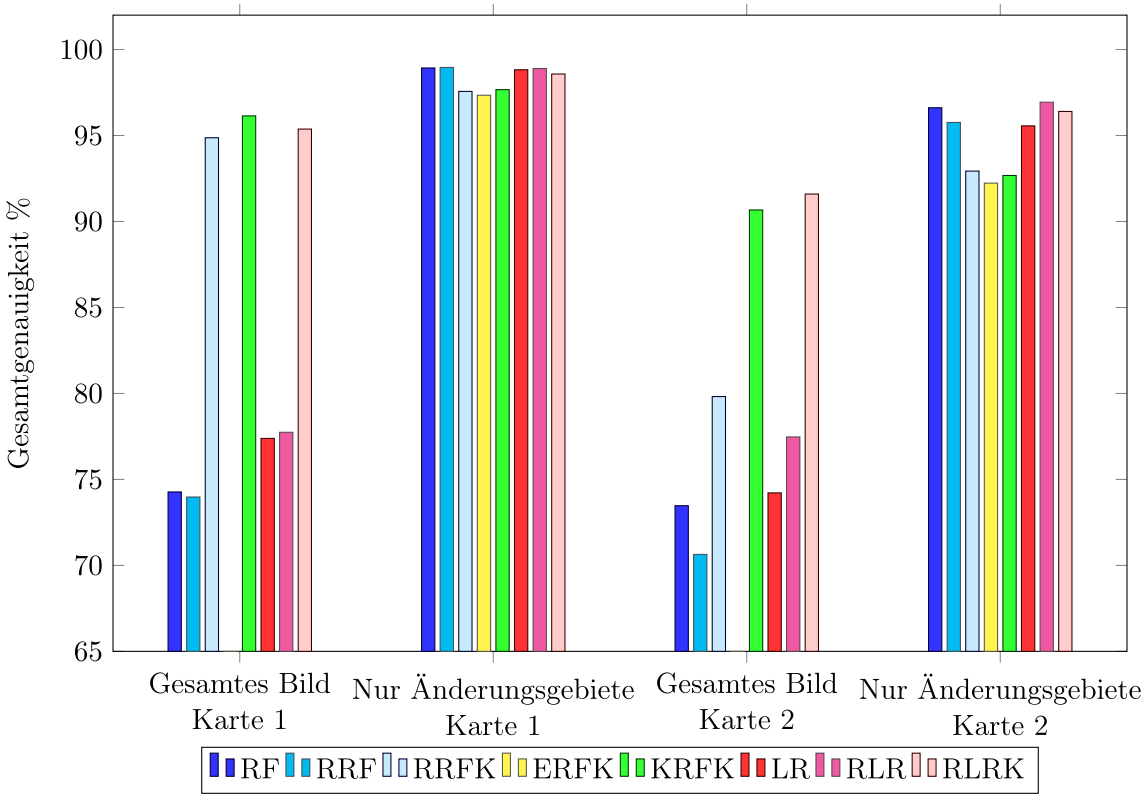


Abbildung 57: Gesamtgenauigkeit [%]. Datensatz: *Herne*.

Frage a) Berücksichtigung der Fehler in den Trainingslabels

Zunächst wird wieder untersucht, ob die Berücksichtigung der Fehler im Training das Klassifikationsergebnis verbessert (Frage a). Bei dem Datensatz *Herne* sinkt, entgegen der Erwartung, die GG bezogen auf das gesamte Bild wenn die Klassifikation mit dem **RRF** anstelle des **RF** erfolgt. Für die Analyse wird die Konfusionsmatrix des **RF** und des **RRF** für das gesamte Gebiet unter Verwendung der Karte 2 (17% Änderungen) betrachtet (s. Tabelle 25). Mit dem

Gebiet		Gesamtes Bild		Nur Änderungen		Änderungsdetektion	
		GG	MF1	GG	MF1	GG	MF1
Karte 1	RF	74,3	73,8	98,9	59,4	74,8	56,3
	RRF	74,0	73,5	**99,0**	**59,9**	74,5	56,0
	RRFK	94,9	94,7	97,6	45,5	95,1	76,6
	ERFK	48,0	40,6	97,3	33,0	50,2	40,4
	KRFK	**96,1**	**96,0**	97,7	50,1	**96,2**	79,9
	LR	77,4	77,0	98,8	59,5	77,8	58,1
	RLR	77,7	77,2	98,9	59,8	78,2	58,6
	RLRK	95,4	95,2	98,6	58,0	95,6	**81,7**
Karte 2	RF	73,5	72,9	96,6	58,3	75,4	69,1
	RRF	70,6	70,1	95,8	55,8	72,8	66,5
	RRFK	79,8	79,9	92,9	42,5	81,7	72,0
	ERFK	41,4	33,3	92,2	32,1	48,2	46,7
	KRFK	90,7	90,5	92,7	47,8	90,9	81,2
	LR	74,2	74,1	95,6	58,2	75,9	68,6
	RLR	77,5	76,9	**96,9**	**60,4**	79,0	72,5
	RLRK	**91,6**	**91,5**	96,4	59,0	**92,5**	**87,1**

Tabelle 23: Genauigkeitsmaße [%]. Datensatz: *Herne*. Der größte Wert pro Gebiet ist fett gedruckt.

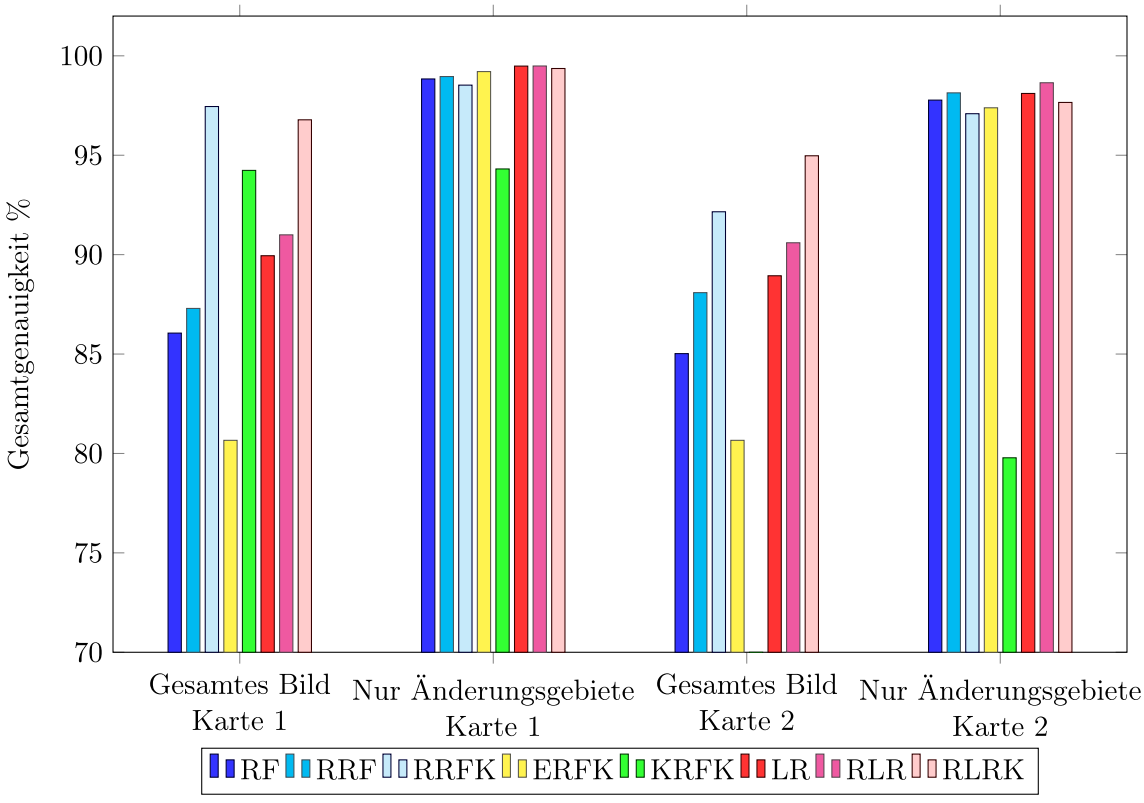


Abbildung 58: Gesamtgenauigkeit [%]. Datensatz: *Husum*.

RF wurden 6,9% der Pixel, die das Label *Boden* in der Referenz haben, als *Wald* klassifiziert; Bei dem **RRF** sind es 9,0%. Die Unterscheidung der beiden Klassen scheint demnach problematisch für beide Klassifikatoren zu sein. Eine visuelle Analyse zeigt, dass die meisten Fehlklassifikationen dieser Art in Bereichen auftreten, in denen der Boden ein ähnliches Er-

Gebiet		Gesamtes Bild		Nur Änderungen		Änderungsdetektion	
		GG	MF1	GG	MF1	GG	MF1
Karte 1	RF	86,1	75,9	98,8	64,3	87,0	76,3
	RRF	87,3	76,0	99,0	64,6	87,9	77,2
	RRFK	**97,4**	**94,5**	98,5	64,1	**97,6**	**93,2**
	ERFK	80,7	29,8	99,2	-	80,9	69,1
	KRFK	94,2	89,8	94,3	52,5	94,3	79,0
	LR	89,9	77,2	**99,5**	**65,7**	90,2	80,5
	RLR	91,0	80,7	**99,5**	**65,7**	91,3	82,1
	RLRK	96,8	92,9	99,4	65,4	97,0	92,3
Karte 2	RF	85,0	74,4	97,8	64,0	87,0	83,3
	RRF	88,1	75,5	98,1	64,7	89,0	85,3
	RRFK	92,2	81,8	97,1	63,6	92,9	89,5
	ERFK	80,7	29,8	97,4	-	81,4	76,8
	KRFK	8,0	12,2	79,8	33,0	9,8	9,4
	LR	88,9	74,7	98,1	64,7	89,5	85,8
	RLR	90,6	79,3	**98,6**	**65,2**	91,1	87,8
	RLRK	**95,0**	**89,1**	97,7	64,1	**95,7**	**93,5**

Tabelle 24: Genauigkeitsmaße [%]. Datensatz: *Husum*. Der größte Wert pro Gebiet ist fett gedruckt.

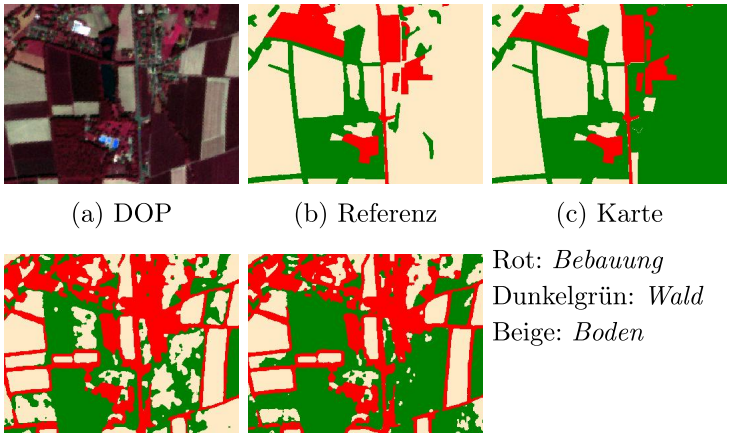
scheinungsbild wie der Wald aufweist. Die Ballungen im Merkmalsraum von Pixeln der Klasse *Wald* und *Boden* sind also nahe beieinander oder überlappen sich. Hinzu kommt, dass 46,7% der Pixel mit fehlerhaftem Label in Karte 2 das Label *Wald* in der Karte aufweisen, obwohl sie der Klasse *Boden* angehören. Eine Fehlklassifikation von Pixeln der Klasse *Boden* mit den entsprechenden Merkmalen als *Wald* ist demnach nachvollziehbar. Wird nun im Training des **RRF** der Belief der vorherigen Iteration berücksichtigt, kann sich die Grenze im Merkmalsraum mit jeder Iteration weiter von der optimalen Grenze entfernen. Somit steigt die Anzahl an Fehlklassifikationen dieser Art.

		RF			**RRF**		
		Referenz			Referenz		
		Bebauung	Wald	Boden	Bebauung	Wald	Boden
Klassifikation	Bebauung	30,9	5,6	10,5	30,2	5,2	11,1
	Wald	1,4	17,1	6,9	1,7	17,7	9,0
	Boden	0,8	1,4	25,5	1,2	1,2	22,7

Tabelle 25: *Konfusionsmatrix* $\mathcal{KM}$ [%]. Datensatz: *Herne*. Grundlage: Gesamtes Bild.

Dieses Problem des Wegdriftens der Grenze tritt bei der **RLR** nicht auf. Wie zu erwarten war, verbessert sich die GG des gesamten Bildes der **RLR** verglichen zu der GG der **LR**.

Bei dem Datensatz *Husum* ist wie bei dem Datensatz *Herne* zu beobachten, dass die Anzahl an Pixeln mit der Klasse *Boden*, die fehlerhaft der Klasse *Wald* zugeordnet werden, in dem Ergebnis des **RRF** größer ist (2,7% aller Pixel für Karte 2) als in dem Ergebnis des **RF** (1,3% aller Pixel für Karte 2). Die fehlerhafte Klassifikation von Pixeln der Klasse *Wald* und *Boden* in die Klasse *Bebauung* ist mit dem **RRF** jedoch deutlich geringer als mit dem **RF**. Die GG und der MF1 Wert bezogen auf das gesamte Bild und bezogen auf die Änderungsgebiete sind



(a) DOP (b) Referenz (c) Karte

(d) **RF** (e) **RRF**

Abbildung 59: Ausschnitt aus *Herne*. Karte 2 (17% Änderungen).

daher bei dem Ergebnis des **RRF** etwas höher als bei dem des **RF**.

Wie bei dem *Herne* Datensatz zeigen die Ergebnisse der **RLR** auch hier eine höhere GG und einen höheren MF1 Wert bezogen auf das gesamte Gebiet als die Ergebnisse der **LR**.

Zusammenfassend kann bei dem **RRF**, durch die Verwendung des Beliefs der vorherigen Iteration, ein leichtes Wegdriften der Grenze im Merkmalsraum zur Unterscheidung der Klassen *Wald* und *Boden* beobachtet werden. Dies führt bei dem Datensatz *Herne* zu einer leichten Reduktion der GG. Bei dem Datensatz *Husum* ist die GG bei dem **RRF** dennoch höher als bei dem **RF**. Die Berücksichtigung der Fehler im Training führt im Allgemeinen aber zu einer Verbesserung, was an dem Vergleich der Ergebnisse der Methoden **LR** und **RLR** zu erkennen ist.

Frage b) Berücksichtigung der Kartenlabels als weitere Beobachtung

Der Vergleich der Ergebnisse des **RRF** und des **RRFK** bzw. der **RLR** und der **RLRK** (Frage b) fällt wie erwartet aus. Sowohl bei dem Datensatz *Herne* als auch bei dem Datensatz *Husum* ist die GG des gesamten Bildes unter Berücksichtigung der Kartenlabels deutlich höher als ohne deren Berücksichtigung. Besonders bei geringer Anzahl an fehlerhaften Labels (Karte 1) ist die Differenz zwischen der Gesamtgenauigkeiten mit bis zu 20,9% sehr groß.

In den Änderungsgebieten ist die GG etwas geringer, wenn das Kartenlabel in der Klassifikation berücksichtigt wird. Auch dies ist bereits bei vorherigen Experimenten, etwa bei jenen zu *Hameln* und *Vaihingen*, beobachtet worden und aus den dort genannten Gründen nachvollziehbar. Die Differenz der GG ist mit maximal 2,9% jedoch relativ gering. Absolut betrachtet ist die GG sehr hoch. In den Änderungsgebieten liegt sie unabhängig von dem Datensatz, der veralteten Karte und dem Basisklassifikator (**RRFK**, **RLRK**) bei über 92,9%.

Zusammenfassend kann gesagt werden, dass die Berücksichtigung der Kartenlabels in der Klassifikation zu einer Verbesserung der GG führt, wenn das gesamte Bild betrachtet wird.

Wie zu erwarten war, sinkt die GG etwas, wenn nur die Pixel in den Änderungsgebieten betrachtet werden.

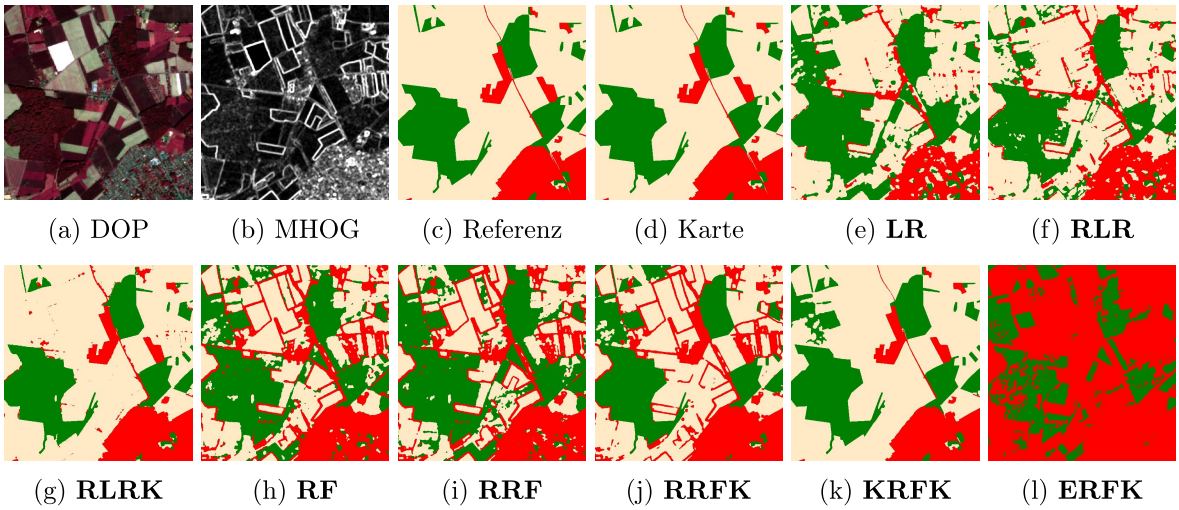
Frage c) Verwendung der Wahrscheinlichkeiten der Klassenzugehörigkeit

Als nächstes wird untersucht, ob das Ersetzen der Zuordnung eines Trainingsbeispieles zu einer Klasse durch die Zuordnung eines Trainingsbeispieles zu allen Klassen mit einer bestimmten Wahrscheinlichkeit zu einem besseren Klassifikationsergebnis führt (Frage c). Hierfür wird zunächst das Klassifikationsergebnis des **RRFK** mit dem des **ERFK** und anschließend mit dem des **KRFK** verglichen.

Die GG des gesamten Bildes unter Verwendung der Methode **ERFK** ist bei beiden Datensätzen deutlich geringer als unter Verwendung der Methode **RRFK**. Der MF1 Wert ist außerdem für *Herne* mit 40,6% (Karte 1, 5% Änderungen) und 33,3% (Karte 2, 17% Änderungen), sowie für *Husum* mit 29,8% (Karte 1, 10% Änderungen) und 29,8% (Karte 2, 20% Änderungen) sehr gering. Bei Betrachtung der Genauigkeitsmaße des **ERFK** für den *Herne* Datensatz in Tabelle B.13 in Anhang B ist zu erkennen, dass die Vollständigkeit der Klasse *Boden* mit 1,2% (Karte 1) bzw. 0,0% (Karte 2) extrem niedrig ist. Die Klasse *Boden* wurde demnach mit dem **ERFK** nicht detektiert. Auch die Vollständigkeit der Klasse *Wald* und die Korrektheit der Klasse *Bebauung* sind relativ gering. Viele Pixel wurden demnach fehlerhaft der Klasse *Bebauung* zugeordnet. Dies ist auch beispielhaft in dem Ausschnitt in Abb. 60 zu erkennen. In Tabelle 25 ist zu sehen, dass bereits in der Initialisierung mit dem **RF** viele Pixel der Klasse *Wald* und *Boden* das Label *Bebauung* erhalten haben. Da bei der Methode **ERFK** die Labels der vorherigen Iteration für das Training verwendet werden, kann es wie auch in diesem Fall vorkommen, dass sich die Grenze im Merkmalsraum mit jeder Iteration weiter von der optimalen Grenze entfernt. Die Genauigkeitsmaße des **ERFK** für den *Husum* Datensatz in Tabelle B.13 in Anhang B zeigen, dass alle Pixel des gesamten Bildes als *Boden* klassifiziert wurden. Auch hier hat sich vermutlich die Entscheidungsgrenze mit jeder Iteration weiter von der optimalen Entscheidungsgrenze entfernt. Die Anzahl der Pixel, die fehlerhaft der Klasse *Boden* zugeordnet wurden, nahmen dementsprechend mit jeder Iteration zu.

Die GG des Klassifikationsergebnisses von dem **KRFK** ist bei dem *Herne* Datensatz etwas höher als die GG des Ergebnisses von dem **RRFK**. Unter Verwendung der Karte 2 (17% Änderungen) beträgt die Differenz 10,9%. Auch die Vollständigkeit der Klasse *Boden* und die Korrektheit der Klasse *Bebauung* sind mit 84,5% und 95,7% deutlich größer bei dem Ergebnis des **KRFK** verglichen zu den Werten 63,1% und 69,8% aus dem Ergebnis des **RRFK**. Ein visueller Vergleich der Ergebnisse zeigt, dass bei dem Ergebnis des **RRFK** an den Rändern der einzelnen Objekte im Bild vermehrt Pixel fehlerhaft der Klasse *Bebauung* zugeordnet werden (s. Abb. 60). Dies ist mit dem Merkmal MHOG zu erklären. Wie an der Bebauung in der unteren rechten Ecke des Ausschnittes in Abb. 60 zu erkennen ist, ist das MHOG Merkmal gut für die Detektion der Klasse *Bebauung* geeignet, da der Wert dieses Merkmals in bebauten Gebieten vergleichsweise groß ist. Da dieses Merkmal auf einer Detektion von Kanten basiert, ist der Wert des Merkmals auch an Objektgrenzen sehr groß. Bei dem Ergebnis des

Klassifikators **KRFK** treten diese Fehlklassifikationen an Objekträndern kaum auf (s. Abb. 60). Dies liegt vermutlich an dem stärkeren Einfluss der Kartenlabels bei dem **KRFK**.



(a) DOP (b) MHOG (c) Referenz (d) Karte (e) **LR** (f) **RLR**

(g) **RLRK** (h) **RF** (i) **RRF** (j) **RRFK** (k) **KRFK** (l) **ERFK**

Abbildung 60: Ausschnitt aus *Herne*. Karte 2 (17% Änderungen).
Rot: *Bebauung*. Dunkelgrün: *Wald*. Beige: *Boden*.

Bei dem Datensatz *Husum* unter Verwendung der Karte 1 (10% Änderungen) ist die GG des gesamten Bildes und in den Änderungsgebieten etwas kleiner, wenn der **KRFK** anstelle des **RRFK** verwendet wird. Auch dies kann durch den stärkeren Einfluss des Kartenlabels bei dem **KRFK** erklärt werden.

Wird hingegen die Karte 2 (20% Änderungen) genutzt, liegt die GG des Klassifikationsergebnisses von **KRFK** bei nur 8,0%. Ein Großteil des Bildes wird hier als *Wald* klassifiziert, was auch die geringe Korrektheit dieser Klasse mit 6,2% verursacht. Eine mögliche Ursache ist das schlechte Verhältnis zwischen Pixeln, die korrekt das Label *Wald* in der Karte aufweisen, und jenen, die nur fälschlicherweise das Label *Wald* in der Karte haben und eigentlich der Klasse *Boden* oder *Bebauung* angehören. Dieses Ungleichgewicht wird mit der Konfusionsmatrix in Tabelle 26 verdeutlicht. Etwa 18,3% der Pixel besitzen fälschlicherweise das Label *Wald* und nur 8,1% der Pixel in der Karte zeigen korrekterweise das Label *Wald*.

Zusammenfassend zeigt sich, dass die Verwendung von Wahrscheinlichkeiten der Klassenzugehörigkeit anstatt einer Zuordnung von nur einem Labels zu einem Trainingsbeispiel das Klassifikationsergebnis deutlich verbessern kann. Die GG der Methoden **ERFK** und **KRFK** sind teilweise deutlich geringer als die GG des **RRFK**. Besonders die GG des **ERFK** bezogen auf das gesamte Bild ist bei beiden Datensätzen sehr gering. Bei dem Datensatz *Husum* ist auch die GG des **KRFK** deutlich kleiner als die GG des **RRFK**.

		Label aus Karte 2		
		Bebauung	*Wald*	*Boden*
Referenz	*Bebauung*	8.6	0.7	1.9
	Wald	0,0	8,1	0,0
	Boden	0,0	17,6	63,0

Tabelle 26: *Konfusionsmatrix* $\mathcal{KM}$ [%]. Datensatz *Husum*, Karte 2 (20% Änderungen). Grundlage: Gesamtes Bild.

Frage d) Vergleich mit der robusten logistischen Regression

Als nächstes wird das Klassifikationsergebnis des **RRF** und des **RRFK** mit dem der **RLR** und der **RLRK** verglichen (Frage d). Bei beiden Datensätzen ist die GG des gesamten Gebietes etwas größer bei dem Ergebnis der **RLR** als bei dem Ergebnis des **RRF**. Auch die **RLRK** liefert, abgesehen von der Klassifikation mit der Karte 1 (10% Änderungen) des Datensatzes *Husum*, eine höhere GG als der **RRFK**.

Bei dem Datensatz *Herne* könnte die geringere Anzahl an Fehlklassifikationen an Objektgrenzen zu der höheren GG führen. Die Fehlklassifikationen an Objektgrenzen wurden zuvor bei dem Ergebnis der **RRFK** beschrieben und treten auch bei den Ergebnissen des **RRF** und **RF** auf. In den Ergebnissen der **LR** und **RLR** sind diese nicht so stark ausgeprägt. Das ist auch in dem Ausschnitt in Abb. 60 zu sehen. Besonders bei dem **RLRK** ist diese Art der Fehlklassifikation, aufgrund des starken Einflusses des Kartenlabels, nicht mehr auszumachen. Diese visuelle Analyse wird durch eine relativ hohe Vollständigkeit der Klasse *Boden* und relativ hohe Korrektheit der Klasse *Bebauung* des Ergebnisses der **RLRK** unterstützt. Auch die Korrektheit der Klasse *Wald* ist bei dem Ergebnis der **RLRK** um 6% (Karte 1) bzw. 15% (Karte 2) größer als bei dem **RRFK**.

Die Unterschiede der Genauigkeitsmaße sind deutlich anders verteilt, wenn der Datensatz *Husum* betrachtet wird. Unter Verwendung der Karte 2 (20% Änderungen) ist hier die Vollständigkeit der Klassen *Bebauung* und *Wald* deutlich größer für das Ergebnis der **RLRK** als für das Ergebnis des **RRFK**.

Zusammenfassend ist zu erkennen, dass die Methoden basierend auf der robusten logistischen Regression sehr gute Ergebnisse liefern. Im Vergleich zu den Ergebnissen der Methoden basierend auf dem robusten Random Forest ist die GG bezogen auf das gesamte Gebiet sogar etwas höher.

Frage e) Einfluss der Anzahl an fehlerhaften Labels

Als letztes wird untersucht ob die Anzahl der Fehler in den Trainingslabels das Klassifikationsergebnis beeinflusst (Frage e). Wie bereits bei den anderen Datensätzen festgestellt wurde, sinkt die GG mit steigender Anzahl der Fehler. Diese Verschlechterung ist bei allen Klassifikatoren zu beobachten. Auch die Verbesserung der GG des gesamten Gebietes durch die Berücksichtigung des Kartenlabels in der Klassifikation sinkt mit steigender Anzahl an fehlerhaften Labels. Die leichte Verringerung der GG in den geänderten Gebieten durch den Einfluss des Kartenlabels nimmt hingegen zu. Dies wurde auch zuvor bei dem Datensatz *Vaihingen* festgestellt.

Zusammenfassend ist auch hier zu beobachten, dass die GG mit steigender Anzahl an Fehlern in den Labels aus der veralteten Karte sinkt.

6.3.5. Diskussion

Die Diskussion orientiert sich ebenfalls an den Fragen aus Kapitel 5.6.3. Für einen besseren Überblick ist in den Tabellen 27 und 28 der Unterschied der GG bezogen auf die Ergebnisse der Methoden **RRF** und **RRFK** stilistisch dargestellt.

			RF	RRFK	ERFK	KRFK	LR	RLR	RLRK
Gesamtes Gebiet	Hameln	N1	+	-	+	+	+	+	+
		N2	+	+	+	+ +	+ +	+ +	+ +
	Las Vegas	1986	- -	- - -	+ +	- -	- -	- - -	- - -
		1991	+	-	-	-	+ + +	-	-
		2000	+ +	+	- -	- -	+	-	- -
	Vaihingen	M1	o	- - -	- - -	- - -	+	-	- - -
		M2	+	- - -	- -	- - -	+ +	+ +	- -
		M3	+	- - -	+ +	- -	+ + +	+ + +	+
	Herne	M1	-	- - -	+ + +	- - -	- -	- -	- - -
		M2	- -	- - -	+ + +	- - -	- -	- - -	- - -
	Husum	M1	+	- - -	+ + +	- - -	- -	- -	- - -
		M2	+ +	- -	+ + +	+ + +	-	- -	- - -
Nur Änderungsgebiete	Hameln	N1	o	+ +	+ +	+ + +	+ +	+ +	+ + +
		N2	+	+ + +	+ + +	+ + +	+ +	+ +	+ + +
	Las Vegas	1986	- - -	-	+	+ +	+ +	+	+ +
		1991	- -	+	+ +	+	+ + +	-	+
		2000	-	-	+	+	+	o	+
	Vaihingen	M1	+	+ +	+	+ +	+	+	+ +
		M2	o	+ +	+	+ +	+ +	+ +	+ + +
		M3	o	+ +	+	+ + +	+ + +	+ +	+ + +
	Herne	M1	o	+	+	+	+	o	+
		M2	-	+ +	+ +	+ +	+	-	-
	Husum	M1	+	+	-	+ +	-	-	-
		M2	+	+	+	+ + +	o	-	+
Änderungsdetektion	Hameln	N1	+	-	+	+ +	+ +	+	+
		N2	+	+	+	+ +	+ +	+ +	+ +
	Las Vegas	1986	- -	- - -	+ +	- -	- -	- - -	- - -
		1991	+	-	-	-	+ + +	-	-
		2000	+ +	-	- -	- -	+	-	- -
	Vaihingen	M1	+	- - -	- - -	- - -	-	-	- - -
		M2	+	- - -	- -	- - -	+ +	+ +	- -
		M3	+	- - -	+	- -	+ + +	+ +	+
	Herne	M1	-	- - -	+ + +	- - -	- -	- -	- - -
		M2	- -	- - -	+ + +	- - -	- -	- - -	- - -
	Husum	M1	+	- - -	+ + +	- - -	- -	- -	- - -
		M2	+ +	- -	+ + +	+ + +	-	- -	- - -

Tabelle 27: Unterschied der Gesamtgenauigkeit bezogen auf das Ergebnis der **RRF**. Kreis (o): Differenz der GG ist kleiner als 0,1%. Ein Zeichen: Differenz der GG ist kleiner als 2%. Zwei Zeichen: Differenz der GG ist kleiner als 5%. Drei Zeichen: Differenz der GG ist größer als 5%. Plus (+): GG der **RRF** ist größer. Minus (-): GG der **RRF** ist kleiner.

			RF	RRF	ERFK	KRFK	LR	RLR	RLRK
Gesamtes Gebiet	Hameln	N1	+ +	+	+	+ +	+ +	+	+
		N2	+	-	+	+ +	+ +	+ +	+ +
	Las Vegas	1986	+ +	+ + +	+ + +	+ +	+ +	+	+
		1991	+ +	+	+	+	+ + +	-	-
		2000	+ +	-	- -	- -	+	-	- -
	Vaihingen	M1	+ + +	+ + +	+ +	+	+ + +	+ + +	+
		M2	+ + +	+ + +	+ +	-	+ + +	+ + +	+ +
		M3	+ + +	+ + +	+ + +	+ +	+ + +	+ + +	+ + +
	Herne	M1	+ + +	+ + +	+ + +	-	+ + +	+ + +	-
		M2	+ + +	+ + +	+ + +	- - -	+ + +	+ +	- - -
	Husum	M1	+ + +	+ + +	+ + +	+ +	+ + +	+ + +	+
		M2	+ + +	+ +	+ + +	+ + +	+ +	+	- -
Nur Änderungsgebiete	Hameln	N1	- -	- -	-	+ +	o	-	+ +
		N2	- - -	- - -	-	+ +	- -	- -	+ +
	Las Vegas	1986	- - -	+	+ +	+ +	+ +	+	+ +
		1991	- -	-	+	+	+ + +	-	-
		2000	-	+	+	+	+	+	+
	Vaihingen	M1	- -	- -	- -	+	-	-	+
		M2	- -	- -	-	+ +	+	-	+ + +
		M3	- -	- -	-	+ + +	+ +	+	+ + +
	Herne	M1	-	-	+	-	-	-	-
		M2	- -	- -	+	+	- -	- -	- -
	Husum	M1	-	-	-	+ +	-	-	-
		M2	-	-	-	+ + +	-	-	-
Änderungsdetektion	Hameln	N1	+ +	+	+	+ +	+ +	+ +	+ +
		N2	+	-	+	+ +	+ +	+ +	+ +
	Las Vegas	1986	+ +	+ + +	+ + +	+ + +	+ +	+	+
		1991	+ +	+	+	+	+ + +	-	-
		2000	+ +	+	- -	-	+	-	-
	Vaihingen	M1	+ + +	+ + +	+ +	+	+ + +	+ + +	+
		M2	+ + +	+ + +	+ + +	+	+ + +	+ + +	+ + +
		M3	+ + +	+ + +	+ + +	+ +	+ + +	+ + +	+ + +
	Herne	M1	+ + +	+ + +	+ + +	-	+ + +	+ + +	-
		M2	+ + +	+ + +	+ + +	- - -	+ + +	+ +	- - -
	Husum	M1	+ + +	+ + +	+ + +	+ +	+ + +	+ + +	+
		M2	+ + +	+ +	+ + +	+ + +	+ +	+	- -

Tabelle 28: Unterschied der Gesamtgenauigkeit bezogen auf das Ergebnis der **RRFK**. Kreis (o): Differenz der GG ist kleiner als 0,1%. Ein Zeichen: Differenz der GG ist kleiner als 2%. Zwei Zeichen: Differenz der GG ist kleiner als 5%. Drei Zeichen: Differenz der GG ist größer als 5%. Plus (+): GG der **RRFK** ist größer. Minus (-): GG der **RRFK** ist kleiner.

Frage a) Berücksichtigung der Fehler in den Trainingslabels

Die Berücksichtigung der Fehler in den Labels beim Training hat in fast allen Experimenten zu einer höheren Gesamtgenauigkeit des gesamten Gebietes geführt. Lediglich bei dem Datensatz *Herne* führte vermutlich die Berücksichtigung des Klassifikationsergebnisses der letzten Iteration zu einer leichten Reduktion der GG. Bei dem Datensatz *Las Vegas* unter Verwendung der Karte von 1986 ist die GG des **RF** ebenfalls etwas größer als die GG des **RRF**. Hier ist das

Klassifikationsergebnis des **RRF** aufgrund der großen Anzahl an fehlerhaften Trainingslabels innerhalb einer Ballung im Merkmalsraum nicht sehr gut.

Auch wenn nur die Änderungsgebiete betrachtet werden, hat sich das Ergebnis bei einigen Experimenten verbessert. Zu erkennen ist dies beispielsweise bei der Klassifikation des Gebietes N2 des Datensatzes *Hameln* oder bei dem Datensatz *Husum*. Obwohl bei dem Training des **RRF** durch die Änderungswahrscheinlichkeiten berücksichtigt wird, ob das Label eines Trainingsbeispiels ggf. fehlerhaft ist, wird bei der Klassifikation nicht zwischen den Pixeln, die in potenziellen Änderungsgebieten liegen und jenen, die sich außerhalb befinden, unterschieden. Indem die fehlerhaften Trainingslabels in dem Training modelliert wurden, verbessert sich somit in einigen Experimenten neben der Gesamtgenauigkeit des gesamten Bildes auch jene in den Änderungsgebieten.

Die Berücksichtigung der Fehler in den Trainingslabels verbessert daher größtenteils das Klassifikationsergebnis.

Frage b) Berücksichtigung der Kartenlabels als weitere Beobachtung

Werden die Kartenlabels in dem CRF bei der Klassifikation berücksichtigt, erhöht sich in den meisten Experimenten die GG des gesamten Bildes. Im Vergleich zu der Klassifikation ohne Berücksichtigung des Kartenlabels ist der Unterschied der GG teilweise sehr groß. Bei den Experimenten mit den Daten aus *Herne* und *Husum* und unter Verwendung der Karte 1 ist die Differenz z.B. über 10%. Auch bei dem *Vaihingen* Datensatz erhöht sich der Median der GG über alle Gebiete um über 10%, wenn die Karte 1, mit den weinigen Änderungen, verwendet wird. In den Änderungsgebieten führt der zusätzliche Einfluss des Kartenlabels meist zu einer leichten Verschlechterung des Klassifikationsergebnisses. Die Abnahme der GG ist jedoch sehr gering, verglichen zu dem prozentualen Anteil an Pixeln mit fehlerhaften Labels in der veralteten Karte, bezogen auf die Gesamtanzahl der Pixel im Bild. Wie zu erwarten ist, nimmt diese Reduktion der GG in den Änderungsgebieten ab und die Verbesserung der GG des gesamten Gebietes zu, wenn die Anzahl an Fehlern in den Labels der veralteten Karte sinkt. Dies konnte bei den Datensätzen *Vaihingen*, *Herne* und *Husum* beobachtet werden. Bei dem Datensatz *Hameln* ist außerdem zu sehen, dass Änderungen an Objektgrenzen aufgrund der Heuristiken zur Detektion der Änderungen schlechter erkannt werden, wenn das Kartenlabel bei der Klassifikation berücksichtigt wird. Dies kann jedoch, abhängig von der Art der Kartenaktualisierung, auch gewünscht sein, z.B. wenn nur neue Gebäude oder Straßen detektiert werden sollen oder wenn von fehlerfreien Objektgrenzen ausgegangen wird.

Zusammenfassend kann festgestellt werden, dass größtenteils das gesamte Klassifikationsergebnis durch die Berücksichtigung der Kartenlabels als weitere Beobachtung verbessert wird. Lediglich in den Änderungsgebieten verschlechtert sich meist das Ergebnis wegen des zusätzlichen Einflusses der Labels aus der Karte. Verglichen zu dem Anteil an Änderungen bezogen auf das gesamte Bild ist die Abnahme jedoch größtenteils sehr gering.

Frage c) Verwendung der Wahrscheinlichkeiten der Klassenzugehörigkeit

Die Verwendung von Wahrscheinlichkeiten der Klassenzugehörigkeit anstatt eines festen Klassenlabels führte ebenfalls bei vielen Experimenten zu einer höheren GG bezogen auf das gesamte Klassifikationsergebnis. Wird das Kartenlabel zum Training verwendet (**KRFK**), ist der Einfluss der Karte sehr groß. Die GG der Pixel in den Änderungsgebieten ist daher in den meisten Experimenten kleiner als die GG unter Verwendung des **RRFK**. Die GG des gesamten Gebietes ist dennoch in nur wenigen Experimenten größer als wenn der **RRFK** verwendet wird. Stammen die Labels zum Training aus dem Ergebnis der vorherigen Iteration (**ERFK**) ist der Einfluss der Karte geringer als bei dem **RRFK**. Werden nur die Pixel in den Änderungsgebieten betrachtet, ist die GG teilweise größer, wenn der **ERFK** verwendet wird. Bezogen auf das gesamte Gebiet ist die GG jedoch in den meisten Experimenten niedriger als die GG des **RRFK**. Außerdem ist in einigen Experimenten zu beobachten, dass die Fehlklassifikationen einer bestimmten Art mit jeder Iteration zunehmen, bis nur noch wenige oder keine Pixel einer bzw. mehrerer Klassen richtig klassifiziert werden.

Zusammenfassend kann gesagt werden, dass sich das gesamte Klassifikationsergebnis meist verbessert, wenn die Wahrscheinlichkeiten der Klassenzugehörigkeit verwendet werden.

Frage d) Vergleich mit der robusten logistischen Regression

Die Frage, ob die Methoden basierend auf dem RRF zu einem besseren Ergebnis führt als die Methoden basierend auf der RLR, kann nicht eindeutig auf Basis der durchgeführten Experimente beantwortet werden.

Bei der Klassifikation niedrig aufgelöster Daten (*Las Vegas*, *Herne*, *Husum*) ist die GG meist größer, wenn die verwendete Methode auf der RLR basiert. Bei den hoch aufgelösten Daten (*Hameln*, *Vaihingen*) hingegen ist die GG der RRF basierten Methoden größer. Der Grund kann, basierend auf den durchgeführten Experimenten, nicht eindeutig genannt werden. Eine Möglichkeit liegt in dem unterschiedlichen Aufbau und somit auch in den verschiedenen Vor- und Nachteilen der Basisklassifikatoren. So kann beispielsweise der RF besser mit mehreren Ballungen im Merkmalsraum umgehen. Dies könnte jedoch auch dazu führen, dass der RF zu viele Ballungen unterscheidet und einigen der Ballungen eine fehlerhafte Klasse zuordnet. Ein weiterer Grund könnten die unterschiedlichen Ansätze zur Modellierungs der Fehler in den Trainingsdaten liefern. Der RRF nutzt ein komplexes Fehlermodell, welches stark auf einigen Heuristiken basiert. Bei dem RLR wird ausschließlich der Zusammenhang zwischen dem Auftreten der Fehler und den Labels modelliert. Vielleicht modelliert das einfache Fehlermodell in der RLR die Fehler bei den Experimenten mit den niedrig aufgelöste Daten besser als das komplexere Fehlermodell in dem RRF.

Zusammenfassend kann festgestellt werden, dass die Methoden basierend auf dem RRF bei den Experimenten mit hoch aufgelösten Daten meist zu besseren Ergebnissen führen.

Frage e) Einfluss der Anzahl an fehlerhaften Labels

Steigt die Anzahl der Fehler in den Trainingslabels sinkt im Allgemeinen die GG des Klassifikationsergebnisses. Diese Beobachtung erfolgte bei allen vier Datensätzen, bei denen mehrere Karten mit unterschiedlicher Anzahl an fehlerhaften Labels zur Verfügung standen. Ist die Anzahl an Fehlern zu groß, wie bei dem Datensatz *Las Vegas* mit der veralteten Karte aus 1986, ist eine Klassifikation mit allen getesteten Methoden stark fehleranfällig. Dennoch ist die GG des gesamten Bildes bei dem Ergebnis des **RRFK** größer als bei dem Ergebnis des **RF**, wenn bei jedem dieser Datensätze die Karte mit den meisten fehlerhaften Pixeln genutzt wird.

Insgesamt ist also ein Anstieg der GG bei sinkender Anzahl an Fehlern in den Trainingslabels festzustellen.

Frage f) Unterschiedliche Auflösung der Daten

In diesem Abschnitt wird darauf eingegangen, ob der vorgestellte Klassifikationsalgorithmus **RRFK** bei hoch und niedrig aufgelösten Datensätzen zu einem besseren Klassifikationsergebnis als der klassische Klassifikator **RF** führt. Wird die GG des gesamten Gebietes betrachtet, ist die GG bei den hoch und niedrig aufgelösten Datensätzen höher, wenn der **RRFK** anstatt des **RF** verwendet wird. In den Änderungsgebieten ist die GG aufgrund des stärkeren Einflusses der Kartenlabels auf das Klassifikationsergebnis bei dem **RRFK** größtenteils etwas geringer. Diese temporale Überanpassung ist jedoch, verglichen zu der ursprünglichen Anzahl an fehlerhaften Labels, sehr gering.

Wie bereit bei Frage d) beschrieben, ist die GG der Klassifikation von niedrig aufgelösten Daten meist größer, wenn die Klassifikationsmethode auf der RLR anstatt des RRF basiert. Bei hoch aufgelösten Daten ist die GG der der RRF basierten Methoden jedoch größer. Mögliche Gründe wurden ebenfalls bei der Beantwortung von Frage d) genannt.

Zusammenfassend kann gesagt werden, dass die GG des gesamten Klassifikationsergebnisses unabhängig von der Auflösung der Daten bei dem **RRFK** höher ist als bei dem **RF**. Die Methoden basierend auf der RLR führen, verglichen zu den Methoden basierend auf dem RRF, bei niedrig aufgelösten Daten meist zu einer höheren GG des Klassifikationsergebnisses.

Fazit

Im Allgemeinen kann aus den Experimenten geschlossen werden, dass die in dieser Arbeit vorgestellten Entwicklungen das Klassifikationsergebnis verbessern.

Besonders der **RRFK** zeigt meist ein deutlich besseres Gesamtergebnis als der **RF**. Die GG von dem Ergebnis der Klassifikation und von der Änderungsdetektion sind meist mehr als 5% größer, wenn der **RRFK** anstelle des **RF** verwendet wird. Auch verglichen zu den Ergebnissen der anderen Methoden ist das Gesamtergebnis des **RRFK** meist besser. Bei den niedrig aufgelösten Daten ist teilweise das Ergebnis der **RLRK** etwas besser.

In den Änderungsgebieten ist die GG bei dem **RF** etwas größer als bei dem **RRFK**, aufgrund des geringeren Einfluss der Kartenlabels auf das Ergebnis. Die Verschlechterung in den Änderungsgebieten ist jedoch im Verhältnis zu der Verbesserung des gesamten Klassifikationsergebnisses und im Verhältnis zu der Anzahl an fehlerhaften Pixeln in der Karte gering.

Die Detektion von Änderungen ist mit dem **RRFK** ebenfalls meist deutlich besser als mit anderen Klassifikatoren.

Problematisch ist das Training des **RRFK**, wenn ein Großteil der Trainingsbeispiele, die eine Ballung im Merkmalsraum bilden, ein fehlerhaftes Labels besitzt. Dann ist jedoch auch das Training mit anderen Methoden fehleranfällig.

Zusammenfassend kann aber gesagt werden, dass das Klassifikationsergebnis bezogen auf das gesamte Bild und auch die Änderungsdetektion bei der Verwendung des neu entwickelten **RRFK** deutlich (meist mehr als 5%) besser ist als bei dem klassischen **RF**.

6.4. Multitemporale Klassifikation

6.4.1. Hameln

Abb. 61 zeigt die GG der Experimente bezogen auf den Datensatz *Hameln*. Tabelle 29 zeigt die GG und den MF1. Weitere Genauigkeitsmaße sind in Tabellen B.19, B.20 und B.21 in Anhang B zu finden.

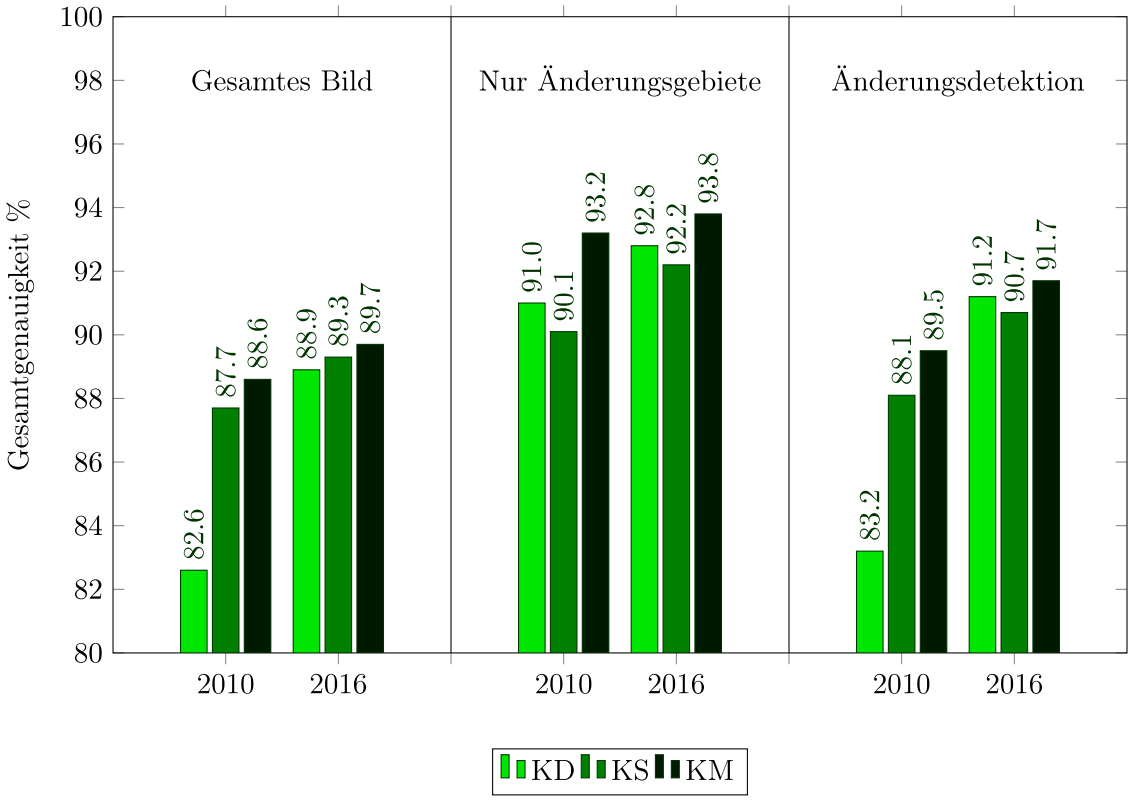


Abbildung 61: Gesamtgenauigkeit [%]. Datensatz: *Hameln*, Multitemporal, Karte: 1955.

Methode	Jahr der Daten	Gesamtes Bild		Nur Änderungen		Änderungsdetektion	
		GG	MF1	GG	MF1	GG	MF1
KD	2010	82,6	70,6	91,0	72,3	83,2	74,5
	2016	88,9	78,7	92,8	79,1	91,2	87,6
KS	2010	87,7	75,3	90,1	69,6	88,1	79,2
	2016	89,3	79,6	92,2	78,0	90,7	86,3
KM	2010	88,6	77,9	93,2	79,3	89,5	83,7
	2016	89,7	80,7	93,8	81,6	91,7	88,4

Tabelle 29: Genauigkeitsmaße [%]. Datensatz: *Hameln*.

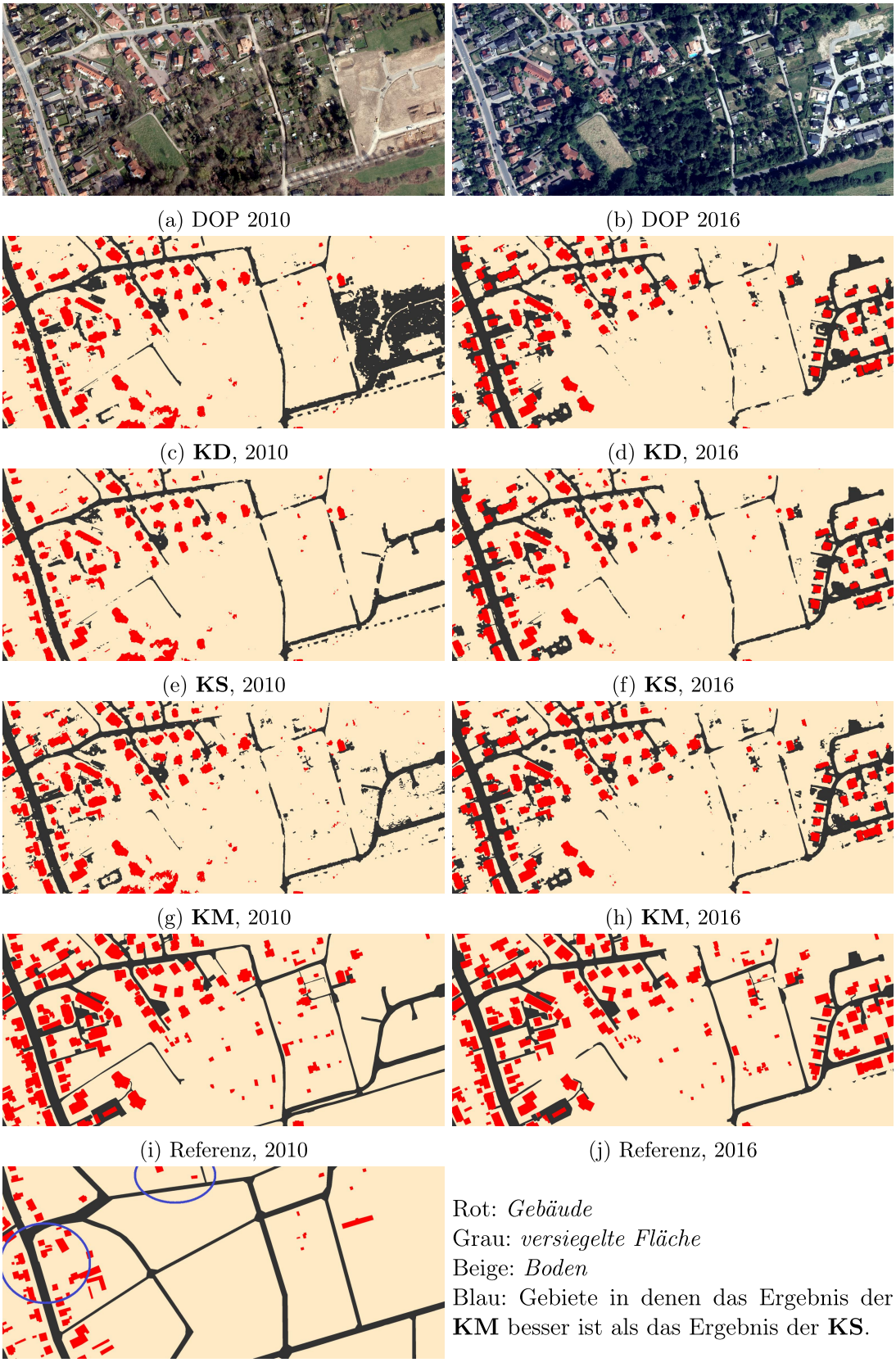
Frage a) Verwendung von veralteten Fernerkundungsdaten

Zunächst wird die direkte Klassifikation **KD** mit der sequenziellen verglichen, um zu untersuchen, ob sich das Klassifikationsergebnis durch die Einbeziehung veralteter Fernerkundungsdaten verbessert (Frage a).

Die GG und der MF1 Wert, bezogen auf das gesamte Bild, sind deutlich geringer bei der direkten Klassifikation als bei der sequenziellen Klassifikation. Bei der Betrachtung weiterer Genauigkeitsmaße ist die deutlich kleinere Korrektheit der Klasse *versiegelte Fläche* von 52,7% bei der direkten Klassifikation zu erkennen, verglichen zu der Korrektheit von 89,2% bei der sequenziellen Methode. Ein visueller Vergleich (Abb. 62) zeigt, dass dies vor allem an dem Neubaugebiet auf der rechten Seite des Bildes liegt. Dieser Bereich ist 2010 zwar noch nicht bebaut, aber das Erscheinungsbild, und somit auch die Merkmale, der Bodenpixel unterscheidet sich stark von denen anderer Pixel der Klasse *Boden*. Demnach ist eine Fehlklassifikation dieser Pixel wahrscheinlich. Der Grund, warum die Pixel bei der **KD** fehlerhaft der Klasse *versiegelte Fläche* und bei der **KS** korrekt der Klasse *Boden* zugeordnet wurden, kann z.B. die Wahl der Trainingsdaten sein. Aufgrund der Größe des Gebietes NA konnten nur 0,01% der Daten zum Training verwendet werden. Die Trainingsbeispiele wurden zufällig ausgewählt, sodass sich vermutlich ein Großteil der verwendeten Trainingsbeispiele der direkten Klassifikation und der sequenziellen Klassifikation unterschieden. Eine weiterer Grund könnten die vielen auf dem Zufall beruhenden Elemente des RF Klassifikators sein.

Die Klassifikationsergebnisse der **KD** und der **KS**, bezogen auf die Daten aus 2016, unterscheiden sich kaum. Die Differenz des MF1 beträgt maximal 1,4%, die der GG nur maximal 0,7%. Wird berücksichtigt, wie stark sich die GG allein aufgrund der Nutzung verschiedener Mengen von Trainingsdaten bzw. der Zufallselemente des RF ändern kann, sind die Unterschiede vernachlässigbar.

Zusammenfassend ist das Ergebnis bei der Klassifikation der aktuellen Daten aus 2016 unter Verwendung der Daten aus 2010 nicht signifikant besser als die direkte Klassifikation der Daten aus 2016.



(a) DOP 2010 (b) DOP 2016

(c) **KD**, 2010 (d) **KD**, 2016

(e) **KS**, 2010 (f) **KS**, 2016

(g) **KM**, 2010 (h) **KM**, 2016

(i) Referenz, 2010 (j) Referenz, 2016

(k) Karte 1955

Rot: *Gebäude*
Grau: *versiegelte Fläche*
Beige: *Boden*
Blau: Gebiete in denen das Ergebnis der **KM** besser ist als das Ergebnis der **KS**.

Abbildung 62: Ausschnitt aus dem Gebiet NA des Datensatzes *Hameln*, in dem sich auch Gebiet N1 befindet, für welchen eine Referenz vorliegt.

Frage b) Integration von temporalem Kontext

Anders sieht es bei dem Vergleich der Ergebnisse der sequenziellen Klassifikation (**KS**) und der multitemporalen Klassifikation (**KM**) aus (Frage b). Bei der multitemporalen Klassifikation werden die Trainingsergebnisse der sequenziellen Klassifikation verwendet. Die Wahl der Trainingsbeispiele und das Trainingsergebnis des RRF sind demnach identisch. Der einzige Unterschied besteht in der Klassifikation mit dem multitemporalen CRF, die ausschließlich bei der **KM** erfolgt.

Die GG der **KM** ist etwas größer als die der **KS**. Der Unterschied bezogen auf das gesamte Gebiet ist mit einer Differenz von 0,8% (Daten aus 2010) bzw. 0,4% (Daten aus 2016) relativ gering. Werden nur die Änderungen betrachtet ist die GG der **KM** mit 93,2% (Daten aus 2010) und 93,8% (Daten aus 2016) deutlich größer als die GG der **KS** mit 90,1% (Daten aus 2010) und 92,2% (Daten aus 2016). Auch die GG der Änderungsdetektion steigt von 88,1% (Daten aus 2010) und 90,7% (Daten aus 2016) mit der **KS** auf 89,5% (Daten aus 2010) und 91,7% (Daten aus 2016) mit der **KM**.

Tabelle 30 zeigt die Differenzen der Vollständigkeit, der Korrektheit und des MF1 Wertes zwischen dem Ergebnis der **KM** und der **KS**. Wird das gesamte Gebiet betrachtet, verbessert sich besonders die Vollständigkeit der Klassen *Gebäude* und *versiegelte Fläche*. In den blau markierten Bereichen in Abb. 62k ist die bessere Detektion der beiden Klassen auch zu erkennen. In den Änderungsgebieten verbessert sich die Vollständigkeit der Klassen *Boden* und *versiegelte Fläche* und die Korrektheit der Klasse *Boden*. Die Korrektheit der Klasse *versiegelte Fläche* nimmt in den Änderungsgebieten und im gesamten Gebiet etwas ab.

Zusammenfassend kann eine Verbesserung des Klassifikationsgebietes durch die Integration von temporalem Kontext festgestellt werden. Die GG steigt dabei sowohl unter Berücksichtigung aller Pixel, als auch in den Änderungsgebieten. Die GG der Änderungsdetektion ist bei der multitemporalen Klassifikation ebenfalls etwas höher.

Insgesamt verbessert die Berücksichtigung von temporalem Kontext demnach bei dem Datensatz *Hameln* die Klassifikation der Daten aus 2010 und 2016.

	Jahr der Daten	Vollständigkeit			Korrektheit			F1		
		Geb.	Bod.	Vers.	Geb.	Bod.	Vers.	Geb.	Bod.	Vers.
Gesamtes Gebiet	2010	9,0	-1,4	7,7	0,1	2,7	-11,6	7,0	0,8	-0,1
	2016	7,7	-0,9	1,1	1,6	2,0	-7,4	6,3	0,6	-3,6
Nur Änderungsgebiete	2010	1,4	6,3	22,9	0,1	12,5	-1,3	0,8	13,3	15,0
	2016	1,2	3,0	5,0	0,1	10,7	-3,9	0,7	9,6	0,7

Tabelle 30: Differenz der Genauigkeitsmaße zwischen **KS** und **KM**: $VG(\mathbf{KM}) - VG(\mathbf{KS})$, $KG(\mathbf{KM}) - KG(\mathbf{KS})$, $F1(\mathbf{KM}) - F1(\mathbf{KS})$. Datensatz: *Hameln*.

6.4.2. Las Vegas

Abb. 63 zeigt die GG der Experimente zur Analyse der multitemporalen Klassifikation bezogen auf den *Las Vegas* Datensatz. Die MF1 Werte der Experimente unterscheiden sich deutlich

stärker als die GG, wenn nur die Pixel in Änderungsgebieten betrachtet werden. In Abb. 64 sind daher die MF1 Werte abgebildet. In Tabelle 31 sind die GG und der MF1 Wert gegenübergestellt. In den Anhang B (Tabellen B.22, B.23 und B.24) sind weitere Genauigkeitswerte zu finden.

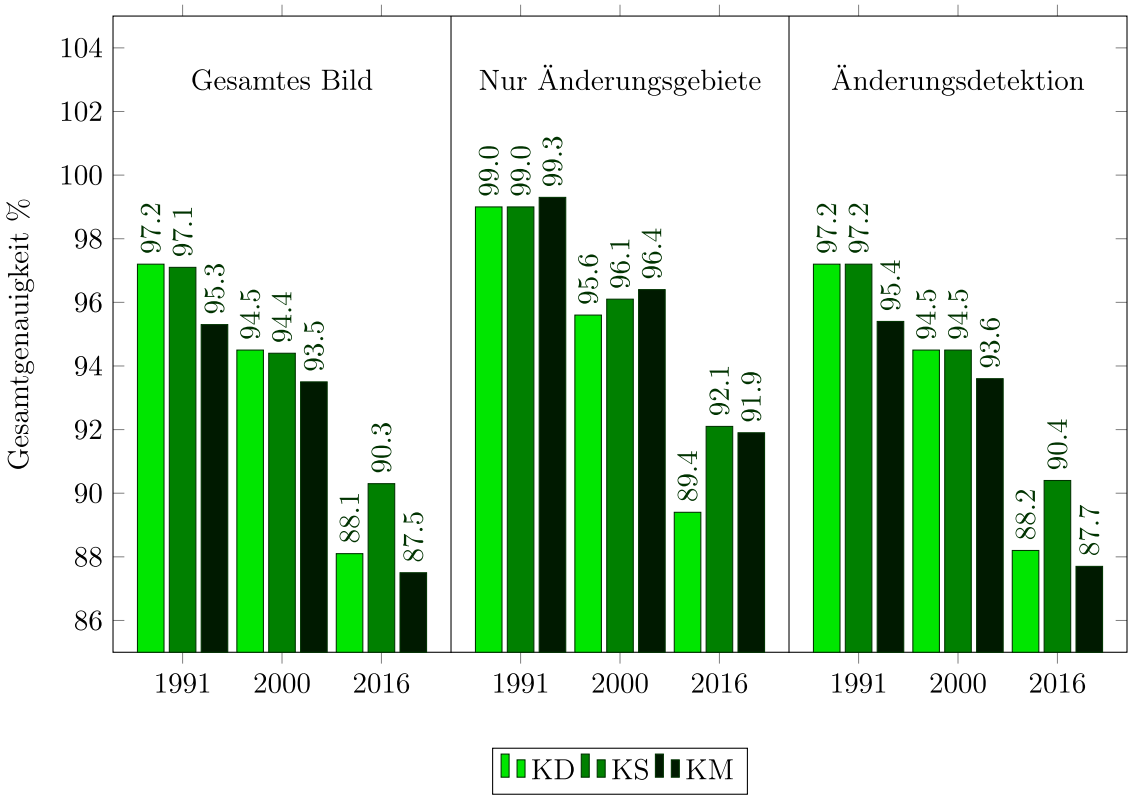


Abbildung 63: Gesamtgenauigkeit [%]. Datensatz: *Las Vegas*, Multitemporal, Karte: 1986.

Methode	Jahr der Daten	Gesamtes Bild		Nur Änderungen		Änderungsdetektion	
		GG	MF1	GG	MF1	GG	MF1
KD	1991	97,2	92,4	99,0	59,0	97,2	73,7
	2000	94,5	90,5	95,6	61,4	94,5	77,9
	2016	88,1	81,2	89,4	52,2	88,2	68,3
KS	1991	97,1	92,2	99,0	59,5	97,2	73,7
	2000	94,4	90,7	96,1	64,4	94,5	79,2
	2016	90,3	85,7	92,1	59,4	90,4	77,3
KM	1991	95,3	88,8	99,3	69,8	95,4	70,1
	2000	93,5	89,4	96,4	80,0	93,6	78,0
	2016	87,5	78,9	91,9	63,1	87,7	73,2

Tabelle 31: Genauigkeitsmaße [%]. Datensatz: *Las Vegas*.

Frage a) Verwendung von veralteten Fernerkundungsdaten

Zunächst wird untersucht, ob das Klassifikationsergebnis einer sequenziellen Klassifikation besser ist, verglichen zu einer direkten Klassifikation (Frage a). Da bei der Klassifikation der Daten aus 1991 noch keine veralteten Fernerkundungsdaten früherer Zeitpunkte vorliegen, ist das Klassifikationsergebnis relativ gleich und die GG ändert sich nur minimal (weniger als

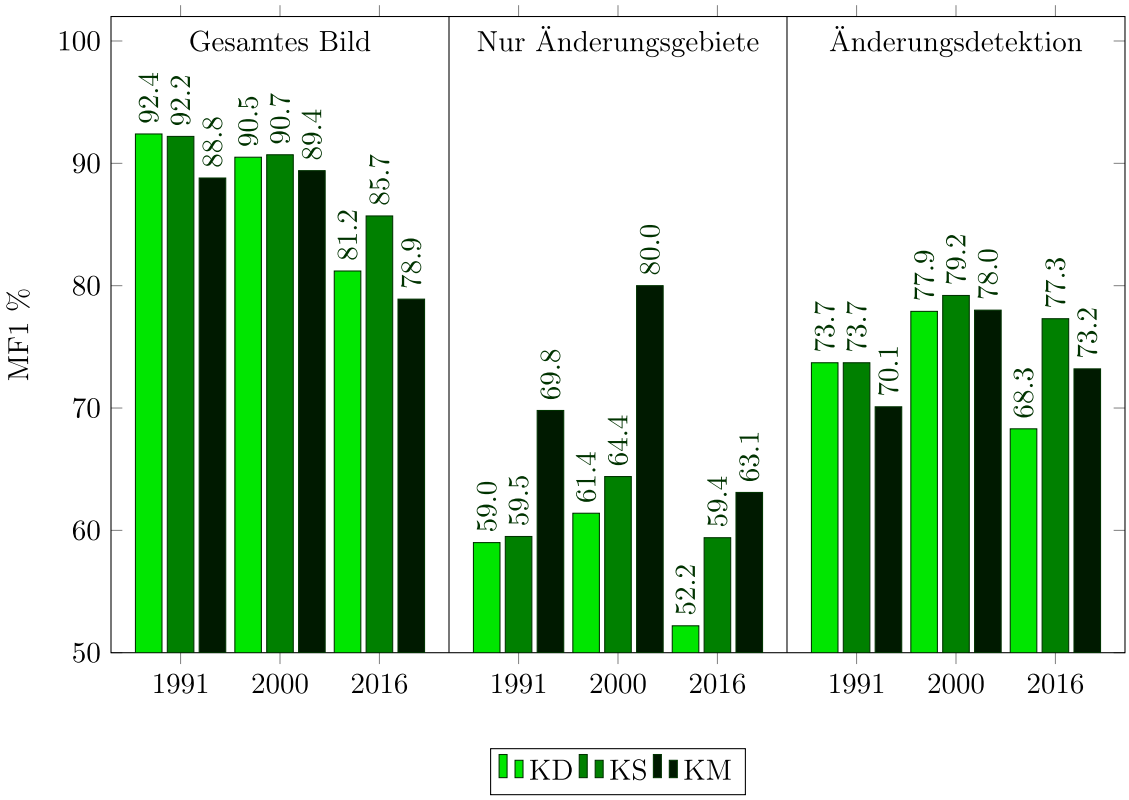


Abbildung 64: MF1 [%]. Datensatz: *Las Vegas*, Multitemporal, Karte: 1986.

0,08%). Bei der Klassifikation der Fernerkundungsdaten aus 2000 und 2016 ist ein Anstieg des MF1 Wertes festzustellen. Besonders wenn nur die Pixel in den Änderungsgebieten betrachtet werden, beträgt die Differenz des MF1 Wertes zwischen dem Ergebnis des **KS** und **KD** 3,0% (Daten aus 2000) bzw. 7,3% (Daten aus 2016). Bei der Betrachtung weiterer Genauigkeitsmaße ist zu erkennen, dass die Vollständigkeit der Klasse *Bebauung* deutlich größer bei der sequenziellen Klassifikation im Vergleich zu der direkten Klassifikation ist. Die Differenz der Vollständigkeit bezogen auf das gesamte Gebiet beträgt 13,2% für die Klassifikation der Daten aus 2016 und 2,9% für die Klassifikation der Daten aus 2000. In den Änderungsgebieten beträgt sie sogar 19,3% (Daten aus 2016) bzw. 5,7% (Daten aus 2000). Abb. 65 zeigt die Ergebnisse der Daten aus 2000 und Abb. 66 die Ergebnisse aus 2016. Vor allem bei den Klassifikationsergebnissen der Daten aus 2016 ist die deutlich bessere Klassifikation der Pixel der Klasse *Bebauung* mit der **KS** als mit der **KD** zu erkennen. Dies liegt vor allem an dem sehr großen Zuwachs der Bebauung. Wie in Kapitel 6.3.2 bereits beschrieben wurde, stellt die direkte Klassifikation der Daten aus 2016 mit der Karte aus 1986 eine besondere Herausforderung dar. Ein Großteil der Pixel, die in 2016 der Klasse *Bebauung* angehören, werden in der Karte aus 1986 noch mit dem Label *Boden* ausgewiesen. Bei der direkten Klassifikation besitzen demnach mehr Trainingsdaten, die in der Referenz der Klasse *Bebauung* angehören, das fehlerhafte Label *Boden* als das korrekte Label. Bei der sequenziellen Klassifikation hingegen werden zunächst die Daten aus 1991 mit den Labels der Karte aus 1986 klassifiziert. Diese Labels werden anschließend für die Klassifikation der Daten aus 2000 verwendet und die daraus resultierenden Labels werden wiederum für die Klassifikation der Daten aus 2016 genutzt. Der Anteil der Trainingsdaten mit der wahren Klasse *Bebauung*, bei denen jedoch fälschlicherweise das Label *Boden* beobachtet wurde, nimmt demnach ab.

Tabelle 32 soll das Ungleichgewicht zwischen den Trainingsdaten der wahren Klasse *Bebauung* mit korrektem und falschem Label verdeutlichen. Hierfür wird in dieser Tabelle sowohl der der Anteil an Pixeln mit der wahren Klasse *Bebauung* und dem beobachtetem Label *Boden*, als auch der Anteil der Pixeln, die der Klasse *Bebauung* angehören und in der Karte das korrekte Label *Bebauung* aufweisen, gezeigt. Der Anteil bezieht sich hierbei auf die gesamte Anzahl an Pixeln. Es ist zu erkennen, dass der Anteil an Pixeln mit dem Label *Boden* in der Karte 1986 und dem Label *Bebauung* in der Referenz aus 2016 mit 14,0% deutlich größer ist als der Anteil von 5,4% an Pixeln, die 1986 und 2016 das Label *Bebauung* besitzen. Werden benachbarte Zeitpunkte betrachtet ist der Anteil an Pixeln, die in der Karte und der Referenz der Klasse *Bebauung* angehören größer als der Anteil der Pixel, die fehlerhaft das Label *Boden* in der Karte vorweisen. Es kann demnach davon ausgegangen werden, dass bei der sequenziellen Klassifikation ein Großteil der Trainingsdaten der Klasse *Bebauung* ein korrektes Label aufweisen. Dennoch ist anzumerken, dass die Zahlen in der Tabelle 32 mit den manuell erstellten Referenzen bestimmt wurden, während bei dem Training das Klassifikationsergebnis der vorherigen Iteration verwendet wird.

Jahr der Karte	Jahr der Daten (Referenz)	Ref.: *Bebauung* Karte: *Bebauung*	Ref.: *Bebauung* Karte: *Boden*
1986	2016	5,4%	14,0%
1986	1991	5,3%	1,9%
1991	2000	7,2%	6,3%
2000	2016	13,4%	6,0%

Tabelle 32: Anteil der Pixel mit dem Label *Bebauung* in der Rerferenz und dem Label *Bebauung* (3. Spalte) bzw. *Boden* (4. Spalte) in der Karte bezogen auf die gesamte Anzahl an Pixeln. Datensatz: *Las Vegas*.

Zusammenfassend kann festgestellt werden, dass sich der MF1 Wert größtenteils verbessert, wenn die Daten aus weiteren Zeitpunkten verwendet werden.

Frage b) Integration von temporalem Kontext

Als nächstes werden die Auswirkungen von temporalem Kontext auf das Klassifikationsergebnis untersucht (Frage b). Bezogen auf das gesamte Gebiet ist ein Abfall der GG zu beobachten, wenn temporaler Kontext bei der Klassifikation berücksichtigt wird. Auch die GG der Änderungsdetektion nimmt etwas ab. Bei den Klassifikationsergebnissen der Daten aller Jahre ist die Korrektheit der Klasse *Bebauung* bezogen auf das gesamte Gebiet gesunken. Auch in den Ergebnissbildern in Abb. 65 und Abb. 66 sind deutlich mehr Pixel zu erkennen, die fehlerhalft als *Babauung* klassifiziert wurden. Dies liegt vermutlich an dem geringeren Einfluss des Kartenlabels. Während bei der direkten und sequenziellen Klassifikation ausschließlich das Kartenlabel, das Label räumlicher Nachbarn und die Daten einen Einfluss auf das Klassifikationsergebnis ausüben, kommt bei der multitemporalen Klassifikation zusätzlich der Einfluss temporaler Nachbarn hinzu. In dem multitemporalen CRF gibt es auch keine Verbindung mehr von den Daten der Jahre 2000 und 2016 zu den beobachteten Kartenlabels. Der direkte Einfluss des Kartenlabels auf das Klassifikationsergebnis ist somit, verglichen zu der direkten

oder sequenziellen Methode, geringer. Bei der multitemporalen Klassifikation der Daten aus 2016 ist neben der Korrektheit der Klasse *Bebauung* auch die Korrektheit der Klasse *Wasser* gesunken. Auch dies liegt an dem geringeren Einfluss des Kartenlabels. Abb. 67 zeigt die Klassifikationsergebnisse eines Ausschnitts der *Las Vegas* Daten. Dort sind deutlich die Fehlklassifikationen der Bodenpixel zu erkennen. Ein visueller Vergleich mit dem DOP zeigt, dass die Fehlklassifikationen besonders in Bereichen auftreten, die aufgrund des Geländes im Schatten liegen. Auch in den monotemporalen Klassifikationen in Kapitel 6.3.2 mit den Methoden ohne oder mit nur geringem Einfluss des Kartenlabels (**RF** und **RRF**) wurden Bodenpixel den Klassen *Bebauung* und *Wasser* zugeordnet.

Neben der Korrektheit ist auch die Vollständigkeit der Klasse *Bebauung* bei der multitemporalen Klassifikation der Daten aus 2016 geringer als bei der sequenziellen. Auch das kann an dem geringeren Einfluss des Kartenlabels liegen. Besonders bei der direkten monotemporalen Klassifikation mit dem **RRF** und dem **RRFK** ist zu erkennen, dass die Pixel der Klasse *Bebauuung*, die bereits in der Karte dieses Label aufweisen, besser erkannt werden, wenn der Einfluss der Karte, wie bei dem **RRFK**, größer ist (s. auch Abb. 49).

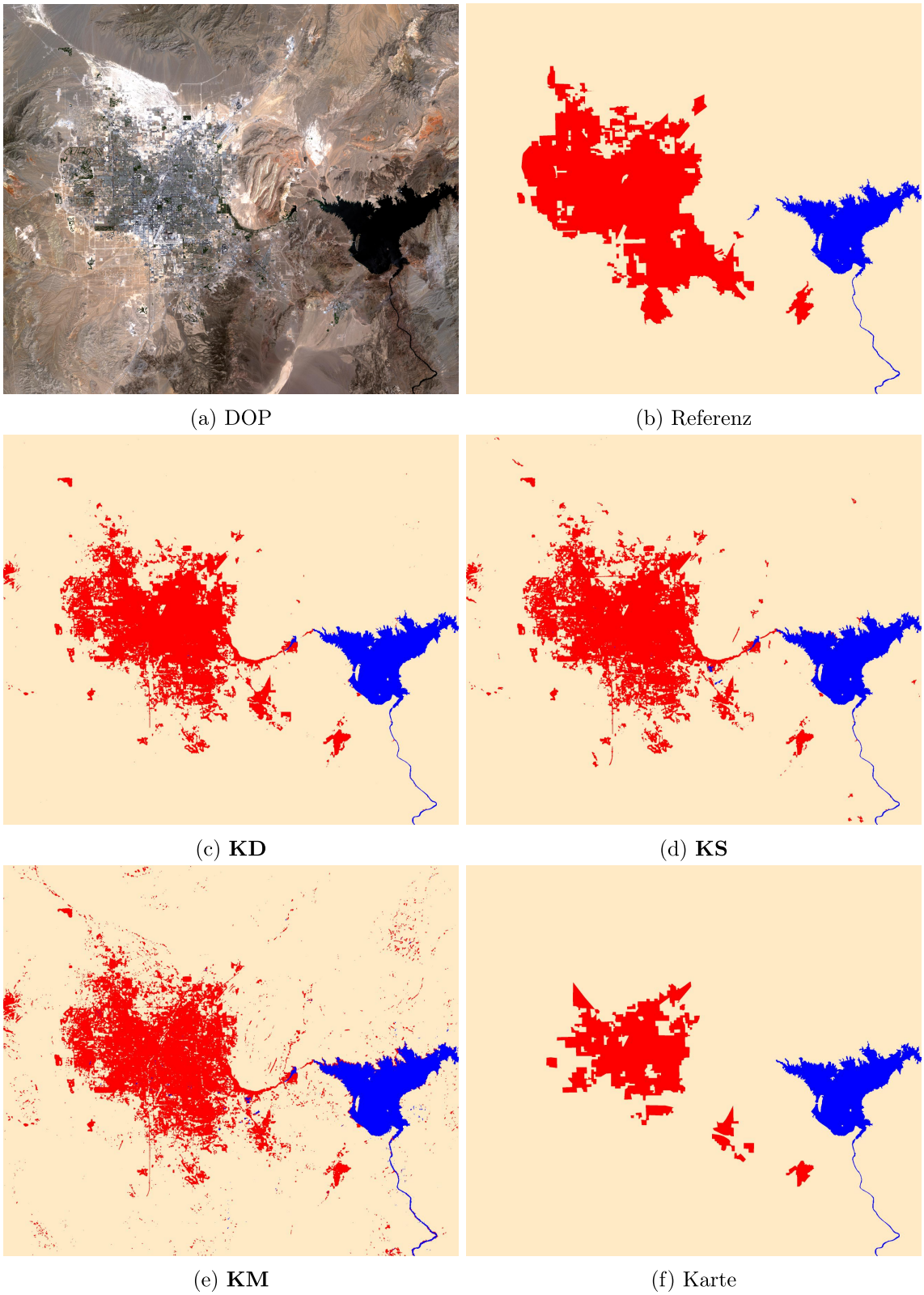
Der MF1 Wert der multitemporalen Klassifikation ist in den Änderungsgebieten hingegen deutlich größer (Abb. 64) als der MF1 Wert der sequenziellen Klassifikation. So steigt der MF1 Wert von 59,5% auf 69,8% (1991), von 64,4% auf 80,0% (2000) und von 59,4% auf 63,1% (2016). Dies liegt vor allem an der besseren Detektion der Klasse *Wasser* in den Änderungsgebieten. So ist die Vollständigkeit und Korrektheit dieser Klasse größer bei den Ergebnissen der **KM** als bei denen der **KS**. Auch dies könnte an dem geringeren Einfluss der Kartenlabels liegen, da die Kartenlabels in den Änderungsgebieten fehlerhaft sind. Jedoch ist die Anzahl der Pixel mit der Klasse *Wasser* in den Änderungsgebieten sehr gering. Die GG ändert sich daher nicht so stark. Hinzu kommt, dass der F1 Wert der Klassen *Boden* und *Bebauuung*, bezogen auf die Pixel in den Änderungsgebieten, bei dem Ergebnis der **KM** für 2016 kleiner ist als bei dem Ergebnis der **KS** für 2016. Die GG in den Änderungsgebieten für das Klassifikationsergebnis der **KM** aus dem Jahr 2016 ist demnach auch etwas geringer als für das Klassifikationsergebnis der **KS** aus dem Jahr 2016.

Zusammenfassend kann festgestellt werden, dass das Klassifikationsergebnis bezogen auf das gesamte Gebiet durch die Hinzunahme von multitemporalem Kontext abnimmt. Dies liegt vermutlich an dem sinkenden Einfluss der Kartenlabels.

Insgesamt kann für den *Las Vegas* Datensatz festgehalten werden, dass die Verwendung von weiteren veralteten Fernerkundungsdaten das Klassifikationsergebnis des aktuellsten Zeitpunktes (2016) deutlich verbessert. Durch die Hinzunahme des temporalen Kontextes wird das Klassifikationsergebnis jedoch nicht weiter verbessert, da die Fehlklassifikationen besonders in den Bereichen ohne Änderungen leicht zunehmen.

6.4.3. Diskussion

Die Diskussion in diesem Kapitel orientiert sich an den Fragen aus Kapitel 5.6.4.

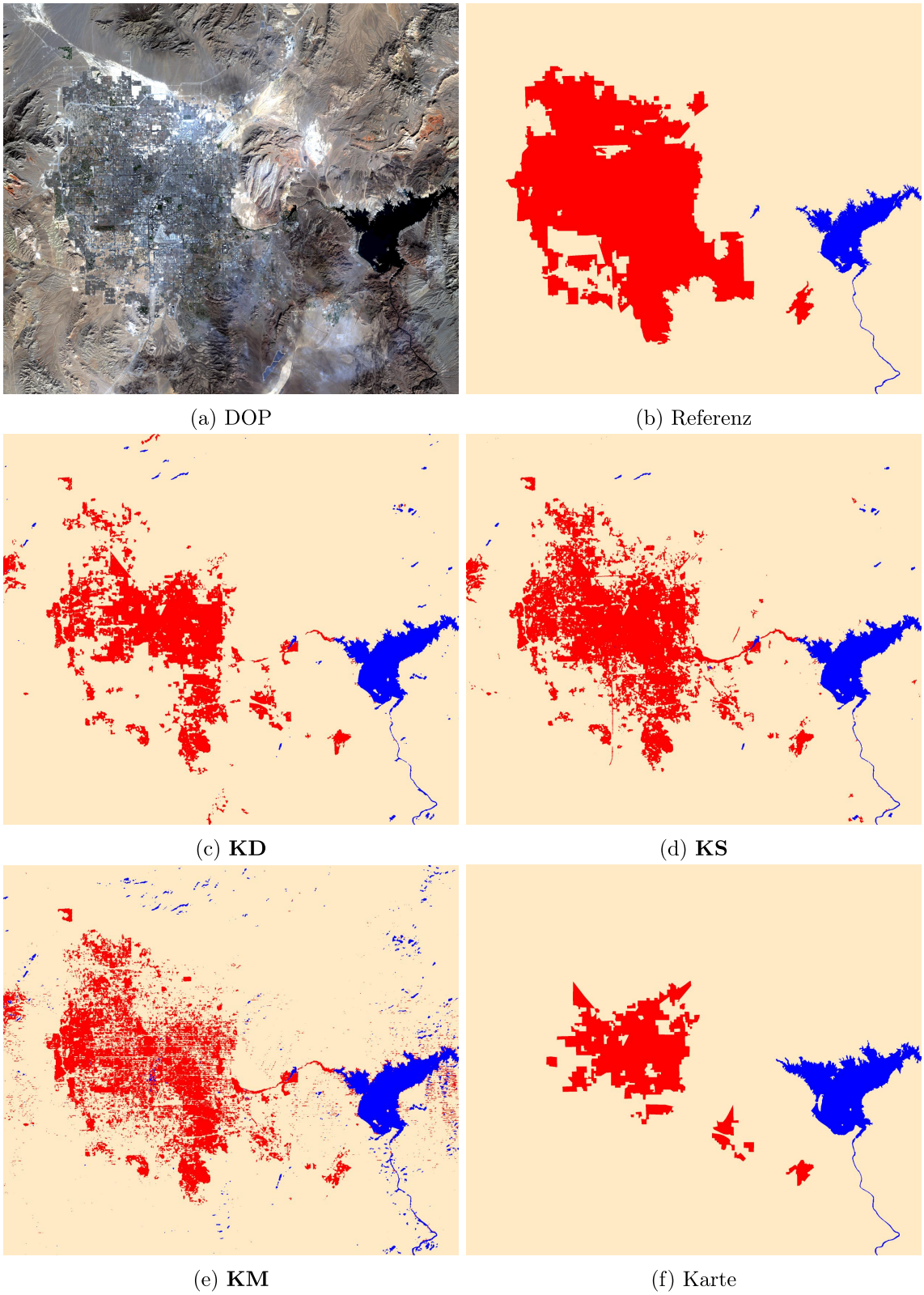


(a) DOP (b) Referenz

(c) **KD** (d) **KS**

(e) **KM** (f) Karte

Abbildung 65: DOP, Referenz und Ergebnisse für das Jahr 2000 des Datensatzes *Las Vegas* und die verwendete Karte aus dem Jahr 1986.
Rot: *Bebauung*. Blau: *Wasser*. Beige: *Boden*.



(a) DOP (b) Referenz

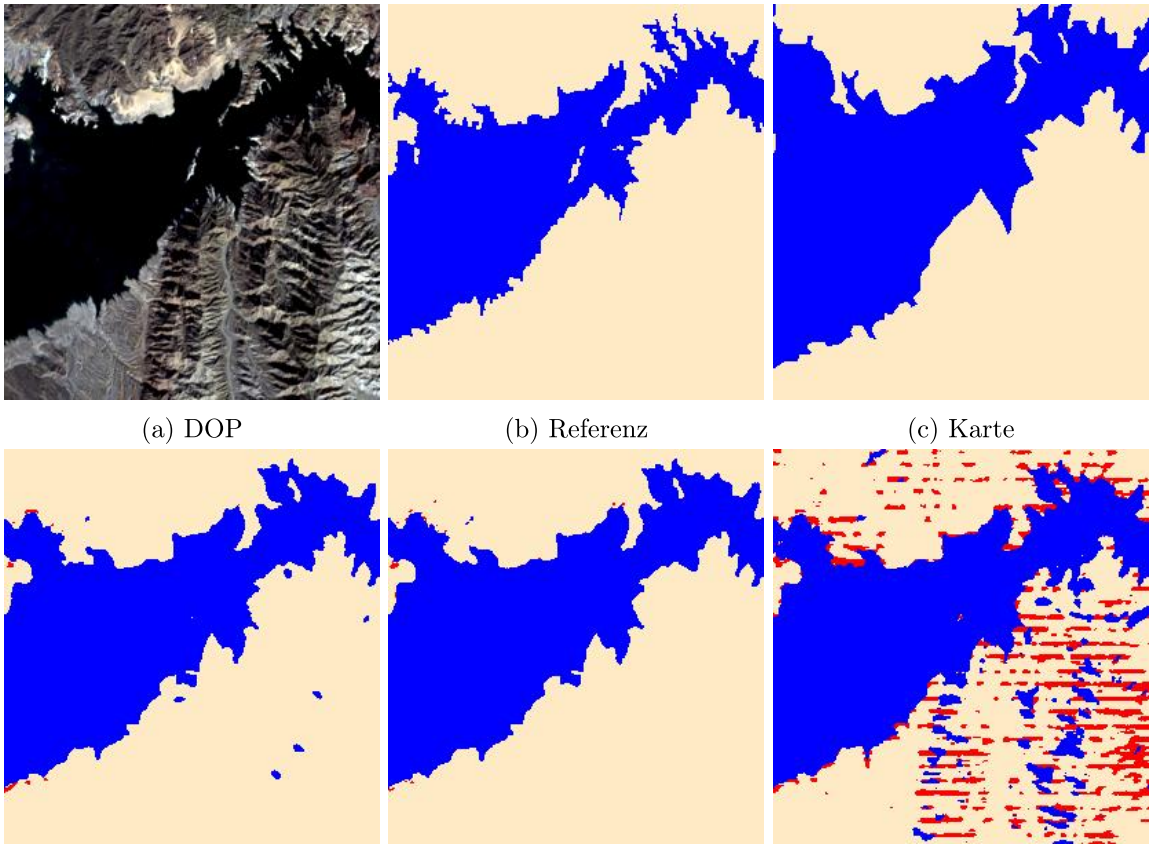
(c) **KD** (d) **KS**

(e) **KM** (f) Karte

Abbildung 66: DOP, Referenz und Ergebnisse für das Jahr 2016 des Datensatzes *Las Vegas* und die verwendete Karte aus dem Jahr 1986.
Rot: *Bebauung*. Blau: *Wasser*. Beige: *Boden*.

Frage a) Verwendung von veralteten Fernerkundungsdaten

Die Experimente mit dem Datensatz *Las Vegas* zeigen, dass die sequenzielle Klassifikation unter Verwendung von veralteten Fernerkundungsdaten besonders bei einer großen Anzahl an



(a) DOP (b) Referenz (c) Karte

(d) **KD** (e) **KS** (f) **KM**

Abbildung 67: Ausschnitt des DOP, der Referenz und der Ergebnisse für das Jahr 2016 aus dem Datensatz *Las Vegas* und der Ausschnitt der Karte aus dem Jahr 1986. Rot: *Bebauung*. Blau: *Wasser*. Beige: *Boden*.

fehlerhaften Pixeln das Ergebnis verbessern kann. Bei dem Ergebnis des Jahres 2016 steigt beispielsweise der MF1 Wert bezogen auf das gesamte Bild um 4,5% und bezogen auf die Änderungsgebiete um 7,2%. Der MF1 Wert der Änderungsdetektion verbessert sich sogar um 9,0%.

Frage b) Integration von temporalem Kontext

Die zusätzliche Modellierung von multitemporalem Kontext führte zu einem geringeren Einfluss der Kartenlabels. Bei dem Datensatz *Las Vegas* sank daher die GG des gesamten Gebietes bei der Integration von multitemporalem Kontext. Die Genauigkeitsmaße unter Verwendung der Pixel in Änderungsgebieten sind jedoch teilweise etwas größer. Bei dem Datensatz *Hameln* stieg die GG sowohl bezogen auf das gesamte Gebiet, als auch bezogen auf die Änderungsgebiete. Der MF1 bezogen auf das gesamte Klassifikationsergebnis aus 2016 verbessert sich beispielsweise um 1,1%. In den Änderungsgebieten steigt der MF1 Wert sogar um 3,6%. Auch der MF1 Wert der Änderungsdetektion wird um 2,1% größer.

Die Integration von temporalem Kontext kann demnach zu der Verbesserung des Klassifikationsergebnisses führen. Besonders in den Änderungsgebieten steigt der MF1 Wert bei beiden

Datensätzen. Werden alle Pixel betrachtet kann der MF1 Wert durch den geringeren Einfluss der Kartenlabels auch sinken.

Fazit

Zusammenfassend kann gesagt werden, dass die Berücksichtigung von multitemporalem Kontext bei dem *Hameln* Datensatz zu einer Verbesserung des Klassifikatuionsergebnisses führte. Das Potenzial einer multitemporalen Klassifikation konnte somit gezeigt werden. Die Experimente mit dem *Las Vegas* Datensatz zeigten jedoch noch Verbesserungsbedarf bei der Umsetzung. Besonders der Einfluss der Kartenlabels war relativ gering. Gegebenenfalls muss auch das temporale Interaktionspotenzial noch angepasst werden, um auch hier eine Verbesserung der Klassifikation zu erzielen. Die Verwendung von veralteten Fernerkundungsdaten durch eine sequenzielle Klassifikation hat jedoch die Klassifikation bei dem *Las Vegas* Datensatz verbessert.

Insgesamt kann festgehalten werden, dass die Verwendung von zusätzlichen Fernerkundungsdaten aus älteren Zeitpunkten das Klassifikationsergebnis verbessern kann, wenn ein Großteil der Trainingsdaten in einer Ballung im Merkmalsraum fehlerhafte Labels besitzt. Dies ist an dem *Las Vegas* Datensatz zu erkennen. Die Integration von temporalem Kontext hat die Klassifikation vor allem in den Änderungsgebieten verbessert. Die multitemporale Klassifikation kann, verglichen zu der sequenziellen Klassifikation, zu einem etwas schlechteren Klassifikationsergebnis führen, da die Kartenlabels dabei einen geringeren Einfluss auf das Ergebnis ausüben. Wie an dem *Hameln* Datensatz zu erkennen ist, kann sich das Ergebnis unter Berücksichtigung aller Pixel jedoch auch verbessern, wenn multitemporaler Kontext integriert wird.

Besonders an dem Datensatz *Hameln* ist also das Potenzial der Verwendung von multitemporalem Kontext zu erkennen. Bei dem *Las Vegas* Datensatz ist eher der Vorteil der Verwendung von Daten aus weiteren Zeitpunkten zu erkennen.

7. Fazit und Ausblick

7.1. Fazit

In dieser Arbeit wurde ein multitemporaler Klassifikator zur Aktualisierung von veralteten Karten vorgestellt, der mit Hilfe der fehlerhaften Labels aus der Karte trainiert wird. Zeit- und kostenintensives manuelles Labeln der Trainingsdaten kann damit vermieden werden.

Die aus der veralteten Karte extrahierten Trainingslabels sind aufgrund der Zeitdifferenz zwischen dem Zeitpunkt der Erstellung der Karte und der Fernerkundungsdaten fehlerhaft. Der verwendete Klassifikator muss demnach mit fehlerhaften Labels im Training umgehen können. Ein passender Klassifikator aus der Literatur, der Fehler in den Trainingslabels berücksichtigt, ist die robuste logistische Regression von Bootkrajang und Kabán [2012]. In dem dort verwendeten Fehlermodell zur Berücksichtigung der Labelfehler im Training wird jedoch nur der Zusammenhang zwischen dem Auftreten eines Labelfehlers und dem Klassenlabel berücksichtigt. Die Modellierung greift einige Aspekte, wie z.B. die Tatsache, dass Pixel, die ein fehlerhaftes Label in der Karte aufweisen, Segmente im Bildraum bilden, nicht auf. In dieser Arbeit wurde daher ein robuster Klassifikator basierend auf dem Random Forest entwickelt, in dem das Fehlermodell auf die Verwendung veralteter Karten zur Extraktion der Trainingslabels angepasst wurde. Dieses Modell ist somit deutlich komplexer aufgebaut verglichen mit dem Modell der robusten logistischen Regression. Der Random Forest Klassifikator wurde so angepasst, dass dieser mit Trainingsbeispielen, denen anstelle einer Klasse alle Klassen mit einer bestimmten Wahrscheinlichkeiten zugeordnet wurden, trainiert werden kann. Diese Wahrscheinlichkeiten der Klassenzugehörigkeit hängen von den Daten und den Klassenlabels ab. Außerdem fließt für jedes Trainingsbeispiel eine Änderungswahrscheinlichkeit mit ein, die angibt, wie wahrscheinlich es ist, dass das Label aus der veralteten Karte nicht mehr aktuell ist. Die Änderungswahrscheinlichkeiten werden in einem iterativen Verfahren bestimmt, in dem auch die Tatsache, dass Änderungen in Segmenten im Bildraum auftreten, berücksichtigt wird.

Neben der Verwendung der Kartenlabels zum Training wurden die Kartenlabels auch als zusätzliche Beobachtung in den Klassifikationsprozess integriert. Die Klassifikation von Objekten, die aufgrund ihrer Merkmale schwierig einer Klasse zuzuordnen sind, soll so verbessert werden. Der Einfluss des Kartenlabels auf das Klassifikationsergebnis ist abhängig davon, ob sich das Label in den aktuellen Daten von dem in der Karte unterscheidet und wird daher für jedes zu klassifizierende Primitiv individuell gewichtet. Das Gewicht wird dabei ähnlich wie die Änderungswahrscheinlichkeiten in einem iterativen Verfahren bestimmt. Ansonsten wird der statistische Zusammenhang zwischen den Kartenlabels und den in der Klassifikation zu bestimmenden Labels mit Hilfe von Übergangswahrscheinlichkeiten modelliert.

Um bei der Klassifikation von Fernerkundungsdaten zu mehreren Zeitpunkten temporalen Kontext zu berücksichtigen, wurde ein multitemporales CRF entwickelt, welches ebenfalls die Kartenlabels als zusätzliche Beobachtung beinhaltet. Der statistische Zusammenhang zwischen den Labels der temporal überlappenden Primitive basiert, wie der Zusammenhang zwischen den zu bestimmenden Labels und den Kartenlabels, auf den Übergangswahrscheinlich-

keiten und einer individuellen Gewichtung jedes Primitivs. Das Training aller erforderlichen Parameter erfolgte durch sequenzielles Training und Klassifikation der Fernerkundungsdaten aller Zeitpunkte.

Die in dieser Arbeit durchgeführte Klassifikation basiert auf Merkmalen, welche vor der Klassifikation aus den Daten extrahiert werden. Da die Anzahl möglicher Merkmale meist zu groß ist, um sie alle in der Klassifikation zu berücksichtigen, werden die Merkmale, die potenziell wichtig für die folgende Klassifikation sind, vorab selektiert. Einige Selektionsmethoden benötigen ebenfalls Trainingsdaten, deren Labels in dieser Arbeit jedoch aus einer veralteten Karte stammen und somit fehlerhaft sind. Obwohl die neue Methode zur Klassifikation den wissenschaftlichen Beitrag dieser Arbeit ausmacht, sollte dennoch gezeigt werden, dass auch für die Vorverarbeitungsschritte, hier die Merkmalsselektion, keine manuell gelabelten Daten benötigt werden. Die Selektion erfolgt mit der Methode aus [Frénay et al., 2014] basierend auf der Transinformation und wurde in den Experimenten ebenfalls analysiert.

In den Experimenten zur Merkmalsselektion konnte gezeigt werden, dass auch mit Fehlern in den Trainingslabels Merkmale selektiert werden, die zu akzeptablen Ergebnissen in der anschließenden Klassifikation führen. Dennoch konnte ein Einfluss der Fehler in den Trainingslabels auf die Merkmalsselektion beobachtet werden.

In den Experimenten zur Klassifikation von Fernerkundungsdaten aus einem Zeitpunkt wurde gezeigt, dass die Ergebnisse der neu entwickelten Klassifikationsmethode (kurz **RRFK**) in den meisten Fällen eine höhere Gesamtgenauigkeit aufweisen, als die Ergebnisse herkömmlicher Klassifikatoren, wie des Random Forests oder der logistischen Regression. Nur vereinzelt werden die Fehler in den Labels der Karte nicht korrekt erkannt und können daher auch nicht kompensiert werden. Dies geschieht beispielsweise, wenn die Anzahl fehlerhaft gelabelter Pixels einer Objektart größer ist, als die Anzahl korrekt gelabelter Pixels. Die Integration des Kartenlabels führt dazu, dass die Gesamtgenauigkeit der Pixel mit geändertem Label etwas abnimmt. Diese Verkleinerung der Gesamtgenauigkeit ist jedoch sehr gering verglichen mit der Verbesserung der Gesamtgenauigkeit aller Pixel und verglichen mit dem Anteil an fehlerhaften Labels in der Karte. Änderungen an Objektgrenzen werden aufgrund der verwendeten Heuristiken für die Bestimmung der Gewichte zur Regulierung des Einflusses der Kartenlabels bei der Klassifikation nicht gut erkannt, wenn die Kartenlabels in die Klassifikation mit einfließen.

Die Experimente zur Klassifikation von Fernerkundungsdaten mehrerer Zeitpunkte anhand des hochaufgelösten Datensatzes *Hameln* zeigten, dass die simultane Klassifikation aller Fernerkundungsdaten unter Berücksichtigung von temporalen Kontext das Klassifikationsergebnis verbessern kann. Bei den Experimenten mit dem Datensatz *Las Vegas* wurde jedoch bei der multitemporalen Klassifikation eine schlechtere Gesamtgenauigkeit beobachtet als bei der sequenziellen Klassifikation. Vor allem der Einfluss der Kartenlabels auf das Ergebnis scheint bei der multitemporalen Klassifikation geringer zu sein.

An dieser Stelle soll noch erwähnt werden, dass selbst wenn die neu entwickelte Methode in einem Experiment nicht zu dem besten Ergebnis führt, die Ergebnisse dennoch eine hohe

Gesamtgenauigkeit aufweisen. Bei anderen Methoden wurde beispielsweise beobachtet, dass die Klassifikation einer oder mehrerer Klassen vollständig fehlgeschlagen ist.

Das Ziel dieser Arbeit aus Kapitel 1.2, die Entwicklung eines multitemporalen Klassifikators zur Aktualisierung von veralteten Karten, der mit den fehlerhaften Labels der Karte trainiert werden kann, wurde somit erfüllt. Auch die einzelnen Aspekte, auf denen der wissenschaftliche Beitrag dieser Arbeit beruht, wurden erläutert und in den Experimenten untersucht. Die Ergebnisse der neu entwickelten Methode sind sehr gut im Vergleich zu Verfahren, die das Auftreten von Fehlern in den Trainingslabels im Training nicht berücksichtigen, und zeigen somit den positiven Beitrag dieser Arbeit.

7.2. Ausblick

Der in dieser Arbeit vorgestellte Prozess zur Aktualisierung veralteter Karten kann in einigen Aspekten noch verbessert werden. Verschiedene Möglichkeiten zur Anpassung des Prozesses werden im Folgenden dargelegt.

Robuste Merkmalsselektion In dieser Arbeit konnte gezeigt werden, dass die Merkmalsselektion durch Fehler in den Trainingsdaten beeinflusst wird. Obwohl bereits Methoden existieren, die fehlerhafte Labels berücksichtigen, wird in diesen ein Fehlermodell verwendet, welches die Fehler, die auftreten, wenn die Labels aus einer veralteten Karte stammen, im Allgemeinen nicht gut beschreibt. In zukünftigen Arbeiten könnte beispielsweise eine robuste Selektionsmethode entwickelt werden, die auch mit Labels aus einer veralteten Karte umgehen kann. So könnte beispielsweise das komplexere Fehlermodell dieser Arbeit in eine Selektionsmethode integriert werden.

Wahrscheinlichkeit der Klassenzugehörigkeit Bei dem neu entwickelten robusten Random Forest werden jedem Trainingsbeispiel anstelle einer bestimmten Klasse alle Klassen mit einer bestimmten Wahrscheinlichkeit zugeordnet. Diese Wahrscheinlichkeiten der Klassenzugehörigkeit werden in dieser Arbeit aus den Kartenlabels, den Beliefs der vorherigen Iteration, den Übergangs- und den Änderungswahrscheinlichkeiten bestimmt. In zukünftigen Arbeiten könnte die Berechnung der Wahrscheinlichkeiten der Klassenzugehörigkeit verbessert werden. Die Tatsache, dass Pixel mit fehlerhaftem Label Segmente im Bildraum bilden, wird beispielsweise noch nicht direkt modelliert. Stattdessen werden mit Hilfe von Heuristiken Segmente aus Pixeln, deren Kartenlabels sich von dem Label aus der Klassifikation unterscheiden, extrahiert. Auch Nachbarschaften im Merkmalsraum könnten mit modelliert werden, da auch die Merkmale der Pixel, die ein neues Objekt repräsentieren, sich ggf. ähneln.

Änderungswahrscheinlichkeit Die Wahrscheinlichkeit, dass sich das Label aus der Karte für ein bestimmtes Pixel von dem wahren (unbekannten) Label unterscheidet, auch Änderungswahrscheinlichkeit genannt, hat einen großen Einfluss auf die Wahrscheinlichkeiten der Klassenzugehörigkeit, die zum Training der neu entwickelten Methode benötigt werden. In dieser

Arbeit basiert die Berechnung der Änderungswahrscheinlichkeiten auf Heuristiken, die wiederum verschiedene Hyperparameter beinhalten. Diese Hyperparameter müssen zur Zeit manuell gesetzt werden, abhängig von der Auflösung der Fernerkundungsdaten und der zu aktualisierenden Objekte in der Karte. Diese Heuristiken könnten in Zukunft durch ein flexibles Modell ersetzt werden, welches beispielsweise verschiedene realistische Szenarien berücksichtigt.

Ballung aus Pixeln mit fehlerhaftem Label Damit Änderungen überhaupt erkannt werden können, muss ein Großteil der Trainingsbeispiele, die eine Ballung im Merkmalsraum bilden, ein korrektes Label besitzen. In der Realität ist dies jedoch nicht immer der Fall. Wie beispielsweise bei dem Datensatz *Las Vegas* könnte der Anteil an Änderungen pro Klasse zu groß sein. Neue Objekte könnten sich außerdem äußerlich von den bereits existierenden Objekten der selben Klasse unterscheiden. In weiteren Arbeiten könnte daher untersucht werden, wie diese Extremfälle berücksichtigt werden können. Ein möglicher Ansatz ist die Berücksichtigung der Distanz zwischen Ballungen im Merkmalsraum. Hierbei liegt die Annahme zugrunde, dass Ballungen aus Trainingsbeispielen der selben Klasse näher zusammen liegen, als Ballungen von unterschiedlichen Klassen. Diese Annahme kann jedoch, abhängig von der Verteilung der Trainingsbeispiele verschiedener Klassen im Merkmalsraum, auch nicht korrekt sein.

Verwendung von Neuronalen Netzen Für semantische Segmentierung, Klassifizierung und verschiedene andere Anwendungen werden vermehrt Convolutional Neural Networks (CNN) [LeCun et al., 1989] genutzt und gelten als aktueller Stand der Technik. Da die Verwendung von CNNs eine große Menge an verfügbaren Trainingsdaten voraussetzt [Drory et al., 2018], sind auch hier alternative Methoden zur Beschaffung der Trainingsdaten von Interesse bzw. ist Toleranz gegenüber fehlerhaften Trainingslabels wünschenswert. Da die Berechnung der Wahrscheinlichkeiten der Klassenzugehörigkeit unabhängig vom Klassifikator ist, könnten sie in zukünftigen Arbeiten in CNNs integriert werden. Das CNN müsste hierfür so angepasst werden, dass die Wahrscheinlichkeiten der Klassenzugehörigkeit die Zuordnung eines Trainingsbeispieles zu einer Klasse ersetzt. Eine Möglichkeit ist die Verwendung einer Loss-Funktion, welche die Wahrscheinlichkeiten der Klassenzugehörigkeit nutzt. Das daraus resultierende robuste CNN könnte anstelle des robusten Random Forests in dem in dieser Arbeit vorgestellten Prozess verwendet werden.

Modellierung des Zusammenhangs zwischen den Kartenlabels und den unbekannten Labels Für die Berücksichtigung des Kartenlabels in dem CRF als weitere Beobachtung wird der statistische Zusammenhang zwischen den Kartenlabels und den unbekannten Labels bestimmt. In dieser Arbeit wird dies mit Übergangswahrscheinlichkeiten realisiert, die den Übergang von einer Klasse zu einer anderen repräsentieren. Diese Übergangswahrscheinlichkeiten sind unabhängig von dem zu klassifizierenden Pixel und beziehen sich auf das gesamte Interessengebiet. Um eine temporale Überanpassung zu vermeiden, ist der Einfluss des Kartenlabels bei jedem zu klassifizierenden Pixel individuell gewichtet. Diese Gewichte basieren auf den selben Heuristiken, und somit Hyperparametern, die auch bei der Bestimmung der Änderungswahrscheinlichkeiten verwendet werden. Die Berücksichtigung der Kartenlabels

könnte in Zukunft auf verschiedene Weise angepasst werden. Zum einen könnte der statistische Zusammenhang zwischen den Kartenlabels und den unbekannten Labels so modelliert werden, dass auch lokale Eigenschaften der klassifizierenden Pixel berücksichtigt werden. Beispielsweise könnte die Tatsache, dass Änderungen in Segmenten auftreten, direkt modelliert werden. Dies könnte auch helfen, die temporale Überanpassung zu reduzieren, sodass die Verwendung von Gewichten nicht mehr notwendig ist. Zum anderen könnte, wenn der statistische Zusammenhang nicht verändert wird, die Bestimmung der Gewichte angepasst werden, sodass diese nicht so stark von zuvor definierten Hyperparametern abhängen. Auch hier könnte räumlicher Kontext im Bildraum direkt modelliert werden.

Anpassung der Übergangswahrscheinlichkeiten im multitemporalen Kontext In dieser Arbeit wurde ein multitemporales CRF für die simultane Klassifikation von Daten aus mehreren Zeitpunkten verwendet. Das temporale Interaktionspotential, welches den statistischen Zusammenhang zwischen den Labels von temporal benachbarten Pixeln beschreibt, wird, wie bei der Berücksichtigung der Kartenlabels, durch Übergangswahrscheinlichkeiten und Gewichte modelliert. Dementsprechend könnte auch die Modellierung des temporalen Interaktionspotentials, wie zuvor beschrieben, angepasst werden.

Gemeinsames Training bei der multitemporalen Klassifikation Mit Hilfe des multitemporalen CRF können die Daten mehrerer Zeitpunkte gemeinsam klassifiziert werden. Das Training zur Bestimmung der Parameter der verschiedenen Potentiale erfolgt jedoch in dieser Arbeit noch sequenziell, sodass temporale Zusammenhänge nicht berücksichtigt werden. Es kann z.B. davon ausgegangen werden, dass ein Pixel die Klasse nicht vermehrt wechselt, besonders wenn relativ kurze Zeitabschnitte betrachtet werden. In einem gemeinsamen Training aller Parameter des multitemporalen CRF könnte dies berücksichtigt werden.

Experimente In dieser Arbeit ist die neu entwickelte Methode an verschiedenen Datensätzen getestet worden. Die Experimente könnten in der Zukunft weiter ausgebaut werden. Es könnten beispielsweise OpenStreetMap Datensätze für die Extraktion der Trainingslabels verwendet werden. Auch die Aktualisierung von Karten mit anderen Klassenstrukturen, wie z.B. dem Baumbestand, könnte in weiteren Experimenten untersucht werden.

A. Anhang - Daten



(a) Karte, 1955, Original



(b) Karte, 1955, Vereinfacht: 3 Klassen



(c) DOP, 2010



(d) DOP, 2016

Abbildung A.1: DOPs und Karten für das Gebiet NA des Datensatzes *Hameln*. Rot: *Gebäude*. Grau: *Versiegelte Fläche*. Beige: *Boden*.



(a) DOP, Gebiet N1

(b) DOP, Gebiet N2

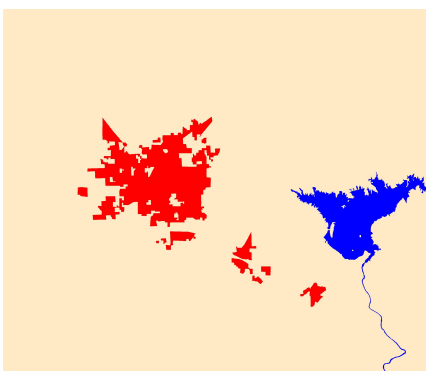
(c) Referenz, Gebiet N1

(d) Referenz, Gebiet N2

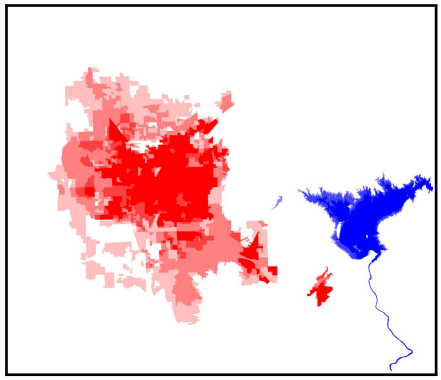
(e) Karte, Gebiet N1

(f) Karte, Gebiet N2

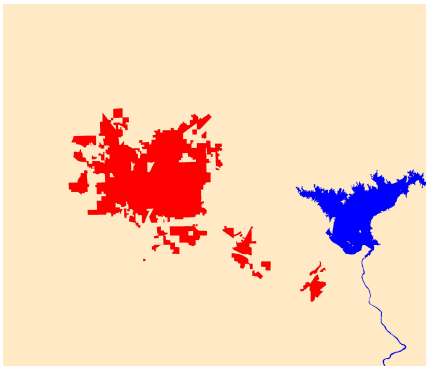
Abbildung A.2: DOP, Referenz und Karte für die Gebiete N1 und N2 des Datensatzes *Hameln*. Das DOP und die Referenz stammen aus dem Jahr 2016 und die Karte aus dem Jahr 2010.
Rot: *Gebäude*. Grau: *Versiegelte Fläche*. Beige: *Boden*.



(a) Karte, 1986



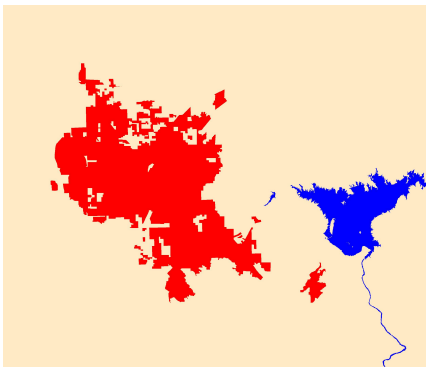
(b) Wachstum der Stadt Las Vegas



(c) Karte/Referenz, 1991



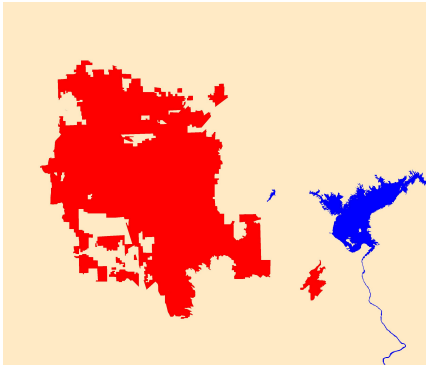
(d) DOP, 1991



(e) Karte/Referenz, 2000



(f) DOP, 2000



(g) Referenz, 2016



(h) DOP, 2016

Abbildung A.3: DOPs, Referenzen und Karte des Datensatzes *Las Vegas*.
Legende der Karten/Ref.: Rot: *Bebauung*. Blau: *Wasser*. Beige: *Boden*.
Legende von Abb. A.3b: Rot: *Bebauung*. Blau: *Wasser*. Weiß: *Boden*. Je dunkler der Farbton in desto mehr Zeitepochen ist die Klasse *Bebauung* bzw. *Wasser*.

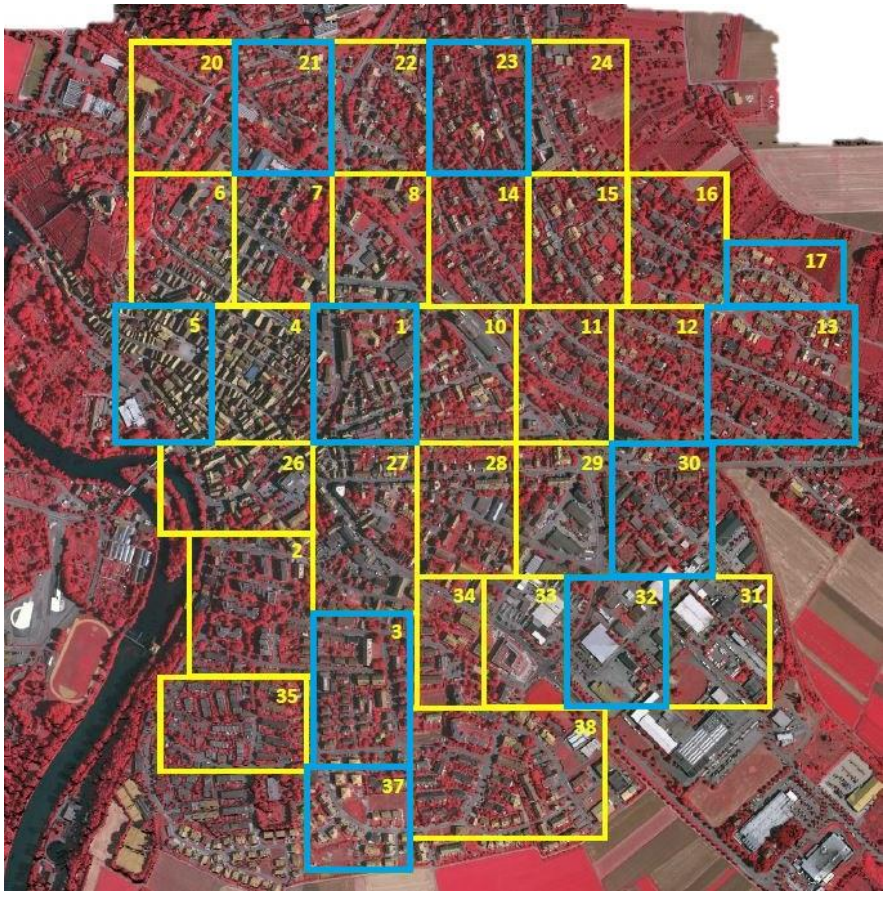
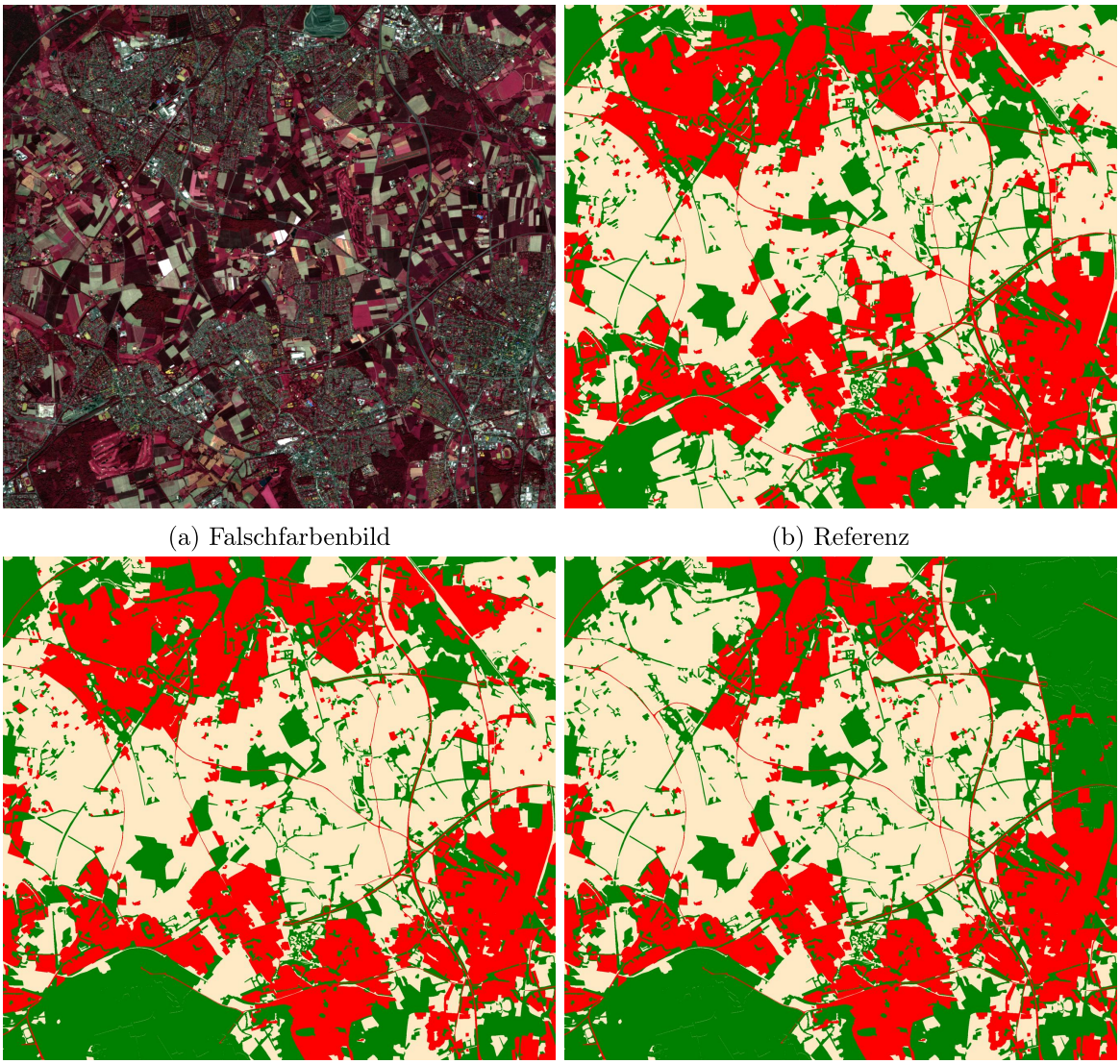


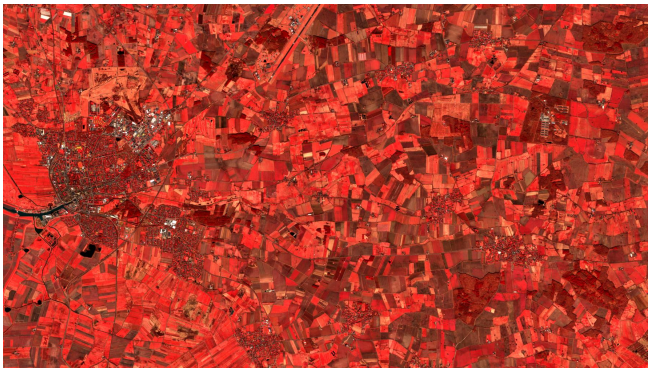
Abbildung A.4: Gebiete aus der ISPRS 2D Labelling Challenge [Wegner et al., 2015]. Die in dieser Arbeit verwendeten Gebiete sind blau umrandet.



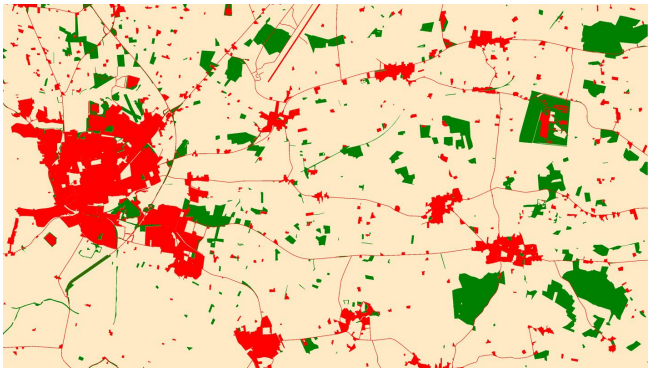
(a) Falschfarbenbild (b) Referenz

(c) Karte 1 (d) Karte 2

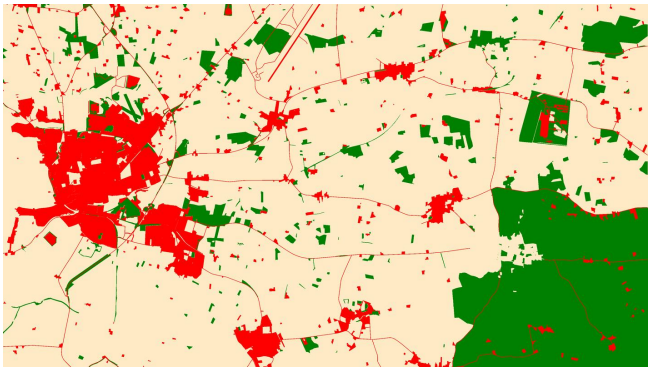
Abbildung A.5: Falschfarbenbild, Referenz und Karten von dem Datensatz *Herne*. Rot: *Bebauung*. Dunkelgrün: *Wald*. Beige: *Boden*.



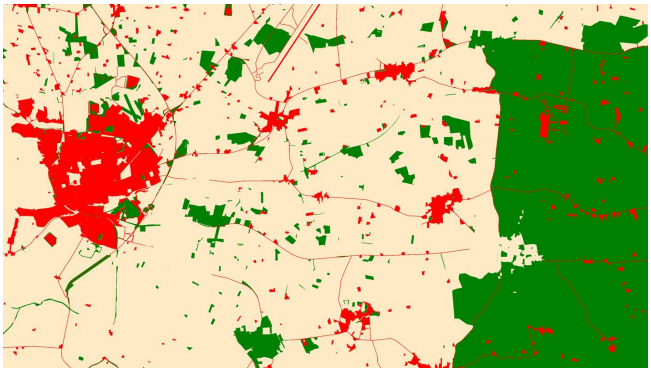
(a) Falschfarbenbild



(b) Referenz



(c) Karte 1



(d) Karte 2

Abbildung A.6: Falschfarbenbild, Referenz und Karten von dem Datensatz *Husum*. Rot: *Bebauung*. Dunkelgrün: *Wald*. Beige: *Boden*.

B. Anhang - Klassifikationsergebnisse

Hameln

		GG	MF1	Vollständigkeit			Korrektheit			F1		
				Geb.	Bod.	Vers.	Geb.	Bod.	Vers.	Geb.	Boden	Vers.
N1	RF	89,6	83,5	86,7	90,5	86,4	84,3	96,9	60,8	85,5	93,6	71,4
	RRF	90,9	84,6	83,9	93,5	81,1	87,8	95,9	66,8	85,8	94,7	73,3
	RRFK	92,3	85,9	75,9	97,5	74,6	91,5	93,3	84,7	83,0	95,4	79,3
	ERFK	90,7	81,7	73,8	98,6	55,3	94,4	90,5	88,2	82,8	94,4	68,0
	KRFK	89,3	82,3	73,0	92,7	83,5	86,0	93,8	67,6	78,9	93,2	74,7
	LR	89,7	79,7	61,4	98,9	57,7	95,6	89,0	91,0	74,8	93,7	70,6
	RLR	90,4	81,9	62,9	98,2	67,3	95,1	90,4	86,6	75,7	94,1	75,7
	RLRK	90,6	82,4	60,8	98,5	69,4	92,4	90,2	93,4	73,3	94,2	79,6
N2	RF	85,4	83,7	85,6	86,2	82,9	90,3	89,6	68,9	87,9	87,9	75,2
	RRF	87,3	85,4	86,3	89,9	80,4	90,6	89,6	75,6	88,4	89,8	77,9
	RRFK	86,8	85,3	77,3	95,6	74,0	91,7	83,4	94,5	83,9	89,1	83,0
	ERFK	86,6	84,2	79,8	97,2	63,9	93,5	82,2	97,6	86,1	89,1	77,2
	KRFK	84,8	83,2	76,8	91,5	75,9	91,5	83,0	82,1	83,5	87,1	78,9
	LR	84,1	81,3	74,6	95,6	62,6	94,2	80,0	88,6	83,3	87,1	73,4
	RLR	84,2	81,4	74,5	95,6	63,2	94,2	80,1	88,6	83,2	87,2	73,8
	RLRK	84,7	82,9	75,9	92,4	73,9	91,9	82,3	83,9	83,1	87,1	78,6

Tabelle B.1: Genauigkeitsmaße [%]. Datensatz: *Hameln*. Grundlage: Gesamtes Bild.

		GG	MF1	Vollständigkeit			Korrektheit			F1		
				Geb.	Bod.	Vers.	Geb.	Bod.	Vers.	Geb.	Bod.	Vers.
N1	RF	97,0	85,2	98,4	85,3	80,0	99,3	83,7	65,9	98,9	84,5	72,3
	RRF	97,0	85,3	98,2	91,7	73,1	99,4	80,5	69,5	98,8	85,8	71,2
	RRFK	94,6	74,5	96,8	88,8	43,8	99,3	60,0	70,1	98,0	71,6	53,9
	ERFK	95,0	75,0	96,9	95,2	39,3	99,6	60,2	80,8	98,2	73,8	52,9
	KRFK	90,4	54,6	96,4	34,9	41,5	98,5	34,2	26,0	97,4	34,5	32,0
	LR	94,6	73,8	96,4	97,3	36,2	99,8	57,0	88,0	98,1	71,9	51,3
	RLR	94,8	76,5	96,4	95,9	45,2	99,8	58,5	84,3	98,1	72,7	58,8
	RLRK	91,6	59,2	94,9	83,8	14,7	99,3	44,7	45,2	97,0	58,3	22,2
N2	RF	93,3	71,3	96,5	58,2	69,1	98,6	48,2	58,8	97,5	52,8	63,5
	RRF	94,1	76,2	96,6	74,9	66,5	98,6	54,2	69,6	97,6	62,9	68,0
	RRFK	87,5	47,7	93,2	55,4	9,6	97,9	22,9	37,1	95,5	32,4	15,3
	ERFK	89,1	48,4	94,2	73,3	3,6	98,5	29,7	40,0	96,3	42,2	6,6
	KRFK	84,9	36,0	93,1	13,2	3,6	97,8	6,4	5,1	95,4	8,6	4,2
	LR	90,5	59,6	94,5	76,1	24,5	99,0	33,7	63,4	96,7	46,7	35,4
	RLR	90,4	59,5	94,5	75,3	24,9	99,0	33,4	62,2	96,7	46,3	35,5
	RLRK	85,0	36,4	92,8	21,5	0,5	97,9	9,7	0,9	95,3	13,4	0,7

Tabelle B.2: Genauigkeitsmaße [%]. Datensatz: *Hameln*. Grundlage: Nur Änderungsgebiete.

		GG	MF1	Vollständigkeit		Korrektheit		F1	
				Änd.	k. Änd.	Änd.	k. Änd.	Änd.	k. Änd.
N1	RF	90,5	83,4	85,6	91,3	63,0	97,4	72,6	94,2
	RRF	91,9	85,4	85,7	92,9	67,7	97,4	75,7	95,1
	RRFK	92,7	84,3	66,5	97,2	80,7	94,4	72,9	95,8
	ERFK	91,0	81,9	68,1	95,0	70,2	94,5	69,1	94,8
	KRFK	89,7	72,8	37,2	98,7	83,5	90,1	51,5	94,2
	LR	89,9	79,6	64,4	94,3	66,0	93,9	65,1	94,1
	RLR	90,6	81,0	66,2	94,8	68,9	94,2	67,5	94,5
	RLRK	90,7	76,3	43,4	98,9	86,7	91,0	57,8	94,8
N2	RF	86,8	77,5	67,7	90,6	59,1	93,3	63,1	91,9
	RRF	88,5	80,4	72,3	91,8	63,9	94,3	67,8	93,0
	RRFK	87,0	66,5	26,3	99,1	86,2	87,0	40,3	92,7
	ERFK	86,7	69,9	35,7	97,0	70,4	88,2	47,4	92,4
	KRFK	84,9	55,0	10,1	99,9	95,6	84,7	18,3	91,7
	LR	84,5	70,1	45,2	92,4	54,4	89,3	49,4	90,8
	RLR	84,5	70,1	44,9	92,5	54,6	89,3	49,2	90,9
	RLRK	84,7	55,1	10,5	99,7	86,2	84,7	18,7	91,6

Tabelle B.3: Genauigkeitsmaße [%]. Datensatz: *Hameln*. Grundlage: Änderungsdetektion.

Las Vegas

Karte		GG	MF1	Vollständigkeit			Korrektheit			F1		
				Bod.	Beb.	Wa.	Bod.	Beb.	Wa.	Bod.	Beb.	Wa.
1986	RF	82,4	80,3	79,5	92,2	98,4	97,7	53,6	75,3	87,7	67,8	85,4
	RRF	78,3	65,6	90,3	27,6	99,2	83,5	45,5	61,3	86,7	34,4	75,8
	RRFK	85,4	78,6	95,2	44,5	98,9	87,4	72,0	82,1	91,1	55,0	89,7
	ERFK	76,1	61,0	90,6	14,9	99,6	81,2	31,8	62,7	85,6	20,3	76,9
	KRFK	80,5	61,7	99,9	0,1	97,5	80,2	94,5	94,5	88,9	0,2	96,0
	LR	80,8	62,2	99,7	2,8	95,7	80,5	95,5	88,7	89,1	5,4	92,1
	RLR	84,5	73,3	99,3	22,9	95,7	83,9	94,2	88,5	91,0	36,9	92,0
	RLRK	85,3	75,1	99,5	26,1	97,8	84,5	97,8	88,4	91,4	41,2	92,9
1991	RF	91,2	86,6	90,5	93,4	98,9	98,3	72,9	72,6	94,2	81,9	83,7
	RRF	92,5	85,9	96,0	77,6	99,1	94,6	87,0	67,6	95,3	82,0	80,3
	RRFK	93,7	90,7	97,9	75,9	98,8	94,2	91,7	87,6	96,1	83,0	92,9
	ERFK	92,9	89,3	98,1	71,4	98,6	93,3	92,1	85,8	95,6	80,4	91,8
	KRFK	93,4	89,4	98,2	73,2	98,7	93,7	93,5	83,1	95,9	82,1	90,3
	LR	81,5	64,9	99,6	6,6	96,2	81,2	94,4	89,8	89,4	12,3	92,9
	RLR	94,1	91,3	97,1	81,9	96,3	95,5	88,5	89,3	96,3	85,0	92,7
	RLRK	94,3	91,1	97,4	81,3	98,4	95,4	90,5	84,9	96,4	85,7	91,1
2000	RF	91,4	87,2	90,1	95,8	98,8	98,9	72,2	74,4	94,3	82,4	84,9
	RRF	93,7	87,3	95,5	85,6	99,2	96,4	86,8	66,5	96,0	86,2	79,6
	RRFK	93,4	91,5	92,5	96,6	98,6	99,1	76,8	88,5	95,7	85,6	93,3
	ERFK	96,1	93,7	98,3	86,8	98,6	96,8	93,9	88,7	97,5	90,2	93,4
	KRFK	96,0	90,7	97,2	90,6	98,9	97,6	92,7	71,6	97,4	91,7	83,0
	LR	92,8	88,9	98,6	68,5	96,4	92,7	94,1	87,8	95,6	79,3	91,9
	RLR	95,1	92,2	97,4	85,8	96,5	96,5	90,1	87,6	96,9	87,9	91,8
	RLRK	95,8	92,6	97,8	87,4	98,5	96,9	92,5	83,9	97,3	89,9	90,6

Tabelle B.4: Genauigkeitsmaße [%]. Datensatz: *Las Vegas*. Grundlage: Gesamtes Bild.

Karte		GG	MF1	Vollständigkeit			Korrektheit			F1		
				Bod.	Beb.	Wa.	Bod.	Beb.	Wa.	Bod.	Beb.	Wa.
1986	RF	98,0	77,1	99,3	90,1	84,9	98,4	96,2	25,7	98,9	93,1	39,4
	RRF	89,3	56,5	99,8	24,7	92,5	89,1	99,0	22,2	94,2	39,5	35,8
	RRFK	89,7	53,2	99,7	28,1	63,7	89,5	98,9	13,0	94,3	43,7	21,5
	ERFK	87,5	49,9	99,8	12,0	93,1	87,5	98,6	21,7	93,2	21,3	35,3
	KRFK	85,9	38,1	99,9	0,0	41,1	86,0	0,0	14,9	92,4	0,0	21,9
	LR	86,1	40,9	99,9	1,5	56,3	86,2	99,6	17,9	92,5	2,9	27,2
	RLR	87,9	48,1	99,8	14,3	55,3	87,7	98,4	17,1	93,4	24,9	26,1
	RLRK	86,1	35,7	99,6	2,7	34,9	86,3	87,1	5,4	92,5	5,1	9,4
1991	RF	98,4	76,8	99,5	90,8	87,3	98,7	97,1	23,9	99,1	93,8	37,5
	RRF	96,4	74,4	99,8	71,6	90,9	96,2	98,9	27,3	98,0	83,1	42,0
	RRFK	95,2	66,2	99,8	62,0	61,7	95,0	99,1	15,6	97,3	76,3	24,9
	ERFK	94,4	67,6	99,8	55,0	70,5	94,1	99,0	23,4	96,9	70,7	35,1
	KRFK	94,5	60,2	99,6	57,8	46,4	94,4	98,9	6,0	96,9	73,0	10,6
	LR	88,2	44,4	99,9	3,6	60,3	88,2	99,1	22,4	93,7	6,9	32,6
	RLR	96,9	72,0	99,8	76,2	60,6	96,8	98,6	21,5	98,2	86,0	31,7
	RLRK	96,0	62,6	99,5	71,3	35,5	96,1	98,6	4,2	97,8	82,7	7,4
2000	RF	98,7	69,5	99,2	91,4	76,5	99,4	90,3	10,3	99,3	90,9	18,2
	RRF	98,0	67,1	99,5	74,6	84,8	98,4	93,8	10,9	99,0	83,1	19,3
	RRFK	98,4	63,8	99,0	90,1	15,7	99,3	86,7	2,3	99,2	88,4	4,0
	ERFK	97,3	59,7	99,7	60,4	23,2	97,5	94,7	3,9	98,6	73,7	6,7
	KRFK	97,4	60,3	99,0	71,9	2,6	98,2	95,7	0,1	98,6	82,1	0,1
	LR	96,8	59,1	99,8	50,7	41,8	96,9	97,1	7,3	98,3	66,6	12,4
	RLR	98,1	64,7	99,7	73,5	42,1	98,3	95,3	7,1	99,0	83,0	12,2
	RLRK	97,3	58,9	99,4	66,2	0,3	97,8	94,7	0,0	98,6	77,9	0,0

Tabelle B.5: Genauigkeitsmaße [%]. Datensatz: *Las Vegas*. Grundlage: Nur Änderungsgebiete.

Karte		GG	MF1	Vollständigkeit		Korrektheit		F1	
				Änd.	k. Änd.	Änd.	k. Änd.	Änd.	k. Änd.
1986	RF	82,8	75,1	90,1	81,6	46,4	97,9	61,2	89,0
	RRF	78,3	58,0	29,3	87,0	28,5	87,4	28,9	87,2
	RRFK	85,5	65,7	31,8	94,9	52,7	88,7	39,7	91,7
	ERFK	76,1	52,0	17,3	86,5	18,5	85,5	17,9	86,0
	KRFK	80,5	48,9	6,2	93,6	14,7	84,9	8,7	89,1
	LR	80,8	49,9	7,6	93,7	17,7	85,2	10,6	89,2
	RLR	84,5	59,4	19,5	96,0	46,2	87,1	27,4	91,3
	RLRK	85,3	52,6	7,5	99,0	58,3	85,8	13,3	92,0
1991	RF	91,5	84,2	90,3	91,7	62,0	98,4	73,5	95,0
	RRF	92,5	83,7	72,6	95,5	70,7	95,9	71,7	95,7
	RRFK	93,7	84,3	62,9	98,3	84,7	94,7	72,2	96,5
	ERFK	92,9	81,8	56,7	98,3	83,6	93,8	67,6	96,0
	KRFK	93,4	82,9	57,6	98,7	87,3	94,0	69,4	96,3
	LR	81,5	50,4	8,8	92,4	14,7	87,2	11,0	89,7
	RLR	94,2	87,2	76,8	96,8	78,4	96,6	77,6	96,7
	RLRK	94,4	86,7	70,0	98,1	84,5	95,6	76,5	96,8
2000	RF	91,9	78,3	89,6	92,0	46,5	99,1	61,2	95,5
	RRF	93,9	80,4	75,8	95,3	55,5	98,1	64,1	96,7
	RRFK	94,1	82,4	87,7	94,6	55,7	99,0	68,1	96,8
	ERFK	96,2	84,0	63,0	98,7	79,0	97,2	70,1	97,9
	KRFK	96,0	83,9	64,1	98,5	76,8	97,3	69,9	97,9
	LR	92,8	74,6	56,7	95,6	49,9	96,6	53,1	96,1
	RLR	95,3	83,5	75,3	96,8	64,6	98,1	69,5	97,4
	RLRK	95,9	83,6	64,8	98,3	74,6	97,3	69,4	97,8

Tabelle B.6: Genauigkeitsmaße [%]. Datensatz: *Las Vegas*. Grundlage: Änderungsdetektion.

Vaihingen

Karte		GG	MF1	F1			
				Geb.	Bäume	n. Veg.	v. Fl.
M1	RF	83,9	81,3	90,8	83,3	71,1	87,3
	RRF	84,0	81,1	90,0	82,8	70,3	86,9
	RRFK	94,7	94,0	97,1	94,3	91,2	96,0
	ERFK	91,5	89,9	92,7	91,4	77,7	91,3
	KRFK	94,3	93,5	97,7	93,9	90,7	95,9
	LR	83,7	82,3	90,5	81,7	71,2	87,8
	RLR	84,3	83,2	91,4	82,1	71,9	88,2
	RLRK	92,8	92,5	96,0	91,3	90,1	94,7
M2	RF	82,1	79,7	89,7	82,0	69,2	85,5
	RRF	82,7	79,9	89,8	82,1	70,1	86,6
	RRFK	89,9	89,9	94,4	91,4	85,3	90,8
	ERFK	85,2	83,7	90,2	84,1	76,3	87,8
	KRFK	90,2	90,6	95,0	90,7	86,5	89,6
	LR	79,8	78,2	87,0	79,3	70,3	80,0
	RLR	80,6	79,3	87,9	80,0	70,1	81,2
	RLRK	86,0	86,7	89,8	87,8	84,6	87,7
M3	RF	82,3	79,0	90,5	81,0	68,3	85,1
	RRF	83,1	80,1	89,7	80,6	69,8	86,2
	RRFK	88,1	87,1	93,4	88,6	80,9	89,8
	ERFK	80,4	79,8	90,6	76,5	71,2	85,9
	KRFK	85,7	84,6	91,4	86,4	80,5	86,6
	LR	75,3	74,2	84,9	77,9	68,9	77,0
	RLR	77,1	77,4	85,5	78,3	68,4	77,7
	RLRK	81,7	81,9	82,4	83,1	81,5	86,6

Tabelle B.7: Genauigkeitsmaße [%]. Datensatz: *Vaihingen*. Grundlage: Gesamtes Bild. Median aller Datensätze.

Karte		GG	MF1	F1			
				Geb.	Bäume	n. Veg.	v. Fl.
M1	RF	97,7	74,2	99,7	80,0	63,5	72,0
	RRF	98,3	74,9	99,7	77,4	62,1	73,4
	RRFK	95,3	58,5	99,6	44,1	38,8	58,3
	ERFK	97,3	65,7	99,6	68,2	42,4	65,1
	KRFK	94,4	46,4	99,3	31,3	30,5	36,8
	LR	97,0	72,3	99,6	69,7	61,8	72,1
	RLR	97,2	72,3	99,6	70,7	63,7	71,7
	RLRK	93,4	52,5	99,0	37,4	35,0	41,1
M2	RF	95,8	77,0	99,4	79,6	69,8	62,9
	RRF	95,7	76,0	99,3	78,6	70,5	65,7
	RRFK	93,6	65,9	99,2	61,4	55,1	48,9
	ERFK	94,6	71,5	99,3	72,8	64,6	66,8
	KRFK	90,8	56,8	99,0	45,3	42,9	32,2
	LR	93,3	67,9	98,7	64,8	66,3	45,0
	RLR	93,7	68,9	98,7	64,3	65,7	47,4
	RLRK	88,1	45,6	97,3	24,3	36,3	14,0
M3	RF	94,5	73,1	99,2	77,5	68,8	59,7
	RRF	94,5	75,7	99,1	77,3	71,4	65,5
	RRFK	91,8	63,5	99,0	62,8	53,3	50,7
	ERFK	93,3	70,5	99,1	70,4	67,2	64,3
	KRFK	86,1	47,4	97,9	42,4	41,9	18,3
	LR	89,3	67,1	97,6	57,4	62,2	39,7
	RLR	91,0	66,7	97,7	60,9	63,0	42,0
	RLRK	83,7	43,4	95,0	20,4	24,9	13,4

Tabelle B.8: Genauigkeitsmaße [%]. Datensatz: *Vaihingen*. Grundlage: Nur Änderungsgebiete. Median aller Datensätze.

Karte		GG	MF1	F1	
				Änd.	k. Änd.
M1	RF	84,6	73,0	56,3	91,0
	RRF	84,9	73,4	56,2	91,0
	RRFK	95,1	85,7	73,8	97,3
	ERFK	92,2	78,6	60,9	95,6
	KRFK	94,7	82,9	69,6	97,1
	LR	85,1	72,9	57,4	91,1
	RLR	85,3	73,2	57,3	91,4
	RLRK	93,3	81,3	67,0	96,3
M2	RF	83,9	77,6	65,3	89,5
	RRF	84,7	78,2	68,3	89,8
	RRFK	92,9	88,5	81,5	95,2
	ERFK	87,8	79,9	69,9	92,0
	KRFK	91,3	85,7	74,7	94,6
	LR	81,9	75,7	60,4	87,5
	RLR	82,5	75,6	61,2	88,0
	RLRK	87,3	76,7	59,2	92,2
M3	RF	84,2	80,1	72,3	89,2
	RRF	84,6	81,0	73,6	89,7
	RRFK	90,6	86,4	79,6	93,6
	ERFK	83,9	80,2	73,6	88,1
	KRFK	86,8	76,2	61,2	91,9
	LR	78,9	76,9	64,6	83,5
	RLR	80,2	77,3	67,5	84,8
	RLRK	83,8	75,4	59,8	89,6

Tabelle B.9: Genauigkeitsmaße [%]. Datensatz: *Vaihingen*. Grundlage: Änderungsdetektion. Median aller Datensätze.

Karte		Volständigkeit				Korrektheit			
		Geb.	Bäume	n. Veg.	v. Fl.	Geb.	Bäume	n. Veg.	v. Fl.
M1	RF	94,6	80,6	73,9	84,2	88,1	82,9	68,1	90,9
	RRF	94	80	71,8	84,7	87,5	84,8	68,2	89,3
	RRFK	98,3	94,5	91	95,5	96,5	95,2	93	97
	ERFK	98,3	91,1	86,8	88,6	91,9	96,5	85,1	97
	KRFK	96,9	94,5	92,1	96,6	98,1	94,9	91,6	95,5
	LR	89,8	78	72,3	88,9	91,5	85,7	69,5	85,7
	RLR	91,4	79,8	72,7	88,5	91,7	84,1	70,3	86,8
	RLRK	93,7	94,4	93,2	96,4	98	92,9	91,9	94,4
M2	RF	94,3	81,1	72,4	83,3	88,4	82,2	67,2	86,5
	RRF	94,3	77,5	72,8	84,9	88	86	67,3	86,4
	RRFK	97,4	90,3	87,8	91,1	92,1	93,6	84,8	93,5
	ERFK	95,7	81,5	77	84,8	90,1	93,4	74,3	90,1
	KRFK	94,4	89	90	95,9	97,5	91	86,5	88,1
	LR	81,6	73,1	68	90,3	93,2	85,4	72,4	73,7
	RLR	83,1	74	70,7	89,2	93,4	83,1	71,2	76,8
	RLRK	83	83,6	87,9	97	98,7	90,3	87	80,1
M3	RF	95,1	81,4	72,5	80,3	88,1	83,2	67,2	86,2
	RRF	94,2	78,1	72,2	84,3	88	85,7	67,5	84,3
	RRFK	96,8	87,2	85,6	89,6	90,8	89,7	82,3	90,8
	ERFK	95,1	83,4	73,2	82,3	88	89,5	76,7	88
	KRFK	86,8	87,3	86,2	94,2	97,7	87,1	81,9	86,3
	LR	76,8	72,3	68	92	94,4	85,3	72,9	70,3
	RLR	79	76,4	67,5	91,5	94,3	79,7	72,3	72
	RLRK	72,8	81,9	78,6	96,6	98,5	83,1	78,3	80

Tabelle B.10: Genauigkeitsmaße [%]. Datensatz: *Vaihingen*. Grundlage: Gesamtes Bild. Median aller Datensätze.

Karte		Volständigkeit				Korrektheit			
		Geb.	Bäume	n. Veg.	v. Fl.	Geb.	Bäume	n. Veg.	v. Fl.
M1	RF	99,8	78,2	70	77,3	99,7	85,1	57,4	71,3
	RRF	99,8	74,7	67,1	79,1	99,8	83,8	54,3	73,9
	RRFK	99,7	32,2	39,7	52,2	99,6	60,5	35	55,5
	ERFK	99,7	62,9	43	63,8	99,6	77,8	50,2	73,4
	KRFK	99,1	36,4	34,4	39	99,6	37,9	29,6	33,9
	LR	99,4	71,6	75,3	84,6	99,9	76	51,7	58,5
	RLR	99,3	72,6	73,6	81,7	99,9	73,8	53,8	58
	RLRK	98,5	39,5	29,6	53,9	99,8	37,4	32,1	30,3
M2	RF	99,5	76,9	68,7	69,8	99,6	82,2	67,2	56
	RRF	99,5	74,7	70,7	79,1	99,7	83,4	67,2	58,7
	RRFK	99,4	56,4	52,2	67,8	99,5	75,3	52,6	44
	ERFK	99,3	62,1	70,8	78	99,7	80,7	61	57,9
	KRFK	98,7	35	51,2	53,9	99,6	65,1	48,1	24,5
	LR	97,5	58,9	67,2	92,3	99,9	81,5	67,7	30
	RLR	97,5	59	72,4	88,5	99,8	80,9	66,9	32,4
	RLRK	94,9	24,2	45,9	60	99,8	36,7	39,4	11,2
M3	RF	99,4	77,9	68,2	65,9	99,4	82,3	66,1	54
	RRF	99,3	71,7	73,7	78,3	99,6	83,4	71,6	56,1
	RRFK	99,2	54,8	53,6	64,7	99,4	72,5	52,6	45,5
	ERFK	99,1	63,3	69,3	68,2	99,4	81,6	66,8	61
	KRFK	96,4	44,7	50,2	43,9	99,6	47,3	42,6	12,3
	LR	95,4	54,7	65,5	88,9	99,8	82,1	67,3	26,3
	RLR	95,6	66,7	66,5	86,6	99,8	79,2	66,7	28,8
	RLRK	92	29,8	36	57,4	99,7	33,3	24,1	7,6

Tabelle B.11: Genauigkeitsmaße [%]. Datensatz: *Vaihingen*. Grundlage: Nur Änderungsgebiete. Median aller Datensätze.

Karte		Vollständigkeit		Korrektheit	
		Änd.	k. Änd.	Änd.	k. Änd.
M1	RF	91,5	84,7	41,3	98,7
	RRF	91,9	84,3	41,5	98,7
	RRFK	64,3	98,7	87,2	95,4
	ERFK	84,9	92,7	51,8	97,9
	KRFK	54,6	99,7	96,2	94,6
	LR	85,4	85,4	42,3	98,1
	RLR	85,9	85,7	42,4	98
	RLRK	55	99,1	88,7	93,5
M2	RF	90,6	83,6	53,7	97,8
	RRF	91,7	83,8	54	97,8
	RRFK	83	96,3	86,5	94,8
	ERFK	91	89,3	59,5	97
	KRFK	61,3	99,4	96,2	90,4
	LR	76,4	83,1	51	93,2
	RLR	79,1	83,6	52,3	93,4
	RLRK	44,7	98,4	88,1	87,7
M3	RF	90,3	83,1	62,1	96,2
	RRF	92,5	83,7	63,6	96,8
	RRFK	74,2	95,3	86,6	92,5
	ERFK	90,2	83,3	66,4	95,8
	KRFK	50,4	99,1	95,9	85,4
	LR	75,5	80,3	56,2	88,8
	RLR	78,4	81	59	90,2
	RLRK	44,7	98,7	89,1	83,1

Tabelle B.12: Genauigkeitsmaße [%]. Datensatz: *Vaihingen*. Grundlage: Änderungsdetektion. Median aller Datensätze.

Herne

Karte		GG	MF1	Vollständigkeit			Korrektheit			F1		
				Beb.	Wald	Bod.	Beb.	Wald	Bod.	Beb.	Wald	Bod.
M1	RF	74,3	73,8	91,5	76,3	59,8	68,4	65,5	92,5	78,3	70,5	72,6
	RRF	74,0	73,5	88,5	74,9	62,3	70,5	64,2	87,8	78,5	69,1	72,8
	RRFK	94,9	94,7	98,6	97,1	90,7	93,7	89,7	99,3	96,1	93,3	94,8
	ERFK	48,0	40,6	99,9	59,9	1,2	45,2	54,7	100,0	62,2	57,2	2,3
	KRFK	96,1	96,0	97,8	99,1	93,2	97,3	90,4	99,0	97,5	94,6	96,0
	LR	77,4	77,0	80,9	76,9	75,0	77,0	70,3	82,5	78,9	73,4	78,6
	RLR	77,7	77,2	83,2	74,5	75,3	75,4	72,6	83,2	79,1	73,6	79,1
	RLRK	95,4	95,2	98,0	92,6	94,9	93,2	96,1	96,7	95,6	94,3	95,8
M2	RF	73,5	72,9	93,4	71,0	59,4	65,7	67,5	92,1	77,2	69,2	72,2
	RRF	70,6	70,1	91,4	73,5	53,0	65,0	62,4	90,5	75,9	67,5	66,8
	RRFK	79,8	79,9	97,9	84,6	63,1	69,8	78,1	98,3	81,5	81,3	76,9
	ERFK	41,4	33,3	100,0	34,4	0,0	38,0	64,2	-	55,0	44,8	0,0
	KRFK	90,7	90,5	93,3	98,1	84,5	95,7	78,6	96,0	94,5	87,3	89,9
	LR	74,2	74,1	71,6	84,5	70,4	81,5	61,7	79,5	76,3	71,3	74,7
	RLR	77,5	76,9	82,4	72,8	76,3	75,3	73,7	81,6	78,7	73,3	78,9
	RLRK	91,6	91,5	94,3	88,1	91,4	89,5	93,3	92,4	91,8	90,7	91,9

Tabelle B.13: Genauigkeitsmaße [%]. Datensatz: *Herne*. Grundlage: Gesamtes Bild.

Karte		GG	MF1	Vollständigkeit			Korrektheit			F1		
				Beb.	Wald	Bod.	Beb.	Wald	Bod.	Beb.	Wald	Bod.
M1	RF	98,9	59,4	99,8	-	66,6	99,5	0,0	95,6	99,6	0,0	78,5
	RRF	99,0	59,9	99,7	-	70,7	99,5	0,0	92,5	99,6	0,0	80,1
	RRFK	97,6	45,5	99,5	-	24,9	99,8	0,0	70,6	99,7	0,0	36,8
	ERFK	97,3	33,0	100,0	-	0,0	97,8	0,0	-	98,9	0,0	0,0
	KRFK	97,7	50,1	99,3	-	39,1	99,9	0,0	71,7	99,6	0,0	50,6
	LR	98,8	59,5	99,5	-	75,3	99,7	0,0	83,0	99,6	0,0	78,9
	RLR	98,9	59,8	99,5	-	75,6	99,6	0,0	84,6	99,6	0,0	79,9
	RLRK	98,6	58,0	99,3	-	70,8	99,8	0,0	78,3	99,6	0,0	74,4
M2	RF	96,6	58,3	99,4	-	63,3	98,2	0,0	95,9	98,8	0,0	76,2
	RRF	95,8	55,8	99,2	-	54,7	98,0	0,0	93,0	98,6	0,0	68,9
	RRFK	92,9	42,5	99,2	-	17,9	98,1	0,0	76,3	98,6	0,0	29,0
	ERFK	92,2	32,1	100,0	-	0,0	93,1	0,0	-	96,4	0,0	0,0
	KRFK	92,7	47,8	97,6	-	34,3	99,8	0,0	64,2	98,7	0,0	44,7
	LR	95,6	58,2	97,3	-	74,8	99,4	0,0	77,8	98,3	0,0	76,3
	RLR	96,9	60,4	98,3	-	80,7	99,0	0,0	84,4	98,7	0,0	82,5
	RLRK	96,4	59,0	98,0	-	77,4	99,4	0,0	79,3	98,7	0,0	78,3

Tabelle B.14: Genauigkeitsmaße [%]. Datensatz: *Herne*. Grundlage: Nur Änderungsgebiete.

Karte		GG	MF1	Vollständigkeit		Korrektheit		F1	
				Änd.	k. Änd.	Änd.	k. Änd.	Änd.	k. Änd.
M1	RF	74,8	56,3	90,6	73,9	16,4	99,3	27,8	84,8
	RRF	74,5	56,0	90,4	73,6	16,2	99,3	27,5	84,5
	RRFK	95,1	76,6	58,2	97,2	53,6	97,6	55,8	97,4
	ERFK	50,2	40,4	91,0	47,8	9,0	98,9	16,3	64,5
	KRFK	96,2	79,9	57,4	98,4	66,8	97,6	61,7	98,0
	LR	77,8	58,1	86,3	77,3	17,7	99,0	29,4	86,8
	RLR	78,2	58,6	88,0	77,6	18,2	99,1	30,2	87,1
	RLRK	95,6	81,7	78,2	96,6	56,6	98,7	65,7	97,7
M2	RF	75,4	69,1	91,3	72,2	39,6	97,6	55,3	83,0
	RRF	72,8	66,5	87,8	69,9	36,8	96,6	51,8	81,1
	RRFK	81,7	72,0	68,6	84,3	46,5	93,1	55,5	88,5
	ERFK	48,2	46,7	94,3	39,0	23,6	97,2	37,7	55,6
	KRFK	90,9	81,2	57,3	97,6	82,6	92,0	67,7	94,7
	LR	75,9	68,6	83,3	74,4	39,4	95,7	53,4	83,7
	RLR	79,0	72,5	90,8	76,6	43,7	97,7	59,0	85,9
	RLRK	92,5	87,1	83,8	94,2	74,4	96,7	78,8	95,4

Tabelle B.15: Genauigkeitsmaße [%]. Datensatz: *Herne*. Grundlage: Änderungsdetektion.

Husum

Karte		GG	MF1	Vollständigkeit			Korrektheit			F1		
				Beb.	Wald	Bod.	Beb.	Wald	Bod.	Beb.	Wald	Bod.
M1	RF	86,1	75,9	87,1	70,8	87,4	49,6	74,0	97,2	63,2	72,4	92,1
	RRF	87,3	76,0	74,5	77,6	90,1	58,5	63,2	95,9	65,5	69,6	92,9
	RRFK	97,4	94,5	92,2	90,5	98,9	99,6	87,8	98,2	95,7	89,1	98,5
	ERFK	80,7	29,8	0,0	0,0	100,0	-	-	80,7	0,0	0,0	89,3
	KRFK	94,2	89,8	97,9	99,9	93,2	99,2	59,5	99,8	98,5	74,6	96,4
	LR	89,9	77,2	59,6	63,6	96,8	79,0	76,4	92,1	67,9	69,4	94,4
	RLR	91,0	80,7	67,4	71,0	96,3	74,9	82,0	93,7	71,0	76,1	95,0
	RLRK	96,8	92,9	89,7	78,1	99,6	97,5	98,7	96,5	93,5	87,2	98,1
M2	RF	85,0	74,4	89,1	63,9	86,6	46,3	79,3	97,1	61,0	70,8	91,6
	RRF	88,1	75,5	62,8	74,5	93,0	66,1	63,6	93,9	64,4	68,6	93,5
	RRFK	92,2	81,8	60,7	62,5	99,5	98,5	93,8	91,6	75,1	75,0	95,4
	ERFK	80,7	29,8	0,0	0,0	100,0	-	-	80,7	0,0	0,0	89,3
	KRFK	8,0	12,2	13,5	73,4	0,7	100,0	6,2	18,7	23,9	11,5	1,3
	LR	88,9	74,7	46,3	72,7	96,5	83,9	68,7	91,3	59,7	70,6	93,9
	RLR	90,6	79,3	58,8	69,4	97,2	78,8	83,6	92,3	67,3	75,8	94,7
	RLRK	95,0	89,1	75,8	74,8	99,7	98,1	97,7	94,5	85,6	84,7	97,0

Tabelle B.16: Genauigkeitsmaße [%]. Datensatz: *Husum*. Grundlage: Gesamtes Bild.

Karte		GG	MF1	Vollständigkeit			Korrektheit			F1		
				Beb.	Wald	Bod.	Beb.	Wald	Bod.	Beb.	Wald	Bod.
M1	RF	98,8	64,3	99,9	-	88,5	99,0	0,0	99,0	99,4	0,0	93,4
	RRF	99,0	64,6	99,8	-	90,9	99,4	0,0	98,0	99,6	0,0	94,3
	RRFK	98,5	64,1	99,4	-	90,4	100,0	0,0	95,2	99,7	0,0	92,7
	ERFK	99,2	-	99,1	-	100,0	100,0	-	92,2	99,6	-	95,9
	KRFK	94,3	52,5	99,8	-	41,1	99,9	0,0	96,9	99,9	0,0	57,7
	LR	99,5	65,7	99,6	-	98,0	99,9	0,0	96,7	99,7	0,0	97,4
	RLR	99,5	65,7	99,7	-	97,4	99,8	0,0	97,2	99,8	0,0	97,3
	RLRK	99,4	65,4	99,5	-	98,5	99,9	0,0	94,9	99,7	0,0	96,7
M2	RF	97,8	64,0	99,7	-	88,7	97,8	0,0	98,6	98,7	0,0	93,4
	RRF	98,1	64,7	98,9	-	94,4	99,3	0,0	95,7	99,1	0,0	95,0
	RRFK	97,1	63,6	96,9	-	97,8	99,9	0,0	87,4	98,4	0,0	92,3
	ERFK	97,4	-	96,8	-	100,0	100,0	-	87,1	98,4	-	93,1
	KRFK	79,8	33,0	96,8	-	0,2	100,0	0,0	23,1	98,4	0,0	0,5
	LR	98,1	64,7	98,2	-	97,5	99,9	0,0	92,8	99,1	0,0	95,1
	RLR	98,6	65,2	98,7	-	98,4	99,8	0,0	94,3	99,2	0,0	96,3
	RLRK	97,7	64,1	97,4	-	99,0	99,9	0,0	89,0	98,6	0,0	93,7

Tabelle B.17: Genauigkeitsmaße [%]. Datensatz: *Husum*. Grundlage: Nur Änderungsgebiete.

Karte		GG	MF1	Vollständigkeit		Korrektheit		F1	
				Änd.	k. Änd.	Änd.	k. Änd.	Änd.	k. Änd.
M1	RF	87,0	76,3	98,1	85,8	43,7	99,7	60,4	92,2
	RRF	87,9	77,2	96,0	87,0	45,4	99,5	61,7	92,8
	RRFK	97,6	93,2	86,6	98,8	89,0	98,5	87,8	98,6
	ERFK	80,9	69,1	94,5	79,4	34,0	99,2	50,0	88,2
	KRFK	94,3	79,0	44,2	99,9	98,4	94,1	61,0	96,9
	LR	90,2	80,5	97,2	89,4	50,7	99,7	66,7	94,2
	RLR	91,3	82,1	97,6	90,6	53,7	99,7	69,3	94,9
	RLRK	97,0	92,3	95,5	97,1	78,9	99,5	86,4	98,3
M2	RF	87,0	83,3	98,6	84,0	61,0	99,6	75,4	91,1
	RRF	89,0	85,3	95,3	87,4	65,8	98,7	77,8	92,7
	RRFK	92,9	89,5	89,1	93,8	78,5	97,1	83,5	95,4
	ERFK	81,4	76,8	90,8	79,0	52,4	97,1	66,4	87,1
	KRFK	9,8	9,4	8,9	10,0	2,5	30,2	3,9	15,0
	LR	89,5	85,8	93,7	88,5	67,4	98,2	78,4	93,1
	RLR	91,1	87,8	95,9	89,9	70,7	98,9	81,4	94,2
	RLRK	95,7	93,5	92,0	96,6	87,4	97,9	89,6	97,3

Tabelle B.18: Genauigkeitsmaße [%]. Datensatz: *Husum*. Grundlage: Änderungsdetektion.

Hameln, Multitemporal

		GG	MF1	Vollständigkeit			Korrektheit			F1		
				Geb.	Bod.	Vers.	Geb.	Bod.	Vers.	Geb.	Boden	Vers.
KD	2010	82,6	70,6	49,5	88,8	73,5	81,4	89,0	52,7	61,6	88,9	61,4
	2016	88,9	78,7	58,6	96,3	74,5	95,5	92,5	64,5	72,6	94,4	69,1
KS	2010	87,7	75,3	45,0	97,5	65,3	81,7	88,0	89,2	58,0	92,5	75,4
	2016	89,3	79,6	55,3	96,8	76,9	92,5	91,3	74,1	69,2	94,0	75,5
KM	2010	88,6	77,9	54,0	96,1	73,0	81,8	90,7	77,6	65,0	93,3	75,3
	2016	89,7	80,7	63,0	95,9	78,0	94,1	93,3	66,7	75,5	94,6	71,9

Tabelle B.19: Genauigkeitsmaße [%]. Datensatz: *Hameln*. Grundlage: Gesamtes Bild.

		GG	MF1	Vollständigkeit			Korrektheit			F1		
				Geb.	Bod.	Vers.	Geb.	Bod.	Vers.	Geb.	Bod.	Vers.
KD	2010	91,0	72,3	93,6	83,3	62,0	99,7	36,6	78,9	96,6	50,8	69,4
	2016	92,8	79,1	94,4	95,4	71,8	99,9	54,8	69,2	97,1	69,7	70,5
KS	2010	90,1	69,6	93,1	89,4	50,0	99,8	34,1	85,4	96,3	49,4	63,1
	2016	92,2	78,0	93,8	92,9	71,6	99,8	47,7	77,1	96,7	63,0	74,2
KM	2010	93,2	79,3	94,5	95,7	72,9	99,9	46,6	84,1	97,1	62,7	78,1
	2016	93,8	81,6	95,0	95,9	76,6	99,9	58,4	73,2	97,4	72,6	74,9

Tabelle B.20: Genauigkeitsmaße [%]. Datensatz: *Hameln*. Grundlage: Nur Änderungsgebiete.

		GG	MF1	Vollständigkeit		Korrektheit		F1	
				Änd.	k. Änd.	Änd.	k. Änd.	Änd.	k. Änd.
KD	2010	83,2	74,5	59,8	89,3	59,7	89,4	59,7	89,4
	2016	91,2	87,6	79,3	94,9	82,7	93,7	81,0	94,3
KS	2010	88,1	79,2	54,2	97,0	82,9	88,9	65,5	92,8
	2016	90,7	86,3	72,6	96,2	85,5	91,9	78,5	94,0
KM	2010	89,5	83,7	71,7	94,2	76,4	92,7	74,0	93,4
	2016	91,7	88,4	82,1	94,6	82,5	94,5	82,3	94,6

Tabelle B.21: Genauigkeitsmaße [%]. Datensatz: *Hameln*. Grundlage: Änderungsdetektion.



(a) **KD**



(b) **KS**



(c) **KM**

Abbildung B.1: Ergebnisse für das Jahr 2000 von Gebiet NA des Datensatzes *Hameln*. Rot: *Gebäude*. Grau: *Versiegelte Fläche*. Beige: *Boden*.



(a) **KD**



(b) **KS**



(c) **KM**

Abbildung B.2: Ergebnisse für das Jahr 2016 von Gebiet NA des Datensatzes *Hameln*. Rot: *Gebäude*. Grau: *Versiegelte Fläche*. Beige: *Boden*.

Las Vegas, Multitemporal

Karte		GG	MF1	Vollständigkeit			Korrektheit			F1		
				Bod.	Beb.	Wa.	Bod.	Beb.	Wa.	Bod.	Beb.	Wa.
KD	1991	97,2	92,4	97,8	89,4	98,9	99,1	79,0	91,1	98,4	83,9	94,8
	2000	94,5	90,5	98,6	68,5	97,4	95,0	89,2	97,4	96,8	77,5	97,4
	2016	88,1	81,2	98,7	43,9	98,8	87,7	93,1	84,6	92,9	59,7	91,1
KS	1991	97,1	92,2	97,7	89,7	99,0	99,1	78,1	91,0	98,4	83,5	94,8
	2000	94,4	90,7	98,1	71,3	97,5	95,4	86,2	96,9	96,7	78,1	97,2
	2016	90,3	85,7	98,3	57,2	98,8	90,3	91,5	87,2	94,1	70,4	92,6
KM	1991	95,3	88,8	95,8	88,0	99,9	99,0	64,6	89,8	97,4	74,5	94,6
	2000	93,5	89,4	96,9	71,0	99,7	95,4	79,9	94,2	96,2	75,2	96,9
	2016	87,5	78,9	97,0	47,9	99,0	88,3	85,5	71,4	92,5	61,4	83,0

Tabelle B.22: Genauigkeitsmaße [%]. Datensatz: *Las Vegas*. Grundlage: Gesamtes Bild.

Karte		GG	MF1	Vollständigkeit			Korrektheit			F1		
				Bod.	Beb.	Wa.	Bod.	Beb.	Wa.	Bod.	Beb.	Wa.
KD	1991	99,0	59,0	99,6	67,1	2,3	99,3	90,6	0,3	99,5	77,1	0,5
	2000	95,6	61,4	99,9	48,2	21,6	95,6	98,9	21,9	97,7	64,8	21,7
	2016	89,4	52,2	99,7	26,1	59,9	89,2	98,3	12,7	94,2	41,3	21,0
KS	1991	99,0	59,5	99,6	68,3	4,8	99,4	90,6	0,5	99,5	77,9	1,0
	2000	96,1	64,4	99,9	53,9	26,2	96,0	98,7	25,0	97,9	69,7	25,6
	2016	92,1	59,4	99,7	45,4	62,3	91,8	98,8	12,2	95,6	62,3	20,4
KM	1991	99,3	69,8	99,7	78,1	97,3	99,6	89,9	15,0	99,6	83,6	26,0
	2000	96,4	80,0	99,9	56,6	92,6	96,3	98,9	56,1	98,1	72,0	69,9
	2016	91,9	63,1	99,7	44,0	88,3	91,6	97,6	20,4	95,5	60,7	33,1

Tabelle B.23: Genauigkeitsmaße [%]. Datensatz: *Las Vegas*. Grundlage: Nur Änderungsgebiete.

Karte		GG	MF1	Vollständigkeit		Korrektheit		F1	
				Änd.	k. Änd.	Änd.	k. Änd.	Änd.	k. Änd.
KD	1991	97,2	73,7	56,5	98,2	43,0	99,0	48,8	98,6
	2000	94,5	77,9	47,4	98,8	77,5	95,4	58,8	97,0
	2016	88,2	68,3	30,0	98,4	77,3	88,8	43,2	93,4
KS	1991	97,2	73,7	58,0	98,1	42,2	99,0	48,9	98,5
	2000	94,5	79,2	52,9	98,2	72,8	95,8	61,3	97,0
	2016	90,4	77,3	47,8	97,9	80,4	91,4	60,0	94,5
KM	1991	95,4	70,1	72,7	96,0	30,2	99,3	42,7	97,6
	2000	93,6	78,0	57,0	96,9	62,2	96,1	59,5	96,5
	2016	87,7	73,2	47,1	94,9	61,8	91,0	53,5	92,9

Tabelle B.24: Genauigkeitsmaße [%]. Datensatz: *Las Vegas*. Grundlage: Änderungsdetektion.

Literatur

Abellán, J., 2013. Ensembles of decision trees based on imprecise probabilities and uncertainty measures. *Information Fusion* 14 (4): 423–430.

Abellán, J. and Moral, S., 2003. Building classification trees using the total uncertainty criterion. *International Journal of Intelligent Systems* 18 (12): 1215–1225.

Abellán, J., Mantas, C. J. and Castellano, J. G., 2017. A random forest approach using imprecise probabilities. *Knowledge-Based Systems* 134: 72–84.

Abellán, J., Mantas, C. J., Castellano, J. G. and Moral-García, S., 2018. Increasing diversity in random forest learning algorithm via imprecise probabilities. *Expert Systems with Applications* 97: 228–243.

Albert, L., Rottensteiner, F. and Heipke, C., 2015. An iterative inference procedure applying conditional random fields for simultaneous classification of land cover and land use. In: *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. II-3/W5, pp. 369–376.

Alcantarilla, P. F., Stent, S., Ros, G., Arroyo, R. and Gherardi, R., 2018. Street-view change detection with deconvolutional networks. *Autonomous Robots* 42 (7): 1301–1322.

An, W. and Liang, M., 2013. Fuzzy support vector machine based on within-class scatter for classification problems with outliers or noises. *Neurocomputing* 110: 101–110.

Arbeitsgemeinschaft der Vermessungsverwaltungen der Länder der Bundesrepublik Deutschland (AdV), 2008. ALKIS-Objektartenkatalog 6.0. Technical Report. http://www.adv-online.de/AAA-Modell/Dokumente-der-GeoInfoDok/, abgerufen am: 2018-08.

Bähr, H.-P. and Vögtle, T., 2005. *Digitale Bildverarbeitung: Anwendung in Photogrammetrie und Fernerkundung und GIS*. 4. Auflage, Wichmann. Heidelberg.

Ban, Y., 2016. *Multitemporal Remote Sensing: Methods and Applications*. Remote Sensing and Digital Image Processing, Springer International Publishing. Cham, Switzerland.

Beigman, E. and Klebanov, B. B., 2009. Learning with annotation noise. In: *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP*, Vol. 1, Association for Computational Linguistics, pp. 280–287.

Bellman, R. E., 1961. *Adaptive control processes: a guided tour*. Princeton University Press. London: Oxford University Press.

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

Bootkrajang, J. and Kabán, A., 2012. Label-noise robust logistic regression and its applications. In: *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, Springer, pp. 143–158.

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

Breiman, L., 1996. Bagging predictors. *Machine Learning* 24 (2): 123–140.

Breiman, L., 2001. Random forests. *Machine Learning* 45 (1): 5–32.

Breiman, L., Friedman, J. H., Olshen, R. A. and Stone, C. J., 1984. *Classification and regression trees*. ChapmanandHall/CRC. London, UK.

Bruzzone, L. and Marconcini, M., 2009. Domain adaptation problems: A dasvm classification technique and a circular validation strategy. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 32 (5): 770–787.

Bruzzone, L. and Persello, C., 2009. A novel context-sensitive semisupervised svm classifier robust to mislabeled training samples. *IEEE Transactions on Geoscience and Remote Sensing* 47 (7): 2142–2154.

Bruzzone, L. and Prieto, D. F., 2000. Automatic analysis of the difference image for unsupervised change detection. *IEEE Transactions on Geoscience and Remote Sensing* 38 (3): 1171–1182.

Bruzzone, L., Prieto, D. F. and Serpico, S. B., 1999. A neural-statistical approach to multitemporal and multisource remote-sensing image classification. *IEEE Transactions on Geoscience and Remote Sensing* 37 (3): 1350–1359.

Burger, W. and Burge, M. J., 2005. *Digitale Bildverarbeitung: Eine algorithmische Einführung mit Java*. 3. Auflage, Springer. Berlin, Heidelberg.

Büschenfeld, T., 2013. Klassifikation von Satellitenbildern unter Ausnutzung von Klassifikationsunsicherheiten. PhD thesis, Fortschritt-Berichte VDI, Reihe 10 Informatik / Kommunikation, Vol. 828, Institute of Information Processing, Leibniz Universität Hannover, Germany.

Büschenfeld, T. and Ostermann, J., 2012. Automatic refinement of training data for classification of satellite imagery. In: *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. I-7, Göttingen: Copernicus GmbH, pp. 117–122.

Celik, T., 2009. Unsupervised change detection in satellite images using principal component analysis and k-means clustering. *IEEE Geoscience and Remote Sensing Letters* 6 (4): 772–776.

Chapelle, O. and Zien, A., 2005. Semi-supervised classification by low density separation. In: *AISTATS Proceedings of the Tenth International Workshop on Artificial Intelligence and Statistics*, pp. 57–64.

Chapelle, O., Weston, J. and Schölkopf, B., 2003. Cluster kernels for semi-supervised learning. In: *Advances in Neural Information Processing Systems 15*, MIT Press, pp. 601–608.

Clement, M. A., Kilsby, C. G. and Moore, P., 2018. Multi-temporal synthetic aperture radar flood mapping using change detection. *Journal of Flood Risk Management* 11 (2): 152–168.

Cover, T. M. and Thomas, J. A., 2012. *Elements of information theory*. John Wiley & Sons. New York (NY), USA.

Cramer, M., 2010. The DGPF-test on digital airborne camera evaluation – overview and test design. *Photogrammetrie-Fernerkundung-Geoinformation* 2010 (2): 73–82.

Dalal, N. and Triggs, B., 2005. Histograms of oriented gradients for human detection. In: *IEEE Conference on Computer Vision and Pattern Recognition*, Vol. 1, pp. 886–893.

Damodaran, B. B., Flamary, R., Seguy, V. and Courty, N., 2020. An entropic optimal transport loss for learning deep neural networks under label noise in remote sensing images. *Computer Vision and Image Understanding* 191: 102863.

Demir, B., Persello, C. and Bruzzone, L., 2010. Batch-mode active-learning methods for the interactive classification of remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing* 49 (3): 1014–1031.

Dempster, A. P., Laird, N. M. and Rubin, D. B., 1977. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)* 39 (1): 1–22.

Denoeux, T., 1995. A k-nearest neighbor classification rule based on dempster-shafer theory. *IEEE Transactions on Systems, Man, and Cybernetics* 25 (5): 804–813.

Drory, A., Avidan, S. and Giryes, R., 2018. On the resistance of neural nets to label noise. *arXiv preprint arXiv:1803.11410*.

Duda, R. O., Hart, P. E., Stork, D. G. et al., 2001. *Pattern Classification*. Vol. 55, 2. Auflage, John Wiley & Sons. New York (NY), USA.

Folleco, A., Khoshgoftaar, T. M., Van Hulse, J. and Bullard, L., 2008. Identifying learners robust to low quality data. In: *IEEE International Conference on Information Reuse and Integration*, pp. 190–195.

Förstner, W., 2013. Graphical models in geodesy and photogrammetry. *Photogrammetrie Fernerkundung Geoinformation* 2013 (4): 255–268.

Frénay, B. and Verleysen, M., 2014. Classification in the presence of label noise: a survey. *IEEE Transactions on Neural Networks and Learning Systems* 25 (5): 845–869.

Frénay, B., Doquire, G. and Verleysen, M., 2014. Estimating mutual information for feature selection in the presence of label noise. *Computational Statistics and Data Analysis* 71: 832–848.

Frey, B. and MacKay, D. J., 1998. A revolution: Belief propagation in graphs with cycles. *Advances in Neural Information Processing Systems 10* pp. 479–485.

Gey, M. H., 2015. *Instrumentelle Analytik und Bioanalytik*. Springer. Berlin, Heidelberg.

Guillaumin, M., Verbeek, J. and Schmid, C., 2010. Multimodal semi-supervised learning for image classification. In: *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 902–909.

Guo, Y. and Schuurmans, D., 2008. Discriminative batch mode active learning. In: *Advances in Neural Information Processing Systems 20*, pp. 593–600.

Hänsch, R. and Hellwich, O., 2017. Random forests. In: *Handbuch der Geodäsie*, 1. Auflage, Springer, pp. 603–644. Berlin.

Haralick, R. M., Shanmugam, K. et al., 1973. Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics* SMC-3 (6): 610–621.

He, S., Chen, H., Zhu, Z., Ward, D. G., Cooper, H. J., Viant, M. R., Heath, J. K. and Yao, X., 2015. Robust twin boosting for feature selection from high-dimensional omics data with label noise. *Information Sciences* 291: 1–18.

Heipke, C., Mayer, H., Wiedemann, C. and Jamet, O., 1997. Evaluation of automatic road extraction. In: *International Archives of Photogrammetry and Remote Sensing*, Vol. XXII-3/4W2, pp. 151–160.

Hoberg, T., Rottensteiner, F., Feitosa, R. Q. and 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.

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

Jia, K., Liang, S., Wei, X., Zhang, L., Yao, Y. and Gao, S., 2014. Automatic land-cover update approach integrating iterative training sample selection and a markov random field model. *Remote Sensing Letters* 5 (2): 148–156.

Jianya, G., Haigang, S., Guorui, M. and Qiming, Z., 2008. A review of multi-temporal remote sensing data change detection algorithms. In: *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. XXXVII-B7, pp. 757–762.

Kaiser, P., Wegner, J. D., Lucchi, A., Jaggi, M., Hofmann, T. and Schindler, K., 2017. Learning aerial image segmentation from online maps. *IEEE Transactions on Geoscience and Remote Sensing* 55 (11): 6054–6068.

Kalai, A. and Servedio, R. A., 2003. Boosting in the presence of noise. In: *Proceedings of the Thirty-Fifth Annual ACM Symposium on Theory of Computing*, ACM, pp. 195–205.

Khalid, S., Khalil, T. and Nasreen, S., 2014. A survey of feature selection and feature extraction techniques in machine learning. In: *Science and Information Conference*, pp. 372–378.

Knudsen, T. and Olsen, B. P., 2003. Automated change detection for updates of digital map databases. *Photogrammetric Engineering and Remote Sensing* 69 (11): 1289–1296.

Kozachenko, L. and Leonenko, N. N., 1987. Sample estimate of the entropy of a random vector. *Problemy Peredachi Informatsii* 23 (2): 9–16.

Kumar, S. and Hebert, M., 2006. Discriminative random fields. *International Journal of Computer Vision* 68 (2): 179–201.

Landesvermessung und Geobasisinformation, 2019a. Digitale Geländemodelle (DGM). https://www.lgln.niedersachsen.de/startseite/geodaten_karten/3dgeobasisdaten/dgm/digitale-gelaendemodelle–dgm-143150.html, abgerufen am: 2019-08-08.

Landesvermessung und Geobasisinformation, 2019b. Digitales Oberflächenmodell (DOM). https://www.lgln.niedersachsen.de/startseite/geodaten_karten/3d_geobasisdaten/dom/digitales-oberflaechenmodell-dom-142876.html, abgerufen am: 2019-08-08.

Lawrence, N. D. and Schölkopf, B., 2001. Estimating a kernel fisher discriminant in the presence of label noise. In: *Proceedings of the 18th International Conference on Machine Learning*, Vol. 1, pp. 306–313.

LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W. and Jackel, L. D., 1989. Backpropagation applied to handwritten zip code recognition. *Neural computation* 1 (4): 541–551.

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

Li, Y., Wessels, L. F., de Ridder, D. and Reinders, M. J., 2007. Classification in the presence of class noise using a probabilistic kernel fisher method. *Pattern Recognition* 40 (12): 3349–3357.

Lillesand, T., Kiefer, R. W. and Chipman, J., 2014. *Remote sensing and image interpretation.* John Wiley & Sons. New York (NY), USA.

Liu, S., Marinelli, D., Bruzzone, L. and Bovolo, F., 2019. A review of change detection in multitemporal hyperspectral images: Current techniques, applications, and challenges. *IEEE Geoscience and Remote Sensing Magazine* 7 (2): 140–158.

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

Maas, A., Rottensteiner, F. and Heipke, C., 2016. Using label noise robust logistic regression for automated updating of topographic geospatial databases. In: *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. III-7, pp. 133–140.

Maggiori, E., Charpiat, G., Tarabalka, Y. and Alliez, P., 2017. Recurrent neural networks to correct satellite image classification maps. *IEEE Transactions on Geoscience and Remote Sensing* 55 (9): 4962–4971.

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

Matasci, G., Volpi, M., Kanevski, M., Bruzzone, L. and Tuia, D., 2015. Semisupervised transfer component analysis for domain adaptation in remote sensing image classification. *IEEE Transactions on Geoscience and Remote Sensing* 53 (7): 3550–3564.

Melgani, F. and Serpico, S. B., 2003. A Markov random field approach to spatiotemporal contextual image classification. *IEEE Transactions on Geoscience and Remote Sensing* 41 (11): 2478–2487.

Mika, S., Ratsch, G., Weston, J., Scholkopf, B. and Mullers, K.-R., 1999. Fisher discriminant analysis with kernels. In: *Neural Networks for Signal Processing IX: Proceedings of*

the 1999 IEEE Signal Processing Society Workshop, pp. 41–48.

Mnih, V. and Hinton, G. E., 2012. Learning to label aerial images from noisy data. In: *Proceedings of the 29th International Conference on Machine Learning*, pp. 567–574.

Nasa, Landsat Science, 2019. Spectral characterization. https://landsat.gsfc.nasa.gov/spectral-characterization/, abgerufen am: 2019-05-23.

Olsen, B. P., Knudsen, T. and Frederiksen, P., 2002. Digital change detection for map database update. In: *International Archives of Photogrammetry Remote Sensing and Spatial Information Sciences*, Vol. XXXIV-2, Natural Resources Canada, pp. 357–364.

Pan, S. J. and Yang, Q., 2009. A survey on transfer learning. *IEEE Transactions on Knowledge and Data engineering* 22 (10): 1345–1359.

Patrini, G., Rozza, A., Menon, A. K., Nock, R. and Qu, L., 2017. Making deep neural networks robust to label noise: A loss correction approach. In: *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2233–2241.

Paul, A., Rottensteiner, F. and Heipke, C., 2016. Iterative re-weighted instance transfer for domain adaptation. In: *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. III-3, pp. 339–346.

Pelletier, C., Valero, S., Inglada, J., Champion, N., Marais Sicre, C. and Dedieu, G., 2017. Effect of training class label noise on classification performances for land cover mapping with satellite image time series. *Remote Sensing* 9 (2): 1–24. Paper 173.

Quinlan, J. R., 1986. Induction of decision trees. *Machine Learning* 1 (1): 81–106.

Quinlan, J. R., 1993. *C4. 5: Programs for machine learning*. Morgan Kaufmann, San Mateo, CA, USA.

Radoux, J. and Defourny, P., 2010. Automated image-to-map discrepancy detection using iterative trimming. *Photogrammetric Engineering and Remote Sensing* 76 (2): 173–181.

Radoux, J., Lamarche, C., Van Bogaert, E., Bontemps, S., Brockmann, C. and Defourny, P., 2014. Automated training sample extraction for global land cover mapping. *Remote Sensing* 6 (5): 3965–3987.

Rolnick, D., Veit, A., Belongie, S. and Shavit, N., 2017. Deep learning is robust to massive label noise. *arXiv preprint arXiv:1705.10694*.

Saha, S., Bovolo, F. and Bruzzone, L., 2019. Unsupervised deep change vector analysis for multiple-change detection in vhr images. *IEEE Transactions on Geoscience and Remote Sensing* 57 (6): 3677–3693.

Sarma, A. and Palmer, D. D., 2004. Context-based speech recognition error detection and correction. In: *Proceedings of Human Language Technology Conference/North America Chapter AACL Annual Meeting*, pp. 85–88.

Schistad Solberg, A. H., Taxt, T. and 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.

Shanab, A. A., Khoshgoftaar, T. M. and Wald, R., 2012. Robustness of threshold-based feature rankers with data sampling on noisy and imbalanced data. In: *Proceedings of the 25th International Florida Artificial Intelligence Research Society Conference*, pp. 92–97.

Shannon, C. E., 1948. A mathematical theory of communication. *Bell System Technical Journal* 27 (3): 379–423.

Stadt Hameln, Abteilung Stadtentwicklung und Planung, 2016. Kartenmaterial der Stadt Hameln aus dem Archiv.

Stefanski, J., Chaskovskyy, O. and Waske, B., 2014. Mapping and monitoring of land use changes in post-soviet western ukraine using remote sensing data. *Applied Geography* 55: 155–164.

Subudhi, B. N., Bovolo, F., Ghosh, A. and Bruzzone, L., 2014. Spatio-contextual fuzzy clustering with markov random field model for change detection in remotely sensed images. *Optics and Laser Technology* 57: 284–292.

Süße, H. and Rodner, E., 2014. *Bildverarbeitung und Objekterkennung*. Springer Vieweg. Wiesbaden.

Szeliski, R., Zabih, R., Scharstein, D., Veksler, O., Kolmogorov, V., Agarwala, A., Tappen, M. and 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.

Tönnies, K. D., 2005. *Grundlagen der Bildverarbeitung*. 1. Auflage, Pearson Studium. München.

Tuia, D., Persello, C. and Bruzzone, L., 2016. Domain adaptation for the classification of remote sensing data: An overview of recent advances. *IEEE Geoscience and Remote Sensing Magazine* 4 (2): 41–57.

Tuia, D., Ratle, F., Pacifici, F., Kanevski, M. F. and Emery, W. J., 2009. Active learning methods for remote sensing image classification. *IEEE Transactions on Geoscience and Remote Sensing* 47 (7): 2218–2232.

Vahdat, A., 2017. Toward robustness against label noise in training deep discriminative neural networks. In: *Advances in Neural Information Processing Systems 30*, pp. 5596–5605.

Vapnik, V., 1995. *The Nature of Statistical Learning Theory*. Springer-Verlag. New York (NY), USA.

Veit, A., Alldrin, N., Chechik, G., Krasin, I., Gupta, A. and Belongie, S. J., 2017. Learning from noisy large-scale datasets with minimal supervision. In: *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 6575–6583.

Vergara, J. R. and Estévez, P. A., 2014. A review of feature selection methods based on mutual information. *Neural Computing and Applications* 24 (1): 175–186.

Vogt, K., Paul, A., Ostermann, J., Rottensteiner, F. and Heipke, C., 2018. Unsupervised source selection for domain adaptation. *Photogrammetric Engineering and Remote Sensing* 84 (5): 249–261.

Wang, M. and Deng, W., 2018. Deep visual domain adaptation: A survey. *Neurocomputing* 312: 135–153.

Wang, Q., Yuan, Z., Du, Q. and Li, X., 2019. GETNET: A general end-to-end 2-D CNN framework for hyperspectral image change detection. *IEEE Transactions on Geoscience and Remote Sensing* 57 (1): 3–13.

Wegner, J., Rottensteiner, F., Gerke, M. and Sohn, G., 2015. The ISPRS 2d labelling challenge. http://www2.isprs.org/commissions/comm3/wg4/semantic-labeling.html, abgerufen am: 2017-04-05.

Wu, T.-F., Lin, C.-J. and Weng, R. C., 2004. Probability estimates for multi-class classification by pairwise coupling. *Journal of Machine Learning Research* 5: 975–1005.

Xiao, T., Xia, T., Yang, Y., Huang, C. and Wang, X., 2015. Learning from massive noisy labeled data for image classification. In: *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2691–2699.

Zhang, W., Rekaya, R. and Bertrand, K., 2005. A method for predicting disease subtypes in presence of misclassification among training samples using gene expression: application to human breast cancer. *Bioinformatics* 22 (3): 317–325.

Zhao, Z., 2017. Classification in the presence of heavy label noise: A Markov chain sampling framework. PhD thesis, Faculty of Applied Sciences, School of Computing Science, Simon Fraser University, Canada.

Zheng, A. and Casari, A., 2019. *Merkmalskonstruktion für Machine Learning: Prinzipien und Techniken der Datenaufbereitung*. O'Reilly, Dpunkt.Verlag GmbH. Heidelberg.

Zhu, X., Wu, X. and Chen, Q., 2003. Eliminating class noise in large datasets. In: *Proceedings of the 20th International Conference on Machine Learning*, pp. 920–927.

Danksagung

An dieser Stelle möchte ich mich bei allen Personen bedanken, die mich in den letzten Jahren bei der Erstellung dieser Arbeit direkt oder indirekt unterstützt haben.

Besonders hervorheben möchte ich meinen Doktorvater Prof. Dr. techn. Franz Rottensteiner für die erstklassige und intensive fachliche Betreuung. Sein Vertrauen und seine Unterstützung haben zum Gelingen dieser Arbeit erheblich beigetragen.

Darüber hinaus möchte ich auch Prof. Dr. Ing. Monika Sester und Prof. Dr. Ing. Stefan Hinz danken, dass sie das Korreferat für diese Dissertation übernommen haben.

Weiterhin möchte ich Prof. Dr. Ing. Christian Heipke für die vielen konstruktiven Gespräche danken.

Zudem möchte ich mich bei meinen Kollegen und Freunden am Institut für Photogrammetrie und Geoinformation (IPI) der Leibniz Universität Hannover für die schöne gemeinsame Zeit und die angenehme Arbeitsatmosphäre bedanken. Besonders mit meinem Zimmerkollegen Andreas gab es viele schöne Diskussionen.

Auch meinem Betreuer Magnús Örn Úlfarsson während des dreimonatigen Forschungsaufenthaltes an der Háskóli Íslands in Reykjavík (Island) und meinen dortigen ehemaligen Kollegen möchte ich für die freundlichen und interessanten Gespräche danken. Sie gaben mir die Gelegenheit ein spannendes fachliches Themenfeld kennenzulernen.

Für die finanzielle Unterstützung des Forschungsprojektes, in dessen Rahmen diese Dissertation entstand, möchte ich der Deutschen Forschungsgemeinschaft (DFG) danken. Ebenfalls für die finanzielle Unterstützung in Form des Caroline Herschel Programms und auch für die stets freundlichen Ansprechpartnerinnen bedanke ich mich beim Hochschulbüro für Chancenvielfalt. Mein Forschungsaufenthalt an der Háskóli Íslands wurde zusätzliche durch die Graduiertenakademie finanziell unterstützt, wofür ich mich ebenfalls bedanken möchte.

Darüber hinaus möchte ich dem Landesamt für Geoinformation und Landesvermessung Niedersachsen (LGLN) und dem Katasteramt Hameln für die Bereitstellung der Daten für den *Hameln* Datensatz danken. Bei der Deutschen Gesellschaft für Photogrammetrie, Fernerkundung und Geoinformation e.V. [Cramer, 2010] möchte ich mich für die Bereitstellung der Daten aus *Vaihingen* bedanken.

Ganz besonders möchte ich meinen Eltern Brigitte und Dr. rer. nat. Reinhard Maas sowie meiner Schwester Tabea für ihre uneingeschränkte Liebe und Unterstützung danken. Auch meinen Freunden möchte ich für die schönen Gespräche und die aufbauenden Worte danken. Einen zusätzlichen Dank möchte ich allen Personen geben, die meine Arbeit korrektur gelesen haben. Abschließend möchte ich mich bei meinem Freund Jens für seine Liebe, seinen Humor und seine Ruhe bedanken, durch die ich immer einen Lichtblick hatte.

Lebenslauf

Persönliche Daten

Name	Alina Elisabeth Maas
Geburtsdatum	14.08.1990
Geburtsort	Hannover, Deutschland

Beruflicher Werdegang

seit 10.2019	**Landesamt für Geoinformation und Landesvermessung Niedersachsen** Fachexpertin für Bildanalyse
07.2018 – 09.2018	**Faculty of Electrical and Computer Engineering, University of Iceland** Wissenschaftliche Mitarbeiterin
06.2015 – 10.2019	**Institut für Photogrammetrie und GeoInformation** Wissenschaftliche Mitarbeiterin
04.2014 – 06.2014	**Firma Robert Bosch GmbH, Car Multimedia** Praktikantin
11.2011 – 03.2014	**Institut für Photogrammetrie und GeoInformation** Hilfswissenschaftliche Mitarbeiterin
10.2009 – 03.2010	**Institut für Kartographie und Geoinformatik** Hilfswissenschaftliche Mitarbeiterin

Ausbildung

10.2012 – 05.2015	**Navigation und Umweltrobotik** (Master, Note: 1,6) Leibniz Universität Hannover
09.2010 – 07.2011	**Surveying and Mapping Science** Newcastle University
10.2008 – 12.2012	**Geodäsie und Geoinformatik** (Bachelor, Note: 1,9) Leibniz Universität Hannover
1996 – 2008	**Schulbildung** (Abitur, Note:2,4) Grundschule Kestnerstraße, Orientierungsstufe Lürstraße, Gymnasium Elsa-Brändström-Schule

Stipendien und Auszeichnungen

03.2016 – 09.2019	**Caroline Herschel Programm** ◦ Förderung des weiblichen Nachwuchses auf dem Weg zur Professur ◦ Hochschulbüro für ChancenVielfalt
01.2019	**ESRI Award for Best Scientific Paper in Geographic Information Systems** ◦ American Society for Photogrammetry and Remote Sensing
07.2018 – 09.2018	**Auslandsstipendium** ◦ Forschungsaufenthalt in Island ◦ Graduiertenakademie
08.2013	**Harbert-Buchpreis** ◦ gute Leistungen in der Abschlussprüfung ◦ DVW Gesellschaft für Geodäsie, Geoinformation und Landmanagement e.V.
02.2013	**Preis des Präsidiums** ◦ beste Absolventin im Bachelorstudium 2012 ◦ Christian-Kuhlemann-Stiftung und der Freundeskreis der LUH
10.2012 – 09.2013	**Erstattung der Studiengebühren** ◦ gute Leistungen im Hauptstudium ◦ Freundeskreis der Leibniz Universität Hannover e.V.
09.2010 – 07.2011	**Erasmus** ◦ Auslandssemester in England ◦ EU-Fördermittel

Veröffentlichungen

begutachtete Zeitschriftenartikel

Maas, A., Rottensteiner, F., Heipke, C. (2019): A label noise tolerant random forest for the classification of remote sensing data based on outdated maps for training. In: Computer Vision and Image Understanding 188, Paper 102782

Maas, A., Rottensteiner, F., Alobeid, A., Heipke, C. (2018): Multitemporal classification under label noise based on outdated maps. In: Photogrammetric Engineering and Remote Sensing 84(5): 263-277

begutachtete Tagungsbeiträge

Maas, A., Rottensteiner, F., Heipke, C. (2017): Classification under label noise using outdated maps. In: ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences IV-1/W1, 215-222

Maas, A., Rottensteiner, F., Heipke, C. (2016): Using label noise robust logistic regression for automated updating of topographic geospatial databases. In: ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences III-7, 133-140

weitere Tagungsbeiträge

Maas, A., Rasti, B., Ulfarsson, M. (2018): Label Noise robust Classification of Hyperspectral Data. In: Proceedings of the 9th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing, Amsterdam, 5p.

Maas, A., Alrajhi, M., Alobeid, A., Heipke C. (2017): Automatic classification of high resolution satellite imagery - a case study for urban areas in the Kingdom of Saudi Arabia. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLII-1-W1, 11-16.