



TECHNISCHE UNIVERSITÄT MÜNCHEN
Institut für Photogrammetrie und Kartographie
Fachgebiet Photogrammetrie und Fernerkundung

Detektion von Fahrzeugen in optischen Satellitenbildern

Jens Leitloff

Vollständiger Abdruck der von der Fakultät für Bauingenieur- und Vermessungswesen der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktor-Ingenieurs (Dr.-Ing.)

genehmigten Dissertation.

Vorsitzender: Univ.-Prof. Dr.-Ing. Fritz Busch

Prüfer der Dissertation:

1. Univ.-Prof. Dr.-Ing. Uwe Stilla
2. Univ.-Prof. Dr.-Ing. habil. Stefan Hinz,
Karlsruher Institut für Technologie (KIT)

Die Dissertation wurde am 22.06.2011 bei der Technischen Universität München eingereicht und durch die Fakultät für Bauingenieur- und Vermessungswesen am 18.09.2011 angenommen.

Kurzfassung

In den vergangenen 40 Jahren ist das Verkehrsaufkommen durch Privatpersonen um mehr als das 3,5-fache gestiegen und der Straßengüterverkehr nahm um ca. 800 % zu. In urbanen Gebieten verursacht der Verkehr bis zu 75 % der Feinstaubbelastung. Daher besteht ein erhöhter Bedarf an intelligenterer Verkehrsüberwachung und -beeinflussung, welchen zunächst eine bessere Modellierung vorausgeht. Die Erfassung der hierfür benötigten Daten erfolgt meist durch terrestrische stationäre Sensoren wie Induktionsschleifen und Brückenkameras. Da diese Systeme nur einen Teil des Straßennetzes abdecken, wird im Rahmen der Arbeit ein neuer Ansatz zur Erfassung verkehrsrelevanter Parameter aus optischen Satellitenbildern entwickelt. Die flächendeckend gewonnenen Daten können als wichtige Ergänzung für vorhandene Verkehrsmodelle genutzt werden.

Um eine möglichst geringe Fehlerrate bei der Fahrzeugerkennung zu erreichen, werden Daten über das Straßennetz aus vorhandenen Geoinformationen genutzt. Diese dienen hauptsächlich zum Ausschluss nicht relevanter Bildregionen. Da Fahrzeuge in Satellitenbildern keine Merkmale wie Front- oder Heckscheiben zeigen, erweisen sich etablierte Verfahren zur Fahrzeugerkennung aus höher aufgelösten Daten als nicht anwendbar.

In dieser Arbeit wird ein neuer Ansatz entwickelt, der sowohl *isolierte* und als auch *in Gruppen auftretende* Fahrzeuge berücksichtigt. Es wird ein *zweistufiges Klassifikationsverfahren* zur Detektion isolierter Fahrzeuge vorgeschlagen. Dieses erreicht eine hohe Zuverlässigkeit. Hypothesen für Fahrzeugreihen werden durch differentialgeometrische Linienextraktion gebildet. Zur Erkennung von Einzelfahrzeugen innerhalb der Fahrzeugreihen werden speziell auf die Problemstellung angepasste *robuste Schätzverfahren* eingehend untersucht. Nach Zusammenführung der Ergebnisse der Mustererkennung und der robusten Parameterschätzung ist es möglich, eine *quantitative Abschätzung* des Verkehrsaufkommens einzelner Straßen zu geben. Erstmals wird für Einzelfahrzeuge automatisch der *Bewegungsstatus* geprüft. Hierfür wird die zeitliche Differenz zwischen den Aufnahmen der einzelnen Spektralkanäle genutzt. Die Initialisierung der verwendeten Methoden erfordert *kein Expertenwissen*. Alle zur Detektion und Bewegungsschätzung verwendeten Parameter werden aus manuell erfassten Trainingsdaten abgeleitet oder durch statistische Testverfahren gewonnen.

Das Verfahren wurde mit QuickBird Bildern eines städtischen Bereichs evaluiert. Während die Vollständigkeit der Ergebnisse bei 65 % liegt, erreicht der Ansatz eine hohe Zuverlässigkeit von durchschnittlich 95 %. Für einige Fahrzeuge konnte ein signifikanter Versatz zwischen den Aufnahmen der Spektralkanäle berechnet werden. Die daraus abgeleitete Geschwindigkeit ist für städtische Gebiete plausibel.

Abstract

Within the last 40 years, the traffic volume of private individuals increased by a factor of 3.5. In the same time, road freight transport has even increased by eight times. The traffic produces 75 % of the particulate air pollution in urban areas. Thus, better modeling of traffic situations is needed to perform intelligent traffic monitoring and control. Currently, traffic research is mostly based on data from stationary sensors like induction loops or bridge mounted cameras. These systems only acquire the traffic flow on parts of the main roads. In this work an approach for automatic vehicle detection from satellite imagery is developed. The obtained traffic data provide useful complement information to current traffic models.

The number of possible false detection can be reduced by restricting the detection methods to road areas, which are obtained from existing geoinformation. In satellite imagery individual parts or substructures of a vehicle can rarely be distinguished. Thus, matured methods for higher resolution imagery cannot be used directly. The newly developed approach distinguishes between *isolated* and *grouped vehicles*. A highly reliable *two-stage classification procedure* for the detection of isolated vehicles is proposed. Hypotheses for vehicle queues are found using differential geometric line extraction. Specialized *robust estimation procedures* are investigated to determine single vehicles within those vehicle queues. The results for isolated and grouped vehicles give a *quantitative estimate* of the traffic volume. For the first time, *movements of isolated vehicles* are estimated automatically. The small time gap between the acquisitions of different spectral channels is utilized for this task. The initialization of the used methods requires *no expert knowledge*. All used parameters are derived from manually acquired training data or statistical test procedures.

The approach is evaluated on QuickBird images of an urban area. While the completeness is about 65 %, the developed approach reaches a high reliability of 95 % on average. A significant displacement of the position in different spectral channels could be observed for some vehicles. The resulting vehicles' speeds are in typical range for urban areas.

Inhaltsverzeichnis

1	Motivation und Aufbau der Arbeit	1
1.1	Motivation	1
1.2	Zielsetzung der Arbeit	2
1.3	Aufbau der Arbeit	3
2	Konzepte zur Bildanalyse und Ansätze zur Fahrzeugextraktion	5
2.1	Konzepte zur Bildanalyse	5
2.1.1	Ziele	5
2.1.2	Verfahren mit impliziter Modellierung	6
2.1.3	Verfahren mit expliziter Modellierung	10
2.1.4	Wahl eines geeigneten Verfahrens	11
2.2	Ansätze zur Fahrzeugextraktion	13
2.2.1	Hochaufgelöste Einzelbilder	14
2.2.2	Niedrigaufgelöste Einzelbilder	18
2.2.3	Schätzung von Fahrzeugbewegungen aus Satellitenbildern	22
2.3	Bewertung bisherigen Ansätze und Schlussfolgerungen	23
3	Fahrzeugmodellierung	27
3.1	Modell für isolierte Fahrzeuge	27
3.2	Modell für Fahrzeugreihen	28
4	Fahrzeugextraktion	31
4.1	Gesamtüberblick der Extraktionsstrategie	31
4.2	Festlegung von Suchbereichen	33
4.3	Isolierte Einzelfahrzeuge	34
4.3.1	Klassifikation der Bildpunkte	35
4.3.2	Klassifikation der gruppierten Bildregionen	45
4.4	Einzelfahrzeuge aus Autoreihen	50
4.4.1	Linienextraktion	50
4.4.2	Grauwertprofile der Fahrzeugreihenypothesen	51

4.4.3	Erkennung einzelner Fahrzeuge	53
4.5	Ergebniszusammenführung	58
4.6	Bewegungsschätzung	60
5	Satellitenbilddaten und Geoinformation	65
5.1	Optische hochauflösende Satellitenbilddaten	65
5.1.1	Satellitenmissionen	65
5.1.2	QuickBird Daten	66
5.1.3	Testdatensätze	68
5.2	Geoinformation	70
5.3	Trainingsdaten	70
6	Experimente und Ergebnisse	73
6.1	Parameterinitialisierung	73
6.1.1	Parameter aus den Trainingsdaten	73
6.1.2	Parameter zur Extraktion von Einzelfahrzeugen aus Fahrzeugreihen	74
6.1.3	Parameter zur Erfassung isolierter Fahrzeuge	75
6.2	Ergebnisse auf den Testdatensätzen	80
6.2.1	Detektionsergebnisse	80
6.2.2	Ergebnisse der Bewegungsschätzung	82
7	Diskussion	87
7.1	Ungenauigkeiten der verwendeten Geoinformationen	87
7.2	AdaBoost Verfahren und Haar-Merkmale	87
7.3	Parameter zur Erfassung von Fahrzeugreihen	88
7.4	Ergebnisse der Fahrzeugerkennung	95
7.5	Schätzung des Bewegungszustandes	97
8	Zusammenfassung und Ausblick	99
8.1	Zusammenfassung	99
8.2	Ausblick	100

Abbildungsverzeichnis

2.1	Ablauf der statistischen Mustererkennung mittels überwachter Klassifikation	7
2.2	Zusammengesetzte Extraktionsstrategie	12
2.3	Wahl des Verfahrens in Abhängigkeit der Komplexität von Bildmaterial und Modell . . .	13
3.1	Erscheinung von Fahrzeugen in optischen Fernerkundungsdaten	28
3.2	Fahrzeug- und Reihenmodell	29
3.3	Verschmelzung einer Fahrzeugreihe bei Reduzierung der Auflösung	30
4.1	Übersicht zur Fahrzeugextraktion	32
4.2	Transformation von ATKIS Daten	34
4.3	Haar-Merkmale	36
4.4	Hauptkomponenten der Trainingsdaten und deren Informationsgehalt	37
4.5	Verwendung des Integralbildes	38
4.6	Training eines Klassifikators mit AdaBoost	40
4.7	Trainingsverlauf des Gentle AdaBoost Algorithmus	42
4.8	Fehlerrate bei unterschiedlicher Knotenanzahl eines Klassifikationsbaumes	44
4.9	Anwendung des Boosting Algorithmus	44
4.10	Trennungsebene und Schlupfvariablen der SVM	46
4.11	Beispiel einer Merkmalsraumtransformation	49
4.12	Erfassung von Fahrzeugreihen	52
4.13	Grauwertprofile von Fahrzeugreihen	53
4.14	Extrahiertes Grauwertprofil	53
4.15	Einflussfunktionen ausgewählter Schätzer	56
4.16	Robuste Schätzung der Parameter der Gauß-Funktion	59
4.17	Ergebnis der Approximation eines Grauwert-Profiles	60
4.18	Farbraumtransformation verschiedener Objekte	61
4.19	Farbiges Fahrzeug in verschiedenen Spektralkanälen	62
4.20	Bestimmung der Fahrzeugposition in den multispektralen Kanälen	63
5.1	QuickBird Satellit und Aufnahmeprinzip	67

5.2	Versatz eines ca. 500 km/h schnellen Flugzeuges	68
5.3	Testdatensätze der Münchner Innenstadt	69
5.4	Objektarten des Basis-DLM	71
5.5	Straßensegmente und im ATKIS definierte Breitenangaben	71
5.6	Referenzdaten für Fahrzeuge	72
6.1	Simulierte Gauß-Funktion	76
6.2	Einzelnes Fahrzeug bei variierender Größe des Bildausschnitts	78
6.3	Testfehler des AdaBoost Trainings	79
6.4	Variation der Knotenanzahl und der Größe des Trainingsmusters	79
6.5	Bestimmung der Parameter einer ν -SVM	80
6.6	Szenen mit hoher Qualität	83
6.7	Szenen mit niedriger Qualität	84
6.8	Ergebnisse der Bewegungsschätzung	85
7.1	Genauigkeit von ATKIS Daten	88
7.2	Fehler durch Ungenauigkeit der Geoinformation	89
7.3	Anwendung des ersten gewählten Haar-Merkmals	90
7.4	Ergebnisse der Linienextraktion	91
7.5	Filterung der Linien bzgl. der Breite	91
7.6	Filterung der Linien bzgl. der Länge	92
7.7	Filterung der Linien bzgl. der Richtung	93
7.8	Ergebnis der gesamten Filterung	93
7.9	Ergebnisse der Detektion	94
7.10	Ergebnisse bei schräg parkenden Fahrzeugen	96
7.11	Ergebnisse bei parallel parkenden Fahrzeugen	96
7.12	Ergebnisse bei fließendem Verkehr	96
7.13	Ergebnisse bei fehlerhaften Geoinformationen	97

Tabellenverzeichnis

4.1	Parameter zur Berechnung der Merkmalsanzahl	36
4.2	Verlauf des Trainings- und Testfehlers sowie der Margin für die Trainingsdaten	42
4.3	Geometrische und radiometrische Eigenschaften zur Klassifikation der Bildsegmente	45
4.4	Verlustfunktionen ausgewählter M-Schätzer	57
4.5	Zusammenführung der Ergebnisse der Teilextraktionen.	61
5.1	Optische Satellitensysteme mit Meter- und Submeterauflösung	66
5.2	Spektralkanäle des QuickBird Sensors	67
6.1	Ermittelte Kenngrößen und Abweichungen aus den Trainingsdaten	74
6.2	Empirische Ermittlung von Bruchpunkten	76
6.3	Evaluierung der Fahrzeugerkennung für die Testgebiete	81
6.4	Zusammenfassende Evaluierung der Fahrzeugerkennung für die Testgebiete	81
7.1	Prozentualer Anteil verschiedener Merkmale am Gesamtklassifikator	89
7.2	Prozentualer Anteil einzelner Filterungen	94

1 Motivation und Aufbau der Arbeit

1.1 Motivation

Verkehr spielt in der heutigen Gesellschaft im alltäglichen Leben eine zentrale Rolle. Eberhard von Kuenhein (Vorstandsvorsitzender der BMW AG von 1970 - 1993) fasste dies in seiner Aussage:

„Mobilität von Menschen und Gütern ist nicht Folge, sondern Grundlage unseres Wohlstands.“

treffend zusammen und beschrieb das Hauptproblem, mit welchem wir uns heute bereits im starken Maße konfrontiert sehen:

„Das Automobil ist so erfolgreich, dass es nur einen wirklichen Feind hat, nämlich sich selbst. Seine massenhafte Verbreitung ist eine Herausforderung an die Zukunft des Straßenverkehrs.“

In den vergangenen Jahrzehnten fand ein rasanter Anstieg des Verkehrsaufkommens statt. Der individuelle Personenkraftverkehr nahm in Deutschland seit 1960 um mehr als das 3,5 - fache zu. Im gleichen Zeitraum stieg der Straßengüterverkehr sogar um 770 % (Deutsches Institut für Wirtschaftsforschung, 2002). Nach Schätzungen des ADAC verbringen Autofahrer jährlich ca. 65 Stunden in Staus, welche größtenteils durch zu hohes Verkehrsaufkommen entstehen. Ein Mehrverbrauch von 12 Mrd. Litern Kraftstoff ist die Folge (Bosserhoff et al., 2004). Insgesamt entfielen 29 % des Gesamtenergieverbrauchs in Deutschland auf den Verkehrssektor. Der Straßenverkehr hatte hieran einen Anteil von 83 %, was etwa 69 Mrd. Liter Kraftstoff entspricht (Bosserhoff et al., 2004; Statistisches Bundesamt, 2006). Der Verkehr verursacht 20 % der klimaschädlichen CO₂ Emissionen. Entgegen dem Kyoto-Protokoll von 1997, welches eine kontinuierliche Minderung des CO₂ Ausstoßes vorsieht, steigerte allein der europäische Verkehrssektor in den vergangenen 15 Jahren seine schädlichen Emissionen um 5 % (Campell, 2007).

Eine Verbesserung der Verkehrsbedingungen, welche auch durch finanzielle Anreize der Europäischen Union gefördert wird, besteht - neben dem Ausbau der vorhandenen Infrastruktur - in der Ergreifung geeigneter Verkehrsmanagementmaßnahmen (Europäische Kommission, 2001). Das Verkehrsmanagement fasst eine Vielzahl an Maßnahmen zusammen, deren Ziel eine optimale Abstimmung von Verkehrsaufkommen und vorhandener Verkehrsinfrastruktur ist. Dadurch könnten punktuelle bzw. periodische Überlastungen trotz steigendem Verkehrsaufkommen gesenkt werden. In diesem Bereich wird der Verkehrstelematik eine wesentliche Bedeutung zur effizienten Nutzung der vorhandenen Verkehrsanlagen zugesprochen

(Bundesministerium für Verkehr, Bau- und Stadtentwicklung, 2006). Die Verfügbarkeit aktueller Verkehrsdaten ist das entscheidende Kriterium für eine erfolgreiche Verkehrsbeeinflussung, weshalb die Erfassung und Verarbeitung von Verkehrsinformationen zeitnah erfolgen muss. Messungen werden heutzutage größtenteils mit Induktionsschleifen durchgeführt, die den hohen Anforderungen an Datenaktualität und -genauigkeit genügen. Allerdings kann durch den punktuellen Erfassungsbereich dieser Systeme das Verkehrsgeschehen nur unvollständig abgebildet werden (Schischmanow, 2005). Im Gegensatz hierzu erfassen Satellitenbilder großflächige Gebiete. Die Verwendung daraus abgeleiteter Daten zur gezielten Verkehrsbeeinflussung ist aufgrund der geringen zeitlichen Auflösung nicht möglich. Aktuelle Studien zeigen aber, dass großflächige Aufnahmen wertvolle Informationen über die Fahrzeuganzahl, Verkehrsdichte und die Ausbreitung bzw. Verdichtung von Fahrzeugkolonnen liefern (Toth et al., 2003). Ebenso ist eine verbesserte Berechnung des durchschnittlichen Tagesverkehrsaufkommens pro Jahr und daraus ableitbarer Größen möglich (McCord et al., 2002; Aurdal et al., 2007; Larsen et al., 2009; Eikvil et al., 2009). Grundlage für die Nutzung der aus Satellitenbildern gewonnenen Daten ist eine möglichst zuverlässige und vollständige Erkennung von Fahrzeugen in diesen Aufnahmen.

1.2 Zielsetzung der Arbeit

Ziel dieser Arbeit ist die Entwicklung und Erprobung geeigneter Verfahren zur Erfassung verkehrsrelevanter Informationen aus optischen Satellitenbildern. Als Grundlage dienen hochauflösende Satellitenbilddaten mit einer Bodenpixelgröße von weniger als einem Meter. Zusätzlich werden Informationen über Position und Ausdehnung von Straßen aus digitalen topographischen Karten verwendet.

Die hohe Fahrzeuganzahl in Satellitenbildszenen erfordert eine vollständige Automatisierung des entwickelten Ansatzes, um eine effiziente Verarbeitung der Bilder zu gewährleisten. Hierzu werden Konzepte des Bildverstehens angewandt, um die abgebildete Verkehrslage zum Zeitpunkt der Aufnahme zu erfassen. Wie in Aurdal et al. (2007) gefordert, erfolgt neben der reinen Lokalisierung einzelner Fahrzeuge eine Abschätzung über deren momentanen Zustand. Es wird zwischen stehendem und fahrendem Verkehr unterschieden. Dafür wird die spezielle Aufnahmeconfiguration des verwendeten Sensors genutzt, welche eine Bewegungsschätzung durch den Versatz zwischen einzelnen Spektralkanälen zulässt.

Die wachsende Verfügbarkeit von Satellitenbildern durch den Start zusätzlicher Systeme führt zu einer höheren zeitlichen Abdeckung, wodurch die hieraus gewonnenen Verkehrsdaten vermehrt Anwendung finden können. In heutigen Systemen für Verkehrsanalysen und -simulationen ist dies kaum möglich. Aufgrund der Erdumlaufbahn derzeitiger Satellitensysteme können gleiche Gebiete nur nach ein bis vier Tagen und zu gleichen Tageszeiten aufgenommen werden, was die Anwendung der gewonnenen Daten erheblich einschränkt. Im Gegensatz zur Verwendung von kontinuierlich erfassten Verkehrsdaten aus Induktionsschleifen, existieren zurzeit noch keine Ansätze, welche die zeitlich beschränkten Verkehrsdaten aus Satellitenbildern nutzen. Eine Übertragung der entwickelten Methoden dieser Arbeit auf Bilddaten mit einer höheren zeitlichen Auflösung ist möglich (Rosenbaum et al., 2010).

1.3 Aufbau der Arbeit

Im zweiten Kapitel dieser Arbeit werden Konzepte zum Aufbau eines Bildanalyse-Systems erläutert. Dies beinhaltet die Vorstellung eines konzeptionellen Rahmens und die Beschreibung einzelner Komponenten. Zusätzlich wird eine Einteilung zur Wahl geeigneter Extraktionsstrategien vorgestellt. Hauptaugenmerk dieses Kapitels liegt auf der Zusammenfassung bisheriger Arbeiten zur Fahrzeugextraktion aus optischen Luft- und Satellitenbilddaten. Die Einteilung der vorgestellten Ansätze erfolgt nach der zugrunde liegenden Datenquelle. Besondere Beachtung finden Arbeiten zur Fahrzeugerkennung aus hochauflösenden optischen Satellitenbildern. Teilaspekte der vorgestellten Ansätze werden anschließend bezüglich ihrer Eignung im Rahmen dieser Arbeit bewertet.

Im dritten Kapitel wird anhand der Bewertung bisheriger Arbeiten ein Modell für Fahrzeuge in Satellitenbildern entwickelt. Das Modell unterscheidet zwischen gruppiert und isoliert auftretenden Fahrzeugen.

Das vierte Kapitel erläutert die entwickelte Extraktionsstrategie für das Fahrzeugmodell und geht auf einzelne Aspekte der Implementierung ein. Zur Detektion isolierter Fahrzeuge werden alle Bildpunkte durch ein Ensemble gewichteter Klassifikationsbäume klassifiziert. Bildpunkte der Fahrzeugklasse werden gruppiert und anhand geometrischer Eigenschaften der entstehenden Bildsegmente validiert. Da gruppierte Fahrzeuge durch dieses Vorgehen häufig nicht separiert werden können, wird anschließend das entwickelte Fahrzeugreihenmodell verwendet. Zunächst werden Linien extrahiert. Diese werden bezüglich der gegebenen Straßenrichtung, Fahrzeuglänge und -breite gefiltert und basierend auf Kollinearitätskriterien fusioniert. Nach der Berechnung geometrischer und radiometrischer Eigenschaften aller Linienpunkte können mit Hilfe einer robusten Parameterschätzung in den Reihen enthaltene Fahrzeuge gefunden werden. Anschließend erfolgt die Fusion der detektierten isolierten und gruppierten Fahrzeuge. Am Ende des Kapitels wird eine Schätzung des Bewegungsstatus isolierter Fahrzeuge durchgeführt. Dabei wird der durch die Sensorgeometrie verursachte Versatz zwischen dem panchromatischen Bild und multispektralen Bildern bestimmt und auf Signifikanz getestet.

Das fünfte Kapitel beinhaltet eine Übersicht der verwendeten Testdaten. Zunächst wird eine Übersicht bisheriger hochauflösender optischer Satellitensysteme gegeben. Anschließend folgt eine genaue Beschreibung des *QuickBird* Satelliten. Zudem wird der Inhalt der verwendeten *digitalen topographischen Karte* (DTK) und deren Genauigkeit beschrieben.

Das sechste Kapitel beinhaltet die Initialisierung aller notwendigen Parameter und verwendeten Verfahren. Zudem erfolgt die numerische Evaluierung des Ansatzes auf den Testdaten.

Im siebten Kapitel werden die erzielten Ergebnisse bewertet. Die Ergebnisse einzelner Verarbeitungsstufen werden genauer untersucht und eine Wertung der durchgeführten Parameterinitialisierung vorgenommen.

Das achte Kapitel fasst die vorgestellte Arbeit zusammen und gibt Vorschläge für weiterführende Untersuchungen.

2 Konzepte zur Bildanalyse und Ansätze zur Fahrzeugextraktion

Das zweite Kapitel ist in drei Abschnitte gegliedert. Im ersten Teil werden allgemeine Grundlagen und Strategien der Bildanalyse erläutert, welche für die Entwicklung des Ansatzes in Kapitel 4 benötigt werden. Es folgt ein Überblick bisheriger Arbeiten auf dem Gebiet der Fahrzeugextraktion aus Luft- und Satellitenbilddaten. Am Ende des Kapitels findet eine zusammenfassende Bewertung bisheriger Ansätze statt.

2.1 Konzepte zur Bildanalyse

2.1.1 Ziele

Die Entwicklung von „Augen und Ohren für den Rechner“ ist seit einigen Jahrzehnten von wissenschaftlichem Interesse (Selfridge, 1955), ohne zunächst eine genaue Vorstellung der auftretenden Probleme zu haben. In ersten Forschungsarbeiten standen zunächst Methoden zur Bildverarbeitung und -aufbereitung im Vordergrund, wobei es sich um reine Bild-zu-Bild Transformationen handelte. Trotz der langen Zeitspanne seit Aufstellung des ehrgeizigen Zieles gibt es bisher kein maschinelles System, welches die menschliche Wahrnehmung befriedigend nachbildet (Neumann, 2003). Dies ist aufgrund der hohen Komplexität des biologischen Sehsystems nicht möglich, da bereits die anfänglichen Verarbeitungsschritte des Gehirns nicht vollständig geklärt sind. Die Ähnlichkeiten von Verfahren der Bildanalyse und natürlichen - im zentralen Nervensystem stattfindenden - Prozessen ist daher meist namentlicher Natur. Ein prominentes und bis heute untersuchtes Beispiel sind *künstliche neuronale Netze*, deren einfachster Vertreter (das Perzeptron) bereits 1958 entworfen wurde (Rosenblatt, 1958).

Ziel des Forschungsgebietes Bildverstehen ist „*die Rekonstruktion und Deutung einer Szene anhand von Bildern*“ (Neumann, 2003). Aus dieser Definition wird ersichtlich, dass ein Bildanalyzesystem zwei wesentliche Aufgaben erfüllen muss. Die Rekonstruktion dient zur Gewinnung von räumlichen und zeitlichen Informationen der dargestellten Szene. Neben der reinen Lokalisierung eines Objektes im Bild kann die Herstellung der Dreidimensionalität aus zweidimensionalen Bildern von Interesse sein. Dies kann über die photogrammetrische Stereoskopie erfolgen, welche eine Rekonstruktion über mehrere Beobachtungen vornimmt. Die abschließende Deutung der rekonstruierten Szene beinhaltet die Zuweisung

einer semantischen Information der im Bild erkannten Objekte.

Durch eine Vielzahl von Berechnungsprozessen wird versucht, dieses Ziel zu erreichen. Die entwickelten Verfahren sind in der Regel stark auf das zu lösende Problem abgestimmt und daher nur zur Erkennung einzelner Objektklassen geeignet. Ein vollständiges Verständnis für die abgebildete Szene ist nicht möglich. Das wichtigste Ziel dieser spezialisierten Anwendungen ist die zuverlässige Identifikation und Lokalisation der gesuchten Objekte. Vor allem in kontrollierten Umgebungen (industrielle Bildverarbeitung) oder mit Hilfe menschlicher Unterstützung (semi-automatisch) existieren eine Vielzahl erfolgreicher Systeme. Weil der Entwurf der verwendeten Methodik stark ergebnisorientiert ist, muss die Anpassung geeigneter Parameter durch Testen auf Beispieldaten erfolgen. Zusätzlich kann die Einhaltung von Nebenbedingungen bezüglich Verarbeitungszeit und Rechnerkapazität notwendig sein, was zu weiteren Einschränkungen bei der Auswahl geeigneter Verfahren führt. An dieser Stelle soll nicht unerwähnt bleiben, dass im Gegensatz zu den beschriebenen spezialisierten Anwendungen, die inhaltsbasierte Bildsuche und die damit verbundene Szenenklassifikation von steigendem Interesse ist (Bosch et al., 2007). Motiviert wird dieses Forschungsgebiet durch die stetig wachsenden Bestände an digitalen Bildinformationen in heutigen Datenbanksystemen (Daschiel, 2004).

Eine Einteilung der Vorgehensweisen zur Erreichung der genannten Ziele kann auf Basis der gewählten Modellierung erfolgen. Die gesuchten Objekte können implizit oder explizit beschrieben werden.

2.1.2 Verfahren mit impliziter Modellierung

Bei der impliziten Modellierung kommen Methoden der statistischen Mustererkennung zur Anwendung. Die gesuchten Objekte werden mit Hilfe eines Merkmalsvektors beschrieben und anhand seiner Lage im Merkmalsraum klassifiziert. Die verwendeten Merkmale basieren dabei entweder auf direkten Eigenschaften der Bildpunkte (Klassifikation von Pixeln) oder auf Charakteristika zuvor segmentierter Bildregionen (Klassifikation von Bildobjekten). Die Beschreibung von Objekten als Merkmalsvektoren stellt dabei in der Regel keine natürliche Abbildung von Realwelteigenschaften dar. Die Parameter zur Charakterisierung einzelner Objektklassen werden bei der überwachten Klassifikation durch Training aus Beispieldaten gewonnen. Im Gegensatz hierzu erfolgt die Partitionierung des Merkmalsraums bei der unüberwachten Klassifikation durch gemessene oder wahrgenommene Ähnlichkeiten innerhalb der Daten. Dies geschieht durch die Zerlegung der vorliegenden gemischten Verteilungen der Merkmalswerte oder durch Verfahren der Cluster Analyse. Eine Schwierigkeit bei der unüberwachten Klassifikation ist, dass die Anzahl der zu unterscheidenden Objektklassen häufig nicht bekannt ist und diese zunächst bestimmt werden muss. Im Rahmen dieser Arbeit findet keine weitere Betrachtung der unüberwachten Klassifikation statt. Als Ergebnis der überwachten Klassifikation wird jeder Bildpunkt einer Objektklasse zugeordnet. Dies kann unter Angabe einer Wahrscheinlichkeit der Klassenzugehörigkeit oder der Zuverlässigkeit der Klassifikation erfolgen. Die Anzahl entwickelter Verfahren (siehe Duda et al. (2001); Bishop (2006); Theodoridis & Koutroumbas (2006)) macht die statistische Mustererkennung zu einer sehr ausgereiften Disziplin mit einer Vielzahl möglicher Anwendungsfelder. Eine umfangreichere Übersicht findet sich in Jain et al. (2000) und Niemann (2003).

2.1. Konzepte zur Bildanalyse

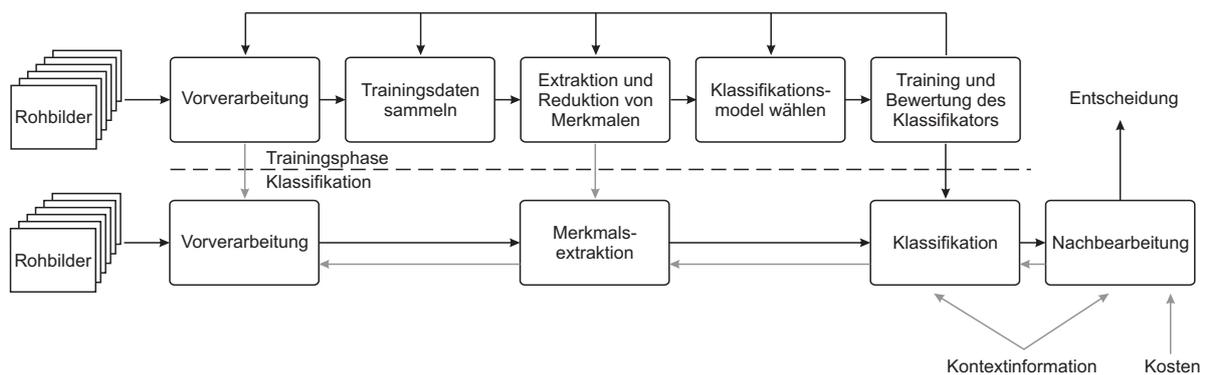


Abbildung 2.1: Ablauf der statistischen Mustererkennung mittels überwachter Klassifikation

Ein Überblick von Teilschritten der statistischen Mustererkennung ist in Abbildung 2.1 dargestellt. Es wird zwischen der Trainings- bzw. Entwurfsphase und der eigentlichen Klassifikation unterschieden. Ähnliche Übersichten mit leichten Variationen werden in Nagao (1984); Fukunaga (1990); Jain et al. (2000); de Sá (2001); Duda et al. (2001); Niemann (2003); Theodoridis & Koutroumbas (2006) vorgestellt. Die folgende Beschreibung einzelner Teilschritte der Klassifikation beschränkt sich auf Klassifikationsprobleme, bei denen eine Objektdetektion gegenüber dem „Rest der Welt“ durchgeführt wird.

Trainingsphase

- **Die Vorverarbeitung** dient - unter Einbringung von sensorspezifischen Korrekturen - zur Entfernung und Verringerung unerwünschter Signalbestandteile. Hierzu gehört die Reduktion von Rauschen, die Behebung von Überstrahlungen oder die Normierung auf einen gewünschten Wertebereich. Eine Reduzierung der signifikanten Bildinformation sollte zu diesem Zeitpunkt nicht stattfinden.
- Im Anschluss muss ein Operateur charakterisierende **Trainingsdaten** für die gesuchten Objektklassen erfassen. Falsch zugeordnete Beispiele können die spätere Klassifikationsleistung erheblich verringern bzw. eine genaue Klassenzuordnung sogar verhindern. Der Einfluss fehlerhafter Trainingsdaten hängt allerdings auch vom verwendeten Klassifikationsalgorithmus ab.

Meist ist eine größere Anzahl an Beispielen für den zu lernenden Hintergrund notwendig, da dieser größere Variationen als die gesuchten Objektklassen aufweisen kann. Um etwa gleich viele Objekt- und Hintergrundbeispiele während des Trainings zu verwenden, werden Bootstrap-Methoden angewandt. Aus allen zur Verfügung stehenden Trainingsdaten des Hintergrundes wird eine Teilmenge zufällig ausgewählt und alle Trainingsphasen durchlaufen. Nach Abschluss der Evaluierung wird der erhaltene Klassifikator auf alle Trainingsdaten des Hintergrundes angewandt, um so Beispiele zu finden, die bis dahin nicht korrekt erkannt werden. Eine Teilmenge dieser falsch klassifizierten Hintergrundbilder wird in der nächsten Iteration des Trainings verwendet. Durch diese Vorgehensweise ist der Klassifikationsalgorithmus in der Lage, eine immer genauer werdende

Unterscheidung zwischen Hintergrund und Vordergrund vorzunehmen.

- Die **Extraktion charakteristischer Merkmale** erfolgt für einzelne Bildpunkte oder für alle Pixel innerhalb eines Suchfensters. Aufgrund der Vielzahl möglicher Merkmale ist es notwendig, bei der Auswahl charakteristischer Merkmale Kenntnisse über die Gestalt der gesuchten Objekte und die geforderten Invarianzen (z.B. gegenüber Rotationen, Maßstabsänderungen, Verdeckungen, projektiven Verzerrungen und Deformationen) mit einfließen zu lassen. Eine Übersicht einiger Merkmale gibt de Sá (2001). Obwohl keine generalisierte Vorgehensweise zur Wahl geeigneter Merkmale existiert, gibt Niemann (2003) eine hilfreiche Übersicht zu deren Wahl bzw. Entwurf. Als eines der grundlegenden Konzepte führt der Autor die heuristische Methode auf, welche „*Merkmale aufgrund von Intuition, Phantasie und Erfahrung...*“ (Niemann, 2003, Seite 165) auswählt. Als Maß für die Aussagekraft einzelner Merkmale können Abhängigkeiten zwischen diesen genutzt werden, welche z.B. durch Anwendung der Karhunen-Loève-Transformation (Jolliffe (2002)) erkennbar werden. Die Bewertung der Merkmale mit statistischen Kenngrößen kann zur analytischen Wahl geeigneter Merkmale herangezogen werden (Niemann, 2003). Bei der Merkmalsberechnung ist zu beachten, dass es zu keinem signifikanten Informationsverlust kommt, so dass die nachfolgenden Schritte nur auf Basis der berechneten Merkmale arbeiten können. Diese müssen sowohl das gesuchte Objekt als auch beliebige Hintergründe gut beschreiben.
- Der Übergang zur anschließenden **Merkmalsreduktion** ist meist fließend und findet nicht immer Beachtung (siehe Duda et al. (2001)) oder wird als ein Prozess verstanden (wie in Fukunaga (1990); Niemann (2003)). Ziel dieses Verarbeitungsschrittes ist, aus den zuvor extrahierten Merkmalen nochmals die charakteristischsten Merkmale auszuwählen. Diese Reduktion ist für eine Verkürzung der Berechnungszeit der Merkmale während der Klassifikation sinnvoll. Das Auffinden der optimalen Zusammenstellung der Merkmale durch eine umfassende Suche kann kombinatorische Schwierigkeiten verursachen. Liegen beispielsweise insgesamt 24 verschiedene Merkmale vor, von denen nur 12 für die weitere Bearbeitung relevant sind, so ergeben sich insgesamt 2,7 Millionen mögliche Merkmalskombinationen (aus Jain et al. (2000)). Daher wurde eine Reihe von Verfahren entwickelt, welche alternative Suchtechniken zum Finden eines reduziert dimensionierten Merkmalsraums anwenden. Obwohl eine Merkmalsreduktion den „Fluch der Dimensionalität“ (Bellman, 1961) im Fall von Wahrscheinlichkeitsdichte-basierten Ansätzen abschwächt, kann dies zu einen Verlust der Fähigkeit zur Klassentrennbarkeit führen.
- Bei der Wahl eines **Klassifikators** stehen viele Verfahren zur Auswahl, deren genaue Beschreibung eine Reihe zum Teil äußerst umfangreicher Fachbücher füllt (siehe Niemann (2003)). Die verschiedenen Ansätze können drei Kategorien zugeordnet werden.
Die erste Gruppe der Klassifikationsverfahren schätzt die Wahrscheinlichkeitsdichten der Merkmale der Trainingsdaten. Bei einer parametrischen Modellierung wird von bekannten Wahrscheinlichkeitsdichtefunktionen ausgegangen und die Momente dieser Dichtefunktionen bestimmt.
Bei der zweiten Gruppe existieren keine Annahmen zur Form der Merkmalsverteilungen. Es wird

2.1. Konzepte zur Bildanalyse

eine nicht-parametrische Schätzung der beliebig gearteten Wahrscheinlichkeitsdichten durchgeführt und anschließend deren beschreibende Parameter berechnet. Damit kann die Bestimmung der Klassenzugehörigkeit eines beliebigen Punktes im Merkmalsraum erfolgen. Zu den nicht-parametrischen Ansätzen gehören auch geometrische Ansätze. Anhand der Trainingsdaten werden Trennungsebenen zwischen Klassen im Merkmalsraum gesucht, um so die verschiedenen Objektklassen voneinander zu unterscheiden. Aufgrund dieser Trennung werden solche Verfahren oft als diskriminativ bezeichnet, wohingegen die erste Gruppe generativ sind.

Die letzte Art an Klassifikationalgorithmen nutzt mehrere Klassifikatoren der zuvor erläuterten Ansätze und kombiniert diese, um so eine gesteigerte Klassifikationsleistung zu erzielen.

Die Wahl eines erfolgversprechenden Klassifikationsalgorithmus ist äußerst schwierig und wird in Praxis daher meist, ähnlich wie die Auswahl geeigneter Merkmale, aufgrund von Erfahrung getroffen (Jain et al., 2000; Niemann, 2003).

- Nach Abschluss des Trainings wird der Klassifikator numerisch evaluiert. Als Maß dient häufig die erreichte Fehlerrate. Es können allerdings auch die Berechnungszeit oder andere Faktoren der Klassifikation zur Bewertung herangezogen werden. Die Evaluierung sollte auf dem Trainingsdatensatz und einem davon unabhängigen Testdatensatz erfolgen, um ein Maß für die Generalisierungsfähigkeit des Klassifikators zu bekommen. Häufig wird eine n-fache Kreuzvalidierung mit variierenden Untermengen des gesamten Datensatzes durchgeführt. Im Fall der n-fachen stratifizierten Kreuzvalidierung, d.h. die Teilmengen für Trainings- und Testdaten unterliegen annähernd gleichen Verteilungen, ist eine sehr genaue Abschätzung der erreichbaren Klassifikationsleistung möglich. Genügt der Klassifikator den geforderten Ansprüchen ist die Trainingsphase abgeschlossen. Andernfalls muss zu einem der vorherigen Arbeitsschritte zurückgesprungen werden und das Training bis zur Bewertung wiederholt werden.

Klassifikation

Während der Klassifikationsphase werden die Eingangsbilder in gleicher Weise vorverarbeitet wie die Bilder der Trainingsdaten. Dies ist notwendig, um beispielsweise gleiche Grauwertbereiche zu garantieren. Anschließend werden die in Trainingsphase verwendeten Merkmale extrahiert. Diese werden mit dem gelernten Klassifikator klassifiziert. Zu diesem Zeitpunkt und in der Nachbearbeitung ist es zudem möglich, zusätzliche Kontextinformationen einzubringen. Dies betrifft vor allem Variationen eines Objektes, welche nicht in den Trainingsdaten enthalten waren. Dadurch ist es möglich, falsch klassifizierte Bildpunkte den richtigen Objekten zuzuordnen.

Obwohl die gegebene Beschreibung eine von der Vorverarbeitung über die Merkmalsextraktion und Klassifikation aufbauende Vorgehensweise darstellt, existieren auch Mustererkennungssysteme, welche gezielte Änderungen an den vorherigen Arbeitsschritten durchführen. Eine schlechte Evaluierung des erlernten Klassifikators kann das Hinzufügen weiterer Merkmale oder Trainingsdaten erfordern.

Während die Durchführung des Trainings nur von der Rechnerkapazität und der benötigten Trainingszeit abhängt, ist die Berechnungszeit während der Klassifikation ein wichtiges Kriterium bei der

Suche nach einem geeigneten Algorithmus und der Wahl der verwendeten Merkmale. Bei zeitkritischen Anwendungen müssen diese Aspekte bereits vor Beginn der Trainingsphase berücksichtigt werden.

2.1.3 Verfahren mit expliziter Modellierung

Verfahren mit expliziter Modellierung setzen Kenntnisse über physikalische Eigenschaften und/oder Geometrie der gesuchten Objekte voraus, welche im Objektmodell erfasst werden. Die Erscheinung von Objekten im Bildraum wird durch das Bildmodell beschrieben. Diese Modelle werden in einer computergerechten Repräsentation dargestellt und durch Entwicklung einer geeigneten Strategie genutzt. Die Extraktionsstrategien können zwischen daten- und modellgetriebener Vorgehensweise unterschieden werden:

- In beiden Fällen finden zunächst notwendige Vorverarbeitungen des Bildmaterials statt.
- Bei der datengetriebenen Verarbeitung werden anschließend geometrische Primitive und deren Relationen extrahiert. Dies erfolgt über die Erkennung von Grauwertdiskontinuitäten oder komplementär über die Segmentierung homogener Bereiche. Dabei stehen verschiedene Verfahren zur Kantenextraktion und Bereichswachstumsalgorithmen zur Verfügung. In natürlichen Szenen können aufgrund der Vielseitigkeit von Objektinteraktionen kaum geschlossene Objektgrenzen erkannt werden. Der Aufwand für die nachfolgenden Verarbeitungsschritte ist daher sehr hoch. Ein Kompromiss zwischen der Zusammenfassung eigenständiger Objekte und der Trennung in schwierig interpretierbare Objektbestandteile ist nur schwer zu finden. Eine erste Reduktion der entstehenden Primitive findet durch deren Gruppierung statt. Dies beinhaltet die Zusammenfassung von Primitiven unter Verwendung geometrischer Relationen. Die initiale Gruppierung kann ohne die Einbringung von spezialisiertem Wissen über die gesuchten Objekte erfolgen. Die Hough-Transformation wird beispielsweise zur Gewinnung gleichgerichteter Elemente verwendet, ohne lokale Nachbarschaften zu berücksichtigen. Die gruppierten Primitive werden zur Konstruktion komplexerer Strukturen verwendet, welche auf Übereinstimmung mit dem Objektmodell überprüft werden.

Die anschließende Erkennung von Objekten erfolgt mit den gebildeten Strukturen. Dabei ist das Vorwissen über die Gestalt und den Kontext der gesuchten Objekte entscheidend. Die Deutung der Objekte muss dabei auf Merkmalen beruhen, welche zum einen aus den Bilddaten gewonnen werden, zum anderen aber auch unabhängig von der konkreten Abbildung und Einflüssen der Bildentstehung sind. Bei der Extraktion zweidimensionaler Objekte vereinfacht sich die Suche nach solchen Merkmalen, da nur eine Rotation in der Ebene möglich ist.

Ein großer Vorteil der datengetriebenen Vorgehensweise besteht darin, dass diese recht robust gegenüber einer unvollständigen Modellierung einzelner Objektteile entworfen werden kann. Ein Objekt kann auch dann bestätigt werden, wenn gefundene Strukturen nicht im Objektmodell enthalten sind, solange die modellierten Objektteile ebenfalls extrahiert wurden. Das Hauptproblem rein datengetriebener Ansätze liegt an den generell vorliegenden Mehrdeutigkeiten der extrahierten

2.1. Konzepte zur Bildanalyse

Primitiven für komplexe Szenen oder Objekte. Die daraus resultierende Kombinatorik der möglichen Gruppierungen kann zu falschen und fehlenden Hypothesen führen (Neumann, 2003).

- Die modellgetriebene Vorgehensweise projiziert zunächst mit Hilfe eines (vereinfachten) Sensormodells das Objektmodell, z.B. in Form eines Drahtgittermodells, in die Bildebene. Die daraus resultierende Repräsentation des Objektes im Bildraum wird im Anschluss in einzelne Objektteile, z.B. Kanten, zerlegt. Dies vereinfacht die abschließende Zuordnung der einzelnen Bestandteile zu entsprechenden Bildregionen. Voraussetzung dieser Vorgehensweise ist eine genaue Objektmodellierung, damit diese mit der vorhandenen Abbildung des gesuchten Objektes möglichst genau übereinstimmt. Im Gegensatz zum datengetriebenen Verfahren zeigt sich die spezifische Beschreibung des Objektes der modellgetriebenen Vorgehensweise robuster gegenüber Störungen des Bildmaterials.

Abschließend beinhaltet die höhere Bilddeutung das Erkennen objektübergreifender Zusammenhänge. Auch wenn eine Anwendung sich auf die reine Objektextraktion beschränkt, können die erkannten Objekte für weiterführende Bildinterpretationen verwendet werden.

Zur Nutzung der Vorteile und Abschwächung der Nachteile beider Vorgehensweisen sind zusammengesetzte Verfahren geeignet. Im Rahmen der Prozessierung findet dabei ein Wechsel zwischen daten- und modellgetriebenen Verfahren statt. Abbildung 2.2 stellt eine solche gemischte Vorgehensweise dar. Zunächst werden ohne die Einbringung spezifischen Wissens geometrische Primitive extrahiert, um einfach erzeugte Hypothesen für einzelne Objektteile - die primären Strukturen - zu gewinnen. Diese werden mit Hilfe des in den Bildraum projizierten detaillierten Objektmodells bestätigt oder abgelehnt und zur anschließenden gezielten Extraktion neuer Primitive genutzt, welche erneut verifiziert oder falsifiziert werden. Dieser Vorgang kann bis zur eindeutigen Erkennung der gesuchten Objekte oder bis zum Ausschöpfen des eingebrachten Wissens fortgeführt werden. Die zusammengesetzte Vorgehensweise entspricht dem Konzept des „Hypothesenbilden-und-Test“. Dabei nimmt der Umfang des eingebrachten Wissens im Rahmen der Verarbeitung stetig zu. In der Anfangsphase der Extraktion fließen sehr allgemeine Objekteigenschaften ein, die noch wenig auf die zu bearbeitende Aufgabe spezifiziert sind, z.B. über die Homogenität der Objektoberfläche. Der sich wiederholende Wechsel zwischen daten- und modellgetriebener Strategie erlaubt zudem die Nutzung von Wissen über erwartete Störungen und die Extrahierbarkeit von Objektteilen.

2.1.4 Wahl eines geeigneten Verfahrens

Die Entscheidung zur Wahl einer geeigneten Vorgehensweise hängt nach Suetens et al. (1992) von der Komplexität des Bildmaterials und des verwendeten Objektmodells ab:

- **Die Komplexität des Bildmaterials** entspricht ungefähr dem Signal-Rausch Verhältnis der digitalen Bilder. Es ist keine direkte Eigenschaft des Bildes sondern abhängig von den gesuchten Objekten. Es kann aber als Eigenschaft des Bildes angesehen werden, da die Gestalt eines gesuchten Objektes auch von der Auflösung des Bildmaterials abhängt. Der Grad der Komplexität lässt sich

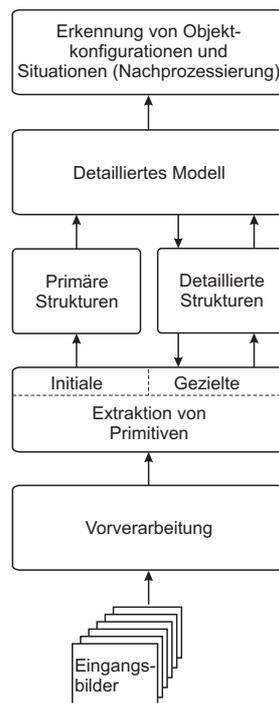


Abbildung 2.2: Zusammengesetzte Extraktionsstrategie

nur schwer anhand des Bildinhaltes abschätzen. Werden die gesuchten Objekte eindeutig und dem Modell entsprechend dargestellt, wie dies bei künstlich generierten Bildern der Fall ist, so wird von einfachen Daten bzw. einer geringen Komplexität gesprochen. Liegen hingegen mehrdeutige Interpretationsmöglichkeiten einzelner Bildbestandteile vor, wird von einer hohen Komplexität der Daten gesprochen. Die Mehrdeutigkeiten werden durch den Inhalt der abgebildeten Szene verursacht. Sowohl eine hohe Anzahl unterschiedlicher Objekte und deren Variabilität als auch unterschiedliche Objektrelationen (z.B. Verdeckungen) erschweren die korrekte Zuordnung von Bildstrukturen und Objekten.

- **Die Komplexität des Modells** wird durch dessen Detaillierungsgrad und die benötigten Verfahren beschrieben, welche eine Zuordnung von Bild- zu Objektinformationen ermöglichen. Wird ein gesuchtes Objekt durch ein einzelnes Formmuster eindeutig beschrieben, liegt eine geringe Komplexität des Modells vor. Ist es notwendig, kleinste Modellteile zusammenzufügen oder diese in hierarchische Relationen zu setzen, kann dies komplexe Datenstrukturen und Suchprozesse erfordern. In diesem Fall würde man von einer hohen Komplexität des Modells ausgehen.

Basierend auf diesem Verständnis von Komplexität können die beschriebenen Vorgehensweisen nach Suetens et al. (1992) wie in Abbildung 2.3 eingeteilt werden. Ist sowohl das verwendete Modell und die zur Verfügung stehende Datengrundlage von geringer Komplexität, können bereits pixelbasierte Objekterkennungen erfolgreich angewandt werden. Hierzu zählen Verfahren mit impliziter Modellierung bei denen Merkmale verwendet werden, welche sich auch direkt aus Werten der Bildpunkte ableiten lassen.

2.2. Ansätze zur Fahrzeugextraktion

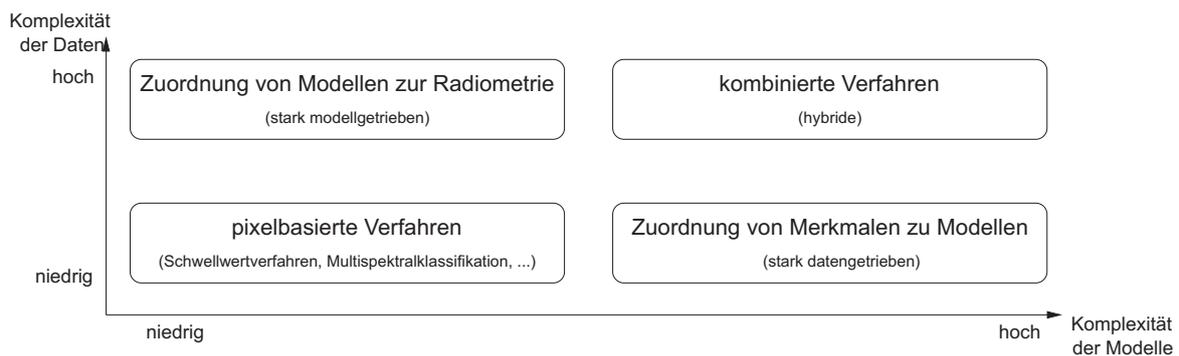


Abbildung 2.3: Wahl des Verfahrens in Abhängigkeit der Komplexität von Bildmaterial und Modell (aus Hinz (2003b), gemäß Suetens et al. (1992) und Mayer (1998))

Sind hingegen die verwendeten Modelle komplex und das vorliegende Bildmaterial einfach, eignen sich stark datengetriebene Verfahren. Die Generierung zuverlässiger Hypothesen durch die geringe Anzahl extrahierbarer Primitive gelingt besonders gut und deren Zuordnung zu komplexeren Modellen ist möglich. Im Gegensatz hierzu führt eine hohe Komplexität der Datengrundlage zur Generierung vieler unsicherer Hypothesen während der Konstruktion und Deutung von Szenenelementen. Sind dabei aber die verwendeten Modelle einfacher Natur, ist eine eindeutige Zuordnung der in den Bildraum projizierten Objektteile meist durchführbar. Daraus wird ersichtlich, dass der Versuch unternommen wird, eine hohe Komplexität am Anfang einer expliziten Modellierung zu vermeiden, um spätere Mehrdeutigkeiten und kombinatorische Schwierigkeiten zu verringern. Liegen komplexe Daten und Modelle vor, werden hybride Verfahren angewandt. Nur durch die gezielte Steuerung der Merkmalsextraktion ist dann der Aufbau höherwertiger Strukturen möglich.

2.2 Ansätze zur Fahrzeugextraktion

Im Folgenden werden für diese Arbeit relevante Entwicklungen auf dem Gebiet der Verkehrserfassung aus Fernerkundungsdaten vorgestellt. Die Erkennung von Fahrzeugen aus Videosequenzen und aus Bildern schräg blickender Kameras ist seit einigen Jahrzehnten intensives Forschungsthema. Erste Pionierarbeiten fanden bereits Anfang der 80er Jahre statt (Dreschler & Nagel, 1981). Die auf diesem Datenmaterial entwickelten Ansätze nutzen meist die Bewegung in Videosequenzen (Cucchiara et al., 2000; Betke et al., 2000; Cucchiara et al., 2000; Haag & Nagel, 2004) oder eine genaue Beschreibung der Objekte (z.B. durch Drahtgittermodelle), welche durch die hohe Auflösung des Bildmaterials gestützt wird (Sullivan et al., 1997; Rajagopalan & Chellappa, 2001). Aufgrund der unterschiedlichen Charakteristika von Videosequenzen im Vergleich zu optischen Luftbildern werden Arbeiten aus diesem Forschungsbereich nicht eingehender betrachtet. Dies trifft ebenfalls auf Arbeiten zu, welche eine Fahrzeugerkennung mit SAR (Ender, 1999; Gierull, 2004; Meyer et al., 2006; Cerutti-Maori et al., 2008), LiDAR (Toth & Grejner-Brzezinska, 2006; Yao et al., 2008b) und thermalen Bildsequenzen (Michaelsen & Stilla, 2001; Kirchhof

& Stilla, 2006) durchführen. Diese Arbeiten müssen anderen Bildeigenschaften Rechnung tragen. Einen Überblick und weiterführende Referenzen hierzu finden sich in Hinz et al. (2006) und Hinz (2008).

Für Sensoren im optischen Bereich erfolgt eine Unterteilung in hochaufgelöste (bis 40 cm) und niedrigaufgelöste Einzelbilder. Diese Grenze beruht hauptsächlich auf der visuellen Erkennung wesentlicher Merkmale bzw. Objektteile eines Fahrzeuges wie Windschutzscheibe oder Motorhaube. Diese Einteilung entspricht meist auch der den Sensor tragenden Plattformen. Mit Ausnahme der vorgestellten Arbeiten von Hinz & Stilla (2006) und Yao et al. (2008a) handelt es bei den niedrigaufgelösten Bildern um Satellitenbilddaten, während die hochaufgelösten Einzelbilder von luftgetragenen Sensoren stammen.

In hochaufgelösten Bildern erlaubt der abgebildete Detailierungsgrad eine komplexe Modellierung von Fahrzeugen. Daher existiert in diesem Bereich eine Vielzahl erfolgreicher Ansätze, deren direkte Übertragung auf Satellitenbildern aufgrund der unterschiedlichen Auflösung nicht erfolgversprechend ist. Trotzdem können Teile der Modellierung und Strategie, wie allgemeine Bildverarbeitungsverfahren (Hintergrundschtzung, Klassifikationsalgorithmen ...), auf niedrigaufgelöstes Bildmaterial angewandt werden. Daher werden in Abschnitt 2.2.1 einige dieser Ansätze betrachtet, bevor in Abschnitt 2.2.2 eine Zusammenfassung von Arbeiten mit niedrigaufgelösten Einzelbildern erfolgt.

Die Vor- und Nachteile der verschiedenen Modellierungen und Strategien bzgl. der Verwendung in dieser Arbeit werden in Abschnitt 2.3 gegenüber gestellt.

2.2.1 Hochaufgelöste Einzelbilder

In Burlina et al. (1997) wird ein zweidimensionales Rechteck als Fahrzeugmodell verwendet. Zur Bildung von Fahrzeughypothesen erfolgt zunächst eine Kantenextraktion mit Hilfe des Canny Algorithmus (Canny, 1986). Optimale Parameter der Filtermaskengröße und der Hystereseschwellwerte werden durch die empirische Untersuchung von Trainingsdaten festgelegt. Im Anschluss erfolgt eine generalisierte Hough-Transformation (Ballard, 1981) des Bildes unter Verwendung der erwarteten Gestalt und Größe des Fahrzeugmodells, wodurch für jeden Bildpunkt eine Wertung bezüglich eines möglichen Fahrzeugmittelpunktes generiert wird. Die Bestätigung erhaltener Hypothesen findet bei ausreichend hoher Wertung durch das Kantenbild statt. Dafür wird die Übereinstimmung zwischen Kantenbild und Fahrzeugmodell berechnet. Überschreitet diese einen festgelegten Minimalwert gilt die Fahrzeughypothese als bestätigt.

Zur Einschränkung des Suchraumes innerhalb des Hough Akkumulators ist eine Verwendung von Zusatzinformationen über Straßen und Parkplätze vorgesehen. In einem weiteren Ansatz dieser Gruppe wird eine Erkennung von Fahrzeugformationen durch Analyse der spektralen Signatur regelmäßig erscheinender Kanten durchgeführt (Burlina & Chellappa, 1997). Dies geschieht durch den Vergleich eines Fourier-transformierten Ausschnittes des Gradientenbilds mit einem aus dem Fahrzeugmodell und Annahmen über die Fahrzeugkonstellation von militärischen Kolonnen erhaltenen Sollspektrums. Ziel dieses Schrittes ist zum einen die geometrische Beschreibung solcher Formationen als auch die Überwachung des Belegungsgrades von Parkplätzen. Aufgrund der Anfälligkeit des Spektrums gegenüber Modellabweichungen im Ortsraum erfordert diese Anwendung allerdings sehr regelmäßige Fahrzeugformationen wie diese meist nur auf Parkplätzen gegeben sind.

2.2. Ansätze zur Fahrzeugextraktion

Auch die Arbeit von Moon et al. (2002) wird durch die Erkennung von Fahrzeugen auf Parkplätzen motiviert. Obwohl ebenfalls von einem einfachen zweidimensionalen Rechteckmodell für die Fahrzeuge ausgegangen wird, erfolgt die Detektion durch Anwendung eines Rechteckoperators. Hierdurch wird die Kantenextraktion und anschließende Gruppierung vereint. Die Größe und Rotation des Rechteckoperators wird variiert und die stärkste Übereinstimmung pro Bildpunkt gewählt. Weitere Verbesserungen der Genauigkeit ergeben sich bei gegebener Fahrzeugrichtung, was die Variationen der rechteckigen Filtermaske auf deren Ausdehnung beschränkt. Durch die frühzeitige Einbringung von Wissen über die Gestalt der Fahrzeuge während der Kantenextraktion zeigt dieses Verfahren eine höhere Unempfindlichkeit gegenüber Rauschen. Durch die unabhängige Betrachtung einzelner Fahrzeuge schlägt das Verfahren allerdings auch bei rechteckigen Objekten an, welche annähernd die Maße von Fahrzeugen besitzen.

Im Gegensatz zu den zuvor beschriebenen Ansätzen, welche Fahrzeuge nur zweidimensional modellieren, erfolgt in der Arbeit von Zhao & Nevatia (2003) die Projektion eines 3D Drahtgittermodell in die Bildebene. Dieses Modell beinhaltet als sicher erkennbare Fahrzeugbestandteile dessen Außenkanten und die Umrisse der Frontscheibe. Als weiteres Merkmal wird der Fahrzeugschatten in Abhängigkeit des dreidimensionalen Modells und des bekannten Sonnenstandes prädiziert. Die Signifikanz einzelner Kanten im Bild wird in Abhängigkeit gerichteter Reflexionseigenschaften der Fahrzeugflächen beschrieben. Als Grundlage dient ein modifiziertes Lambert-Modell, dessen Parameter aus Beispieldaten und dem zu bearbeitenden Bild geschätzt werden. Die Erkennung von Fahrzeugen erfolgt nach Parametrisierung und Quantifizierung der projizierten Kanten mit Hilfe eines Bayes'schen Netzes. Die notwendigen bedingten Wahrscheinlichkeitsdichten werden aus Trainingsdaten gewonnen. Zudem findet eine Variation der Größe des Fahrzeugmodells statt, wobei die finale Entscheidung zu Gunsten der Hypothese mit der größten Wahrscheinlichkeit getroffen wird. Eine Beachtung der Rotation findet innerhalb der Hypothesenbildung nicht statt, da die Bilder zuvor entlang ihrer Hauptkantenrichtung gedreht werden. Dies erfolgt unter der Annahme, dass urbane Gebiete in erster Linie durch Straßen und daran ausgerichtete Häuser geprägt sind und so viele gleichgerichtete Strukturen eine zuverlässige Berechnung dieser Hauptkantenrichtung ermöglichen.

Die Verwendung eines Drahtgittermodells wird auch in Hinz (2004) forciert. Eine große Erweiterung dieser Arbeit besteht im Aufbau eines hierarchisches Fahrzeugmodells dessen einfachste Repräsentation ein Quader ist. Hiervon ausgehend werden spezialisiertere Modelle abgeleitet, welche einzelne Fahrzeugbestandteile beinhalten. Zusätzlich findet eine Adaption des in den Bildraum projizierten dreidimensionalen Modells und dessen Schattens statt, indem die erwartete Erkennbarkeit einzelner Modell- und Schattenkanten berücksichtigt wird. So wird im Falle dunkler Fahrzeuge eine geringe Erkennbarkeit von Windschutzscheibe und Dach angenommen. Bei hellen Fahrzeugen werden hingegen in diesem Bereich starke Kontrastunterschiede erwartet. Die Erkennung von Fahrzeugen findet im Rahmen einer stark modellgetriebenen Suchstrategie auf dem Gradientenbild statt. In der Arbeit wird ein Fahrzeugreihenmodell

vorgelegt, welches zu einer von der Fahrzeugerkennung unabhängigen Bildung von Hypothesen für Fahrzeugkolonnen führt. Diese Hypothesen werden anschließend mit den zuvor detektierten Fahrzeugen fusioniert, indem ausschließlich die Kolonnenhypothesen erhalten bleiben, welche zumindest teilweise durch Fahrzeuge abgedeckt werden. Sollte innerhalb der verifizierten Kolonnen ausreichend Platz für weitere Fahrzeuge bestehen, wird auf diese Bereiche ein einfacher Blobdetektor angewandt. Dadurch können zusätzliche Fahrzeuge erkannt werden, welche zuvor dem komplexeren Drahtgittermodell nicht entsprochen haben. Auch ohne die Verwendung von zusätzlichen Geoinformationen erreicht der Ansatz einen hohen Grad an Zuverlässigkeit und brauchbare Vollständigkeit. Dies ist unter anderem der konsequenten Verknüpfung einer strengen Modellierung einzelner Objekte und der Ausnutzung von Wissen über vorhandene Relationen zwischen diesen Objekten zu verdanken.

Die bisher vorgestellten Arbeiten verwendeten eine explizite Modellierung für Fahrzeuge, auch wenn diese in ihrer Detailgenauigkeit stark variieren. Die folgenden Arbeiten konzentrieren sich dagegen mehr auf eine implizite Modellierung und anschließende Klassifikation der Bildpunkte.

In der Arbeit von Liu et al. (1999) werden gemittelte Grauwerte, Gradientenstärken und mittlere Gradientenrichtungen als Merkmale verwendet. Deren Berechnung erfolgt innerhalb einer rechteckigen Suchmaske, welche in einen inneren Fahrzeugbereich, einen äußeren Straßenbereich und den dazwischen liegenden Randbereich unterteilt ist. Durch Anwendung einer Hauptkomponenten-Transformation wird der Merkmalsraum auf die drei ersten Hauptachsen reduziert und im Anschluss - basierend auf dem Entropiemaß - nacheinander quantisiert. Die Klassifikation erfolgt durch iterative Anwendung eines Bayes-Klassifikators, dessen bedingte Wahrscheinlichkeitsdichten aus Trainingsdaten gelernt werden. Da der beschriebene Ansatz hauptsächlich auf geparkte Fahrzeuge angewandt wird, erfolgt zusätzlich die Schätzung der A-priori Wahrscheinlichkeiten für das Auftreten benachbarter Fahrzeuge, welche im Laufe der iterativen Detektion aktualisiert werden. Als Ergebnis der Klassifikation wird die Wahrscheinlichkeit der Zugehörigkeit zur Fahrzeugklasse jedes Bildpunktes ausgegeben. Im Falle überlappender Detektionen wird die Hypothese mit der höchsten Wahrscheinlichkeit angenommen. Der Vergleich mit den Arbeiten in Burlina et al. (1997) und Burlina & Chellappa (1997) zeigt eine deutliche Verbesserung dieses Ansatzes, welche auch der Einbringen von Vorwissen - in Form der A-priori Wahrscheinlichkeiten - zu verdanken ist.

In der Arbeit von Ruskoné et al. (1996) findet ein hierarchisches Modell Anwendung. Nach dem Training eines künstlichen Neuronalen Netzes erfolgt die Klassifikation aller Bildpunkte. Da es sich bei dem Ergebnis der Klassifikation nur um ein Maß für die Wahrscheinlichkeit des Vorhandenseins eines Fahrzeuges handelt, wird zunächst eine Normierung der erhaltenen Werte durchgeführt und diese in eine Wahrscheinlichkeit überführt. Somit ist die unterste Stufe - der Pixellevel - erreicht. Durch Anwendung von Hystereseschwellen werden Bildpunkte mit ähnlichen Wahrscheinlichkeiten gruppiert. Die Ergebnisse dieses Verarbeitungsschrittes repräsentieren den Fahrzeuglevel. Aufgrund der hohen Fehleralarmrate findet eine zusätzliche Validierung der erhaltenen Fahrzeughypothesen statt. Hierfür werden mindestens zwei aufeinander folgende Fahrzeuge als Linie modelliert. Durch die Zusammenfassung gleichgerich-

2.2. Ansätze zur Fahrzeugextraktion

teter Linien, welche zusätzlich mit der gegebenen Straßenrichtung übereinstimmen müssen, kann so eine bessere Abstimmung zwischen Detektion- und Fehleralarmrate gefunden werden. Die erzeugten Fahrzeugreihen sind Bestandteile des Linienlevels, aus welchen durch Analyse der erhaltenen Strukturen der „Meta-Linien“ Level erstet. Dieser trifft zusätzlich die Entscheidung, ob es sich um Fahrzeuge auf einer Straße oder innerhalb eines Parkplatzes handelt, wodurch das Vorhandensein einer Straße oder eines Parkplatzes bestätigt werden soll. Es ist anzumerken, dass der vorliegende Ansatz nur bis zum Pixellevel evaluiert wird. Die Verbesserungen durch die Einführung höherer (semantischer) Level sind daher nur Vermutungen. Zudem ist es fragwürdig, wie sinnvoll die Bestätigung der Existenz einer Straße ist, wenn deren Richtung bereits in den vorherigen Arbeitsschritten genutzt wird und somit Informationen über dieses Objekt schon zur Verfügung stehen.

Ein Ansatz, welcher mit Hilfe Neuronaler Netze eine Fahrzeugerkennung in hochgelösten Luftbildern durchführt, ist in Spiegel (2000) beschrieben. Das Hauptaugenmerk dieser Arbeit liegt auf der umfangreichen Untersuchung möglicher Parameterkombinationen und der Variation von Trainingsdaten zur Erzeugung der Klassifikatoren. Mit einem numerisch optimierten Parametersatz konnte vereinzelt eine Extraktionsqualität von bis zu 90 % erreicht werden, sofern vor der Fahrzeugerkennung eine Einschränkung der Interessengebiete stattfindet und der bearbeitete Bildausschnitt hauptsächlich helle Fahrzeuge enthält. Trotz der hohen Auflösung der verwendeten Bilder lag die Qualität der erreichten Ergebnisse ohne Verwendung von Vorinformation teils deutlich unter 70 %. Dies ist in erster Linie auf die verwendeten Merkmale und die Anzahl der Trainingsdaten zurückzuführen. Als Merkmale zum Training der Klassifikatoren werden entweder direkte Grauwerte der Eingangsbilder oder Werte aus radiometrisch vorverarbeiteten Bildern verwendet, wodurch die Erfassung geometrischer Grundstrukturen wie Linien, Kanten oder Punkte nur schlecht möglich ist. Die Anzahl der verwendeten Trainingsbilder wird in der Arbeit mit 18 angegeben, welche sich in 9 (helle) Fahrzeuge und 9 Beispiele der Gegenklasse aufteilt. Durch Variation der Orientierung und Verschiebungen innerhalb der Suchfensters werden hieraus zwar über 2500 Bildausschnitte geniert, diese spiegeln trotzdem nicht ausreichend die gesuchten Objekte und deren Gegenklasse wieder.

Ziel des Ansatzes von Grabner et al. (2008) ist die gezielte und zeitsparende Erfassung von Trainingsdaten, welche im Rahmen jeder überwachten Klassifikation benötigt werden und deren Bereitstellung daher einen erheblichen Arbeitsaufwand verursacht. Als Klassifikator kommt der ursprüngliche AdaBoost Ansatz nach Freund & Schapire (1997) unter Verwendung von schnell berechenbaren und zum Teil rotationsinvarianten Merkmalen zum Einsatz. Auf Grundlage eines gegebenen Trainingsbeispiels wird der zuvor mit Zufallsmerkmalen initialisierte Klassifikator neu berechnet und die Klassifikation auf einem Bildausschnitt durchgeführt. Aus der daraus entstehen hohen Anzahl an falschen und fehlenden Detektionen wählt ein Operateur das nächste Trainingsbeispiel. Der Klassifikator wird neu berechnet und der Bildausschnitt (oder ein anderer) erneut klassifiziert. Durch die manuelle Auswahl falscher oder fehlender Fahrzeuge können intuitiv die repräsentativsten Beispiele erfasst werden. Im Gegensatz zur

einmaligen Erfassung eines umfangreichen Trainingsdatensatzes kann so die Erfassung von Beispielen vermieden werden, welche bereits durch den Klassifikator der richtigen Objektklasse zugeordnet werden. Nach Beendigung des Trainings erfolgt eine Klassifikation unter Verwendung verschiedener Bildausrichtungen. Die aus den unterschiedlichen Rotationen erhaltenen Konfidenzwerte jedes Bildpunktes werden abschließend durch einen nicht parametrischen Clusteralgorithmus zusammengefasst. Die Evaluierung erfolgt mit und ohne die Nutzung von Straßeninformationen. Es treten zum Teil gravierende Unterschiede der erreichten Zuverlässigkeit von bis zu 20 % auf. Trotz der sehr hohen Auflösung von 8 cm bis 10 cm wird daher eine vorherige Einschränkung des Suchbereiches empfohlen.

2.2.2 Niedrigaufgelöste Einzelbilder

Bereits vor der Verfügbarkeit optischer Satellitenbilder mit einer Auflösung im Meter- bzw. Submeterbereich wurden erste Studien zur Nutzung daraus zu gewinnender Verkehrsdaten durchgeführt. Ein Schwerpunkt der Arbeit von McCord et al. (1995) liegt in der Untersuchung der Potentiale und Kosten für die Ableitung von Verkehrsparametern aus dem für 1999 geplanten IKONOS Satelliten. Eine Detektion von Fahrzeugen findet auf eingescannten und auflösungsreduzierten Luftbildern von Highway-Ausschnitten statt. Durch Anwendung eines manuell gewählten Schwellwertes entstehen sog. „Klumpen“, deren Ausdehnung und Anzahl enthaltener Bildpunkten als diskriminative Merkmale verwendet werden. Es wird ein Verfahren vorgestellt, welches zunächst - basierend auf dem Förstner Operator (Förstner & Gülch, 1987) - Schattenbereiche von Fahrzeugen findet und diese durch ein Wachstumsverfahren vergrößert. Eine Filterung der erhaltenen Ergebnisse erfolgt mit geometrischen Eigenschaften der erhaltenen Schattenregionen. Parallel hierzu werden Fahrzeughypothesen, zum einem durch die Gruppierung markanter Kanten und zum anderen durch Anwendung automatisch berechneter Schwellwerte gebildet. Die Zusammenführung dieser drei Informationen erfolgt innerhalb eines Entscheidungsbaumes, welcher systematisch Regionen ausschließt die den geforderten geometrischen Eigenschaften nicht entsprechen. Außerdem findet eine Kontrolle auf Übereinstimmung zur Straßenrichtung statt. Auf wenig komplexen Highway-Szenen können dadurch gute Detektionsraten mit hoher Zuverlässigkeit erreicht werden. Für Aufnahmen mit einer simulierten Bodenpixelgröße von einem Meter werden bis zu 94 % der Fahrzeuge bei einer Fehlerrate von unter 10 % erkannt.

Ein anderer Ansatz dieser Forschergruppe wird in McCord & Merry (2000) vorgestellt. Die Erkennung basiert auf der Unterscheidung von stationären und dynamischen Bildpunkten. Erstere stehen für den Straßenbelag, während die zweite Gruppe Fahrzeuge repräsentiert. Hierzu setzen die Autoren ein Hintergrundbild der zu bearbeitenden Szene ohne Fahrzeuge voraus. Zur Bildung des Differenzbildes wird vom zu bearbeitenden Bild ein transformiertes (d.h. radiometrisch angepasstes) Hintergrundbild subtrahiert. Die dabei verwendeten Transformationsparameter werden iterativ durch Schätzung der posteriori Wahrscheinlichkeiten für dynamische und stationäre Bildpunkte bestimmt, bis diese Wahrscheinlichkeiten konvergieren. In späteren Arbeiten (Sharma, 2002; Sharma et al., 2006) wird dieses Verfahren als „Baye-

2.2. Ansätze zur Fahrzeugextraktion

sian Background Transformation“ (BBT) vorgestellt. In der Nachbearbeitung werden manuell gesetzte Schwellwerte auf die erhaltenen Wahrscheinlichkeiten angewandt und so Ansammlungen von dynamischen Pixeln bestimmt. Die Evaluierung auf gescannten Luftbildern ergibt eine deutliche Steigerung der Leistungsfähigkeit des Ansatzes im Vergleich zu reinen Schwellwertverfahren. Allerdings stellt die Verfügbarkeit der benötigten Hintergrundbilder eine enorme Einschränkung des vorgestellten Verfahrens dar.

Neben der BBT werden in Sharma et al. (2006) zwei weitere Verfahren zur Fahrzeugerkennung vorgestellt. Die erste Methode verwendet ebenfalls ein zuvor erzeugtes Hintergrundbild, welches keine Fahrzeuge mehr beinhaltet. Aus diesem und dem zu untersuchenden Bild werden anschließend vier verschiedene Texturkanäle erzeugt. Die Texturkanäle werden einer Hauptkomponententransformation unterzogen, wobei eine der entstehenden Komponenten heuristisch zur weiteren Bearbeitung ausgewählt wird. Meist wird die dritte oder vierte Hauptkomponente verwendet, da die ersten beiden überwiegend den Einfluss der Straße wiedergeben. Die ausgewählte Komponente wird durch Anwendung geeigneter Schwellwerte binarisiert und mehreren morphologischen Filterungen zur Beseitigung kleinerer Störungen unterzogen. Im daraus erhaltenen Ergebnisbild werden Blobs, welche vorgegebene geometrische Kriterien erfüllen, als Fahrzeuge deklariert.

Das zweite vorgestellte Verfahren kommt ohne das schwierig zu erzeugende Hintergrundbild aus. Zunächst wird das Gradientenbild erzeugt und die dazu gehörende Grauwertmatrix berechnet, welche in eine Wahrscheinlichkeit für das Zusammentreffen zweier Gradientenstärken transformiert wird. Nach Anwendung eines Schwellwertes auf das Gradientenbild wird für das erhaltene Binärbild ebenfalls die Grauwertmatrix berechnet und transformiert. Aus den erhaltenen beiden Wahrscheinlichkeiten wird nun ein Maß für die relative Entropie bestimmt. Durch Variation des Schwellwertes findet eine Minimierung dieses Entropiemaßes statt. Der resultierende Schwellwert wird erneut auf das Gradientenbild angewandt. Nach Anwendung geeigneter morphologischer Operatoren werden wie zuvor die erhaltenen Blobs als Fahrzeuge deklariert, sofern sie geometrischen Vorgaben entsprechen. Eine zusammenfassende Evaluierung aller vorgestellten Verfahren zeigt, dass das BBT-Verfahren Detektionsraten von über 95 % bei einer Fehlerrate von weniger als 2 % liefert. Zudem zeigt dieser Ansatz die größte Zuverlässigkeit gegenüber störenden Einflüssen. Es bleibt allerdings anzumerken, dass auch bei diesen Arbeiten einfache Highway-Szenen verwendet werden und die vorgenommene numerische Evaluierung auf einer geringen Anzahl von Referenzfahrzeugen beruht.

In Alba-Flores et al. (2004) und Alba-Flores (2005) werden automatisch bestimmte Schwellwerte zur Fahrzeugerkennung verwendet. Diese werden durch das Verfahren nach Otsu (1979) für eine separate Detektion von hellen und dunklen Fahrzeugen bestimmt. Die Otsu Methode bestimmt durch eine vollständige Suche die Schwellwerte zur Binarisierung eines Bildes, so dass die Varianzen der Grauwerte innerhalb der entstehenden zwei Bildbereiche minimal werden. Abschließend werden die jeweils erhaltenen Regionen anhand ihrer Größe in PKWs und LKWs unterteilt. Das vorgestellte Verfahren erzielt auf einfachen

Highwayszenen gute Ergebnisse. Allerdings geben die Autoren zu bedenken, dass im Falle komplexerer Szenen oder bei starkem Verkehrsaufkommen die Leistungsfähigkeit des Algorithmus nachlässt, da eng stehende Fahrzeuge zusammengefasst werden.

Ein Verfahren zur Erkennung gruppierter Fahrzeuge wird in Leitloff et al. (2006b) vorgestellt. Zunächst werden Fahrzeugkolonnen im Rahmen einer Linienerkennung im auflösungsreduzierten Bild extrahiert. Die Breite der erhaltenen Linien wird anschließend subpixelgenau bestimmt. Obwohl die Lücke zwischen gruppierten Fahrzeugen aufgrund der geringen Auflösung nicht sichtbar ist, zeigen sich wiederholende Muster in der berechneten Breite. So nimmt diese am Übergang zwischen aufeinander folgenden Fahrzeugen signifikant ab, wodurch eine Trennung einzelner Fahrzeuge innerhalb der Kolonne möglich ist. Die daraus erhaltenen Hypothesen werden anschließend durch eine robuste Schätzung des Kontrastes zur Umgebung validiert. Aufgrund der fehlenden Modellierung isoliert stehender Fahrzeuge erreicht dieser Ansatz insgesamt nur eine geringe Detektionsrate. Die Zuverlässigkeit der Erkennung einzelner Fahrzeuge innerhalb der Reihen ist hingegen sehr hoch.

Ein weiterer Ansatz der sich auf die Erkennung von Fahrzeugkolonnen konzentriert, ist in Hinz & Stilla (2006) zu finden. Obwohl thermale Videobilder mit einer Auflösung von 1 m verwendet werden, zeigen Fahrzeuge auf diesen Bildern ähnliche Charakteristiken wie auf optischen Satellitenbildern. Wegen des verwendeten Sensors ist allerdings eine Bestimmung des Rauschanteils vor der eigentlichen Prozessierung notwendig. Die Extraktion von Fahrzeugen erfolgt anschließend mit Hilfe eines differential-geometrischen Blobdetektors. Die recht einfache Modellierung führt zu einer Reihe falscher Fahrzeughypothesen. Daher werden die Parameter der extrahierten Ellipsen als Unsicherheiten in eine modifizierte Hough-Transformation für Linien eingebracht. Hierdurch ist es möglich, im Akkumulatorbild Reihen von mindestens zwei Fahrzeugen zu erkennen und diese Hypothesen zu bestätigen. Eine Erweiterung dieses Ansatz durch Hinzunahme der Bewegung zwischen den Videobildern findet in Yao et al. (2008a) statt. Dadurch können auch isoliert stehende Fahrzeuge erkannt werden, sofern eine Extraktion einzelner Blobs über die Bildsequenz hinweg erfolgt.

Die bisher vorgestellten Arbeiten nutzen explizite Modelle zur Beschreibung und Detektion von Fahrzeugen. Im Gegensatz dazu wird in den folgenden Arbeiten eine implizite Modellierung mit Hilfe von Trainingsdaten durchgeführt.

In Gerhardinger et al. (2005) wird die Software *Feature Analyst* zur Erkennung von Fahrzeugen verwendet, wobei die Detektion auf einem manuell erfassten Straßenlayer begrenzt wird. Da die Software aus den gegebenen Trainingsdaten automatisch das implizite Modell lernt, werden keine genauen Angaben zur Art der verwendeten Merkmale und des Klassifikators getroffen. Aufgrund der sehr genauen Straßendaten erreicht der Ansatz eine hohe Zuverlässigkeit und Vollständigkeit.

2.2. Ansätze zur Fahrzeugextraktion

In Jin & Davis (2007) werden spezielle neuronale Netze, sogenannte *Morphological shared-weight neural networks*, zur Klassifikation verwendet. Eine Besonderheit der verwendeten Methode besteht darin, dass neben dem Training der Abbildungsfunktion des neuronalen Netzes eine Optimierung der Merkmalsberechnung stattfindet. Hierfür werden die Trainingsbilder mit verschiedenen Strukturelementen (Hit-Miss Operatoren) morphologisch gefiltert. Die daraus erhaltenen Bilder dienen als Merkmale für das Training des neuronalen Netzes. Abhängig von den erzielten Klassifikationsergebnissen werden die Strukturelemente modifiziert und das neuronale Netz mit den neuen Merkmalen trainiert. Dadurch ist es möglich, eine sehr gute Abstimmung zwischen den verwendeten Merkmalen und dem Lernalgorithmus zu erreichen. Das Ergebnis der Klassifikation ist ein Konfidenzbild, auf welches ein zuvor gewählter Schwellwert angewandt wird, um Fahrzeugregionen zu erhalten. Der Vergleich mit einem Blobdetektor nach Lindeberg (1998) zeigt, dass sich zwar kaum Verbesserungen bezüglich der Vollständigkeit ergeben, die implizite Modellierung aber robuster gegenüber falschen Detektionen ist. Die gleiche Vorgehensweise findet in Zheng et al. (2006) Anwendung. Auch in diesem Fall werden gute Ergebnisse erzielt, allerdings berichten die Autoren, dass bei starkem Verkehrsaufkommen und auf Parkplätzen Fahrzeuge fälschlicherweise zusammengefasst werden, was zu einer deutlichen Reduktion der Vollständigkeit führt.

Ein weiterer Ansatz dieser Forschergruppe wird in Zheng & Li (2007) und Zheng & Hu (2008) vorgestellt. Aus den Trainingsdaten werden durch Berechnung der Ähnlichkeiten (entspricht annähernd der Korrelation) die repräsentativsten Beispiele gesucht. Die Klassifikation findet dann durch Berechnung der Ähnlichkeit des zu bearbeitenden Bildausschnittes mit diesen Beispielen statt. Tritt eine Ähnlichkeit über einem zuvor gewählten Schwellwert auf, wird der Mittelpunkt des Bildausschnitts als Fahrzeug deklariert. Zur Entfernung zu kleiner Regionen wird das Ergebnis abschließend morphologisch gefiltert. Im Vergleich zur Verwendung neuronaler Netze (Zheng et al., 2006) werden ähnlich gute Ergebnisse erzielt. Allerdings ist auch in diesem Fall eine Trennung eng zusammen stehender Fahrzeuge nicht möglich.

Der in Eikvil et al. (2009) vorgestellte Ansatz konzentriert sich im Gegensatz zu den meisten anderen Arbeiten auf diesem Gebiet nicht ausschließlich auf einfache Autobahnszenen. Nach der Beschränkung auf Interessengebiete durch Geoinformationen segmentieren die Autoren die Eingangsbilder zunächst in potentielle Regionen von Vegetation, Straßen und Fahrzeuge. Letzteres geschieht durch die Berechnung geeigneter Schwellwerte nach der Otsu Methode innerhalb der als Straßen angenommenen Bildbereiche. Die genaue Klassenzuordnung wird in einem zweistufigen Verfahren durchgeführt. Nach der Berechnung geometrischer und radiometrischer Eigenschaften für alle segmentierten Regionen wird ein Großteil der nicht zur Fahrzeugklasse gehörigen Regionen, mit einem manuell erstellten regelbasierten Klassifikator von der weiteren Prozessierung ausgeschlossen. Aus den verbleibenden Bildbereichen werden mit Hilfe eines nicht linearen Klassifikator (trainiert durch quadratische Diskriminanzanalyse) und den zuvor bereits aufgeführten Merkmale die zur Fahrzeugklasse gehörigen Bildbereiche bestimmt. Die Evaluierung der erhaltenen Ergebnisse wird auf zwei unabhängig voneinander erfassten Referenzdatensätzen durchgeführt, welche nur zu 80 % überstimmen. Die Detektionsqualität beträgt bis zu 83 % bei einem der Referenzda-

tensätze erreicht. Da die Referenzdatensätze allerdings Unstimmigkeiten aufweisen, ist eine endgültige Wertung der erreichten Ergebnisse nur schwer möglich.

Einen ebenfalls auf der Klassifikation zuvor segmentierter Bildbereiche basierenden Ansatz verfolgen die Autoren in Larsen et al. (2009). Die Segmentierung erfolgt in diesen Fall durch die mehrfache Bestimmung von Schwellwerten für helle und dunkle Regionen nach der Otsu Methode. Durch die Verwendung multispektraler Informationen werden neben den gesuchten Fahrzeugen auch Flächen mit Vegetationen und die hierdurch hervorgerufenen Schattenbereiche bestimmt. Dadurch ist es den Autoren möglich, Bereiche in denen eine Vielzahl von falschen Detektionen auftreten, von der weiteren Verarbeitung auszuschließen. Anschließend werden wie in der Arbeit von Eikvil et al. (2009) Merkmale der verbleibenden Regionen berechnet, auf deren Basis eine Klassifikation mit der Maximum-Likelihood-Methode durchgeführt wird. Die Autoren geben im Rahmen ihrer numerischen Evaluierung an, dass ihr Ansatz dazu neigt, die tatsächliche Anzahl von Fahrzeugen teils signifikant zu unterschätzen. Als Ursache hierfür wird in erster Linie der Segmentierungsschritt angesehen, in dem ein hoher Anteil der Fahrzeuge aufgrund zu schwachen Kontrastes zur Fahrbahn nicht korrekt segmentiert wird und in der weiteren Bearbeitung fehlt. Die Autoren kommen daher zum Schluss, dass die Anwendung von Schwellwerten auf Grauwerte in diesen Fällen ungeeignet erscheint und für den operationellen Betrieb weitere Modifikationen des Ansatzes notwendig sind.

2.2.3 Schätzung von Fahrzeugbewegungen aus Satellitenbildern

Optische Satellitensysteme verwenden Zeilenkameras zur Aufnahme der Bilddaten. Die Sensorgeometrie dieser Systeme kann zu unterschiedlichen Positionen eines bewegten Objektes in den einzelnen Spektralkanälen führen. Daher werden einige Arbeiten zur Schätzung von Fahrzeugbewegungen aus diesem Bildmaterial vorgestellt.

Erste Untersuchungen zur Bestimmung der Versatzes von Objekten und ihrer daraus resultierenden Geschwindigkeit wurden in Etaya et al. (2004) durchgeführt. Die Erfassung der unterschiedlichen Positionen im panchromatischen und in den multispektralen Kanälen erfolgte manuell durch Überlagerung dieser Bilder. Das gleiche Prinzip der manuellen Messung wenden die Autoren auf SPOT Bilder an (Etaya et al., 2005), um Wellenbewegungen nach dem verheerenden Nordsumatra Erbeben 2005 zu erfassen. Trotz der geringen Bodenpixelgröße dieses Bildmaterials von 10 m können eindeutige Wellenbewegungen erkannt werden, da der zeitliche Abstand zwischen den unterschiedlichen Kanälen etwa 2,7 s beträgt.

Zhang & Xiong (2006) und Xiong & Zhang (2008) konzentrieren sich auf die genaue Koregistrierung der zeitlich versetzten Bilder. Die Autoren führen eine Verbesserung des Sensormodells mit Hilfe von manuell gemessenen Verknüpfungspunkten und eines hoch genauen digitalen Höhenmodells durch. Die Positionen der Fahrzeuge in diesen hoch genau koregistrierten Bildern wird anschließend manuell erfasst und mit Hilfe der bekannten Zeitdifferenz die Geschwindigkeit bestimmt. Aufgrund fehlender

2.3. Bewertung bisherigen Ansätze und Schlussfolgerungen

Referenzdaten sind die erhaltenen Ergebnisse nicht kontrollierbar. Eine numerische Evaluierung wird nicht durchgeführt. Es ist allerdings auffällig, dass die Mehrheit der erfassten Fahrzeuge über der erlaubten Höchstgeschwindigkeit des abgebildeten Gebietes fahren und die Genauigkeit der bestimmten Geschwindigkeiten daher anzuzweifeln sind.

Auch Pesaresi et al. (2007) bzw. Pesaresi et al. (2008) führen eine genaue Koregistrierung der Spektralkanäle über Passpunkte durch. Anschließend werden Fahrzeuge in beiden Bildern grob markiert. Um die Positionen werden Blob-ähnliche Strukturen extrahiert und durch morphologische Filterung geglättet. Die Bestimmung der Mittelpunkte der erhaltenen Regionen liefert eine genauere Position der Fahrzeuge, aus denen Versatz und Geschwindigkeit bestimmt werden.

Die Integration der Bewegungsschätzung in einen automatischen Ansatz erfolgt in Leitloff et al. (2007b) auf QuickBird Bildern. Nach der Detektion von Fahrzeugen im panchromatischen Kanal, werden die Fahrzeuge und Teile ihrer Umgebung als Bildvorlage ausgeschnitten und auf die Bodenpixelgröße der multispektralen Kanäle reduziert. Nach der Transformation des Farbraums der multispektralen Kanäle wird die Übereinstimmung zur Bildvorlage berechnet, wobei meist ein eindeutiges Maximum gefunden werden kann. Die genaue Position des Maximums und des Fahrzeuges im panchromatischen Kanal wird durch Parameterschätzung bestimmt. Mit dem Positionsfehler aus der Parameterschätzung erfolgt abschließend ein Test auf die Signifikanz einer Bewegung.

Die Bestimmung von Fahrzeugbewegungen ist nicht nur aus dem Versatz zwischen den Spektralkanälen möglich. Einige Satelliten (z.B. IKONOS) sind in der Lage ein Gebiet in Vor- und Rückblick während eines Überfluges aufzunehmen. Damit wäre es möglich Fahrzeuge in beiden Aufnahmen zu erfassen (McCord et al., 1998). Da die Zeitdifferenz zwischen Vor- und Rückblick jedoch über 10 s liegt und ist eine zuverlässige Wiedererkennung von Fahrzeugen kaum möglich.

2.3 Bewertung bisherigen Ansätze und Schlussfolgerungen

In diesem Abschnitt sollen Teilaspekte der vorgestellten Ansätze auf ihre Nutzung zur Erfassung von Fahrzeugen aus Satellitenbildern bewertet werden. Dafür wird die Verwendung von zusätzlichen Informationen aus Geoinformationssystem, die Art der Modellierung und Verfahren im Rahmen der Extraktionsstrategie betrachtet.

Die bisherigen Arbeiten zur Erfassung verkehrsrelevanter Informationen aus Satellitenbildern nutzen Geoinformationen zur Reduzierung der Bilder auf Interessengebiete. Dies ist zur Reduktion falscher Detektionen sinnvoll. Eines der wichtigsten Argumente zur Nutzung externer Geoinformationen ist, dass die spätere Verwendung der gewonnenen Verkehrsinformationen nur in Bereichen sinnvoll ist, deren Nutzung dem Anwender bekannt ist und an dessen Beobachtung er deshalb Interesse hat. Im Gegensatz

2.3. Bewertung bisherigen Ansätze und Schlussfolgerungen

hierzu stehen einige Arbeiten mit luftgetragenen Bildsensoren, welche das Vorhandensein von Fahrzeugen zur Extraktion und/oder Validierung von Straßen nutzen (Hinz, 2003a). In diesen Fällen ist das primäre Ziel allerdings nicht, in einer vollständigen Erkennung von Fahrzeugen zu sehen. Es genügt bereits eine geringe Anzahl zuverlässiger Fahrzeugdetektionen, um Rückschlüsse auf die Existenz einer Straße zu ziehen. Auffällig bei den in Abschnitten 2.2.2 und 2.2.3 vorgestellten Arbeiten ist die Genauigkeit der verwendeten Straßeninformationen. So umfassen die dargestellten Interessengebiete ausschließlich Bildpunkte mit Straßen. Dies erscheint aufgrund der erreichbaren Georeferenzierung von Satellitenbildern sowie von im GIS enthaltenen Fehlern sehr optimistisch. Bei der Entwicklung automatischer Verfahren sollten potentielle Ungenauigkeiten in diesem Bereich daher berücksichtigt werden.

Bezüglich der verwendeten Modellierungen können aus bisherigen Arbeiten folgende Schlussfolgerungen gezogen werden:

- Bei Verwendung von **Drahtgittermodellen** (Hinz, 2004; Zhao & Nevatia, 2003) werden auch ohne die Nutzung von Vorabinformationen gute Ergebnisse erzielt. Allerdings ist dies nur durch die hohe Auflösung des Bildmaterials und deutlich hervortretende Objektteile möglich. Diese Voraussetzungen sind in Satellitenbildern nicht erfüllt, wodurch dieser Modellierungsansatz keine Anwendung in dieser Arbeit findet.
- Isoliert erscheinende Fahrzeuge sollten durch **implizite Modellierung** erfasst werden, da deutlich bessere Ergebnisse (Gerhardinger et al., 2005; Zheng et al., 2006; Jin & Davis, 2007; Zheng & Li, 2007; Zheng & Hu, 2008; Grabner et al., 2008) als durch Anwendung von Schwellwertverfahren (Alba-Flores, 2005; Alba-Flores et al., 2004; McCord & Merry, 2000; McCord et al., 1995; Sharma et al., 2006) erzielt werden, sofern die verwendeten Merkmale auf die Fahrzeugerkennung abgestimmt und in ausreichender Anzahl vorhanden sind. Eine Einschränkung der Verfahren liegt in der Erkennung gruppierter Fahrzeuge, da diese nur als zusammengesetzte Struktur erfasst werden und eine korrekte Separierung einzelner Fahrzeuge verhindert.
- Ein weiterer Schwachpunkt bisheriger Verfahren zeigt die Gruppierung der Ergebnisse von aussehensbasierten Klassifikationen. Die Verwendung manueller Schwellwerte auf zuvor segmentierte Bildregionen ist die Hauptfehlerquelle in Jin & Davis (2007). Die segmentierten Bildregionen sollten wie in Eikvil et al. (2009) und Larsen et al. (2009) demonstriert, ebenfalls durch eine **Bildobjekt-Klassifikation** bewertet werden.
- Im Falle gruppierter Fahrzeuge zeigen **globale Konfigurationsmodelle** deutliche Vorteile. Regelmäßig auftretende Muster können zur Detektion (Burlina & Chellappa, 1997; Leitloff et al., 2006b), Reduktion der Fehlerrate und gezielten Betätigung schwacher Fahrzeughypothesen (Ruskoné et al., 1996; Liu et al., 1999; Hinz, 2004) verwendet werden.

Die angewandten Extraktionsstrategien bisheriger Arbeiten sind sehr vielseitig und beschreiben keine

2.3. Bewertung bisherigen Ansätze und Schlussfolgerungen

allgemein gültige Vorgehensweise. Es lassen sich aber einige Schlussfolgerungen für eine erfolgversprechende Strategie ziehen:

- Bei der Verwendung von impliziten Modellen ist ein Vorgehen nach Abschnitt 2.1.2 durchzuführen. Die verwendeten Klassifikationsalgorithmen müssen in der Lage sein, die vorliegenden Trainingsdaten ausreichend zu unterscheiden.
- Die Erkennung sollte in einem zweistufigen Verfahren erfolgen. Zunächst sollte eine pixel-basierte Klassifikation die Klassenzugehörigkeit einzelner Bildpunkte bestimmen. Die Gruppierung benachbarter Bildpunkte gleicher Klassenzugehörigkeit sollte ebenfalls mittels überwachter Klassifikationsverfahren stattfinden, da eine Zusammenführung basierend auf Schwellwertverfahren meist unzureichende Ergebnisse liefert.
- Bei der Umsetzung globaler Modelle eignen sich hybride Extraktionsstrategien. Hierdurch können Primitive als Hypothesen für Fahrzeugreihen erkannt und gemäß den Modellanforderungen bearbeitet werden. Anschließend ist eine gezielte Erfassung von Eigenschaften dieser Strukturen möglich. Auf Basis dieser Eigenschaften kann die Extraktion darin enthaltener Einzelfahrzeuge erfolgen.

2.3. Bewertung bisherigen Ansätze und Schlussfolgerungen

3 Fahrzeugmodellierung

In diesem Kapitel wird das Modell zur Erkennung von Fahrzeugen aus optischen Satellitenbildern vorgestellt. In Abschnitt 2.3 wurde gezeigt, dass eine zweistufige Modellierung vorteilhaft ist. Es wird daher ein Modell zur Erkennung isoliert stehender Fahrzeuge und in ein Modell zur Erfassung von Fahrzeugreihen entwickelt. Eine Fahrzeugreihe muss aus mindestens zwei hintereinander auftretende Fahrzeuge bestehen, die aufgrund des geringen Abstands zueinander nicht getrennt voneinander erscheinen. Diese Trennbarkeit ist nur schwer quantifizierbar, so dass Autos sowohl als isolierte wie auch als gruppierte Fahrzeuge erfasst werden können. Eine Zusammenführung der Ergebnisse dieser zunächst unabhängig voneinander betrachteten Modelle ist notwendig und erfolgt im Rahmen der Extraktion.

3.1 Modell für isolierte Fahrzeuge

In Jensen & Cowen (1999) werden minimale Anforderungen an die geometrische Auflösung von Fernerkundungsdaten für Anwendungen in urbanen Gebieten gestellt. Für eine zuverlässige Erfassung von Fahrzeugen wird eine minimale Bodenpixelgröße von 25 cm bis 50 cm gefordert. Bei dieser Bodenpixelgröße treten Strukturen von Fahrzeugen wie Heck- und Frontscheibe deutlich hervor (Abbildung 3.1a) und ermöglichen deren explizite Modellierung. In Satellitenaufnahmen sind derartige Bestandteile eines Autos nicht mehr erkennbar. Fahrzeuge erscheinen nur als Blobs unterschiedlicher Ausdehnung, Orientierung und Kontrast zu ihrer Umgebung (Abbildung 3.1b). Wie die Ergebnisse einiger Arbeiten aus Abschnitt 2.2.2 zeigen, gestaltet sich die Wahl geeigneter Parameter für diese Merkmale sehr schwierig. Daher sind die Ergebnisse von Segmentierungsverfahren, die auf absoluten Schwellwerten beruhen, oft unzureichend oder stark abhängig von Änderungen der Beleuchtung. In diesen Situationen sind Verfahren mit impliziter Modellierung deutlich zuverlässiger, sofern mögliche Variationen in den verwendeten Trainingsdaten vorkommen.

In dieser Arbeit wird eine implizite Modellierung für isoliert stehende Fahrzeuge angewandt. Die verwendeten Trainingsdaten müssen möglichst viele im Bild vorkommende Variationen von Fahrzeugen und der Gegenklasse enthalten. Für die verwendeten Trainingsbeispiele müssen Merkmale berechnet werden, welche die blobhafte Struktur widerspiegeln. Diese sollten also in der Lage sein, geometrische Formen wie Linien, Kanten und Flächen zu repräsentieren. Im Zuge der Klassifikation mittels impliziter Modellierung wird jedes Pixel einer Klasse zugeordnet. Benachbarte Bildpunkte der gleichen Klasse werden zu geschlossenen Strukturen zusammengefasst. Isoliert auftretende Fahrzeuge können detektiert

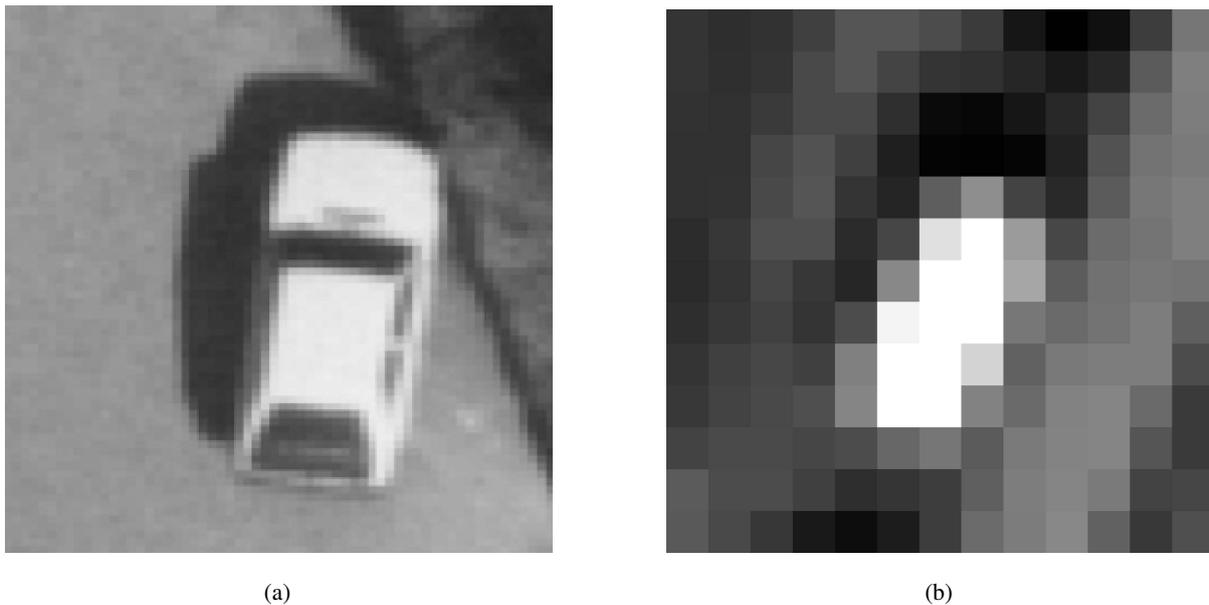


Abbildung 3.1: Erscheinung von Fahrzeugen in optischen Fernerkundungsdaten. (a) Hochaufgelöstes Luftbild mit 8 cm Bodenpixelgröße (aus Hinz (2004)), (b) Isoliertes Fahrzeug bei 60 cm Bodenpixelgröße

werden, wenn:

- eine ausreichend hohe Anzahl variierender Trainingsdaten für Fahrzeuge und deren Gegenklasse(n) dem Trainingsalgorithmus übergeben werden
- die blobhafte Struktur erfassende Merkmale verwendet werden
- die klassifizierten Bildpunkte sich zu zusammenhängenden Bereichen gruppieren lassen

3.2 Modell für Fahrzeugreihen

Die Modellierung von Fahrzeugreihen in Satellitenbilddaten, erfolgt in Anlehnung an die Arbeiten von Ruskoné et al. (1996), Michaelsen & Stilla (2001) und Hinz (2004). Es werden mindestens zwei aufeinander folgende Fahrzeuge als Linie modelliert. Abbildung 3.3 verdeutlicht diesen Sachverhalt anhand synthetischer Bilder einer Fahrzeugreihe. In der höchsten Auflösung (Abbildung 3.3a) zeigen die einzelnen Fahrzeuge deutlich voneinander getrennte Umriss. Diese entstehen durch Filterung mit einem Laplace Operator und der anschließenden Bestimmung der Nulldurchgänge des gefilterten Bildes. Außerdem wird eine Mittellinie für jedes Fahrzeug bestimmt. Deren Extraktion erfolgt nach dem Ansatz von Steger (1998), welcher in Abschnitt 4.4.1 genauer erläutert wird. Durch kontinuierliche Reduzierung der Auflösung „verschmelzen“ die einzelnen Fahrzeuge und bilden in der niedrigsten Auflösungsstufe (Abbildung 3.3d) eine zusammengehörige Struktur mit einer gemeinsamen Mittellinie und einem geschlossenen Umriss. Im Gegensatz zu den in Ruskoné et al. (1996), Michaelsen & Stilla (2001) und Hinz (2004) erläuterten Vorgehensweisen, welche einzelne Fahrzeughypothesen zu Linien

3.2. Modell für Fahrzeugreihen

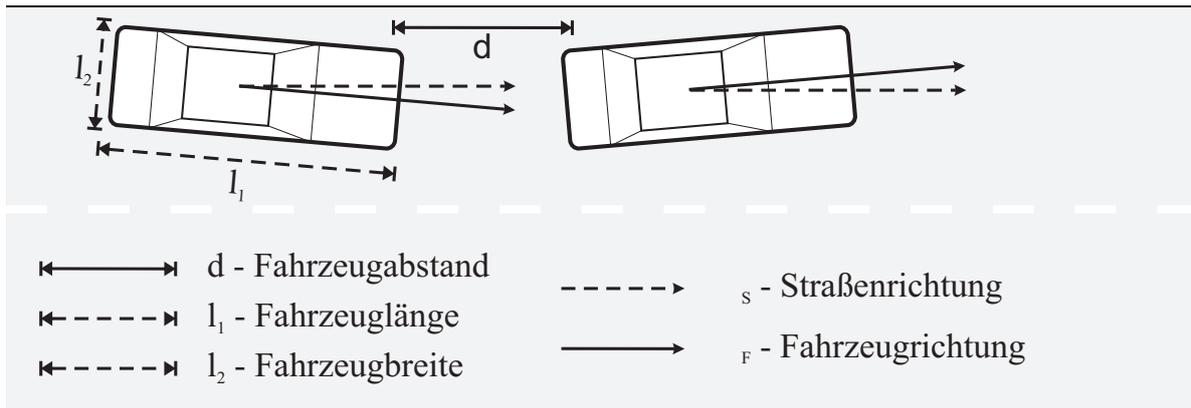


Abbildung 3.2: Fahrzeug- und Reihenmodell

gruppieren, wird in der vorliegenden Arbeit eine „Grob-zu-Fein“ Modellierung angewandt. Ein einzelnes Fahrzeug wird durch seine Länge l_1 , Breite l_2 und Richtung ϕ_F bestimmt. Abbildung 3.2 zeigt außerdem den Abstand d einzelner Fahrzeuge zueinander und die aus Geoinformationen ableitbare Straßenrichtung ϕ_S . Die als Linien modellierter Fahrzeugreihen, müssen die folgenden Anforderungen erfüllen:

- **Zutreffende Breite:** Nur Linien einer vorgegeben Breite werden erfasst.
- **Ausreichende Länge:** Nur Linien mit einer Länge von mindestens zweifacher Fahrzeuglänge, werden als Fahrzeugreihen betrachtet.
- **Übereinstimmung zur Straßenrichtung:** Die Differenz zwischen Fahrzeug- und Straßenrichtung darf einen vorgegeben Wert nicht überschreiten.

Zudem können kollineare Linien fusioniert werden, wenn folgende Bedingungen eingehalten werden:

- **Geringe Krümmung:** Kollineare Linien müssen annähernd gleiche Richtung aufweisen.
- **Geringer Abstand in Fahrzeugrichtung:** Kollineare Linien dürfen in Fahrzeugrichtung maximal eine Fahrzeuglänge auseinander liegen.
- **Geringer Abstand quer zur Fahrzeugrichtung:** Der Versatz von Linien darf quer zur Fahrzeugrichtung maximal die Hälfte der Fahrzeugbreite betragen.

Die Wahl geeigneter Werte für die in Abbildung 3.2 dargestellten Parameter erfolgt im sechsten Kapitel.

Fahrzeuge innerhalb dieser Reihen kennzeichnen sich durch die Variation des Kontrastes zur Umgebung und der Breite der Linien. Einzelne Fahrzeuge werden entlang der Fahrzeugreihe durch eine geeignete Funktion approximiert. Die Parameter der approximierenden Funktion müssen mit dem Kontrast und der Breite eines Einzelfahrzeuges korrelieren und im Rahmen einer Parameterschätzung bestimmbar sein.

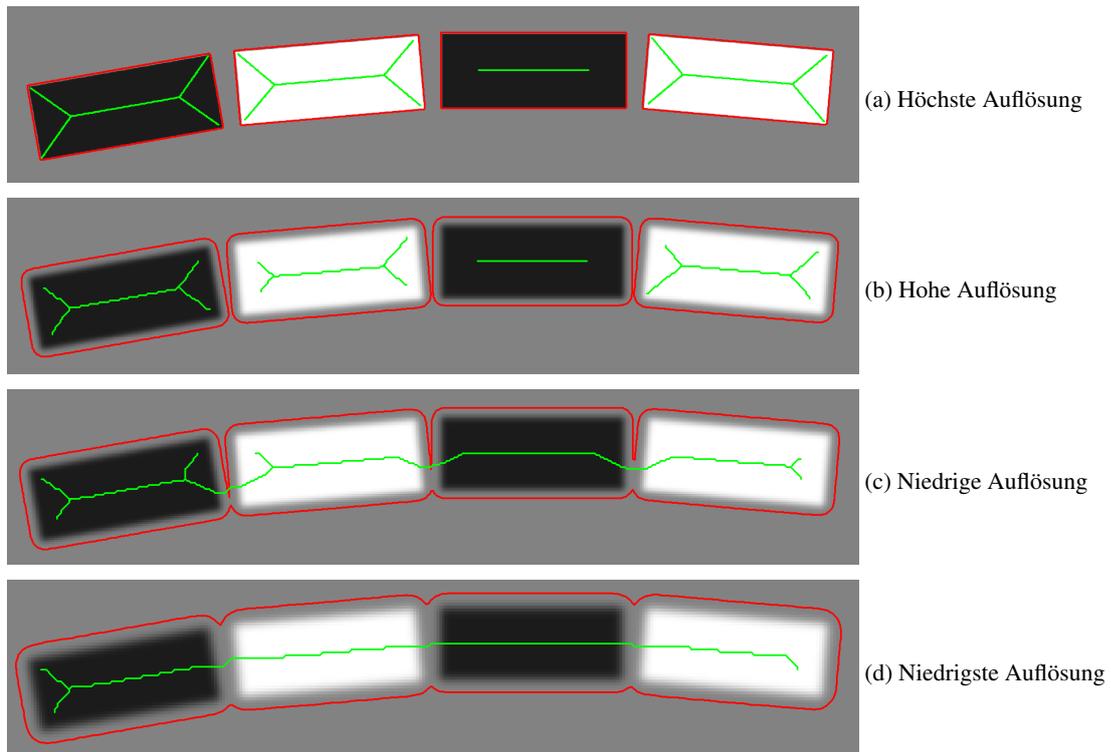


Abbildung 3.3: Verschmelzung einer Fahrzeugreihe bei Reduzierung der Auflösung. Mit dem Laplace Operator detektierte Kanten (rot) und durch Linienextraktion nach Steger (1998) gefundene Mittellinien (grün)

4 Fahrzeugextraktion

Ziel der Extraktion ist die computergerechte Umsetzung des im dritten Kapitel entwickelten Fahrzeugmodells. Zunächst wird eine Gesamtübersicht des implementierten Bildanalyse-Systems gegeben (Abschnitt 4.1). Anschließend wird die Beschränkung des Suchbereichs durch Verwendung zusätzlicher Geoinformationen beschrieben (Abschnitt 4.2). Der weitere Aufbau des Kapitels orientiert sich an der Modellierung für isolierte und gruppierte Fahrzeuge. In Abschnitt 4.3 werden die zur Erkennung von isolierten Fahrzeugen implementierten Verfahren erläutert. Abschnitt 4.4 beschreibt die Extraktion von Linien und die anschließende Erkennung von einzelnen Fahrzeugen aus diesen Hypothesen für Fahrzeugreihen. Die Zusammenführung der Ergebnisse aus beiden Modellierungen wird im Abschnitt 4.5 eingehend dargestellt. Am Ende des Kapitels wird für isolierte Fahrzeuge eine Schätzung des Bewegungszustandes vorgenommen.

4.1 Gesamtüberblick der Extraktionsstrategie

Die Fahrzeugextraktion gliedert sich in die drei Phasen der Vorverarbeitung, der Extraktion von Einzelfahrzeugen und der Nachbearbeitung. Abbildung 4.1 zeigt die beschriebenen Teilschritte und deren jeweilige Ergebnisse, welche in der nachfolgenden Prozessierung verwendet werden. Als Eingabedaten dienen hochaufgelöste Satellitenbilder mit einer Bodenpixelgröße unter einem Meter. Außerdem werden Geoinformationen über die Lage und Breite von Straßen vorausgesetzt. Liegen die Geoinformationen im gleichen Koordinatensystem wie die Satellitenbilder vor, kann auf die im Abschnitt 4.2 beschriebene Koordinatentransformation verzichtet werden.

In der Vorverarbeitung findet zunächst die Beschränkung auf Interessengebiete statt. Hierfür werden Straßendaten aus zusätzlichen Geoinformationen benötigt. Oft liegen diese Geoinformationen nicht im Koordinatensystem der Satellitenbilddaten vor. In diesen Fällen ist eine Transformation der Straßendaten notwendig. Diese Transformation wird am Beispiel der Überführung von Gauß-Krüger Koordinaten in das World Geodetic System 1984 (WGS84) System erläutert. Die anschließenden Bearbeitungsschritte finden ausschließlich auf den durch die Geoinformationen definierten Verkehrsflächen statt.

Die Trainingsphase wird zeitlich unabhängig von der Fahrzeugerkennung durchgeführt. Daher ist sie in der Abbildung 4.1 nur schwach dargestellt. Die verwendeten Lernalgorithmen werden im Rahmen der Anwendung der resultierenden Klassifikationsparameter beschrieben.

Die Extraktion von Einzelfahrzeugen beginnt mit der Klassifikation aller Bildpunkte. Die berechneten

4.1. Gesamtüberblick der Extraktionsstrategie

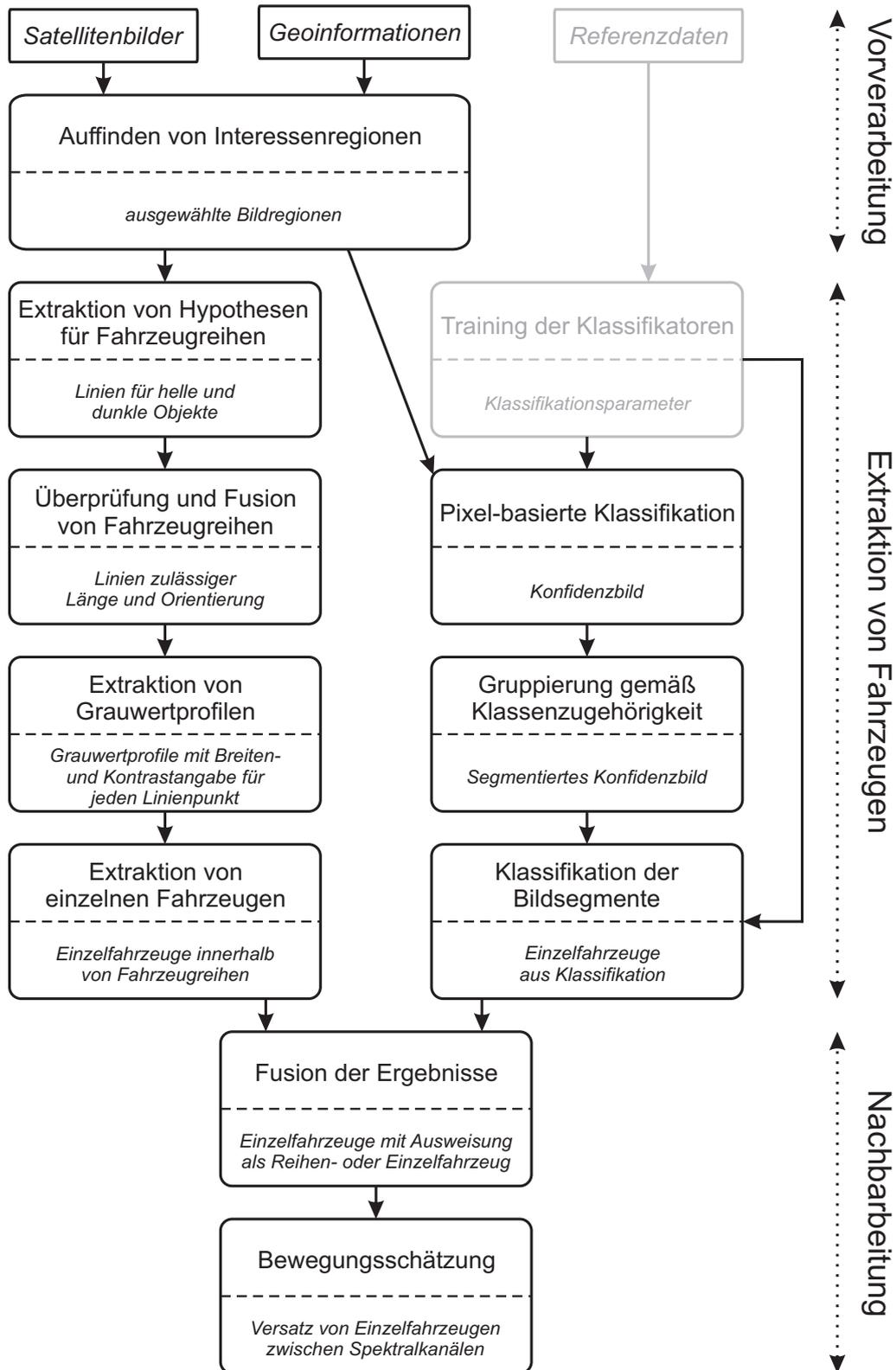


Abbildung 4.1: Übersicht zur Fahrzeugextraktion

4.2. Festlegung von Suchbereichen

Konfidenzwerte nehmen für Nicht-Fahrzeuge einen minimalen negativen Wert von -1 und für Fahrzeuge einen maximalen positiven Wert von $+1$ an. Der gesamte Wertebereich der berechneten Konfidenzwerte ist daher $F = [-1, +1]$. Benachbarte Bildpunkte (P_i, P_j) gleicher Klasse ($sign(F_{P_i}) = sign(F_{P_j})$) werden zu Segmenten zusammengefasst. Für diese Segmente werden geometrische Eigenschaften bestimmt. Zusätzlich erfolgt die Berechnung radiometrischer Eigenschaften innerhalb der Segmente. Die berechneten Merkmale der Bildsegmente werden durch einen weiteren trainierten Klassifikator bewertet, wodurch eine abschließende Klassifikation der Segmente in Fahrzeuge und Nicht-Fahrzeuge möglich ist.

Zur Umsetzung des globalen Fahrzeugreihenmodelles werden zunächst Linien in Abhängigkeit ihrer Breite und des Kontrastes extrahiert. Diese werden bezüglich der Modellanforderungen evaluiert und die verbleibenden Hypothesen für Fahrzeugreihen auf Basis von geometrischen Relationen zusammengefasst. Entlang dieser Linien werden Grauwertprofile berechnet. Diese dienen zur Erkennung von Einzelfahrzeugen innerhalb der Fahrzeugreihen.

Die Zusammenführung der Ergebnisse der zunächst getrennt stattfindenden Modellumsetzungen wird in Abschnitt 4.5 beschrieben. Dabei können Ergebnisse der Teilmodellierung für Einzelfahrzeuge auch Fahrzeugreihen zugeordnet werden.

Für isoliert verbleibende Fahrzeuge wird eine Schätzung des momentanen Bewegungszustandes durchgeführt (Abschnitt 4.6). Neben der Unterscheidung zwischen bewegten und unbewegten Fahrzeugen werden aus dem Versatz zwischen dem panchromatischen Kanal und den multispektralen Kanälen die Geschwindigkeit und deren mittlerer Fehler ermittelt. Eine Evaluierung dieser Werte kann aufgrund fehlender Referenzdaten nur auf Basis von Plausibilitätskriterien erfolgen.

4.2 Festlegung von Suchbereichen

Liegen Geoinformationen und Bildmaterial in unterschiedlichen Koordinatensystemen vor, müssen die Daten in ein einheitliches Koordinatensystem überführt werden. Die Vorgehensweise wird für die Transformation von Gauß-Krüger Koordinaten in UTM-Koordinaten (*Universal Transverse Mercator*) erläutert. Beide Abbildungen sind transversale konforme Zylinderabbildungen, allerdings haben diese ein unterschiedliches geodätisches Datum und verschiedene Meridianstreifenbreiten. Das Gauß-Krüger Koordinatensystem mit dem Rauenbergdatum ist zurzeit noch Standard für amtliche Geoinformationen in Deutschland. Das Rauenbergdatum ist Grundlage für das Deutsche Hauptdreiecksnetz (DHDN). Satellitenbilder werden hingegen meist mit UTM-Projektion im global definiertem WGS84-System (*World Geodetic System 1984*) ausgeliefert. Dieses Bezugssystem kann als identisch mit dem *Europäischen Terrestrischen Referenzsystem 1989* (ETRS89) angesehen werden.

Zur Umrechnung der ebenen konformen Koordinaten bieten sich verschiedene Vorgehensweisen an, welche in Abbildung 4.2 dargestellt sind. So könnte die Überführung von ebenen Koordinaten aus unterschiedlichen Projektionen, direkt durch die Verwendung korrespondierender Punkte mit Hilfe linearer oder affiner Transformationen durchgeführt werden. Diese Vorgehensweise ist aufgrund der verbleibenden Klaffungen zwischen den korrespondierenden Punkten nicht empfehlenswert. Die mathematisch

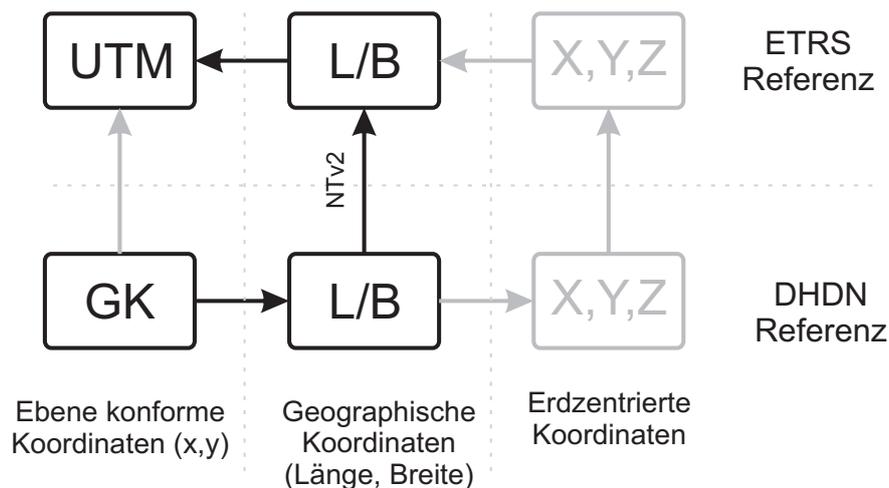


Abbildung 4.2: Transformation der ATKIS Daten gemäß AdV (2009)

korrekte Vorgehensweise sieht zunächst die Transformation der ebenen in geographische Koordinaten vor. Anschließend müssen diese Polarkoordinaten in erdzentrierte kartesische Koordinaten umgerechnet werden, welche dann immer noch das ursprüngliche geodätische Datum besitzen. Durch Anwendung einer 7-Parameter Helmert Transformation kann der Übergang zwischen den Referenzsystemen erfolgen. Anschließend findet eine Rücktransformation der kartesischen in geographische Koordinaten und deren entsprechende Projektion in ebene Koordinaten statt.

Eine weitere Möglichkeit zur Transformation der ebenen Koordinaten besteht durch den direkten Bezugssystemwechsel über geographische Koordinaten, wie dieser in Abbildung 4.2 hervorgehoben ist. Gemäß der Anweisung zur „Bundeseinheitlichen Transformation für AKTIS - (BeTA2007)“ (AdV, 2009) wird der Datumsübergang durch in einem Gitter vorliegende Shift-Werte realisiert. Dies sind achsparallele Versatzwerte für die geographische Länge und Breite, welche kostenfrei von der Arbeitsgemeinschaft für Vermessungsverwaltungen (AdV) und dem Informationssystem CRS-EU (<http://www.crs-geo.eu/BeTA2007>) zur Verfügung gestellt werden. Die Transformation findet unter Verwendung dieser Shiftwerte mittels des NTV2 (*National Transformation Version 2*) Ansatzes (Junkins & Farley, 1995) statt. Ein Vorteil dieser Vorgehensweise ist, dass keine Höheninformationen verwendet werden, wie diese zur Umrechnung der Polar- in erdzentrierte kartesische Koordinaten notwendig wären.

4.3 Isolierte Einzelfahrzeuge

Die Erkennung isoliert stehender Fahrzeuge, erfolgt durch zweifache Anwendung geeigneter Klassifikatoren. In Abbildung 4.1 ist dargestellt, dass zunächst alle Bildpunkte des Suchbereichs klassifiziert werden. Das berechnete Konfidenzbild wird segmentiert und die erhaltenen Bildsegmente erneut klassifiziert. Das Ergebnis dieses Prozessierungsschrittes sind die gesuchten isolierten Fahrzeuge. Die Wahl der Klassifikationsalgorithmen wurde unter Berücksichtigung der gesuchten Objekte und der verwendeten Merkmale

4.3. Isolierte Einzelfahrzeuge

gewählt.

Das Training und die Anwendung der Klassifikatoren folgt der in Abschnitt 2.1.2 vorgestellten allgemeinen Vorgehensweise.

4.3.1 Klassifikation der Bildpunkte

Merkmalsberechnung

Das Training der Klassifikatoren erfordert zunächst die Auswahl und anschließende Berechnung von geeigneten Merkmalen. Obwohl die Wahl von Merkmalen keinen Beschränkungen unterliegt (Niemann, 2003), müssen diese, auf das zu bearbeitende Klassifikationsproblem abgestimmt sein. Bei der Fahrzeugerkennung aus Satellitenbildern sollten für eine erfolversprechende Durchführung folgende Kriterien erfüllt werden:

- Die Merkmale müssen eine Unterscheidung des Aussehens von Fahrzeugen und Nicht-Fahrzeugen erlauben. Diese sollten also in der Lage sein, die Blob-Struktur von Fahrzeugen und beispielsweise linienhaften Strukturen, wie Fahrbahnrande, zu beschreiben. Um dies sicherzustellen, ist die Berechnung einer überrepräsentativen Menge an Merkmalen, aus welchen anschließend die Markantesten ausgewählt werden können, sinnvoll.
- Aufgrund der Menge an zu klassifizierenden Bildpunkten im Fall großflächiger Satellitenaufnahmen muss die Berechnung der Merkmale sehr effizient erfolgen.
- Eine erfolgreiche Anwendung der Merkmale zur Objekterkennung sollte bereits demonstriert worden sein.

Eine besonders erfolgreiche Merkmalsgruppe zur Objekterkennung sind die Haar-Wavelets (Haar, 1910) ähnliche Blockmerkmale, welche eine allgemeine Beschreibung von Objekten bieten (Papageorgiou et al., 1998; Papageorgiou & Poggio, 2000). Diese berechnen Grauwertunterschiede verschiedener Bildregionen. Obwohl es keine Beschränkung der Größe und gegenseitigen Lage der Bildregionen zueinander gibt, hat sich eine Reihe von Standardmerkmalen dieses Typus durchgesetzt (Lienhart & Maydt, 2002; Bartlett et al., 2003; Viola & Jones, 2004; Grabner et al., 2008). Diese umfassen Merkmale zur Erkennung von Kanten (Abbildungen 4.3a bis 4.3d), Linien (Abbildungen 4.3e bis 4.3l und 4.3o) und flächenhaften Objekten (Abbildungen 4.3m und 4.3n).

Die Berechnung der Merkmalswerte erfolgt durch Bildung der Differenz zwischen der Summe aller Bildpunkte in dunklen Regionen und der Summe aller Bildpunkte in hellen Regionen. Diese Grauwertunterschiede werden für alle möglichen Kombinationen von Größe und Lage innerhalb eines Suchfensters bestimmt, welches um den zu bearbeitenden Bildpunkt gelegt wird. Die Anzahl der möglichen Kombinationen kann gemäß Lienhart et al. (2003) für achsenparallele Merkmale berechnet werden durch

$$X \cdot Y \cdot \left(B + 1 - b \cdot \frac{X + 1}{2} \right) \cdot \left(H + 1 - h \cdot \frac{Y + 1}{2} \right) \quad (4.1)$$

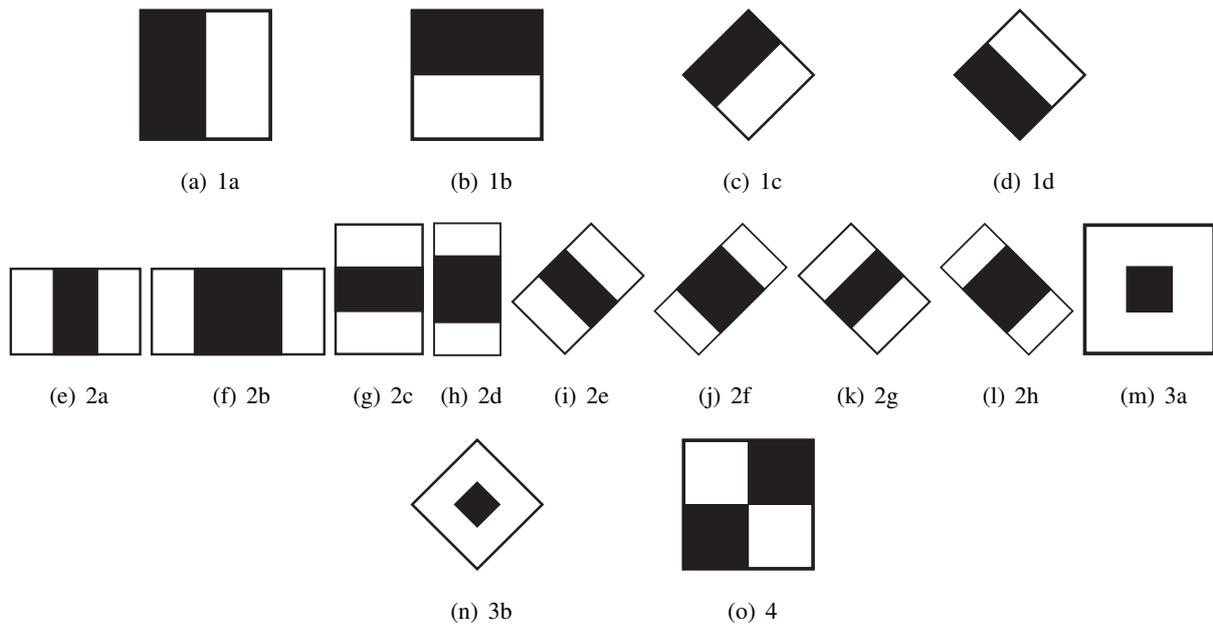


Abbildung 4.3: Haar-Merkmale

Merkmal	1a	1b	1c	1d	2a	2b	2c	2d	2e	2f	2g	2h	3a	3b	4
<i>b</i>	2	1	2	1	3	4	1	1	1	1	3	4	3	3	2
<i>h</i>	1	2	1	2	1	1	3	4	3	4	1	1	3	3	2

Tabelle 4.1: Parameter zur Berechnung der Merkmalsanzahl

mit den maximal möglichen Skalierungsfaktoren $X = B/b$ und $Y = H/h$. Die Größe des Suchfensters ist durch die Breite B und die Höhe H gegeben. Die Anzahl möglicher um 45° rotierter Merkmale ergibt sich aus

$$X \cdot Y \cdot \left(B + 1 - (b+h) \cdot \frac{X+1}{2} \right) \cdot \left(H + 1 - (b+h) \cdot \frac{Y+1}{2} \right) \quad (4.2)$$

mit $X = B/(b+h-1)$ und $Y = H/(b+h-1)$. In Abhängigkeit der Merkmale gelten die in Tabelle 4.1 aufgeführten Werte für w und h . Die gute Eignung dieser Merkmale wird ersichtlich, wenn Trainingsdaten der Fahrzeugerkennung einer Hauptkomponentenanalyse unterzogen werden. Als deren Ergebnis erhält man für die einzelnen Hauptkomponenten eine, der Größe des Suchfensters entsprechenden, Transformationsmatrix. Abbildung 4.4 zeigt die Visualisierung dieser Transformationsmatrizen und gibt deren Informationsgehalt I_σ an, welcher dem Anteil der Hauptkomponenten an der gesamten in den Trainingsdaten enthaltenen Varianz entspricht und proportional zu den korrespondierenden Eigenwerten der Hauptkomponenten ist. Dunkle Grauwerte stehen für niedrigere und helle Grauwerte für höhere Werte der Transformationsmatrix. Bei Betrachtung der ersten Hauptkomponente wird die Blob-hafte Struktur der Fahrzeuge gut sichtbar, da der innere Bereich dieser Transformationsmatrix deutlich kleinere Werte als die Ränder zeigt. Wandelt man die in Abbildung 4.4 dargestellten Hauptkomponenten in Binärbilder um, zeigen sich große Ähnlichkeiten zu den in Abbildung 4.3 dargestellten Haar-Merkmalen. Besonders

4.3. Isolierte Einzelfahrzeuge

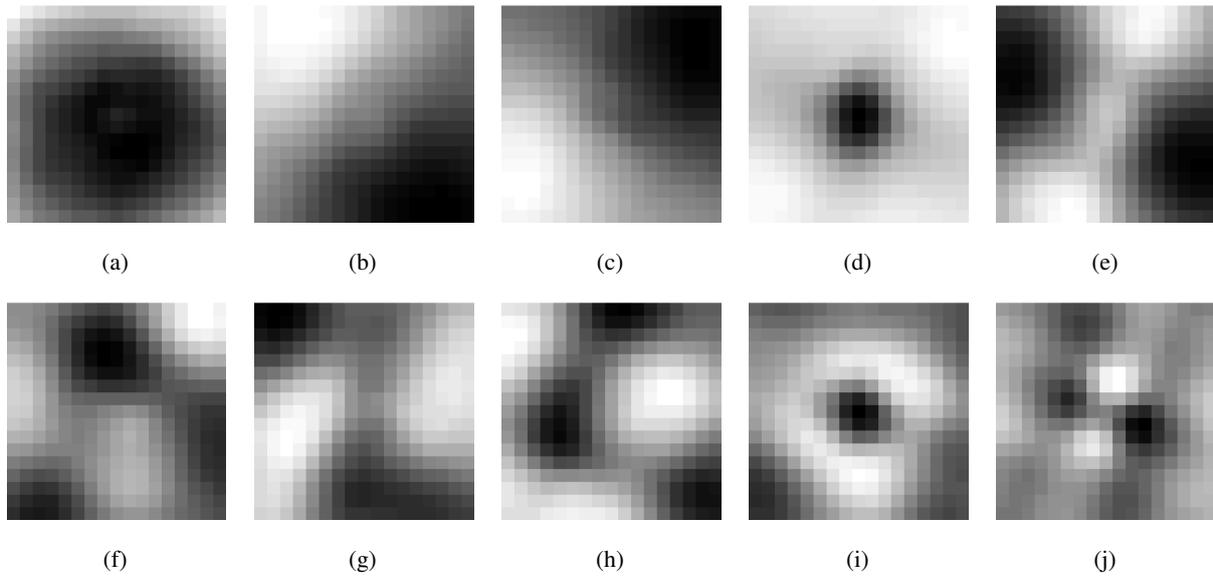


Abbildung 4.4: Hauptkomponenten (HK_i) der Trainingsdaten und deren Informationsgehalt (I_σ). (a) HK_1 , $I_\sigma = 54,0\%$, (b) HK_2 , $I_\sigma = 8,7\%$, (c) HK_3 , $I_\sigma = 8,3\%$, (d) HK_4 , $I_\sigma = 4,3\%$, (e) HK_5 , $I_\sigma = 3,3\%$, (f) HK_6 , $I_\sigma = 1,8\%$, (g) HK_7 , $I_\sigma = 1,7\%$, (h) HK_8 , $I_\sigma = 1,5\%$, (i) HK_9 , $I_\sigma = 1,3\%$, (j) HK_{10} , $I_\sigma = 1,0\%$

deutlich wird dies bei den ersten fünf Hauptkomponenten, welche direkt durch die vorgestellten Haar-Merkmale approximierbar sind, während die weiteren Hauptkomponenten durch Kombination mehrerer Haar-Merkmale dargestellt werden könnten.

Ein weiterer entscheidender Vorteil von Haar-Merkmalen ist deren effiziente Berechnung durch Verwendung von Integralbildern (Viola & Jones, 2001, 2004). Die Integralbilder I haben die gleichen Ausmaße wie die entsprechenden Eingabebilder i . Der Wert eines Punktes $P(x, y)$ im Integralbild berechnet sich durch Summierung aller Grauwerte der Bildpunkte innerhalb des rechteckigen Bereichs, welcher durch den Ursprung und den betrachteten Punkt aufgespannt wird:

$$I(u, v) = \sum_{x=1}^u \sum_{y=1}^v i(x, y) \quad \forall \quad u = 1, \dots, H \quad \wedge \quad v = 1, \dots, B \quad (4.3)$$

H entspricht der Höhe des Bildes und B der Breite. Die Berechnung der Summe aller Grauwerte innerhalb eines Rechtecks, definiert durch die Punkte $P_1(x_1, y_1)$ bis $P_4(x_4, y_4)$ (Abbildung 4.5a), des Eingabebildes kann nach einmaliger Bildung des entsprechenden Integralbildes unabhängig von der Größe des Rechtecks in konstanter Zeit durchgeführt werden:

$$\sum_{x=x_1}^{x_4} \sum_{y=y_1}^{y_4} i(x, y) = I(x_4, y_4) + I(x_1, y_1) - I(x_2, y_2) - I(x_3, y_3) \quad (4.4)$$

Mit Ausnahme des Haar-Merkmals 4 lassen sich alle in Abbildung 4.3 dargestellten Merkmale in beliebiger Größe durch vier Additionen, vier Subtraktionen und maximal einer Multiplikation berechnen, wobei nur 6 bzw. 8 Zugriffe auf das Integralbild notwendig sind. Für Haar-Merkmal 3a (Abbildung 4.5b) erfolgt exemplarisch die Berechnung der Grauwertdifferenz Δg durch:

$$\Delta g = I(x_4, y_4) + I(x_1, y_1) - I(x_2, y_2) - I(x_3, y_3) - 2 \cdot (I(x_8, y_8) + I(x_5, y_5) - I(x_6, y_6) - I(x_7, y_7)) \quad (4.5)$$

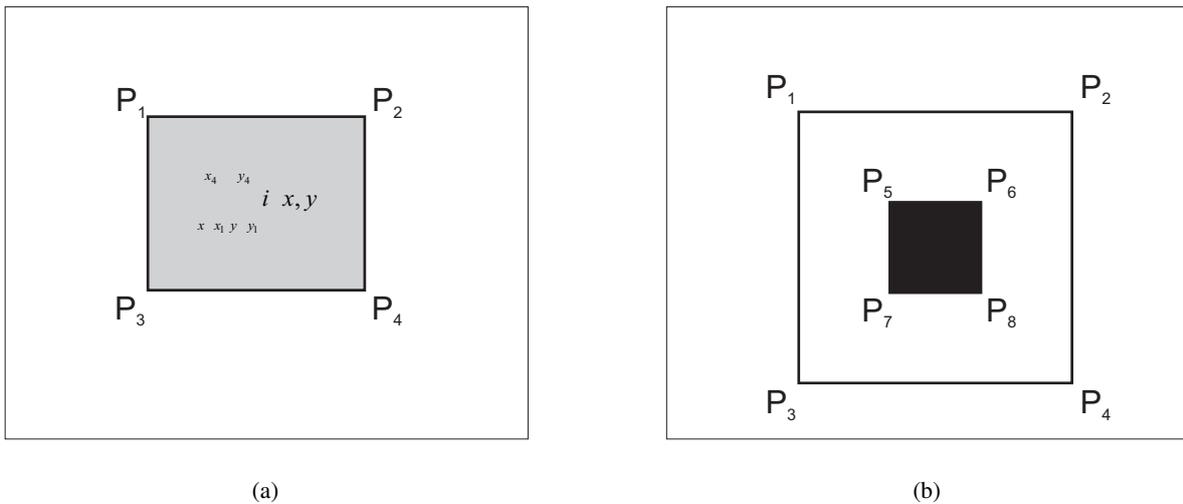


Abbildung 4.5: Verwendung des Integralbildes. (a) Berechnung der Summe aller Grauwerte innerhalb eines Rechtecks, (b) Punkte zur Berechnung der Grauwertdifferenz für das Haar-Merkmal 3a

Die Berechnung der um 45° rotierten Merkmale erfolgt mit einem rotierten Integralbildes I_r (Lienhart et al., 2003).

Bevor die Haar-Merkmale zum Training der Klassifikatoren verwendet werden, erfolgt eine Kontrastanpassung der Merkmalswerte mittels Varianznormalisierung der Grauwerte innerhalb des Suchfensters. Hierfür werden alle Werte des Integralbildes mit Hilfe des Mittelwerts μ und der Standardabweichung σ des Suchfensters transformiert:

$$\bar{I}(x, y) = \frac{I(x, y) - \mu}{\sigma} \quad (4.6)$$

Der Mittelwert für ein Suchfenster kann einfach mit Hilfe des bereits vorliegenden Integralbildes bestimmt werden. Die Bestimmung der Varianz erfordert hingegen die Berechnung eines quadrierten Integralbildes I^2 :

$$I^2(x, y) = \sum_{x=1}^H \sum_{y=1}^B (i(x, y))^2 \quad (4.7)$$

Damit kann die Varianz ebenfalls durch vier Zugriffe auf das quadrierte Integralbild berechnet werden

$$\sigma^2 = \frac{1}{b \cdot h} \cdot (I^2(x_4, y_4) + I^2(x_1, y_1) - I^2(x_2, y_2) - I^2(x_3, y_3) - \mu^2) \quad (4.8)$$

Die Größe des Suchfensters ist durch die Breite b und die Höhe h definiert. Durch Anwendung dieser Vorgehensweise können variierende Beleuchtungssituationen im Bildmaterial besser ausgeglichen werden, da die Merkmale innerhalb aller Suchfenster einen einheitlichen Wertebereich besitzen.

Gentle AdaBoost

Auf Basis der für die Trainingsbeispiele berechneten Merkmale können eine Reihe verschiedener Klassifikationsalgorithmen trainiert werden. Die schnelle Berechnung einzelner Haar-Merkmale ist bei Verwendung von Integralbildern möglich und wurde im vorherigen Abschnitt darlegt. Allerdings wäre die

4.3. Isolierte Einzelfahrzeuge

Verwendung aller möglichen Merkmale, deren hohe Anzahl sich gemäß der Formeln 4.1 und 4.2 ergibt, trotz der zuvor beschriebenen Prozedur äußerst ineffizient. Ein erfolgversprechender Klassifikationsansatz muss daher die zur Verfügung stehende Merkmalsmenge deutlich reduzieren.

Eine Übersicht zur Erzeugung einer geeigneten Untermenge von Merkmalen ist in Niemann (2003, Kap. 3.9.1) aufgeführt. Ein großer Nachteil der dort vorgestellten Verfahren ist jedoch, dass dem Lernverfahren nicht mehr sämtliche Merkmale zur Verfügung stehen. Es besteht das Risiko, dass eine nicht optimale Untermenge zum Training herangezogen wird. In dieser Arbeit wird ein AdaBoost Verfahren (Freund & Schapire, 1997, 1999) verwendet. Die Auswahl der geeignetsten Merkmale erfolgt im Rahmen des Trainings. Eine vorherige Merkmalsreduktion ist nicht notwendig.

AdaBoost (engl. *Adaptive Boosting*) gehört zur Gruppe der Boosting Verfahren. Die ursprüngliche Idee zu diesen Ensemble Lernverfahren wurde in Kearns (1988) vorgestellt. Dort wurde die Frage aufgeworfen, ob eine Reihe schwacher Lerner eine hohe Lernbarkeit impliziert, also einen starken Lerner erzeugen kann. Die Definition eines starken Lernverfahren erfolgt im Rahmen des *Wahrscheinlich Annähernd Richtigen Lernen* Konzeptes (Valiant, 1984). Demnach muss ein Lernalgorithmus in der Lage sein, mit einer angemessenen Anzahl zufällig gewählter Trainingsbeispiele einen Klassifikator zu lernen, der mit einer Wahrscheinlichkeit von $1 - \delta$ auf ungesehenen Daten eine Fehlerrate von maximal ε erzeugt. Für ε und für die Wahrscheinlichkeit einen solchen Klassifikator nicht zu erzeugen δ , gilt $0 < \varepsilon < 0,5$ bzw. $0 < \delta < 0,5$. Dies bedeutet, dass auch für sehr kleine Werte von ε eine Wahrscheinlichkeit über 50 % bestehen muss, diese Fehlerrate einzuhalten. Im Gegensatz hierzu muss ein schwacher Lerner nur ein etwas besseres Ergebnis als ein Zufallsprozess liefern. Die Fehlerrate muss also kleiner als 50 %, aber nicht beliebig klein sein.

In Friedman et al. (2000) wurde gezeigt, dass der AdaBoost Algorithmus eine Minimierung der exponentiellen Kostenfunktion

$$J = \sum_{i=1}^N \exp(-y_i \cdot F(\mathbf{x}_i)) \quad (4.9)$$

für die Trainingsdaten $(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_N, y_N)$ durchführt. Die einzelnen Merkmale der Trainingsdaten liegen als Vektoren \mathbf{x}_i vor. Die Klassenzugehörigkeit ist mit $y_i \in \{-1, +1\}$ kodiert. Da eine direkte Optimierung der Kostenfunktion nur schwer möglich ist, erfolgt die Berechnung des Ensemble Klassifikators $F(\mathbf{x})$ iterativ. Im Rahmen der Iteration findet eine Vergrößerung des Abstandes, der sogenannten Margin $y \cdot F(\mathbf{x})$, der Trainingsdaten von der Trennungsebene statt. AdaBoost zählt daher zu den *Large Margin Classifier*. Die allgemeine Vorgehensweise ist in Abbildung 4.6 dargestellt.

Variationen des AdaBoost Algorithmus zeichnen sich durch unterschiedliche Berechnung des Klassifikationsfehlers der schwachen Lerner und unterschiedliche Berechnung der Neugewichtung aus. Die am häufigsten verwendeten Varianten sind:

- Discrete AdaBoost (Freund & Schapire, 1995, 1997): Es wird der schwache Klassifikator berechnet, der den kleinsten Klassifikationsfehler $\mathbf{e}^t = \sum_{i=1}^N w_i^t \cdot |f^t(\mathbf{x}_i) - y_i|$ mit $f^t \in \{-1, +1\}$ auf den gewichteten Trainingsdaten erzeugt. Es wird also nur eine binäre Entscheidung über die Klassenzugehörigkeit getroffen. Die Aktualisierung der Gewichte erfolgt durch $w_i^{t+1} = w_i^t \cdot \exp(-y_i \cdot \alpha^t \cdot f^t(\mathbf{x}_i))$ mit

- Gegeben sind N Trainingsdaten: $(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_N, y_N)$ mit $y_i \in \{-1, +1\}$
- Initialisierung die Gewichte der Trainingsdaten: $w_i^0 = \frac{1}{N}$
- Iteration über $t = 1, \dots, T$
 1. Normalisierung der Gewichte: $w_i^t = \frac{w_i^{t-1}}{\sum_{i=1}^N w_i^{t-1}}$
 2. Training eines schwachen Klassifikator in Abhängigkeit der gewichteten Trainingsdaten:
 $f^t = f^t(w^t, \mathbf{x}, y)$
 3. Aktualisierung der Gewichte in Abhängigkeit der bisherigen Gewichte, des schwachen Klassifikators und der Klassenzugehörigkeit: $w_i^{t+1} = w_i^{t+1}(w_i^t, f^t, y)$
 4. Aktualisierung des Ensemble Klassifikators: $F^t(\mathbf{x}) = F^{t-1}(\mathbf{x}) + f^t(\mathbf{x})$
- Ausgabe des finalen Ensemble Klassifikator: $F^T(\mathbf{x})$

Abbildung 4.6: Training eines Klassifikators mit AdaBoost

$$\alpha^t = \frac{1}{2} \ln \left(\frac{1 - \varepsilon^t}{\varepsilon^t} \right).$$

- Real AdaBoost (Schapire & Singer, 1999): Es wird wieder der schwache Klassifikator mit dem kleinsten Klassifikationsfehler berechnet. Allerdings wird das reellwertige Ergebnis verwendet $f^t \in [-1, \dots, +1]$. Die Klassenzugehörigkeit ergibt sich aus dem Vorzeichen $\text{sign}(f^t)$, während der Betrag $|f^t|$ als Gütemaß der Klassifikation betrachtet werden kann. Zudem erfolgt eine andere Aktualisierung der Gewichte $w_i^{t+1} = w_i^t \cdot \exp(-y_i \cdot f^{t,j}(\mathbf{x}_i))$, wobei $f^{t,j}$ mit Hilfe der Logit Transformation der a posteriori Wahrscheinlichkeit berechnet wird $f^{t,j}(\mathbf{x}_i) = \frac{1}{2} \ln \frac{P(y=1|\mathbf{x}^j)}{P(y=-1|\mathbf{x}^j)}$. Hierbei stellt \mathbf{x}^j die durch den schwachen Klassifikator gebildete Untermenge j der Trainingsdaten dar. Basiert der schwache Klassifikator auf einfacher Schwellwertbildung, wird die Trainingsmenge in zwei Untermengen $j = \{1, 2\}$ geteilt. Für neuronale Netze als schwachem Klassifikator würde die Anzahl der Untermengen gleich der Anzahl der Einheiten der Ausgabeschicht sein. Die a posteriori Wahrscheinlichkeiten für eine Untermenge können durch Summierung und Normalisierung der entsprechenden Gewichte der Untermenge gebildet werden:

$$P(y = 1|\mathbf{x}^j) = \frac{\sum_{i=1 \forall \mathbf{x}_i \in \mathbf{x}^j \wedge y_i=1}^N w_i^t}{\sum_{i=1 \forall \mathbf{x}_i \in \mathbf{x}^j}^N w_i^t} \quad \text{und} \quad P(y = -1|\mathbf{x}^j) = \frac{\sum_{i=1 \forall \mathbf{x}_i \in \mathbf{x}^j \wedge y_i=-1}^N w_i^t}{\sum_{i=1 \forall \mathbf{x}_i \in \mathbf{x}^j}^N w_i^t} \quad (4.10)$$

- Gentle AdaBoost (Friedman et al., 2000): Auch bei dieser AdaBoost Variante erzeugt der schwache Klassifikator ein reellwertiges Ergebnis. Allerdings wird zur Bildung des schwachen Lernalters der quadratische Klassifikationsfehler $\varepsilon^t = \sum_{i=1}^N w_i^t (y_i - f^t(\mathbf{x}_i))^2$ minimiert. Die Aktualisierung der Gewichte erfolgt wie bei Real AdaBoost, wobei $f^{t,j}(\mathbf{x}_i)$ definiert ist durch $f^{t,j}(\mathbf{x}_i) = P(y = 1|\mathbf{x}^j) - P(y = -1|\mathbf{x}^j)$. Die a posteriori Wahrscheinlichkeiten berechnen sich gemäß der Formel 4.10.

4.3. Isolierte Einzelfahrzeuge

Durch Minimierung des quadratischen Klassifikationsfehlers und Aktualisierung der Gewichte mit Hilfe der Differenz der a posteriori Wahrscheinlichkeit wird eine Newton-Iteration entlang der monoton fallenden Kostenfunktion 4.9 durchgeführt (Friedman et al., 2000). Vergleicht man $f^{t,j}$ von Real und Gentle AdaBoost ist erkennbar, dass die Logit-Transformation besonders große Werte annimmt, wenn $P(y = 1|\mathbf{x}^j)$ gegen 1 strebt. Falsch klassifizierte Beispiele erhalten in diesem Fall ein deutlich höheres Gewicht als beim Gentle AdaBoost Algorithmus, welcher solchen Ausreißern weniger Beachtung schenkt. Dies ist der Grund für die Bezeichnung *gentle*, da die Neugewichtung *behutsamer* erfolgt.

Wegen der vorsichtigeren Gewichtung falsch klassifizierter Trainingsdaten wird in dieser Arbeit der Gentle AdaBoost Algorithmus verwendet. Es hat sich gezeigt, dass eine fehlerfreie Erfassung von Trainingsdaten aus Satellitenbildern kaum möglich ist, da Fahrzeuge nicht immer eindeutig identifiziert werden können. Zudem sind in den Trainingsdaten Beispiele enthalten, welche nicht repräsentativ für die Fahrzeugklasse sind, da sie einen zu geringen Kontrast zur Umgebung aufweisen und die Blob-hafte Struktur nicht hervortritt. Discrete und Real AdaBoost könnten diesen falschen bzw. schlechten Beispielen im Verlauf der Iteration, eine immer höhere Gewichtung geben und so die Generalisierungsfähigkeit des finalen Ensemble Klassifikators einschränken.

Die gute Generalisierungsfähigkeit von Gentle AdaBoost ist in Abbildung 4.7a erkennbar. Es wurde ein Klassifikationsbaum mit drei Knoten als schwacher Klassifikator verwendet. Nach weniger als 50 Iterationen ist der Fehler auf den Trainingsdaten minimal. Obwohl man vermuten könnte, dass weitere Iterationen zu einer Überanpassung des Klassifikators führen und damit der Fehler auf dem Testdatensatz ansteigt, ist dieser Effekt nicht zu beobachten. Stattdessen sinkt die Fehlerrate für den Testdatensatz im Verlauf der Iterationen. Auch nach korrekter Klassifikation aller Trainingsdaten wird die Margin $y \cdot F(\mathbf{x})$ weiterhin maximiert. Tabelle 4.2 verdeutlicht diesen Zusammenhang. Der Trainingsfehler beträgt nach 50 Iteration 0,0 % und der Testfehler noch 2,8 %. Im weiteren Verlauf bleibt der Trainingsfehler bei 0,0 % und der Testfehler sinkt nach 300 Iterationen auf 1,3 %. Während nach 5 Iterationen noch 10,5 % der Trainingsdaten eine Margin kleiner als 0,25 haben, sinkt deren Anzahl im Verlauf des Trainings auf 0,0 %. Die minimale Margin beträgt nach 300 Iterationen 0,31. Der Abstand der Trainingsdaten zur Trennebene wächst stetig. Sehr gut erkennbar wird dies auch bei Betrachtung der kumulierten Margin-Verteilung in Abbildung 4.7b. Die kontinuierliche Vergrößerung des Abstands der Trainingsdaten zur Trennungsebene wird deutlich, da sich die kumulierte Verteilung der Margin im Verlauf der Iterationen in den positiven Bereich verschiebt. Bei sehr stark fehlerbehafteter Trainingsdaten - Rauschen von über 10 % - neigt auch AdaBoost zur Überanpassung (Dietterich, 2000; Rätsch et al., 2001). Dieses Verhalten kann durch Validierung mit einem von den Trainingsdaten unabhängigen Testdatensatz erkannt werden.

Schwacher Klassifikator

AdaBoost ist ein Meta-Lernverfahren und benötigt ein Lernalgorithmus, der den geforderten Klassifikationsfehler minimiert. Hierfür sind Algorithmen geeignet, welche gewichtete Trainingsdaten verarbeiten

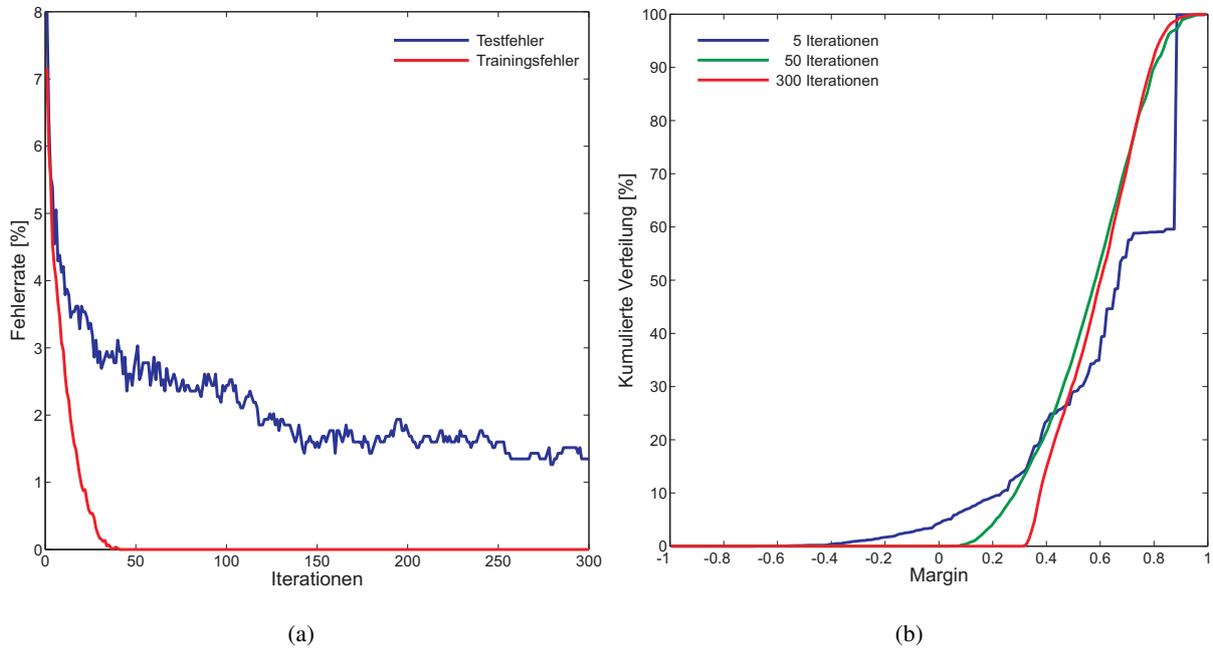


Abbildung 4.7: Trainingsverlauf des Gentle AdaBoost Algorithmus bei Verwendung eines Klassifikationsbaumes mit drei Knoten. (a) Trainingsfehler (rot) und Testfehler (blau), (b) Kumulierte Margin-Verteilung nach 5 Iterationen (blau), 50 Iterationen (grün) und 300 Iterationen (rot) des Ensemble Trainings

	Iterationen					
	5	25	50	100	200	300
Trainingsfehler	4,2 %	0,5 %	0,0 %	0,0 %	0,0 %	0,0 %
Testfehler	4,5 %	3,4 %	2,9 %	2,4 %	1,8 %	1,3 %
$y \cdot F(\mathbf{x}) < 0,25$	10,5 %	10,2 %	7,1 %	1,7 %	0,1 %	0,0 %
$\min(y \cdot F(\mathbf{x}))$	-0,67	-0,11	0,04	0,20	0,24	0,31

Tabelle 4.2: Verlauf des Trainings- und Testfehlers sowie der Margin für die Trainingsdaten

4.3. Isolierte Einzelfahrzeuge

können. Häufig finden Schwellwertklassifikatoren, Klassifikationsbäume oder neuronale Netze Anwendung.

In dieser Arbeit werden Klassifikationsbäume (Breiman et al., 1984) verwendet, da diese Vorteile gegenüber anderen schwachen Lernern haben. Die einzelnen Knoten der Klassifikationsbäume stellen Schwellwerte auf jeweils ein Merkmal dar. Daher müssen während der Anwendung der Klassifikatoren nur die Merkmale berechnet werden, welche in den Knoten der Klassifikationsbäume vorkommen. Dies führt zu einer deutlichen Reduktion der Merkmalsanzahl. Werden beispielsweise Bäume mit 5 Knoten über 100 Iterationen trainiert, müssen während der Klassifikation maximal 500 Merkmale berechnet werden. Da einige besonders diskriminative Merkmale im Verlauf des iterativen Trainings mehrmals ausgewählt werden, fällt die Merkmalsanzahl oft geringer aus. Einige Autoren geben aber zu bedenken, dass ein einzelner schwacher Lerner nur eine begrenzte Klassifikationsleistung haben sollte, da andernfalls die Generalisierungsfähigkeit des Ensemble Klassifikators vom schwachen Lerner beeinträchtigt wird (Friedman et al., 2000; Lienhart & Maydt, 2002; Lienhart et al., 2003). Die Tiefe von Klassifikationsbäumen sollte beschränkt werden, da sich diese ansonsten zu stark an die Trainingsdaten anpassen. Eine günstig gewählte Knotenanzahl steigert hingegen die Klassifikationsleistung des Ensemble Klassifikators signifikant gegenüber der Verwendung von einfachen Schwellwertklassifikatoren. Zur Verdeutlichung findet in Abbildung 4.8 eine Gegenüberstellung der Trainings- und Testfehler bei variierender Knotenanzahl statt. Aus Abbildung 4.8a wird deutlich, dass eine höhere Knotenanzahl (4 und 7 Knoten) eine bessere Anpassung des Klassifikators an die Trainingsdaten zulässt, da in diesen Fällen weniger als 50 Iterationen benötigt werden, um den Trainingsfehler auf sein Minimum zu reduzieren. Beim einfachen Schwellwertklassifikator (1 Knoten) sind hierfür circa 100 Iterationen notwendig. Bei Verwendung von Klassifikationsbäumen können sowohl Abhängigkeiten zwischen Merkmalen ausgenutzt werden, als auch multimodale und sich überlappende Verteilungen der Merkmalswerte besser getrennt werden.

Die Validierung des Ensemble Klassifikators in Abbildung 4.8b zeigt, dass keine beliebig hohe Anzahl von Knoten verwendet werden sollte. Im dargestellten Fall weist sogar ein einfacher Schwellwertklassifikator eine geringere Fehlerrate auf ungesehenen Daten aus, als ein Klassifikator mit einer offensichtlich zu hoch gewählten Knotenanzahl. Im dargestellten Beispiel erreicht die beste Generalisierungsfähigkeit ein Ensemble Klassifikator, welcher Klassifikationsbäume mit 4 Knoten verwendet. Eine umfangreiche Evaluierung zur Wahl der optimalen Konfiguration des schwachen Lerners findet in Abschnitt 6.1.3 statt.

Gruppierung Klassifikationsergebnisse

Das Ergebnis der Klassifikation aller Bildpunkte ist ein Konfidenzbild. Das Vorzeichen des Konfidenzwertes entspricht der Klassenzugehörigkeit des Bildpunktes. Der Betrag dient als Gütemaß dieser Entscheidung. Abbildung 4.9b zeigt das Konfidenzbild für den QuickBird Ausschnitt in Abbildung 4.9a. Die Klassenzugehörigkeit ist durch die Farbe und der Betrag als entsprechende Farbintensität dargestellt. Die Gruppierung benachbarter Bildpunkte gleicher Klassenzugehörigkeit erfolgt durch Bestimmung der Nulldurchgänge des Konfidenzbildes (Abbildung 4.9c). Da ein Teil der resultierenden Bildregionen keine Fahrzeuge sind, ist eine abschließende Klassifikation der Bildsegmente notwendig.

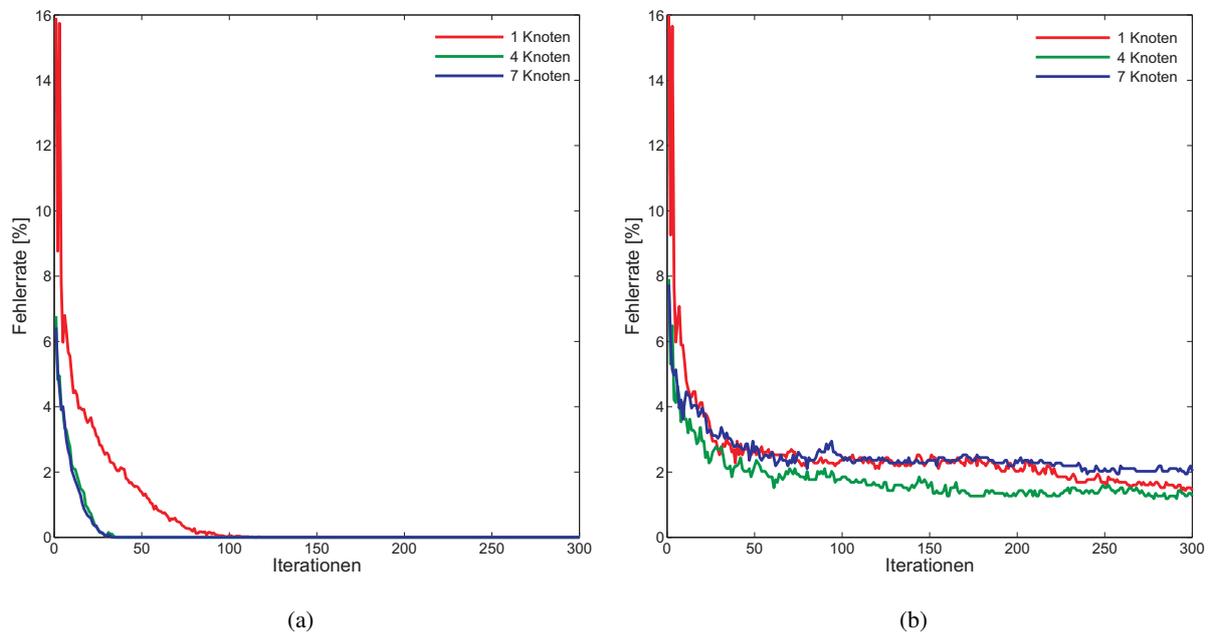


Abbildung 4.8: Fehlerrate bei unterschiedlicher Knotenanzahl eines Klassifikationsbaumes. (a) Trainingsfehler, (b) Testfehler

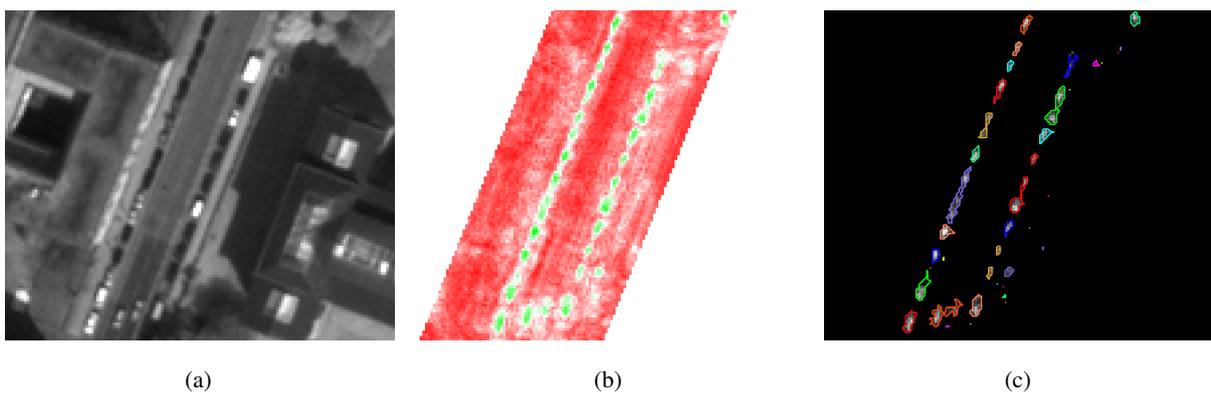


Abbildung 4.9: Anwendung des Boosting Algorithmus. (a) QuickBird Ausschnitt, (b) Konfidenzbild (grün - Fahrzeugklasse, weiß - nicht klassifiziert, rot - Fahrzeuggegenklasse), (c) Nulldurchgänge des Konfidenzbildes

4.3. Isolierte Einzelfahrzeuge

4.3.2 Klassifikation der gruppierten Bildregionen

Zur Klassifikation der gruppierten Bildregionen werden sowohl geometrische als auch radiometrische Eigenschaften der darin enthaltenen Bildpunkte verwendet. Die Berechnung der radiometrischen Eigenschaften aus der Grauwertematrix erfolgt für das Konfidenzbild, im panchromatischen Kanal und den drei multispektralen Kanälen. Durch die Verwendung der Farbinformation können in an Straßen angrenzenden Gebieten Fehldetektionen verringert werden. Die verwendeten Merkmale sind in Tabelle 4.3 aufgeführt.

Geometriemerkmale (aus Steger et al. (2008))	Radiometriemerkmale (aus Haralick et al. (1973))
Halbachsen der kleinsten umschließenden Ellipse	Mittelwert und Standardabweichung
Fläche	Entropie
Rundheit	Homogenität
Rechteckigkeit	Energie
Kompaktheit	Korrelation
Exzentrizität	Kontrast

Tabelle 4.3: Geometrische und radiometrische Eigenschaften zur Klassifikation der Bildsegmente

Die Klassifikation der Bildsegmente wird mit einer *Support Vector Machine* (SVM) (Vapnik, 1982; Burges, 1998; Cristianini & Shawe-Taylor, 2000; Schölkopf & Smola, 2001) durchgeführt. Im Gegensatz zum AdaBoost Verfahren können SVMs besser mit nicht linear trennbaren und fehlerhaften Trainingsdaten umgehen.

Lineare Support Vector Machines (SVM)

Die SVM konstruiert zur Klassifikation der Trainingsdaten eine Trennungsebene im Merkmalsraum. Die Trainingsdaten liegen als Merkmalsvektoren $\mathbf{x}_i, i = 1, \dots, N$ vor. N entspricht der Anzahl der Trainingsdaten. Die Klassenzugehörigkeit ist durch $y_i \in \{-1, +1\}$ definiert, da SVM nur binäre Entscheidungen treffen können. Sind die Trainingsdaten vollständig linear trennbar, ist die Trennungsebene (siehe Abbildung 4.10a) in Abhängigkeit ihres Normalenvektors \mathbf{w} und des senkrechten Abstands b zum Ursprung des Merkmalsraums gegeben:

$$f(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + b = 0 \quad (4.11)$$

Unendlich viele Trennungsebenen erfüllen diese Bedingung. Daher wird die Trennungsebene mit der größten Margin m bevorzugt. Die Trainingsdaten, welche der Trennungsebene am nächsten liegen, sind die Support Vektoren. Ihr Abstand zur Trennungsebene wird als 1 definiert, wodurch sich die zu maximierende Margin ergibt (siehe Abbildung 4.10a):

$$m = \frac{1}{\|\mathbf{w}\|} + \frac{1}{\|\mathbf{w}\|} = \frac{2}{\|\mathbf{w}\|} \quad (4.12)$$

Während des Trainings ist folgendes Optimierungsproblem, in Abhängigkeit von den Parametern der

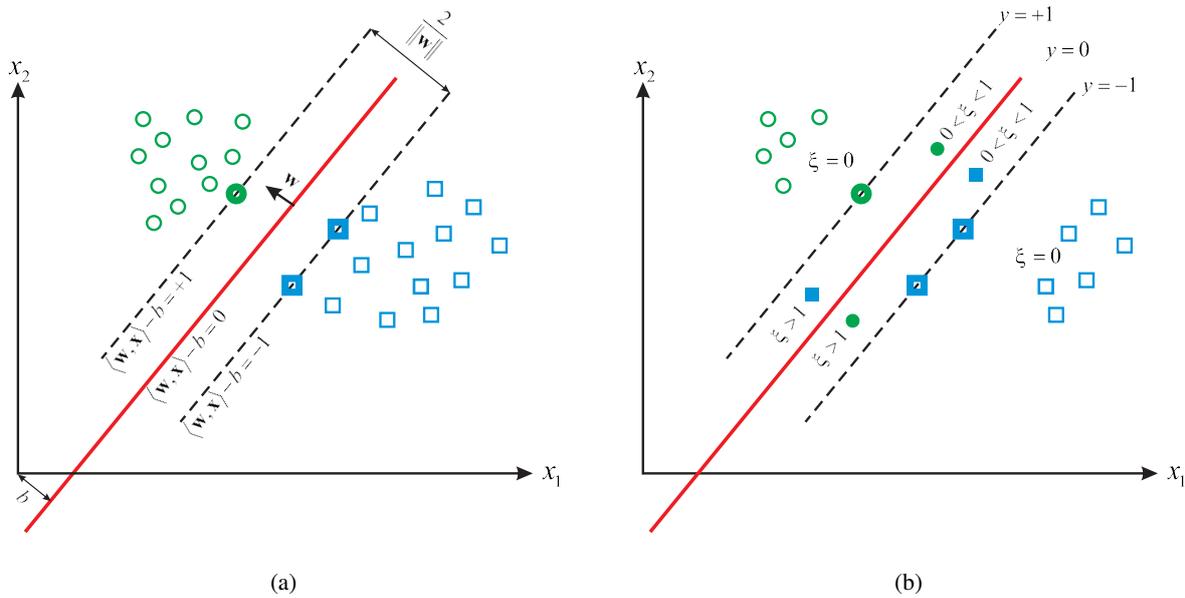


Abbildung 4.10: Trennungsebene und Schlupfvariablen der SVM. (a) Trennungsebene (rot) und Margin bei linearer Trennbarkeit, Support Vektoren der jeweiligen Klassen fett hervorgehoben, (b) Slackvariablen der SVM, Support Vektoren (fett), falsch klassifizierte Beispiele (ausgefüllt)

Trennungsebene zu lösen:

$$\min_{\mathbf{w}, b} \left(J(\mathbf{w}, b) = \frac{1}{2} \|\mathbf{w}\|^2 \right) \text{ mit } y_i (\mathbf{w}^T \mathbf{x}_i - b) \geq 1, \forall i = 1, \dots, N \quad (4.13)$$

Die Minimierung der Norm in 4.13 führt zur geforderten Maximierung von m aus Gleichung 4.12. Da eine gleichzeitige Lösung für die beiden gesuchten Parameter \mathbf{w} und b nicht durch effiziente Verfahren möglich ist, wird das Problem mit Hilfe von Lagrange-Multiplikatoren α_i umformuliert:

$$L(\mathbf{w}, b, \alpha) = \frac{1}{2} \mathbf{w}^T \mathbf{w} - \sum_{i=1}^N \alpha_i (y_i (\mathbf{w}^T \mathbf{x}_i + b) - 1) \quad (4.14)$$

Die zur Optimalität der Lösung notwendigen Karush-Kuhn-Tucker-Bedingungen lauten:

$$\frac{\partial L(\mathbf{w}, b, \alpha)}{\partial \mathbf{w}} = 0 \quad (4.15)$$

$$\frac{\partial L(\mathbf{w}, b, \alpha)}{\partial b} = 0 \quad (4.16)$$

$$\alpha_i \geq 0, \forall i = 1, \dots, N \quad (4.17)$$

$$\alpha_i (y_i (\mathbf{w}^T \mathbf{x}_i + b) - 1) = 0, \forall i = 1, \dots, N \quad (4.18)$$

Durch Bildung der partiellen Ableitungen 4.15 und 4.16 von 4.14 ergeben sich die Bedingungen:

$$\mathbf{w} = \sum_{i=1}^N \alpha_i y_i \mathbf{x}_i \quad (4.19)$$

4.3. Isolierte Einzelfahrzeuge

$$\sum_{i=1}^N \alpha_i y_i = 0 \quad (4.20)$$

Werden diese partiellen Ableitungen wieder in 4.14 eingesetzt, erhält man die während des Trainings zu lösende duale Form des Optimierungsproblems:

$$\max_{\alpha} \left(L(\alpha) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j (\mathbf{x}_i^T \mathbf{x}_j) \right) \text{ unter Einhaltung von 4.20 und 4.17} \quad (4.21)$$

Da zur Lösung von Gleichung 4.21 die Ungleichung 4.17 eingehalten werden muss, finden Verfahren der quadratischen Programmierung (Smola, 1998; Boyd & Vandenberghe, 2009) oder der *Sequential Minimal Optimization* Algorithmus (Platt, 1998) Anwendung. Eine Übersicht zur Wahl eines geeigneten Lösungsverfahrens in Abhängigkeit der vorliegenden Trainingsdaten geben Schölkopf & Smola (2001, Kap. 10). Für den Großteil der Trainingsdaten nimmt α_i den Wert 0 an. Bei Trainingsdaten mit $\alpha_i > 0$ handelt es sich um die gesuchten Support Vektoren, aus denen sich die gesuchten Parameter \mathbf{w} und b des ursprünglichen Optimierungsproblems aus Gleichung 4.13 berechnen lassen. Hierfür ist \mathbf{w} bereits als Linearkombination in Gleichung 4.19 gegeben. Die Berechnung von b kann durch Umstellung von Gleichung 4.18 für einen beliebigen Support Vektor erfolgen. In der Praxis wird meist das arithmetische Mittel über alle Support Vektoren gebildet. Die Klassifikation neuer Merkmalsvektoren \mathbf{x} findet mit diesen Parametern durch

$$f(\mathbf{x}) = \text{sign}(\mathbf{w}^T \mathbf{x} + b) \quad (4.22)$$

statt.

Soft Margins für SVM

In diesem Abschnitt soll die Erweiterung der SVM für nicht linear trennbare Daten vorgestellt werden. Diese Erweiterung ist sinnvoll, da nicht immer absolut fehlerfreie und die gesuchte Klasse gut repräsentierende Trainingsdaten vorausgesetzt werden können. Für jedes Trainingsbeispiel wird eine Schlupfvariable ξ_i eingeführt. Die Schlupfvariablen können drei unterschiedliche Werte annehmen (siehe Abbildung 4.10b):

- $\xi_i = 0$: für korrekt klassifizierbare Trainingsdaten
- $0 < \xi_i \leq 1$: für korrekt klassifizierbare Trainingsdaten, welche zwischen Support Vektoren und Trennungsebene liegen
- $\xi_i > 1$: für falsch klassifizierte Trainingsdaten

In Anlehnung an Gleichung 4.13 ist das neue Optimierungsproblem gegeben durch:

$$\min_{\mathbf{w}, b} \left(J(\mathbf{w}, b) = \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^N I(\xi_i) \right) \text{ mit } I(\xi_i) = \begin{cases} 1, \forall \xi_i > 0 \\ 0, \forall \xi_i = 0 \end{cases} \quad (4.23)$$

Die Konstante C spiegelt den Kompromiss zwischen dem Erhalt eines kleinen Trainingsfehlers auf der einen Seite und einer möglichst großen Margin auf der anderen Seite wider. Bei der Überführung in die duale Form entfallen die Schlupfvariablen und es ändert sich ausschließlich eine der einzuhaltenden Bedingungen. So muss Gleichung 4.21 unter Einhaltung der Bedingung 4.20 und mit

$$0 \leq \alpha_i \leq C, \forall i = 1, \dots, N \quad (4.24)$$

gelöst werden. Der Parameter C wenig intuitiv ist und kann nur schwer im Rahmen des Training festgelegt werden. Daher wurde C in der Arbeit von Schölkopf et al. (2000) durch den Parameter ν ersetzt, was zur folgenden dualen Form führt:

$$\max_{\alpha} \left(L(\alpha) = -\frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j (\mathbf{x}_i^T \mathbf{x}_j) \right) \quad (4.25)$$

Hierfür gelten die Bedingungen:

$$\begin{aligned} 0 &\leq \alpha_i \leq \frac{1}{N}, \\ \sum_{i=1}^N y_i \alpha_i &= 0, \\ \sum_{i=1}^N \alpha_i &\geq \nu \end{aligned} \quad (4.26)$$

Der Vorteil dieser Darstellung ist, dass ν dem prozentualen Anteil der Trainingsbeispiele entspricht, welche falsch klassifiziert werden. Hierdurch ist die Festlegung von ν vor dem Trainings leichter möglich.

Nichtlineare SVM

Mit der bisherigen Darstellung der SVM kann nur eine lineare Trennung der Trainingsdaten vorgenommen werden. Zur Erweiterung auf nichtlineare Klassifikationsprobleme findet eine Abbildung der Merkmalsvektoren \mathbf{x}_i in einen höher dimensional Merkmalsraum statt, in welchem eine lineare Trennbarkeit möglich ist. Abbildung 4.11 zeigt eine Überführung exemplarisch. Die Daten der roten und blauen Klasse können im Ausgangsmerkmalsraum aus Abbildung 4.11a nur durch einen Kreis und damit nicht linear voneinander getrennt werden. Die Merkmale werden deshalb durch ein Polynom 2. Grades von zweidimensionalen in einen dreidimensionalen Merkmalsraum transformiert (siehe Abbildung 4.11b). Die transformierten Merkmalsvektoren $\Phi(\mathbf{x}_i)$ sind anschließend durch die in Abbildung 4.11c dargestellte Ebene linear trennbar. Eine explizite Transformation ist bei diesem Beispiel möglich. Das verwendete Polynom 2. Grades fügt nur eine zusätzliche Dimension zum ursprünglichen Merkmalsraum hinzu. Weil die Dimensionalität aber exponentiell mit dem Grad des Polynoms steigt, ist eine explizite Berechnung der transformierten Merkmale für Polynome höheren Grades kaum möglich. In Cortes & Vapnik (1995) wurde daher die Verwendung des *Kernel Tricks* (Aizerman et al., 1964) für SVM eingeführt. Der Kernel ist das Skalarprodukt zweier transformierter Merkmalsvektoren:

$$k(\mathbf{x}_i, \mathbf{x}_j) = \Phi(\mathbf{x}_i)^T \Phi(\mathbf{x}_j) \quad (4.27)$$

Für bestimmte Kernelfunktionen ist es nicht notwendig, die ursprünglichen Merkmalsvektoren explizit in einen höheren Merkmalsraum zu transformieren. Stattdessen wird nur das Ergebnis des Skalarprodukts

4.3. Isolierte Einzelfahrzeuge

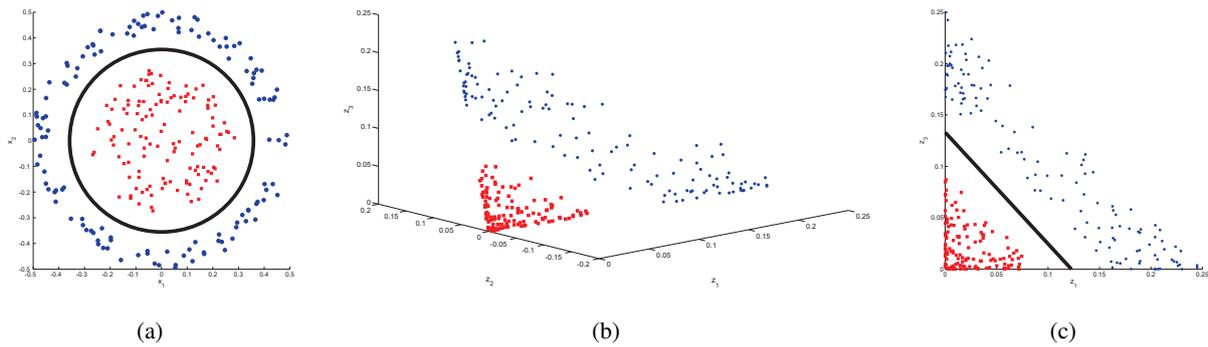


Abbildung 4.11: Beispiel einer Merkmalsraumtransformation für ein Zwei-Klassen Problem. (a) Merkmalswerte im zweidimensionalen Merkmalsraum, (b) Schrägansicht der Merkmalswerte im transformierten Merkmalsraum, (c) Seitenansicht der Merkmalswerte im transformierten Merkmalsraum und der als Linie dargestellten linearen Trennungsebene

transformiert. Für ein Polynom n -ten Grades erfolgt die Berechnung des Kernels durch:

$$k(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i^T \mathbf{x}_j)^n \quad (4.28)$$

Ein weiterer und in dieser Arbeit verwendeter Kernel sind sogenannte gaußsche radiale Basisfunktionen (RBF-Kernel):

$$k(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma^2}\right) \quad (4.29)$$

Der Kernelparameter σ muss vor dem Training definiert werden und kann durch Kreuzvalidierung bestimmt werden. Setzt man den Kernel in Gleichung 4.25 ein, lautet das zu lösende Optimierungsproblem:

$$\max_{\alpha} \left(L(\alpha) = -\frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j k(\mathbf{x}_i, \mathbf{x}_j) \right) \quad (4.30)$$

Die Bedingungen aus Gleichung 4.26 gelten weiterhin.

Mehrklassenprobleme

Eine strikte Unterteilung in Fahrzeugklasse und Gegenklasse führt im vorliegenden Fall zu einer signifikanten Anzahl falscher und fehlender Fahrzeuge. Es können Mehrfachdetektionen für ein Fahrzeug auftreten, wenn mehrere Bildsegmente innerhalb des Fahrzeuges auftreten. Daher wird eine zusätzliche Klasse mit „wahrscheinlichen/unsicheren“ Fahrzeugen eingeführt. Die finale Zuweisung der hierin enthaltenen Hypothesen geschieht durch ein Hystereseverfahren. Es wird überprüft, ob im Umfeld einer unsicheren Hypothese eine sichere Hypothese gefunden wird. Trifft dies zu, wird das unsichere Fahrzeug verworfen. Andernfalls wird die Hypothese als Fahrzeug bestätigt.

Die Erfassung von Trainingsdaten für die SVM erfolgt automatisch. Zunächst wird mit den manuell erfassten Trainingsdaten der AdaBoost Klassifikator gelernt. Diese Trainingsdaten werden anschließend mit dem AdaBoost Klassifikator klassifiziert. Die resultierenden Konfidenzbilder werden nach Abschnitt

4.3.1 segmentiert. Treten innerhalb eines Fahrzeuges mehrere Bildregionen auf, werden alle Bildregionen der Menge unsicherer Fahrzeuge hinzugefügt, die dem Fahrzeugmittelpunkt nicht am nächsten liegen. Bildregionen außerhalb eines Fahrzeuges bilden den Trainingsdatensatz der Fahrzeuggegenklasse. Die verbleibenden Bildregionen dienen als Beispiele der Fahrzeugklasse.

Da SVM nur binäre Entscheidungen treffen, kann zu deren Erweiterung eine der beiden folgenden Trainingsstrategien angewandt werden, welche das Multiklassenproblem in Zwei-Klassen-Probleme umwandelt:

one-versus-all : Für jede Klasse wird ein Klassifikator erzeugt, der die betrachtete Klasse gegenüber der Menge aller verbleibenden Klassen trennt. Die Anzahl erzeugter Klassifikatoren entspricht der Klassenanzahl. Ein Merkmalsvektor wird der Klasse zugeordnet, bei welcher das Maximum des absoluten Betrages der Klassifikationsgleichung (Formel 4.22) erreicht wird.

one-versus-one : Es wird eine gegenseitige Abgrenzung aller Klassen untereinander durchgeführt. Die Anzahl aufzustellender Klassifikatoren steigt quadratisch mit der Anzahl der Klassen. So sind im Fall von n -Klassen genau $n(n-1)/2$ Klassifikatoren aufzustellen. Bei einer hohen Anzahl zu unterscheidender Klassen ist diese Vorgehensweise nicht empfehlenswert, da sowohl das Training als auch die spätere Anwendung einer sehr hohen Anzahl von Klassifikatoren mehr Zeit in Anspruch nimmt. Ein Merkmalsvektor wird der Klasse zugeschrieben, in welche er am häufigsten fällt.

In dieser Arbeit wird die one-versus-one Strategie benutzt, da diese bei drei Klassen eine höhere Geschwindigkeit und kleinere Klassifikationsfehler erwarten lässt. Die Überlegenheit der one-versus-one Strategie konnte für verschiedene Problemstellungen in Hsu & Lin (2002) beobachtet werden. Im Vergleich zur one-versus-all Strategie wird zwar ein Klassifikator mehr benötigt, allerdings fallen auch die zu unterscheidenden Klassen kleiner aus, wodurch weniger Support Vektoren zu deren Trennung benötigt werden.

4.4 Einzelfahrzeuge aus Autoreihen

Wesentliches Merkmal des Modells für Fahrzeugreihen ist die Variation der Breite und des Kontrastes zur Umgebung. Der implementierte Ansatz folgt einem „grob nach fein“ Schema. Dafür müssen zunächst Hypothesen für Fahrzeugreihen extrahiert werden. Diese werden auf Übereinstimmung zum Fahrzeugreihenmodell geprüft und anschließend auf Basis von Kollinearität fusioniert. Entlang der erhaltenen Linien wird ein Grauwertprofil berechnet, das zur Erkennung einzelner Fahrzeuge innerhalb einer Autoreihe verwendet wird.

4.4.1 Linienextraktion

Die Extraktion von Linien erfolgt nach dem Ansatz von Steger (1998). Zunächst werden die ersten und zweiten partiellen Ableitungen des Bildes durch Faltung mit entsprechenden Ableitungen der Filtermaske einer gaußschen Glättung berechnet. Damit wird die Richtung der maximalen Krümmung für

4.4. Einzelfahrzeuge aus Autoreihen

jeden Bildpunkt bestimmt. Durch Approximation mit einer Parabel wird überprüft, welche Bildpunkte entlang dieser Richtung ein Maximum (helle Linien) oder Minimum (dunkle Linien) aufweisen. Weisen Bildpunkte ein solches Extremum auf, werden diese zur Liste potentieller Linienpunkte hinzugefügt. Da die gesuchten Linien senkrecht zur Richtung der maximalen Krümmung verlaufen, sind neben den subpixelgenauen Positionen auch deren Orientierungen bekannt. Zur Bildung von Pixelketten werden die Linienpunkte ausgewählt, deren Krümmung über einem gegebenen oberen Schwellwert liegen. Ausgehend von diesen Startpunkten werden benachbarte Linienpunkte mit korrespondierender Richtung der Pixelkette hinzugefügt, sofern deren Krümmung über einem gegebenen unteren Schwellwert liegt. Linienpunkte deren Krümmung größer als der untere Schwellwert ist, aber nicht mit einer Pixelkette verbunden werden konnten, finden keine weitere Beachtung. Die Linienextraktion ist von drei Parametern abhängig, die wie folgt berechnet werden:

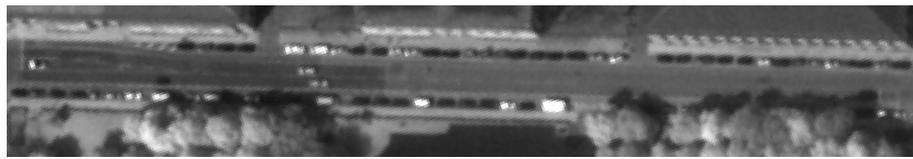
- Glättungsfaktor der gaußschen Filtermaske $\sigma = \frac{w}{\sqrt{3}}$
- Unterer Schwellwert für Krümmung $t_u = 2 \cdot c_u \cdot \frac{w}{\sqrt{2 \cdot \pi \cdot \sigma^3}} e^{-\frac{1}{2} \left(\frac{w}{\sigma}\right)^2}$
- Oberer Schwellwert für Krümmung $t_o = 2 \cdot c_o \cdot \frac{w}{\sqrt{2 \cdot \pi \cdot \sigma^3}} e^{-\frac{1}{2} \left(\frac{w}{\sigma}\right)^2}$

Die halbe Fahrzeugbreite w entspricht $w = l_2/2$, während c_u der untere und c_o der obere Schwellwert für den Kontrast der Fahrzeugreihen zur Umgebung ist. Die Initialisierung dieser Parameter wird in Abschnitt 6.1.2 beschrieben.

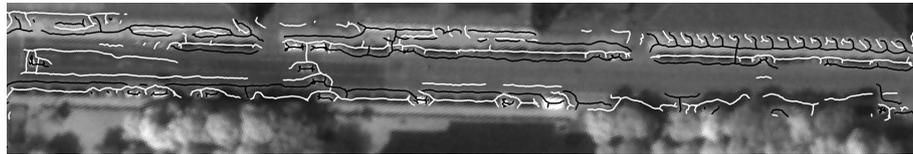
Ein Ergebnis der Linienextraktion ist in Abbildung 4.12b dargestellt. Obwohl die vom Modell geforderte Breite zur Berechnung des Glättungsfaktors σ verwendet wurde, weisen einige Linien eine zu hohe oder zu geringe Breite auf. Diese sind in Abbildung 4.12c rot dargestellt. Die Linien werden außerdem auf ausreichende Länge überprüft. Rot dargestellte Linien in Abbildung 4.12d erfüllen dieses Kriterium nicht. Das letzte geforderte Kriterium für Fahrzeugreihen ist deren Übereinstimmung zur Straßenrichtung. Viele extrahierte Linien verlaufen nicht annähernd parallel zur Straßenrichtung (rote Linien in Abbildung 4.12e). Nur Linien welche alle Kriterien erfüllen, werden beibehalten. Die Fusion der verbleibenden Linien erfolgt mit den Bedingungen aus Abschnitt 3.2. Abbildung 4.12f zeigt das Ergebnis nach Überprüfung des Modells für Fahrzeugreihen für helle und dunkle Linien. Das Ergebnis der Fusion dieser Linien ist in Abbildung 4.12g dargestellt. Die Wahl geeigneter Parameter wird in Abschnitt 6.1.2 gegeben.

4.4.2 Grauwertprofile der Fahrzeugreihenypothesen

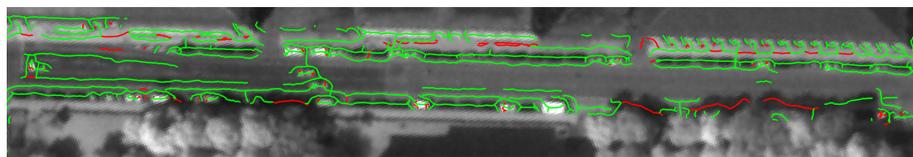
Fahrzeuge innerhalb einer Reihe zeigen Variationen der Breite und des Kontrastes zur Umgebung. Daher würde sich die Bestimmung dieser Parameter entlang der extrahierten Linien wie in Abbildung 4.13a dargestellt anbieten. Als Breite der Grauwertprofile wurde der zweifache Wert der anzunehmenden Fahrzeugbreite w verwendet. Die Extraktion der Umriss einer Fahrzeugreihe kann bei schwachem Kontrast nicht immer zuverlässig durchgeführt werden. Die Auswertung der resultierenden Breiten- und Kontrastprofile (Abbildung 4.13b) durch heuristische Maximumssuche (Hinz et al., 2005; Leitloff et al.,



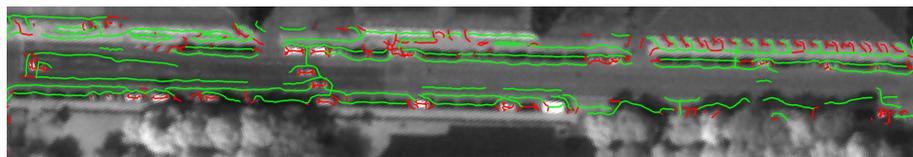
(a)



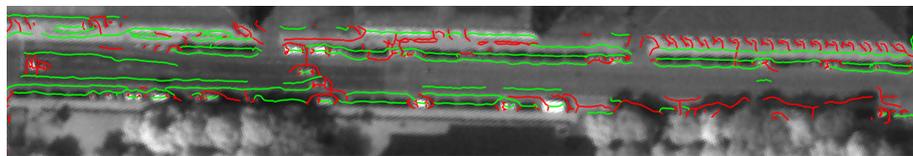
(b)



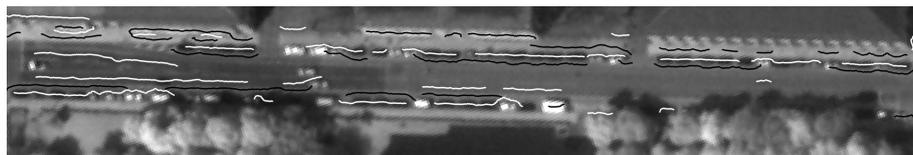
(c)



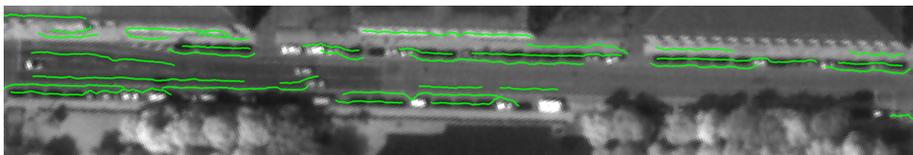
(d)



(e)



(f)



(g)

Abbildung 4.12: Erfassung von Fahrzeugreihen. (a) Originalausschnitt, (b) Initiale Liniextraktion für helle und dunkle Linien, (c) Filterung bzgl. Breite, (d) Filterung bzgl. Länge, (e) Filterung bzgl. Richtung, (f) gefilterte Linien, (g) fusionierte Linien

4.4. Einzelfahrzeuge aus Autoreihen

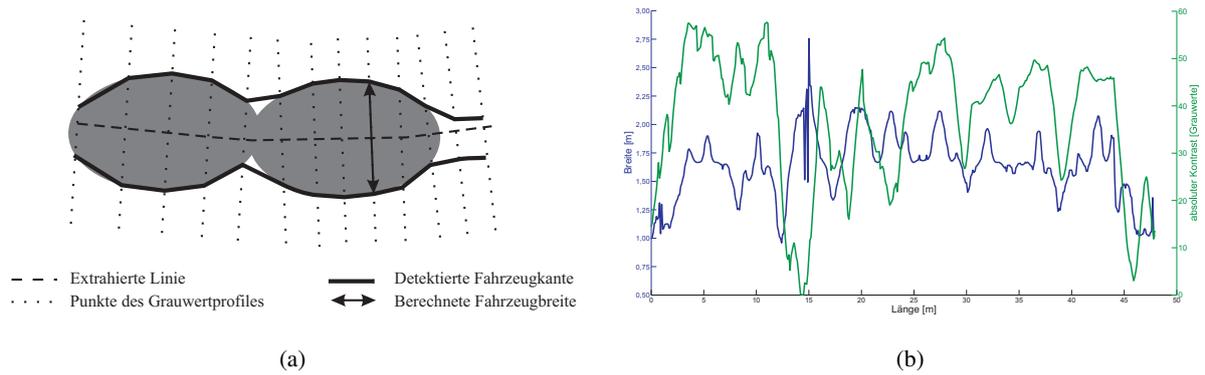


Abbildung 4.13: Grauwertprofile von Fahrzeugreihen. (a) Erfassung der Grauwertprofile und Bestimmung der Fahrzeugbreite, (b) Breiten- und Kontrastprofil

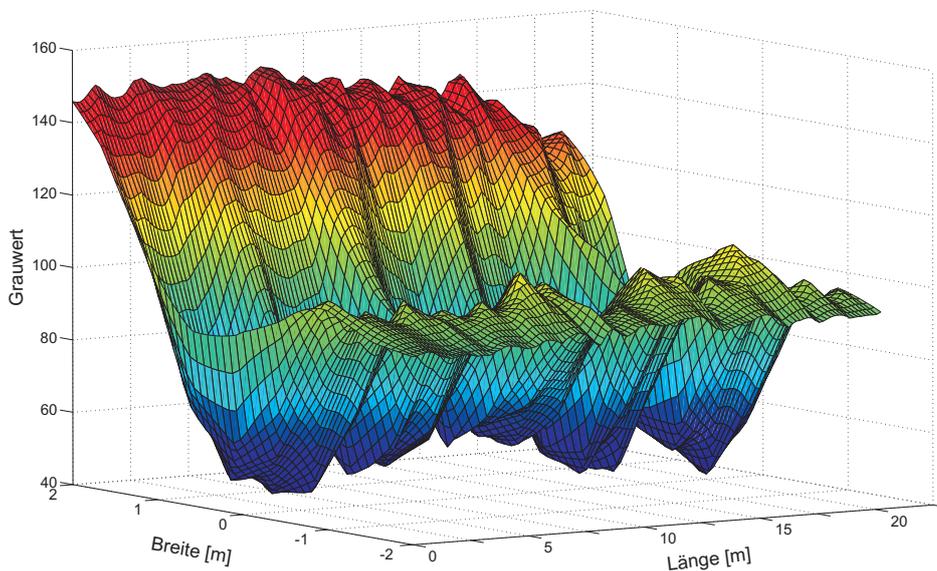


Abbildung 4.14: Extrahiertes Grauwertprofil

2005c,b,a, 2006b) oder Schätzung der Parameter einer approximierenden Funktion (Leitloff et al., 2006a, 2007a, 2010) ist oft fehlerhaft.

In dieser Arbeit wird deshalb eine direkte Auswertung der Grauwertprofile (Abbildung 4.14) durchgeführt.

4.4.3 Erkennung einzelner Fahrzeuge

Einzelne Fahrzeuge innerhalb der Grauwertprofile werden durch eine zweidimensionale Gauß-Funktion approximiert:

$$g(x,y) = a_G \cdot \exp \left(- \left(\frac{(x-x_0)^2}{2 \cdot \sigma_x^2} + \frac{(y-y_0)^2}{2 \cdot \sigma_y^2} \right) \right) \quad (4.31)$$

Diese Funktion wird durch die folgenden Parameter bestimmt:

- x, y : Position entlang bzw. quer zum extrahierten Profil
- $g(x, y)$: Wert des Profils an der Stelle x, y , proportional zum Kontrast
- x_0, y_0 : Position des Maximums der Gauß-Funktion
- σ_x, σ_y : Zur Ausdehnung der Gauß-Funktion äquivalente Größen in x - bzw. y -Richtung, welche proportional zur Länge bzw. Breite der Fahrzeuge sind. Die Halbwertsbreite $FWHM$ in x - oder y -Richtung berechnet sich durch $FWHM_{x \wedge y} = 2 \cdot \sqrt{2 \cdot \ln 2} \cdot \sigma_{x \wedge y}$.
- a_G : Wert des Maximums der Gauß-Funktion

Aus dem Profil gegebene Größen sind $g(x, y)$, x und y . Die Bestimmung der verbleibenden Größen erfolgt durch eine modifizierte Ausgleichung nach vermittelten Beobachtungen. Das Grundprinzip der vermittelten Ausgleichung ist die Darstellung der fehlerbehafteten Beobachtungen \mathbf{b} als Funktion der ausgeglichenen gesuchten Parameter $\hat{\mathbf{x}}$:

$$\mathbf{b} + \hat{\mathbf{v}} = f(\hat{\mathbf{x}}) \quad (4.32)$$

wobei $\hat{\mathbf{v}}$ Schätzwerte für die Verbesserungen der Beobachtungen sind. Voraussetzung zur Lösung dieses Gleichungssystems ist, dass die Anzahl der Beobachtungen höher ist als die Anzahl der gesuchte Parameter und somit redundante Beobachtungen zur Verfügung stehen. Da die verwendete Gauß-Glocke einen nicht linearen Zusammenhang darstellt, existiert keine analytische Lösung dieses Gleichungssystems durch die vermittelnde Ausgleichung. Die Funktion f wird an Näherungswerten \mathbf{x}_0 der gesuchten Parameter durch eine Taylor-Reihe erster Ordnung

$$f(\mathbf{x}_0 + \Delta\hat{\mathbf{x}}) = f(\mathbf{x}_0) + \frac{\delta f}{\delta \mathbf{x}_0} \cdot \Delta\hat{\mathbf{x}} \quad (4.33)$$

approximiert. Diese Approximation ist linear und der Zuschlag $\Delta\hat{\mathbf{x}}$ kann mit dem Newton-Verfahren iterativ berechnet werden:

$$\Delta\hat{\mathbf{x}} = (\mathbf{A}^T \cdot \mathbf{P} \cdot \mathbf{A})^{-1} \cdot \mathbf{A}^T \cdot \mathbf{P} \cdot (\mathbf{b} - f(\mathbf{x}_0)) \quad (4.34)$$

In der Jacobi-Matrix \mathbf{A} sind die berechneten Funktionswerte der partiellen Ableitungen an der Stelle \mathbf{x}_0 enthalten. Die Matrix \mathbf{P} dient zur Gewichtung der einzelnen Beobachtungen und ist im Allgemeinen indirekt proportional zu den Varianzen und Kovarianzen der Beobachtungen. Die gesuchten Parameter berechnen sich durch:

$$\hat{\mathbf{x}} = \mathbf{x}_0 + \Delta\hat{\mathbf{x}} \quad (4.35)$$

Bis zur Konvergenz des Verfahrens werden in jeder Iteration die Näherungswerte der aktuellen Iteration durch die berechneten Unbekannten der vorherigen Iteration ersetzt. Ziel dieser Vorgehensweise ist die Minimierung der Verlustfunktion

$$\rho(\hat{\mathbf{v}}) = \mathbf{v}^T \cdot \mathbf{P} \cdot \mathbf{v} \quad (4.36)$$

Beim Grauwertprofil wird davon ausgegangen, dass alle Beobachtungen gleich genau und unabhängig voneinander sind. Die Gewichtsmatrix entspricht daher einer Einheitsmatrix. Das Newton Verfahren hat nur einen kleinen Konvergenzbereich. Deshalb erfolgt eine Stabilisierung nach der Methode von

4.4. Einzelfahrzeuge aus Autoreihen

Levenberg-Marquardt (Marquardt, 1963). Das Normalgleichungssystem $\mathbf{N} = \mathbf{A}^T \cdot \mathbf{P} \cdot \mathbf{A}$ wird regularisiert, indem die Diagonale von \mathbf{N} in Abhängigkeit des Dämpfungsfaktor λ neu bestimmt wird:

$$\mathbf{N}_{LM} = \mathbf{A}^T \cdot \mathbf{P} \cdot \mathbf{A} + \lambda \cdot \text{diag}(\mathbf{A}^T \cdot \mathbf{P} \cdot \mathbf{A}) \quad (4.37)$$

Bei großen Werten für λ entspricht das Verfahren einem Gradientenabstieg entlang der Verlustfunktion $\rho(\hat{\mathbf{v}})$, da die Unbekanntenzuschläge unabhängig voneinander bestimmt werden. Strebt λ gegen 0 wird $\Delta\hat{\mathbf{x}}$ nach der Methode der kleinsten Quadrate berechnet. Durch geschickte Wahl von λ kann zwischen der langsamer aber sicher konvergierenden Methode des steilsten Abstiegs und dem schnell konvergierenden Newton-Verfahren gewechselt werden. Zu Beginn wird der Dämpfungsfaktor mit einem sehr kleinen Wert ($\lambda = 10^{-3}$) initialisiert und die Unbekanntenzuschläge berechnet. In der darauf folgenden Iterationen wird überprüft, ob $\Delta\hat{\mathbf{x}}$ größer oder kleiner geworden ist. Steigt $\Delta\hat{\mathbf{x}}$ muss davon ausgegangen werden, dass der Konvergenzbereich noch nicht erreicht wurde. Daher wird λ um den Faktor 10 erhöht. Sinkt $\Delta\hat{\mathbf{x}}$ bei aufeinander folgenden Iterationen, wird λ um den Faktor 10 verkleinert.

Die Gauß-Funktion stellt nur eine einfache Approximation für Fahrzeuge dar. Zusätzlich kann die Extraktion der Grauwertprofile durch Bildrauschen verfälscht werden. Die resultierenden Fehler müssen bei der Ausgleichung berücksichtigt werden. Diese Forderung wurde besonders deutlich durch Hampel et al. (1986, Kap. 1.4.) gestellt:

„Most importantly: Any way of treating outliers which is not totally inappropriate, prevents the worst. [...] Totally inappropriate are nonrobust computer program (e.g., least squares!) without any built-in checks ... “

In dieser Arbeit werden M-Schätzer nach Huber (1964) verwendet. Die gegenüber Ausreißern sehr empfindliche Verlustfunktion der kleinsten Quadrate Minimierung $\sum_i v_i^2 \rightarrow \min$ wird durch eine symmetrische, positiv definite Funktion mit eindeutigem Minimum bei null ersetzt:

$$\sum_i \rho(v_i) \rightarrow \min \quad (4.38)$$

Zur Minimierung dieser Funktion werden die partiellen Ableitungen nach den m gesuchten Parametern gebildet und gleich Null gesetzt:

$$\sum_i \psi(v_i) \frac{\partial v_i}{\partial x_j} = 0, j = 1, \dots, m \quad (4.39)$$

$\psi(v_i) = \partial \rho(v_i) / \partial v_i$ sind die Einflussfunktionen der Beobachtungsverbesserungen. Definiert man nun eine geeignete Gewichtsfunktion

$$w(v_i) = \frac{\psi(v_i)}{v_i} \quad (4.40)$$

lässt sich Formel 4.38 als ein kleinstes Quadrate Problem mit iterativer Neugewichtung (Holland & Welsch, 1977) darstellen:

$$\sum_i w(v_i^{k-1}) v_i^2 \rightarrow \min \quad (4.41)$$

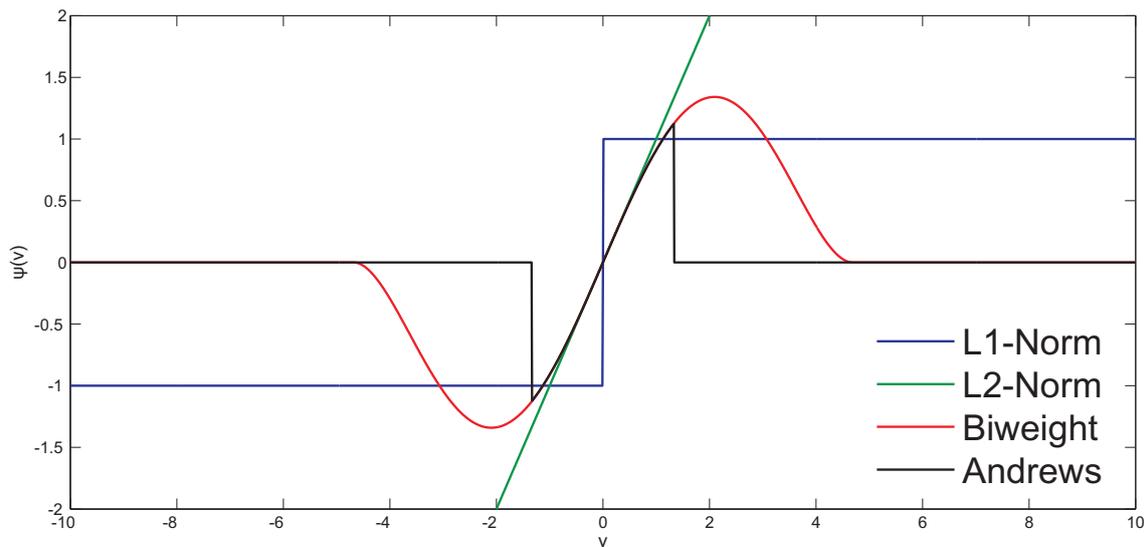


Abbildung 4.15: Einflussfunktionen ausgewählter Schätzer

Die Gewichte w_i werden in Abhängigkeit der Verbesserungen der vorherigen Iteration v_i^{k-1} berechnet. Eine Übersicht möglicher Verlustfunktionen ist in Tabelle 4.4 gegeben. Diese können in hart und schwach abweisende Schätzer unterteilt werden. Hart abweisende Schätzer besitzen einen konstanten, von den Verbesserungen unabhängigen Anteil ihrer Verlustfunktion. Dies führt dazu, dass die Gewichtsfunktion in diesen Intervallen Null ist und die entsprechenden (stark fehlerbehafteten) Beobachtungen keinen Einfluss auf die gesuchten Größen haben. Zu diesen Schätzern gehören der Andrews-, Biweight-, Talwar-, Hampel- und der modifizierte Huber-Schätzer. Der Unterschied der Verlustfunktionen aus Tabelle 4.4 zur kleinsten Quadrate Methode ist die stärkere Beschränkung der Einflussfunktion, wodurch grob fehlerbehaftete Beobachtungen keinen oder nur einen geringen Einfluss auf die Schätzung der gesuchten Parameter haben. Abbildung 4.15 zeigt diese Beschränkungen für einige Verlustfunktionen im Vergleich zur kleinsten Quadrate Schätzung (L2-Norm).

Einige Schätzer benötigen sogenannte Tuningkonstanten zur Festlegung der Intervalle. In der Tabelle sind die Standardwerte der jeweiligen Quelle angegeben. Diese setzen normiert, normalverteilte Verbesserungen voraus. Diese Voraussetzung ist im Allgemeinen nicht erfüllt und erfordert eine Normierung der Verbesserungen in Abhängigkeit ihrer Standardabweichung:

$$v_i \rightarrow \frac{v_i}{\sigma_v} \quad (4.42)$$

Die empirische Standardabweichung ist ebenfalls durch fehlerbehaftete Beobachtungen beeinflusst und wird deshalb robust mit dem Median der absoluten Abweichung bestimmt:

$$\sigma_v = \frac{\text{median} |\mathbf{v} - \text{median}(\mathbf{v})|}{0,6745} \quad (4.43)$$

Die robuste Parameterschätzung unter Verwendung des Levenberg-Marquardt Verfahrens wird für separate Intervalle entlang der Grauwertprofile durchgeführt. Die Bestimmung der Intervallgrenzen erfolgt durch:

4.4. Einzelfahrzeuge aus Autoreihen

Schätzer (Quelle)	Verlustfunktion $\rho(v)$ (Tuningkonstanten)
Andrews (Andrews et al. (1972))	$\rho(v) = \begin{cases} a^2 \cdot (1 - \cos(\frac{v}{a})) & \forall v \leq \pi \cdot a \\ 2 \cdot a^2 & \forall v > \pi \cdot a \end{cases}$ ($a = 1,339$)
Biweight (Beaton & Tukey (1974))	$\rho(v) = \begin{cases} \frac{1}{2} a^2 \cdot \left(1 - \left(1 - \frac{v^2}{a^2}\right)^3\right) & \forall v \leq a \\ \frac{1}{2} a^2 & \forall v > a \end{cases}$ ($a = 4,685$)
Talwar (Hinich & Talwar (1975))	$\rho(v) = \begin{cases} \frac{1}{2} \cdot v^2 & \forall v \leq a \\ \frac{1}{2} \cdot a^2 & \forall v > a \end{cases}$ ($a = 2,795$)
Cauchy (Holland & Welsch (1977))	$\rho(v) = \frac{1}{2} \cdot a^2 \cdot \log\left(1 + \left(\frac{v}{a}\right)^2\right)$ ($a = 2,385$)
Welsch (Dennis & Welsch (1978))	$\rho(v) = \frac{1}{2} a^2 \cdot \left(1 - e^{-(v/a)^2}\right)$ ($a = 2,985$)
Huber (Huber (1964))	$\rho(v) = \begin{cases} \frac{1}{2} \cdot v^2 & \forall v < a \\ a \cdot v - \frac{1}{2} \cdot a^2 & \forall v \geq a \end{cases}$ ($a = 1,345$)
modifizierter Huber (Huber (1977))	$\rho(v) = \begin{cases} \frac{1}{2} \cdot v^2 & \forall v \leq a \\ a \cdot v - \frac{1}{2} \cdot a^2 & \forall a < v \leq b \\ a \cdot b - \frac{1}{2} \cdot a^2 & \forall v > b \end{cases}$ ($a = 1,5 \quad b = 3,0$)
Hampel (Hampel (1980))	$\rho(v) = \begin{cases} \frac{1}{2} \cdot v^2 & \forall v < a \\ a \cdot v - \frac{1}{2} \cdot a^2 & \forall a \leq v < b \\ a \cdot b - \frac{1}{2} a^2 + \frac{1}{2} a \cdot (c - b) \cdot \left(1 - \left(\frac{c - v }{c - b}\right)^2\right) & \forall b \leq v < c \\ a \cdot b - \frac{1}{2} a^2 + \frac{1}{2} a \cdot (c - b) & \forall v \geq c \end{cases}$ ($a = 2,0 \quad b = 4,0 \quad c = 8,0$)
Logistic (Holland & Welsch (1977))	$\rho(v) = a^2 \cdot \log(\cosh(\frac{v}{a}))$ ($a = 1,205$)
Fair (Fair (1974))	$\rho(v) = a^2 \cdot \left(\frac{ v }{a} - \log\left(1 + \frac{ v }{a}\right)\right)$ ($a = 1,4$)
L1-Norm (Jäger et al. (2005))	$\rho(v) = v $

Tabelle 4.4: Verlustfunktionen ausgewählter M-Schätzer

1. Berechnung der maximal möglichen Fahrzeuganzahl innerhalb eines Profils auf Basis einer minimal anzunehmenden Fahrzeuglänge l_{\min}
2. Glättung des Profils durch diskreten Gaußfilter
3. Bestimmung von Grauwert-Plateaus entlang der Mittellinie des geglätteten Profils
4. Falls Anzahl der Plateau-Punkte größer als Anzahl möglicher Fahrzeuge Rückkehr zu Punkt 2
5. Festlegung der Intervallgrenzen durch Mittelpunkt zwischen aufeinander folgenden Grauwert-Plateaus bzw. Anfangs- und Endpunkt des Profils

Für die resultierenden Intervalle werden geeignete Näherungswerte der gesuchten Parameter der Gauß-Funktion bestimmt. Diese werden anschließend durch das Levenberg-Marquardt Verfahren iterativ verbessert. Auf Basis ihrer Verbesserungen findet eine Neugewichtung der Beobachtungen durch die Gewichtsfunktion eines M-Schätzer statt, bis die iterativ neugewichtete Ausgleichung konvergiert. Der vollständige Algorithmus ist in Abbildung 4.16 gegeben.

Die geschätzten Parameter und deren Fehler dienen als Merkmale für eine Klassifikation mit einer Support Vector Machine (siehe Abschnitt 4.3.2). Aus den Trainingsdaten des AdaBoost Verfahrens werden Grauwertprofile von Fahrzeugen bestimmt. Die Ausgleichung dieser Grauwertprofile erzeugt die Trainingsdaten für eine Ein-Klassen-SVM (Schölkopf et al., 2001). Da bei der Klassifikation von geschätzten Parametern und Fehlern keine Gegenklasse existiert, kann keine Abschätzung des Rauschanteils ν erfolgen. Dieser Wert muss vor Beginn des Trainings festgelegt werden und entspricht der Irrtumswahrscheinlichkeit, dass eine Fahrzeughypothese abgelehnt wird, obwohl diese richtig ist. In dieser Arbeit wurde $\nu = 0,001$ verwendet. Abbildung 4.17b zeigt exemplarisch die mittels eines L1-Norm Schätzers angepassten Gauß-Funktionen an das in Abbildung 4.17a dargestellte Profil.

4.5 Ergebniszusammenführung

Die Zusammenführung der Ergebnisse für isolierte und gruppierte Fahrzeuge ist notwendig, sofern ein Fahrzeug durch beide Verfahren erfasst wurde oder die Verfahren unterschiedliche Ergebnisse für das gleiche Gebiet liefern.

In Tabelle 4.5 sind mögliche Ergebniskombinationen und deren Handhabung zusammengefasst. Anhand eines Beispiels soll die Vorgehensweise genauer erläutert werden.

Wird in einem Bildausschnitt ein Fahrzeug durch die Klassifikation ($N_K = 1$) detektiert und mehrere Fahrzeuge aus den Grauwertprofilen ($N_R > 1$) bestimmt, gilt:

- Als Fahrzeuge F werden die Hypothesen aus den Fahrzeugreihen H_R akzeptiert. Die Hypothese aus der Klassifikation isolierter Fahrzeuge H_K wird verworfen.
- Die Positionen P der Fahrzeuge sind durch die geschätzten Parameter x_0 und y_0 gegeben. Die Mittelpunktskoordinaten x_M und y_M des klassifizierten Bildsegments werden verworfen.

4.5. Ergebniszusammenführung

1. Gegeben sind die gleich genauen Beobachtungen $g(x, y)$ an der fehlerfreien Position x, y
2. Gesucht sind die Parameter $x_0, y_0, \sigma_x, \sigma_y$ und a_G
3. Wiederhole bis Profil durchlaufen
 - (a) Bestimme Näherungswerte $\mathbf{x}_0 = [x_0, y_0, \sigma_x, \sigma_y, a_G]^T$ für Intervall:

x_0, y_0 : Aus nicht zentrierten Momenten 1. Ordnung $M_{pq} = \sum_x \sum_y x^p \cdot y^q \cdot g(x, y)$ ergibt sich
 $x_0 = M_{10}/M_{00}$ bzw. $y_0 = M_{01}/M_{00}$

σ_x, σ_y : Aus zentralen Momenten 2. Ordnung ergibt sich $\sigma_x = \sqrt{M_{20} - x_0 \cdot M_{10}}$ bzw.
 $\sigma_y = \sqrt{M_{02} - y_0 \cdot M_{01}}$

a_G : Aus Maximalwert innerhalb des Intervalls ergibt sich $a_G = \max(g(x, y))$
 - (b) Initialisiere $\lambda = 10^{-3}$
 - (c) Wiederhole über j bis $\max |\Delta \hat{\mathbf{x}}| < 10^{-5}$
 - i. Berechne \mathbf{A} und $f(\mathbf{x}_0)$ für Näherungswerte
 - ii. Berechne $\Delta \hat{\mathbf{x}}$ mit Normalgleichungssystem nach Formel 4.37 für $\lambda_1 = \lambda$ und $\lambda_2 = \lambda_1 \cdot 10$
 - iii. Falls $\sum_i (v_i(\lambda_1))^2 < \sum_i (v_i(\lambda_2))^2$ setze $\lambda = \lambda_1/10$ und $\mathbf{x}_0^{j+1} = \mathbf{x}_0^j + \Delta \mathbf{x}_{\lambda_1}^j$, sonst setze $\lambda = \lambda_2$ und verwerfe $\Delta \mathbf{x}_{\lambda_2}^j$
 - (d) Berechne Verbesserungen nach Formel 4.32
 - (e) Wiederhole über j bis $\max |\Delta \hat{\mathbf{x}}| < 10^{-5}$
 - i. Berechne neue Gewichtung $w(\mathbf{v}^{j-1})$ der Beobachtungen in Abhängigkeit der Verbesserungen der vorherigen Iteration und des gewählten Schätzers
 - ii. Berechne \mathbf{A} und $f(\mathbf{x}_0)$ für Näherungswerte
 - iii. Berechne $\Delta \hat{\mathbf{x}}$ nach Formel 4.34
 - iv. Berechne Verbesserungen nach Formel 4.32

Abbildung 4.16: Robuste Schätzung der Parameter der Gauß-Funktion

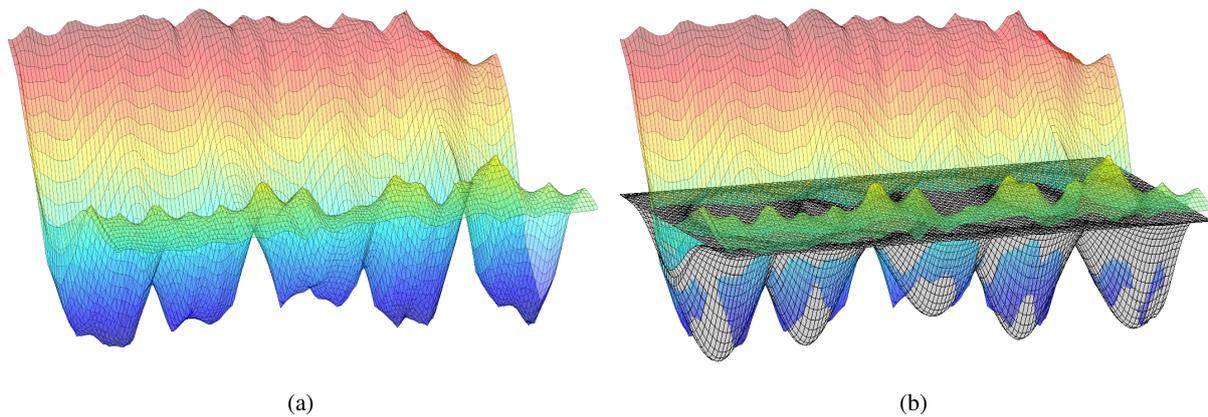


Abbildung 4.17: Ergebnis der Approximation eines Grauwert-Profiles. (a) das zu approximierende Grauwertprofil, (b) an das Grauwertprofil angepasste Gauß-Funktionen

- Die Fahrzeuge werden als aus der Reihe extrahiert markiert. Ihre Kategorie K entspricht daher „gruppiert“ und nicht „isoliert“.

Zusammenfassend ist anzumerken, dass beim Auftreten einer Fahrzeughypothese aus einer Fahrzeugreihe dieser mehr Bedeutung zugemessen wird und daher immer akzeptiert wird. Gleiches gilt für die Position eines Fahrzeuges aus der Parameterschätzung.

4.6 Bewegungsschätzung

Optische Satellitenbildsysteme bestehen meist aus Linienscannern. Die räumliche Anordnung der CCD-Elemente führt zu einer zeitlichen Differenz bei der Aufnahme einzelner Spektralkanäle. Die genaue Zeitdifferenz sollte Bestandteil der mit den Bilddaten gelieferten Metainformationen sein. Bewegte Objekte werden durch den unterschiedlichen Aufnahmezeitpunkt an verschiedenen Positionen in den Spektralkanälen abgebildet. Dies kann zur Erkennung der Bewegung dieser Objekte genutzt werden.

Die Bewegungsschätzung wird nur für „isoliert“ markierte Fahrzeuge vorgenommen. Bei zu Reihen gruppierten Fahrzeugen handelt es sich in den meisten Fällen um stehende oder nur sehr langsam fahrende Fahrzeuge, welche keinen signifikanten Versatz zwischen den Spektralkanälen zeigen. Die folgende Bewegungsschätzung setzt eine ausreichende Zeitdifferenz zwischen dem panchromatischen und den multispektralen Kanälen voraus.

Die Position eines Fahrzeuges im panchromatischen Kanal ist durch die Klassifikation oder die Parameterschätzung bekannt. Zur Bestimmung des Versatzes eines Fahrzeuges wird dessen Position in den multispektralen Kanälen benötigt. Um ein Suchverfahren nicht in allen Farbkanälen durchführen zu müssen, werden die RGB-Bilder in entsprechende Intensitäts-, Farbton- und Sättigungsbilder transformiert. Abbildung 4.18 zeigt die Verteilung der Farbton- und Sättigungswerte für verschiedene Objekte. Der Farbton ist als Winkel zur Abszissenachse und die Sättigung als Abstand zum Ursprung definiert. Aus den Abbildungen 4.18a und 4.18b wird deutlich, dass die farbigen Fahrzeuge einen großen Einfluss auf die

4.6. Bewegungsschätzung

N_R	N_K		
	0	1	> 1
0		$F = H_K$ $P = x_M, y_M$ $K = \text{isoliert}$	$F = H_K$ $P = x_M, y_M$ $K = \text{isoliert}$
	1	$F = H_R$ $P = x_0, y_0$ $K = \text{gruppiert}$	$F = H_R \wedge H_K$ $P = x_0, y_0$ $K = \text{gruppiert}$
> 1	$F = H_R$ $P = x_0, y_0$ $K = \text{gruppiert}$	$F = H_R$ $P = x_0, y_0$ $K = \text{gruppiert}$	$\left. \begin{array}{l} F = H_R \wedge H_K \\ F = H_R \end{array} \right\} \begin{array}{l} N_R = N_K \\ N_R \neq N_K \end{array}$ $P = x_0, y_0$ $K = \text{gruppiert}$

Tabelle 4.5: Zusammenführung der Ergebnisse der Teilextraktionen.

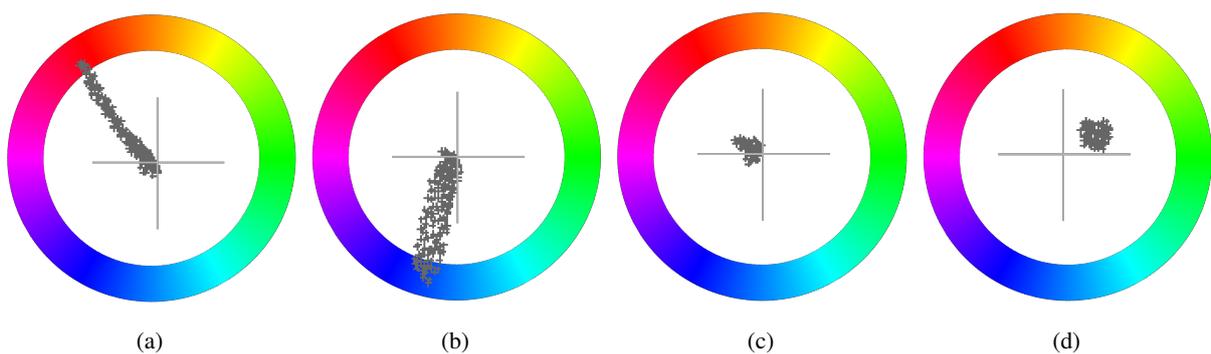


Abbildung 4.18: Farbraumtransformation verschiedener Objekte, Kodierung des Farbtons als Winkel zur Abszissenachse und Sättigung als Abstand zum Ursprung. (a) rotes Auto auf Straßenuntergrund, (b) blaues Fahrzeug auf Straßenuntergrund, (c) Straße, (d) Rasenfläche

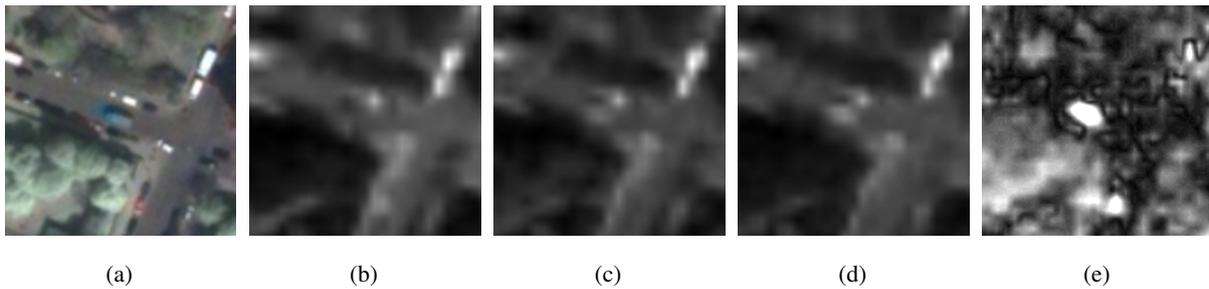


Abbildung 4.19: Farbiges Fahrzeug in verschiedenen Spektralkanälen. (a) Panchromatischer und RGB-Kanäle, (b) roter Kanal, (c) grüner Kanal, (d) blauer Kanal, (e) Sättigungsbild

Sättigungswerte haben, während graue Straßen nur eine geringe Sättigung zeigen (Abbildung 4.18c). Auch in Abbildung 4.19 ist dies erkennbar. So treten das blaue und das rote Fahrzeug aus Abbildung 4.19a in den RGB-Kanälen (Abbildungen 4.19b bis 4.19d) nur schwach hervor, besitzen aber hohe Sättigungswerte in Abbildung 4.19e.

Zum Auffinden eines detektierten Fahrzeuges im Sättigungsbild wird das Subpixel-genaue Matchingverfahren nach Steger (2001) angewandt. Zunächst wird ein Ausschnitt um die im panchromatischen Kanal bekannte Fahrzeugposition erzeugt. Der Bildausschnitt wird in der niedrigeren Bodenpixelgröße des Sättigungsbilds neu berechnet. Anschließend erfolgt die Berechnung der Gradientenrichtungen für jeden Bildpunkt des Bildausschnitts und innerhalb eines beschränkten Suchbereichs des Sättigungsbilds. Die Größe des Suchbereichs wird aus der bekannten Fahrzeugposition im panchromatischen Bild und einer maximal anzunehmenden Geschwindigkeit bestimmt. Die maximal anzunehmende Geschwindigkeit wurde sehr großzügig gewählt und beträgt in innerstädtischen Bereichen 100 km/h . Außerhalb urbaner Gebiete muss diese den dort auftretenden Fahrzeuggeschwindigkeiten angepasst werden. Mit den Gradientenrichtungen wird für jeden Bildpunkt des beschränkten Sättigungsbilds ein Ähnlichkeitsmaß zu den Gradientenrichtungen der Bildvorlage berechnet. Das für Abbildung 4.20a resultierende Ähnlichkeitsbild des Matchingverfahrens ist in Abbildung 4.20b dargestellt. Die Nutzung von Richtungsinformationen, also Winkelmaßen, macht den Ansatz invariant gegenüber dem sehr unterschiedlich ausfallenden Kontrast des panchromatischen Kanals und des Sättigungsbildes. Das Maximum des Ähnlichkeitsbildes dient als Position des Fahrzeuges in den multispektralen Kanälen, sofern der aus der Maximalgeschwindigkeit resultierende Versatz nicht überschritten wird.

In Abbildung 4.20c ist die Fahrzeugposition im panchromatischen (grünes Kreuz) und im Sättigungsbild (rotes Kreuz) sichtbar. Entlang der grünen bzw. roten Linie werden Grauwertprofile im panchromatischen und im Sättigungsbild extrahiert (siehe Abschnitt 4.4.2). Die Richtung dieser Linien entspricht der aus Geoinformationen bekannten Straßenrichtung. Mit den Grauwertprofilen erfolgt eine Verfeinerung beider Fahrzeugpositionen und die Bestimmung des Positionsfehler durch Anwendung der in Abschnitt 4.4.3 vorgestellten Parameterschätzung.

Durch Fehlerfortpflanzung wird der Fehler des berechneten Versatzes bestimmt. Das Ergebnis eines t-Tests, mit einer Irrtumswahrscheinlichkeit von 5 %, entscheidet über die Signifikanz des Versatzes. Für

4.6. Bewegungsschätzung

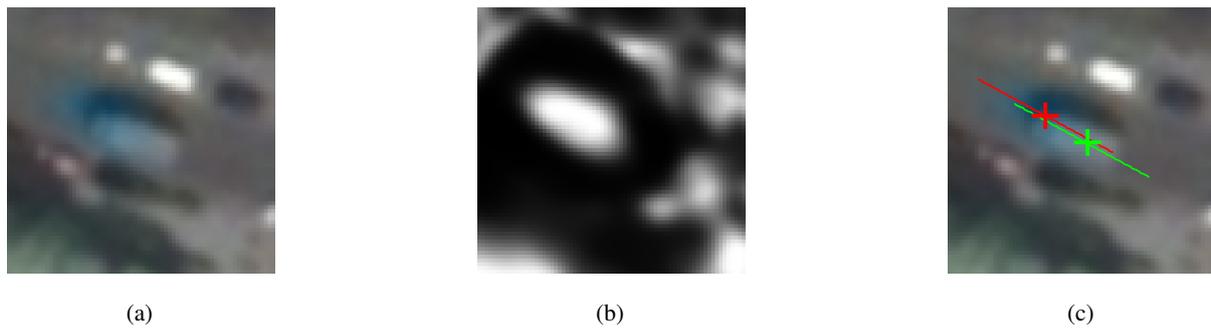


Abbildung 4.20: Bestimmung der Fahrzeugposition in den multispektralen Kanälen. (a) Panchromatischer und RGB-Kanäle, (b) Ähnlichkeitswerte aus Matchingverfahren, helle Werte entsprechen hoher Ähnlichkeit, (c) Position (Kreuz) des Fahrzeuges im panchromatischen (grün) und multispektralen (rot) Kanälen

einen zeitlichen Versatz von 0,2 Sekunden zwischen dem panchromatischen Kanal und den multispektralen Kanälen wurde in Leitloff et al. (2007b) ein empirischer Mittelwert von 1,2 m zur Validierung einer Bewegung bestimmt. Dies entspricht etwa der Hälfte der Bodenpixelaufösung der verwendeten multispektralen Daten. Die dem entsprechende Mindestgeschwindigkeit wurde mit 20 km/h angegeben.

5 Satellitenbilddaten und Geoinformation

5.1 Optische hochauflösende Satellitenbilddaten

In diesem Abschnitt werden bisherige und zukünftige optische Satellitenmissionen mit einer Bodenpixelgröße von weniger als einem Meter vorgestellt. Auf den QuickBird Satelliten wird gesondert eingegangen, da die Testdatensätze zur Evaluierung des entwickelten Ansatzes von dessen Bildsystem stammen. Weiterführende Übersichten für hochauflösende Satellitensysteme geben Jacobsen (2005) und Jacobsen (2011).

5.1.1 Satellitenmissionen

Die ersten Bilddaten mit einer Bodenpixelgröße im 1-Meter- bzw. Submeterbereich standen nach dem erfolgreichen Start des *IKONOS* Satelliten im Jahr 1999 zur Verfügung. Die Bodenpixelgröße des panchromatischen Kanals in Nadirrichtung wird mit 82 cm angegeben. Bilder werden im blauen, grünen, roten und nahen infraroten Spektralband mit einer Bodenpixelgröße von bis zu 3,2 m aufgenommen. Bei Verwendung seitlich aufgenommener Bilder können mit *IKONOS* Gebiete innerhalb von 3 bis 5 Tagen wiederholt erfasst werden.

2001 folgte der *QuickBird* Satellit mit einer Bodenpixelgröße von bis zu 61 cm im panchromatischen Kanal und 2,4 m in den multispektralen Kanälen. Eine genaue Beschreibung dieses Satellitensystems erfolgt in Abschnitt 5.1.2. Der *OrbView-3* Satellit startete im Jahr 2003 und lieferte panchromatische Bilder mit einer Bodenpixelgröße von einem Meter und multispektrale Bilder von vier Metern Bodenpixelgröße. Aufgrund technischer Fehler des Kamerasystems fiel dieser 2007 aus.

Im gleichen Jahr startete der *WorldView-1* Satellit, der im Gegensatz zu bisherigen Systemen ausschließlich panchromatische Bilder erfasst. Die Bilddaten haben eine bis dahin kommerziell nicht verfügbare Bodenpixelgröße von bis zu 50 cm bei einer Aufnahmebreite von 17,6 km. Zudem ist die Genauigkeit der Georeferenzierung im Vergleich zu vorherigen Systemen gestiegen. Beträgt die horizontale Genauigkeit ohne Kontrollpunkte bei *IKONOS* und *QuickBird* noch über 20 m, so erreicht *WorldView-1* bereits eine horizontale Genauigkeit von 6,5 m.

2007 startete ein weiterer ebenfalls rein panchromatischer Sensor mit dem *CARTOSAT-2* Satelliten. Dieser erzeugt Bilder mit einer Bodenpixelgröße von 80 cm in Nadirrichtung. 2008 folgte der Erdbeobachtungssatellit *GeoEye-1*, welcher Bilder mit einer Bodenpixelgröße von 41 cm im panchromatischen

Satellit	Betriebszeit	panchromatische Bodenpixelgröße	multispektrale Bodenpixelgröße
IKONOS	seit 1999	82 cm	3,20 m
QuickBird	seit 2001	61 cm	2,40 m
OrbView-3	2003 bis 2007	100 cm	4,00 m
WorldView-1	seit 2007	50 cm	
CARTOSAT-2	seit 2007	80 cm	
GeoEye-1	seit 2008	41 cm	1,65 m
WorldView-2	seit 2009	46 cm	1,84 m

Tabelle 5.1: Optische Satellitensysteme mit Meter- und Submeterauflösung

und 1,65 m in den multispektralen Kanälen erfasst. Allerdings ist die betreibende Firma *GeoEye Inc.* aufgrund ihrer erteilten Betriebslaubnis verpflichtet, eine Reduzierung der Bodenpixelgröße auf 50 cm vorzunehmen (GeoEye, Inc., 2010).

Im Herbst 2009 startete der *WorldView-2* Satellit, dessen panchromatische Aufnahmen eine Bodenpixelgröße von bis zu 46 cm besitzen. Eine Neuerung des Systems ist die erweiterte Anzahl hochaufgelöster Spektralkanäle. Neben den Standardfarbkanälen verfügt *WorldView-2* über vier weitere Spektralkanäle, welche neue Anwendungsgebiete erschließen sollen. Die Bodenpixelgröße aller Spektralkanäle beträgt in Nadirrichtung 1,84 m. Die Verbindung von *QuickBird*, *WorldView-1* und *WorldView-2* versetzt die Betreiberfirma *DigitalGlobe Inc.* in die Lage, ein Gebiet in weniger als einem Tag wiederholt aufzunehmen. Für die kommenden zwei bis drei Jahre ist mit dem Satelliten *GeoEye-2* ein System in Planung, welches die bisher erreichten Auflösungen weit übertreffen könnte.

Eine zusammenfassende Übersicht von bisherigen Satellitensystemen mit einer Bodenpixelgröße von mindestens einem Meter in Tabelle 5.1 gegeben.

5.1.2 QuickBird Daten

QuickBird (Abbildung 5.1a) ist ein hochauflösender optischer Erdbeobachtungssatellit der Firma DigitalGlobe der am 18. Oktober 2001 vom Vandenberg Luftwaffenstützpunkt (Kalifornien, USA) startete. Als Bildsensor dient das *Ball Global Imaging System 2000* der *Ball Aerospace & Technologies Corporation*. In einem Orbit von 450 km beträgt die Bodenpixelgröße des panchromatischen Kanals in Nadirrichtung 61 cm und bei einem Seitenblick von 25 Grad 72 cm. Mit einer Neigung von 30 Grad gegenüber der Nadirrichtung wird eine Streifenbreite von 16,5 km erfasst. Neben dem panchromatischen Kanal werden die in Tabelle 5.2 aufgeführten Spektralbänder mit einer Bodenpixelgröße von bis zu 2,44 m erfasst. Wie bei den meisten Satellitenbildsystemen handelt es sich bei dem *Ball Global Imaging System 2000* um einen Pushbroom Scanner (Gupta & Hartley, 1997; Jacobsen, 2005). Quer zur Flugrichtung sind eine oder mehrere CCD Zeilen angeordnet. Abbildung 5.1b verdeutlicht eine solche Aufnahmegeometrie. Durch die Vorwärtsbewegung des Satelliten kann QuickBird Bildstreifen von bis zu 165 km Länge aufnehmen.

5.1. Optische hochauflösende Satellitenbilddaten

Kanal	Spektralbereich	Bodenpixelgröße in Nadirrichtung	Bodenpixelgröße bei 25° Blickwinkel
panchromatisch	450 - 900 <i>nm</i>	61 <i>cm</i>	72 <i>cm</i>
blau	450 - 520 <i>nm</i>	244 <i>cm</i>	288 <i>cm</i>
grün	520 - 600 <i>nm</i>	244 <i>cm</i>	288 <i>cm</i>
rot	630 - 690 <i>nm</i>	244 <i>cm</i>	288 <i>cm</i>
nahes infrarot	760 - 900 <i>nm</i>	244 <i>cm</i>	288 <i>cm</i>

Tabelle 5.2: Spektralkanäle des QuickBird Sensors

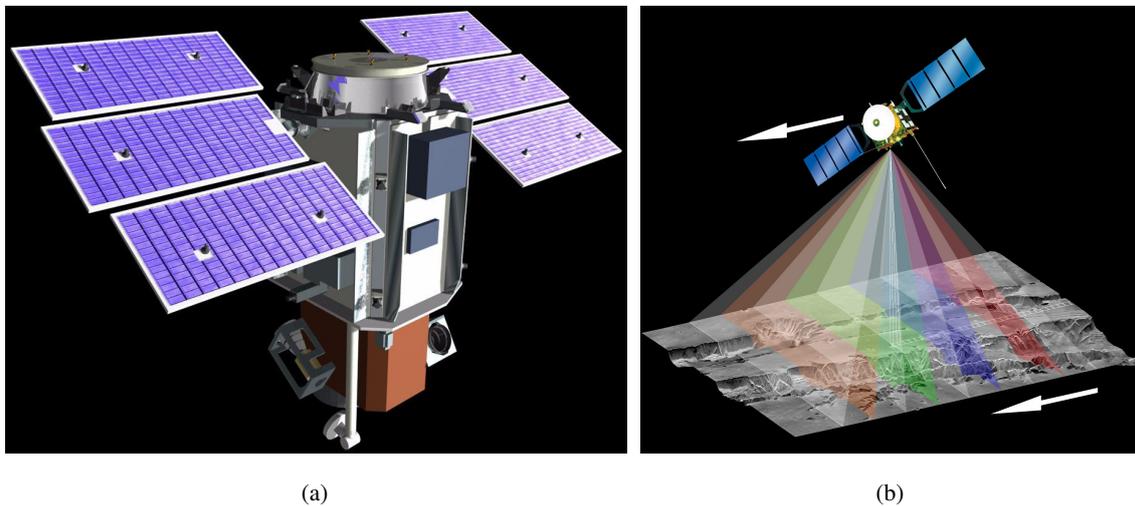


Abbildung 5.1: QuickBird Satellit und Aufnahmeprinzip. (a) QuickBird Satellit (Bild: DigitalGlobe), (b) Prinzip eines CCD Linien Scanners am Beispiel der hochauflösenden Stereokamera (HRSC) (Bild: DLR)



Abbildung 5.2: Versatz eines ca. 500 km/h schnellen Flugzeuges (Bild: Eurimage (2010))

Das panchromatische CCD-Zeilen-Element besteht aus 64 Sensorelementen quer zur Flugrichtung. Die Aufnahme erfolgt unter Verwendung der *transfer delay and integration* Technik, wobei die aufgenommene Ladung einer CCD-Zeile auf die nachfolgende CCD-Zeile übertragen wird. Das Rauschen einzelner CCD-Elemente hat daher nur einen geringen Anteil an der integrierten Ladung aller CCD-Zeilen. Für QuickBird Aufnahmen wird im Regelfall über 13 CCD-Zeilen integriert, was zur radiometrischen Auflösung von 11 Bit im panchromatischen Kanal führt. Die multispektralen Kanäle werden jeweils durch eine Zeile erfasst.

Die räumliche Anordnung des panchromatischen und der multispektralen Elemente verursacht einen zeitlichen Versatz von circa 0,2 Sekunden. Abbildung 5.2 verdeutlicht den Effekt der getrennten Aufnahme der CCD-Zeilen Elemente für eine QuickBird Aufnahme. Mit Hilfe der Distanz lässt sich unter Verwendung der bekannten Zeitdifferenz und einer angenommenen Flughöhe des Flugzeuges, dessen Geschwindigkeit mit ca. 500 km/h angeben.

5.1.3 Testdatensätze

Die Evaluierung des entwickelten Ansatzes erfolgt unter anderem für die in Abbildung 5.3 dargestellten Ausschnitte einer QuickBird-Szene. Diese wurde im April 2003 mit 20 % Wolkenbedeckung und einem Seitenblickwinkel von $10,5^\circ$ aufgenommen. Die Bildqualität ist als „ausreichend“ deklariert und entspricht damit der untersten ausgelieferten Qualitätsstufe. Die mittlere Bodenpixelgröße des verwendeten Datenmaterials beträgt im panchromatischen Kanal 63 cm und für die multispektralen Kanäle 2,5 m. Die Bilder liegen als *Standard Imagery Product* vor (DigitalGlobe, 2009). Hierzu wurden die aufgenommenen Daten radiometrisch, sensorspezifisch sowie geometrisch korrigiert. Für dieses Produkt Level wird ein grobes digitales Höhenmodell im Rahmen der kartographischen Projektion in das UTM-System verwendet. Für die gesamte Szene wurde eine einheitliche Geländehöhe von 562 m angenommen. Die resultierenden Korrekturen fallen gering aus und das vorliegende Bildmaterial wird nicht als orthorektifiziert sondern lediglich als georektifiziert eingestuft (DigitalGlobe, 2009). Die absolute Lageposition besitzt einen mittleren quadratischen Fehler von 14 m und weicht mit 90 prozentiger Wahrscheinlichkeit um weniger als 23 m von der tatsächlichen Lage ab.

5.1. Optische hochauflösende Satellitenbilddaten

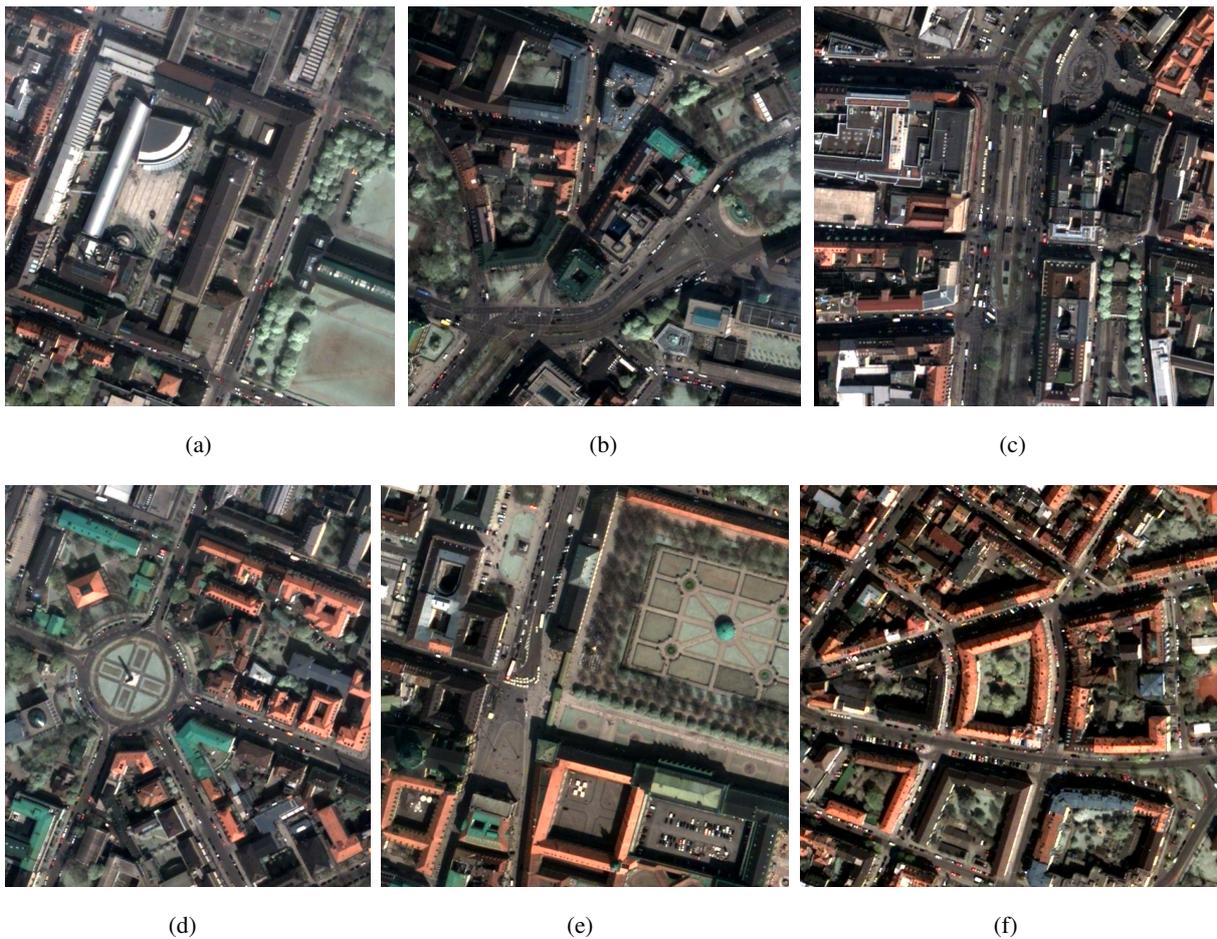


Abbildung 5.3: Testdatensätze der Münchner Innenstadt. a) Testgebiet 1, b) Testgebiet 2, c) Testgebiet 3, d) Testgebiet 4, e) Testgebiet 6, f) Testgebiet 7

5.2 Geoinformation

Die Verwendung von zusätzlichen Geoinformationen ist zur Verringerung falscher Detektionen sinnvoll (siehe Abschnitt 2.3) und wurde in allen vorgestellten Arbeiten zur Fahrzeugerkennung aus optischen Satellitenbilddaten genutzt (siehe Abschnitt 2.2.2).

Eine deutschlandweite Beschreibung topographischer Objekte ist im Digitalen Landschaftsmodell (DLM) des Amtlichen Topographisch-Kartographischen Informationssystem (ATKIS) realisiert. Der höchste Detaillierungsgrad ist im Basis-DLM gegeben, welches aus dem Inhalt der amtlichen topographischen Karte 1:25.000 abgeleitet wird. Diese Daten unterscheiden verschiedene Objektbereiche wie Siedlung, Verkehr, Vegetation und Gewässer. In dieser Arbeit wird der Objektbereich Verkehr mit seiner Objektgruppe Straßenverkehr und den darin enthaltenen Objektarten verwendet. Abbildung 5.4 zeigt exemplarisch die für den innerstädtischen Verkehr relevanten Objektarten:

- 3101(*Straße*) - Zu dieser Objektart gehören Straßen deren Fahrtrichtungen nicht physisch voneinander getrennt sind. Autobahnausfahrten werden ebenfalls dieser Objektart zugeordnet.
- 3104 (*Straße (komplex)*) - Diese Objektart dient zur Beschreibung größerer Straßen und findet vor allem bei Autobahnen Anwendung. Zudem besitzt diese Objektart keine eigenen Attribute zur Angabe der Straßenbreite, da Objekte dieser Art durch Zusammenfassung von Objekten der Objektarten 3101, 3105 und 3106 gebildet werden.
- 3105 (*Straßenkörper*) - *Straßenkörper* beschreiben dem allgemeinen Verkehr dienende Flächen und werden zur Bildung der Objektart 3104 erfasst.
- 3106 (*Fahrbahn*) - Die Objektart beinhaltet voneinander getrennte Fahrspuren, wie diese auf Autobahnen und größeren Bundesstraßen zu finden sind. Objekte dieser Art werden zu *Straßenkörpern* zusammengefasst.

Die Beschreibung dieser Straßenobjekte erfolgt durch Angabe der Mittelachse in Form von Linien und deren mittleren Breite. Sofern die Breite einer Straße erfasst wurde, wird diese direkt in Dezimetern oder durch 3 *m*-Intervalle definiert. Für Objekte mit fehlender Breitenangabe wird in dieser Arbeit eine Mindestbreite von 6 *m* angenommen. Dies ist notwendig, da die Breite eines Großteils von Straßen-segmenten nicht definiert ist. Abbildung 5.5 zeigt drei Testgebiete mit überlagerten Straßensegmenten. Jedes Straßensegment ist in einer unterschiedlichen Farbe dargestellt. Für Segmente ohne die Angabe einer Breite wurde in der Abbildung eine Breite von einem Meter verwendet, wodurch diese sehr dünn erscheinen. Aus der Abbildung wird außerdem deutlich, dass die im ATKIS erfassten Segmente keine vollständigen Straßen abdecken, sondern nur Verbindungen zwischen Kreuzungspunkten sind.

5.3 Trainingsdaten

Zur Durchführung der überwachten Klassifikationen werden Trainingsdaten benötigt. Aus Beispielszenen wurden über 1200 Fahrzeuge erfasst (siehe Abbildung 5.6). Zu den erfassten Eigenschaften der Trainingsdaten gehören die Position, Orientierung und Ausdehnung. Außerdem wurden in den Beispielszenen

5.3. Trainingsdaten

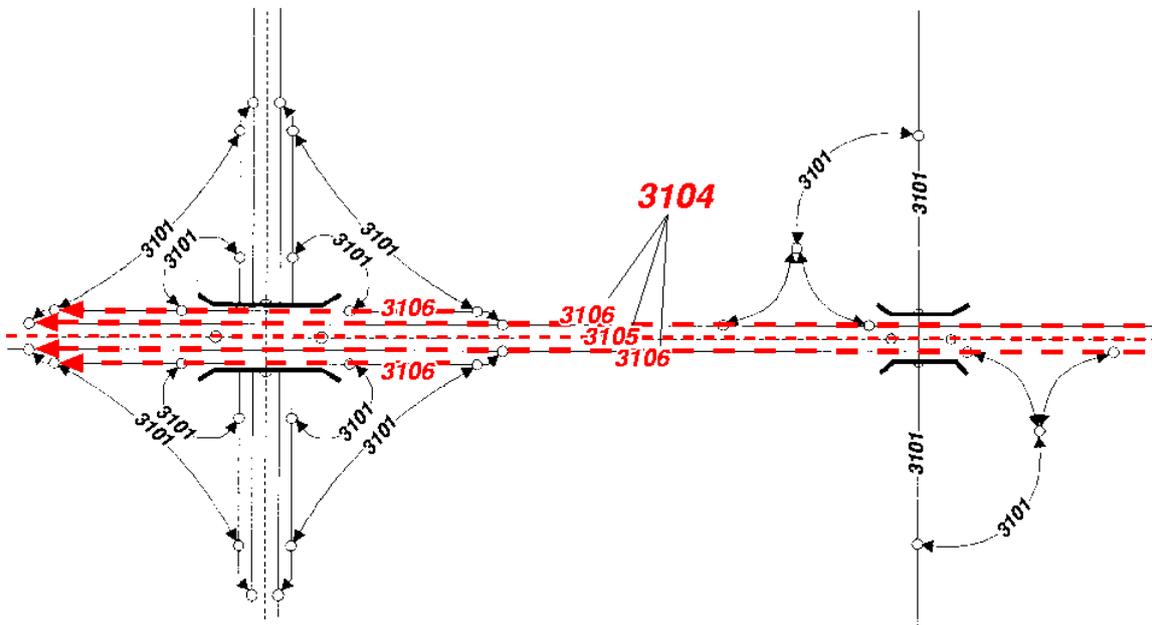


Abbildung 5.4: Objektarten des Basis-DLM (aus AdV (2003)) im Bereich einer Autobahnauffahrt

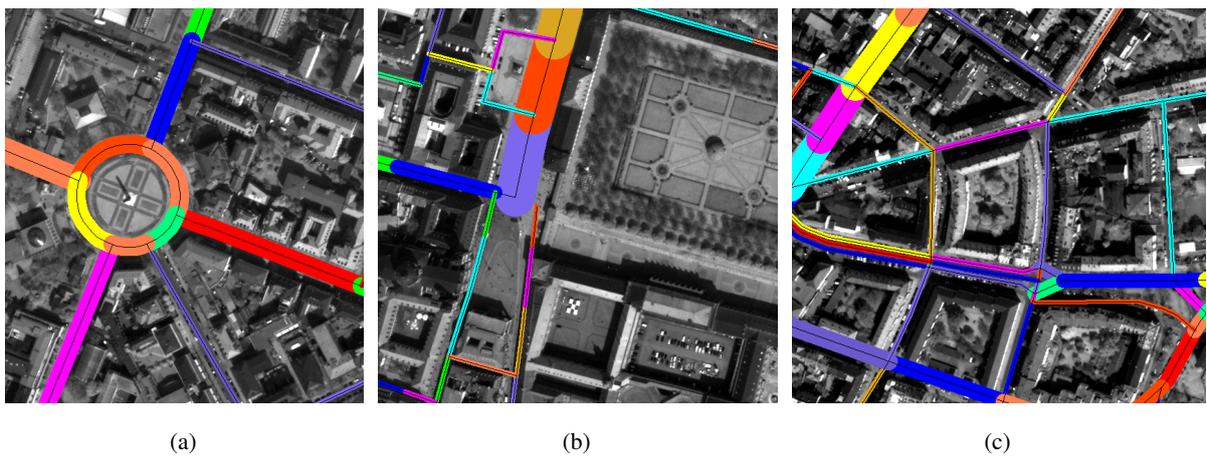


Abbildung 5.5: Straßensegmente und im ATKIS definierte Breitenangaben



Abbildung 5.6: Referenzdaten für helle (weiße Umrandung), dunkle (schwarze Umrandung) und graue (graue Umrandung) Fahrzeuge.

Flächen ausgewiesen, in denen keine Fahrzeuge vorhanden sind. Deren Auswahl beschränkte sich auf Straßenflächen und an Straßen angrenzende Gebiete, da durch die Bestimmung von Interessenregionen nur in diesen Bereichen die Klassifikation stattfindet. Innerhalb der ausgewiesenen Flächen wurden zufällig Bildausschnitte als Beispiele für die Fahrzeuggegenklasse selektiert.

6 Experimente und Ergebnisse

In diesem Kapitel wird der entwickelte Ansatz evaluiert. Zunächst wird in Abschnitt 6.1 die Initialisierung verwendeter Parameter beschrieben. Diese werden entweder aus den manuell erfassten Trainingsdaten abgeleitet oder durch Kreuzvalidierung bestimmt.

Anschließend werden in Abschnitt 6.2 einige Ergebnisse auf den Testdaten gezeigt und die numerische Evaluierung aller Testszenen gegeben.

6.1 Parameterinitialisierung

In Abschnitt 6.1.1 werden die aus den Trainingsdaten direkt ableitbaren Parameter berechnet. Damit können die Parameter der Linienextraktion (Abschnitt 6.1.2) und der robusten Parameterschätzung (Abschnitt 6.1.2) bestimmt werden. Die Initialisierung des AdaBoost Algorithmus (Abschnitt 6.1.3) sowie der Schlupf- und Kernelvariablen der SVM (Abschnitt 6.1.3) erfolgt durch Kreuzvalidierung.

6.1.1 Parameter aus den Trainingsdaten

Die Trainingsdaten besitzen folgende Eigenschaften, welche als Vektoren vorliegen:

- Koordinaten des Mittelpunktes \mathbf{x} und \mathbf{y}
- Fahrzeuglänge \mathbf{l}_1 und -breite \mathbf{l}_2
- Fahrzeugorientierung ϕ_F

Für die Größen \mathbf{l}_1 , \mathbf{l}_2 und ϕ_F wird aus den Trainingsdaten jeweils der Median berechnet und eine robuste Schätzung der Standardabweichung mittels des Medians der absoluten Abweichung durchgeführt (Formel 4.43). Über die Koordinaten des Fahrzeugmittelpunktes erfolgt die Zuordnung einer Straßenrichtung ϕ_S , wodurch sich die absolute Richtungsdifferenz zwischen ϕ_F und ϕ_S ergibt:

$$\Delta\phi = |\phi_F - \phi_S|$$

Außerdem wird für jedes Fahrzeug der Kontrast \mathbf{c} zur Umgebung berechnet. Dieser entspricht der absoluten Differenz zwischen dem Mittelwert aller Grauwerte innerhalb eines Fahrzeugs und dem Mittelwert aller Grauwerte des Straßensegments. Die Berechnung des Medians und der robusten Standardabweichung erfolgt im Fall des Kontrastes unabhängig für helle und dunkle Fahrzeuge, da deren Extraktion getrennt voneinander durchgeführt wird. Tabelle 6.1 fasst die ermittelten Mediane und robusten Standardabweichungen zusammen.

Parameter	Median	Standardabweichung
Fahrzeugbreite I_1	$\bar{I}_1 = 2,06m$	$\sigma_{I_1} = 0,21m$
Fahrzeuglänge I_2	$\bar{I}_2 = 3,67m$	$\sigma_{I_2} = 0,41m$
Richtungsdifferenz $\Delta\phi$	$\bar{\Delta\phi} = 3,41^\circ$	$\sigma_{\Delta\phi} = 3,36^\circ$
Kontrast (helle Fahrzeuge) c_h	$\bar{c}_h = 52$	$\sigma_{c_h} = 29$
Kontrast (dunkle Fahrzeuge) c_d	$\bar{c}_d = 32$	$\sigma_{c_d} = 19$

Tabelle 6.1: Ermittelte Kenngrößen und Abweichungen aus den Trainingsdaten

6.1.2 Parameter zur Extraktion von Einzelfahrzeugen aus Fahrzeugreihen

Parameterwahl für Linienextraktion und -filterung

Die Extraktion für helle und dunkle Linien erfolgt getrennt. Für beide Verarbeitungsschritte müssen obere und untere Schwellwerte für den Kontrast der Linienextraktion bestimmt werden. Diese ergeben sich aus den zuvor berechneten Kontrastwerten \bar{c}_h bzw. \bar{c}_d und deren Standardabweichung. Für helle Fahrzeuge gilt:

$$c_u = \bar{c}_h - \sigma_{c_h} = 23$$

$$c_o = \bar{c}_h = 59$$

Dunkle Fahrzeuge zeigen aufgrund der meist grau erscheinenden Straßen einen geringeren Kontrast:

$$c_u = \bar{c}_d - \sigma_{c_d} = 13$$

$$c_o = \bar{c}_d = 32$$

Mit diesen Kontrastwerten können Schwellwerte für die Krümmung t_u und t_o bestimmt werden (siehe Abschnitt 4.4.1). Der Glättungsfaktor σ der Linienextraktion berechnet sich in beiden Fällen aus der ermittelten Fahrzeugbreite von $\bar{I}_1 = 2,06m$. Die Linienextraktion für helle Fahrzeuge verwendet demnach folgende (gerundete) Schwellwerte:

$$\sigma = 1,0 \quad t_u = 7,2 \quad t_o = 18,5$$

Für dunkle Linien gilt entsprechend:

$$\sigma = 1,0 \quad t_u = 4,0 \quad t_o = 10,0$$

Die extrahierten Linien werden anhand ihrer Richtung, Breite und Länge auf Übereinstimmung zum Fahrzeugreihenmodell überprüft. Als Schwellwerte für diese Eigenschaften wird jeweils die Summe eines Parameters und dessen Standardabweichung aus den Trainingsdaten angesetzt. Die zulässige Richtungs-differenz ist durch $\bar{\Delta\phi} + \sigma_{\Delta\phi} = 6,77^\circ$ gegeben. Die extrahierten Linien müssen eine minimale Länge von $\bar{I}_2 + \sigma_{I_2} = 4,08m$ aufweisen und dürfen eine Breite von $\bar{I}_1 + \sigma_{I_1} = 2,27m$ nicht überschreiten. Die Fusion kollinearere Linien erfordert ebenfalls die Einhaltung einer maximalen Richtungs-differenz. Diese entspricht dem zweifachen Wert der zulässigen Richtungs-differenz zur Straße, um so auch gegensätz-lich ausschlagende Linien zu vereinigen. Die zulässige Richtungs-differenz kollinearere Linien beträgt

6.1. Parameterinitialisierung

daher $13,54^\circ$. Linien werden nur fusioniert, wenn der Abstand in Fahrzeurichtung eine Fahrzeuglänge $\bar{l}_2 = 3,67m$ nicht überschreitet. Quer zur Fahrzeurichtung beträgt der zugelassene Abstand die halbe Fahrzeugbreite $l_1/2 = 1,03m$.

Konfiguration der robusten Parameterschätzung

Vor der Erkennung von Einzelfahrzeugen aus Reihen müssen Intervallgrenzen innerhalb der Grauwertprofile definiert werden. Hierfür ist eine minimal anzunehmende Fahrzeuglänge l_{min} erforderlich. Diese wird durch die ermittelte Fahrzeuglänge $\bar{l}_2 = 3,67m$ abzüglich deren Standardabweichung $\sigma_{l_2} = 0,41m$ festgesetzt und beträgt $l_{min} = 3,26m$.

Die Wahl eines geeigneten M-Schätzers erfolgt durch empirische Ermittlung von Bruchpunkten. Zunächst werden Werte für eine Gauß-Glocke durch die Formel 4.31 erzeugt. In Abhängigkeit der Amplitude a_G der Gauß-Funktion wird auf einen variierenden Prozentsatz dieser Werte ein gleichförmig verteiltes Rauschen addiert. Der Wertebereich des Rauschens $a_R(t)$ mit dem Rauschniveau t ergibt sich durch $a_R(t) \in [0, \dots, t \cdot a_G]$. Abbildung 6.1 zeigt die verwendete Gauß-Funktion mit und ohne additivem Rauschen, bei einem Rauschniveau von $t = 0,25$ und einen Rauschanteil von 30 %. Zur Ermittlung des Bruchpunktes wird der Rauschanteil solange reduziert, bis die geschätzten Parameter der verrauschten Funktion mit denen der vorgegeben (simulierten) Funktion übereinstimmen. Für jedes t wurden zehn unabhängige Versuche durchgeführt und deren Mittelwerte berechnet. Die hierdurch erhaltenen empirischen Bruchpunkte sind in Abhängigkeit des Rauschniveaus in Tabelle 6.2 dargestellt. Das Ergebnis des jeweils besten Schätzers ist hervorgehoben. Der Logistic-Schätzers konnte nur bei rauschfreien Daten eine Übereinstimmung von geschätzten und Sollparametern im Bereich von 10^{-5} liefern. Daher wird er in der Tabelle nicht aufgeführt.

Aus Tabelle 6.2 wird erkennbar, dass einige Schätzer einen annähernd konstanten Bruchpunkt besitzen, der unabhängig von Rauschniveau ist. Dies trifft vor allem auf die hart abweisenden M-Schätzer zu. Der sehr niedrige Bruchpunkt des Andrews Schätzer bei $t = 5,00$ wird durch eine ungenaue Initialisierung der gesuchten Parameter durch das Levenberg-Marquardt Verfahren verursacht. Ein Großteil der fehlerfreien Beobachtungen wird bei hohem Rauschanteil durch die sehr restriktive Gewichtung des Schätzers in der ersten Iteration verworfen. Die Ausgleichung konvergiert in diesem Fall mit sehr hohen Varianzen. Basierend auf der empirischen Ermittlung der Bruchpunkte findet die Approximation der extrahierten Grauwertprofile durch einem M-Schätzer mit L1-Norm statt, da dieser insgesamt betrachtet, die günstigsten Bruchpunkte aufweist. Zudem müssen keine, für das vorliegende Problem eventuell ungeeignete, Tuningkonstanten (siehe Tabelle 4.4) verwendet werden.

6.1.3 Parameter zur Erfassung isolierter Fahrzeuge

Initialisierung des AdaBoost Algorithmus

Eine positive Eigenschaft des AdaBoost Algorithmus ist dessen Parameterfreiheit. Dies bedeutet, dass zu Beginn des Verfahrens keine Angaben zu dessen Konfiguration getroffen werden müssen. Der Algorithmus

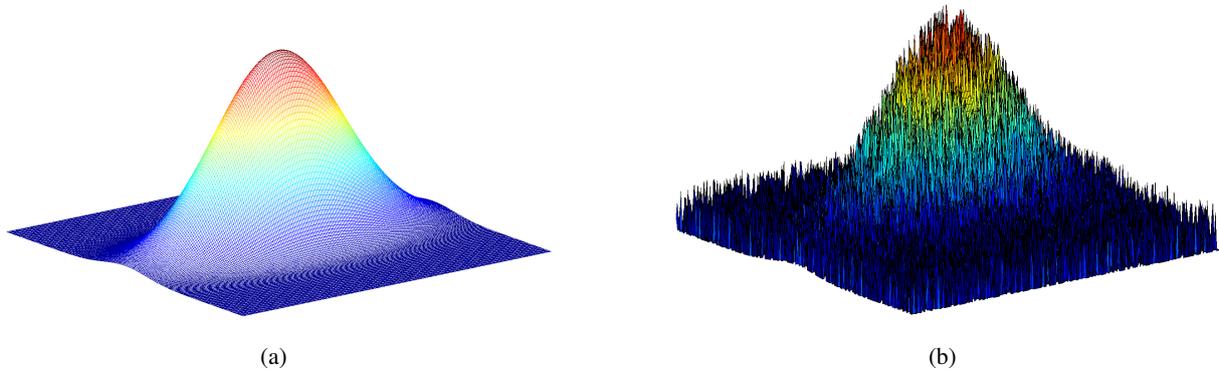


Abbildung 6.1: Simulierte Gauß-Funktion (a) mit und (b) ohne Rauschen

Schätzer	Vielfaches von a_R						
	$t = 0,25$	$t = 0,50$	$t = 1,00$	$t = 1,50$	$t = 2,00$	$t = 2,50$	$t = 5,00$
Andrews	0,48	0,49	0,46	0,45	0,46	0,46	0,17
Biweight	0,43	0,41	0,37	0,38	0,39	0,40	0,40
Talwar	0,46	0,46	0,40	0,41	0,42	0,43	0,43
Cauchy	0,42	0,43	0,38	0,36	0,37	0,35	0,27
Welsch	0,45	0,47	0,38	0,43	0,38	0,40	0,38
Huber	0,40	0,39	0,37	0,27	0,27	0,00	0,00
mod. Huber	0,47	0,47	0,42	0,43	0,44	0,44	0,41
Hampel	0,42	0,43	0,37	0,37	0,39	0,37	0,38
Fair	0,29	0,32	0,23	0,18	0,22	0,20	0,13
L1-Norm	0,47	0,48	0,47	0,47	0,47	0,47	0,47

Tabelle 6.2: Empirische Ermittlung der Bruchpunkte bei einer Rauschamplitude von $a_R \in [0, \dots, t \cdot a_G]$ in Abhängigkeit der Amplitude a_G einer simulierten Gauß-Glocke. Spaltenweise ist der jeweils beste Bruchpunkt einer Rauschamplitude hervorgehoben.

6.1. Parameterinitialisierung

adaptiert sich an die vorliegenden Daten. Allerdings beeinflussen die Größe des Bildausschnittes zur Berechnung der Haar-Merkmale und die Knotenanzahl der Klassifikationsbäume die erreichbare Fehlerrate der Klassifikation.

Die Größe der Bildausschnitte

Die Größe der Bildausschnitte zur Berechnung der Haar-Merkmale wird als konstant angenommen, da die Bodenpixelgröße des Bildes bekannt ist und unverändert bleibt. Im verwendeten Bildmaterial können einzelne Fahrzeugbestandteile nicht erkannt werden (siehe Abbildung 6.2a). Zur Beschreibung der Fahrzeugform ist es daher notwendig, einen Teil ihrer Umgebung in das Trainingsmuster mit einzubeziehen. Wird die Größe des Bildausschnitts zu klein gewählt, ist der Lernalgorithmus nicht in der Lage, die gesuchten Objekte von anderen homogenen Flächen zu unterscheiden (Abbildung 6.2a). In Abbildung 6.3a erzeugt die kleinste gewählte Größe für das Trainingsmuster auch den tendenziell größten Testfehler.

Werden zu große Bildbereiche als Trainingsmuster gewählt, können die Haar-Merkmale anderer markanter Strukturen den Trainingsdatensatz stark beeinflussen. Abbildungen 6.2b und 6.2c zeigen größer gewählte Bildausschnitte, in denen neben dem gesuchten Fahrzeug auch Bestandteile des angrenzenden Gebäudes und weiterer Fahrzeuge zur Heterogenität des Ausschnittes beitragen. Auch für diesen Fall ist aus Abbildung 6.3a ersichtlich, dass bei der maximal gewählten Größe der Trainingsmuster nicht der minimale Testfehler aller Größenvariationen erreicht wird. In Abhängigkeit einer maximal auszunehmenden Fahrzeuglänge von sechs Metern wird die Größe der Bildausschnitte zwischen 11 bis 17 Bildpunkten variiert. Die Trainingsmuster sind quadratisch, da Fahrzeuge in den Trainingsdaten mit beliebiger Orientierung vorkommen und hierdurch deren vollständige Erfassung garantiert wird. Abbildung 6.3a zeigt die numerische Evaluierung für alle untersuchten Größen des Trainingsmusters.

Anzahl an Knotenpunkten

Wie bereits in der Arbeit von Friedman et al. (2000) erwähnt und in Lienhart et al. (2003) eingehender untersucht, ist bei Verwendung von CART als schwachem Klassifikationsalgorithmus die Knotenanzahl entscheidend für den Trainingsprozess. Die Anzahl sollte nicht willkürlich gesetzt werden. Große Klassifikationsbäume können Abhängigkeiten zwischen Merkmalen besser erfassen, was zur schnelleren Reduzierung des Testfehlers führt. In Abbildung 6.3b haben Klassifikationsbäume mit mehr als einem Knoten zunächst einen deutlich niedrigeren Fehler als ein Schwellwertklassifikator. Findet aber eine Überanpassung an die Trainingsdaten innerhalb eines Baumes statt, beeinträchtigt dies die Generalisierungsfähigkeit des AdaBoost Algorithmus erheblich. Auch dieser Effekt ist in Abbildung 6.3b deutlich zu sehen. Der Klassifikator mit der höchsten gewählten Knotenanzahl hat nach 15 Iterationen den größten Testfehler. Der Schwellwertklassifikator (ein Knoten) zeigt hingegen ähnlich gute Ergebnisse wie Klassifikationsbäume mit drei oder fünf Knoten. Insgesamt wurden CART mit einem bis sechs Knoten untersucht. Abbildung 6.3b zeigt die Testfehler dieser Untersuchung.

Die optimale Wahl von Knotenanzahl und Größe des Bildausschnittes wird durch gleichzeitige Variati-

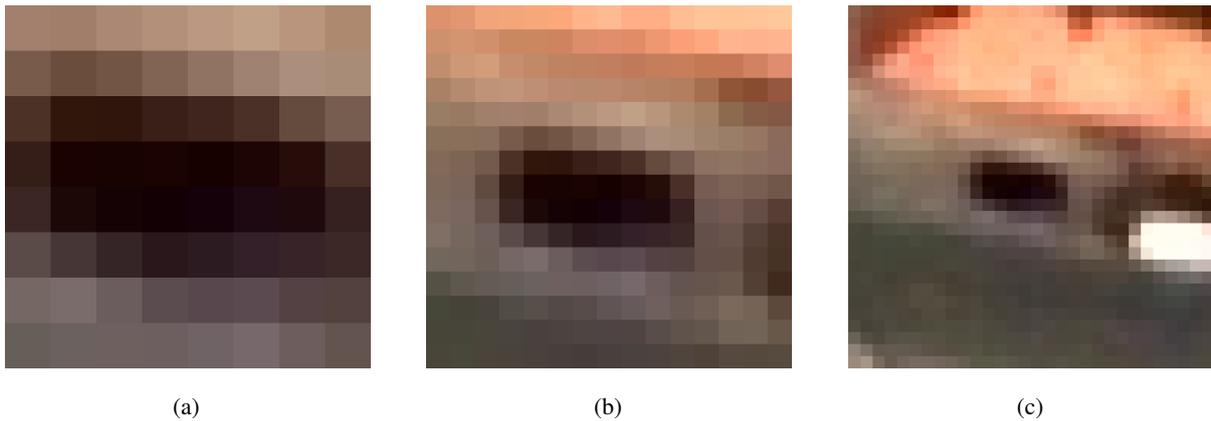


Abbildung 6.2: Einzelnes Fahrzeug bei variierender Größe des Bildausschnitts. (a) 8 mal 8 Bildpunkte, (b) 15 mal 15 Bildpunkte, (c) 30 mal 30 Bildpunkte

on beider Parameter untersucht (Abbildung 6.4). Der gesamte zur Verfügung stehende Trainingsdatensatz wird zufällig in fünf gleich große Teilmengen zerlegt. Das Training wird mit vier Teilmengen durchgeführt und der berechnete Klassifikator auf der verbleibenden Teilmenge validiert. Auf eine Kreuzvalidierung wurde aufgrund der sehr langen Trainingszeit pro Parameterkombination verzichtet. Der minimale Testfehler wird bei einem Klassifikationsbaum mit 4 Knoten und einer Größe der Trainingsmuster von 13 mal 13 Pixel erreicht.

Parameter der Support-Vector-Machine

Die ν -SVM und der verwendete RBF-Kernel hängen von zwei Parametern ab, welche die Fehlerrate der Klassifikation bestimmen:

1. dem Rauschanteil innerhalb der Trainingsdaten ν
2. dem Kernelparameter σ , der die Breite der gaußschen RBF bestimmt (siehe Formel 4.29)

Die Bestimmung dieser Parameter erfolgt durch fünffache Kreuzvalidierung. Der Trainingsdatensatz wird zufällig in fünf große Teilmengen zerlegt. Mit vier Teilmengen erfolgt das Training der SVM für einen Parametersatz von ν und σ . Die resultierende SVM wird zur Bestimmung der Fehlerrate auf der verbleibenden Teilmenge verwendet. Dieser Vorgang wird fünf Mal mit permutierenden Teilmengen wiederholt. Die Fehlerrate eines Parametersatzes entspricht dem Mittelwert der mit unterschiedlichen Teilmengen bestimmten Fehlerraten. Der Parameter ν wurde im Bereich von 0,0005 bis 0,0150 und σ zwischen 0,01 und 0,300 variiert. Abbildung 6.5 zeigt die erreichten Fehlerraten. Die geringste Fehlerrate von 0,074 bzw. 7,4 % wurde erreicht bei $\nu = 0,0145$ und $\sigma = 0,030$. Aus der Abbildung wird deutlich, dass die Variation von ν nur einen geringen Einfluss auf die Fehlerrate hat. Die Variation des Kernelparameters beeinflusst die Fehlerrate deutlich. Dies ist verständlich, da durch diesen Parameter die Transformation des Merkmalsraums bestimmt wird.

6.1. Parameterinitialisierung

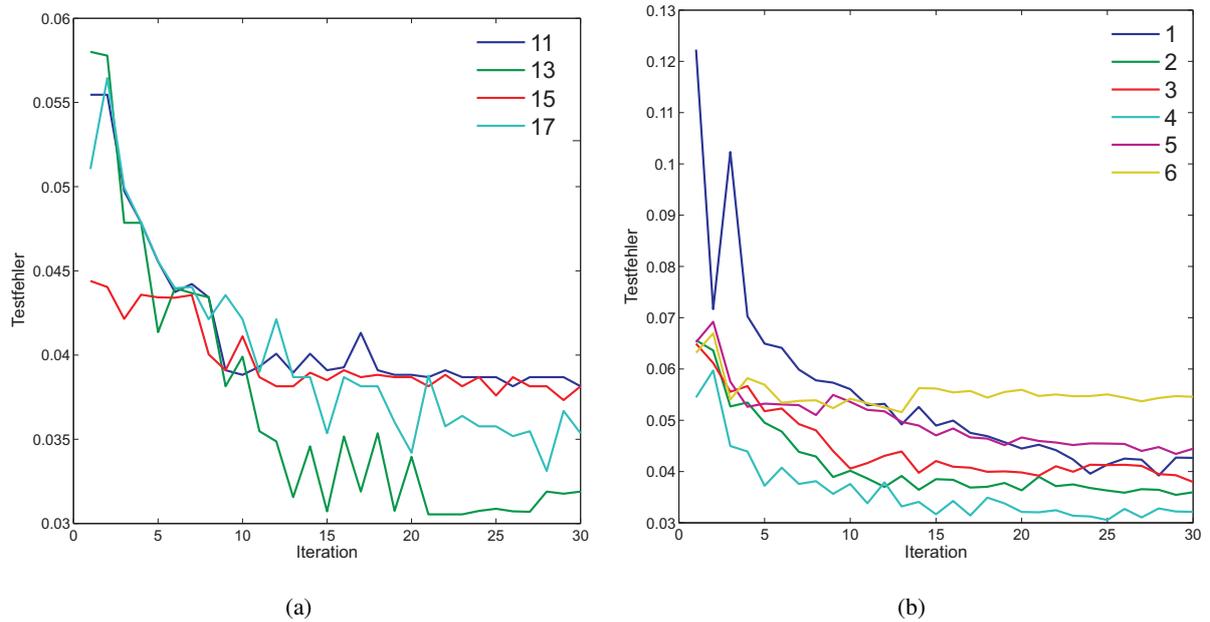


Abbildung 6.3: Testfehler im Verlauf des AdaBoost Trainings. (a) Variation der Größe des Trainingsmusters, jede Farbe entspricht einer anderen Größe des Trainingsmusters, (b) Variation der Anzahl von Knotenpunkten, jede Farbe entspricht einer anderen Knotenanzahl

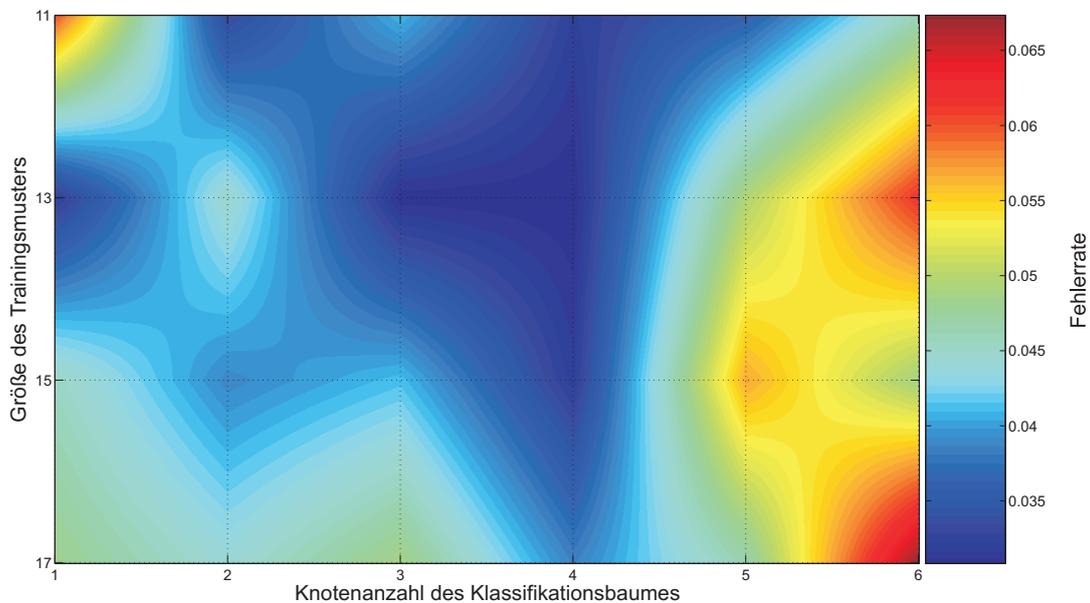
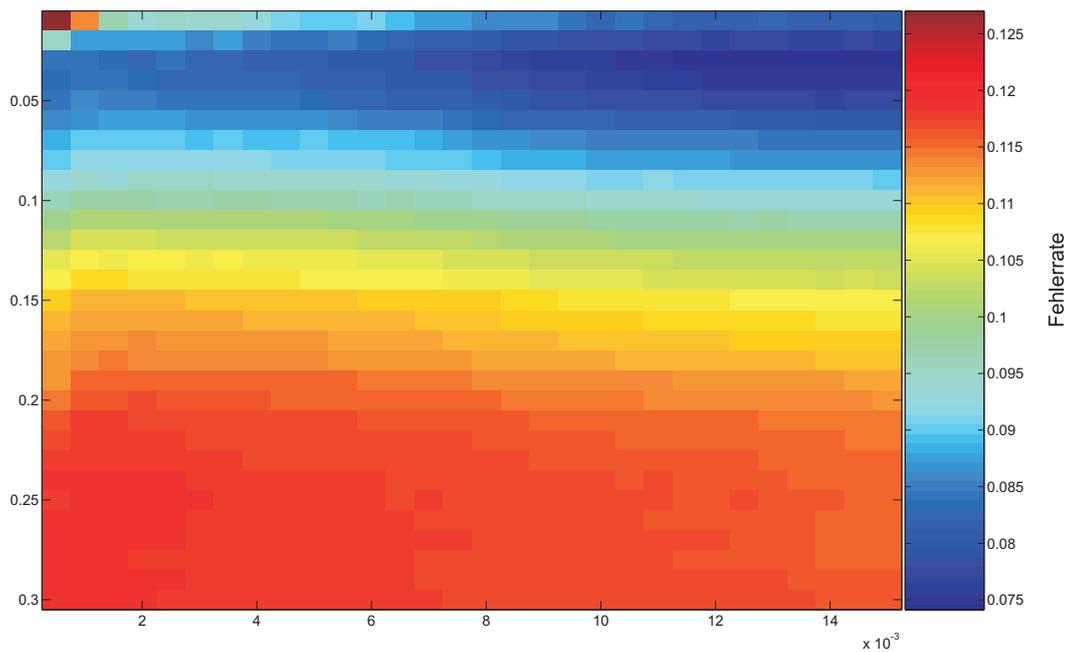


Abbildung 6.4: Testfehler bei gleichzeitiger Veränderung der Knotenanzahl und der Größe des Trainingsmusters nach 100 Iterationen

Abbildung 6.5: Bestimmung der Parameter einer ν -SVM durch fünffache Kreuzvalidierung

6.2 Ergebnisse auf den Testdatensätzen

6.2.1 Detektionsergebnisse

Die Bewertung der erzielten Ergebnisse erfolgt mit den Kriterien Zuverlässigkeit, Vollständigkeit und Qualität. Diese sind wie folgt definiert:

- **Zuverlässigkeit:** Spiegelt das Verhältnis zwischen richtigen und falschen Detektionen wider. Im Fall des optimalen Wertes von 100 % lägen keine falschen und mindestens eine richtige Detektion vor. Die Berechnung erfolgt durch

$$\text{Zuverlässigkeit} = \frac{\text{richtige Detektionen}}{\text{richtige} + \text{falsche Detektionen}} \cdot 100 [\%]$$

- **Vollständigkeit:** Entspricht dem prozentualen Anteil der Detektionen an den insgesamt in den Referenzdaten enthaltenen Fahrzeugen. Der optimale Wert von 100 % wird erreicht, wenn keines der in den Referenzdaten enthaltenen Fahrzeuge fehlt. Die Vollständigkeit wird berechnet durch

$$\text{Vollständigkeit} = \frac{\text{richtige Detektionen}}{\text{richtige Detektionen} + \text{fehlende Fahrzeuge}} \cdot 100 [\%]$$

- **Qualität:** Dies ist das strengste Bewertungskriterium, da sowohl falsche Detektionen als auch fehlende Fahrzeuge die Qualität negativ beeinflussen. Der optimale Wert von 100 % wird nur erreicht, wenn keine falschen Detektionen vorliegen und alle in den Referenzdaten enthaltenen Fahrzeuge erkannt wurden. Die Berechnung erfolgt durch

$$\text{Qualität} = \frac{\text{richtige Detektionen}}{\text{richtige Detektionen} + \text{falsche Detektionen} + \text{fehlende Fahrzeuge}} \cdot 100 [\%]$$

6.2. Ergebnisse auf den Testdatensätzen

Richtige Detektionen werden als *True Positives* (TP), falsche Detektionen als *False Positives* (FP) und fehlende Detektionen als *False Negatives* (FN) bezeichnet. Eine getrennte numerische Evaluierung für die exklusive Erfassung von Fahrzeugen aus Reihen und isolierter Fahrzeuge ist in Tabelle 6.3 gegeben. Als weitere bewertete Gruppen sind in dieser Tabelle beidseitig detektierte Fahrzeuge aufgelistet, welche sowohl aus Reihen als auch durch die Klassifikation erfasst wurden. Die Berechnung der Zuverlässigkeit geschieht aus den jeweils aufgelisteten Werten für richtige und falsche Detektionen. Die Vollständigkeit bezieht sich auf die Gesamtanzahl der im Bild enthaltenen Fahrzeuge. Würden alle Fahrzeuge in einem Bild detektiert werden, entspräche die Summe der Vollständigkeit für Fahrzeugreihen, isolierte und beidseitig detektierte Fahrzeuge 100 %.

Testgebiet	Fahrzeugreihen				isolierte Fahrzeuge				beidseitig detektierte Fahrzeuge			
	TP	FP	Vollst.	Zuverl.	TP	FP	Vollst.	Zuverl.	TP	FP	Vollst.	Zuverl.
1	60	1	36,1%	98,4%	11	2	6,6%	84,6%	77	1	46,4%	98,7%
2	14	0	9,3%	100,0%	34	10	22,5%	77,3%	58	2	38,4%	96,7%
3	25	1	14,6%	96,2%	54	11	31,6%	83,1%	53	3	31,0%	94,6%
4	29	0	16,7%	100,0%	22	5	12,6%	81,5%	81	1	46,6%	98,8%
5	15	0	11,8%	100,0%	6	1	4,7%	85,7%	63	0	49,6%	100,0%
6	13	0	10,2%	100,0%	38	1	29,9%	97,4%	18	2	14,2%	90,0%
7	23	0	6,3%	100,0%	73	4	19,8%	94,8 %	71	4	19,3 %	94,7 %
Gesamt	179	2	13,9 %	98,9 %	238	34	18,5 %	87,5 %	421	13	32,8 %	97,0 %

Tabelle 6.3: Evaluierung der Fahrzeugerkennung für die Testgebiete. Unterteilung für exklusiv aus Fahrzeugreihen, isolierte und beiseitig detektierte Fahrzeuge

Die Ergebnisse des gesamten Ansatzes sind in Tabelle 6.4 aufgeführt. Hierzu wurden die entsprechende Werte für TP und FP aus Tabelle 6.3 aufsummiert und die Anzahl fehlender Fahrzeuge bestimmt. Eine ausführliche Diskussion dieser numerischen Evaluierung findet im folgenden Kapitel statt.

Testgebiet	TP	FP	FN	Zuverlässigkeit	Vollständigkeit	Qualität
1	148	4	18	97,4 %	89,2 %	87,1 %
2	106	12	45	89,8 %	70,2 %	65,0 %
3	132	15	39	89,8 %	77,2 %	70,1 %
4	132	6	42	95,7 %	75,9 %	73,3 %
5	84	1	43	98,8 %	66,1 %	65,6 %
6	69	3	58	95,8 %	54,3 %	53,1 %
7	167	8	201	95,4 %	45,3 %	44,4 %
Gesamt	838	49	446	94,5 %	65,3 %	62,9 %

Tabelle 6.4: Zusammenfassende Evaluierung der Fahrzeugerkennung für die Testgebiete

Abbildung 6.6 zeigt die beiden Szenen mit der höchsten erreichten Qualität, während in Abbildung 6.7 Ergebnisse mit niedriger Vollständigkeit und daraus resultierender geringer Qualität abgebildet sind. In diesen Abbildungen sind die richtigen Detektionen durch grüne und fehlende Fahrzeuge mit roten Rechtecken gekennzeichnet. Falsche Detektionen sind durch gelbe Kreise markiert.

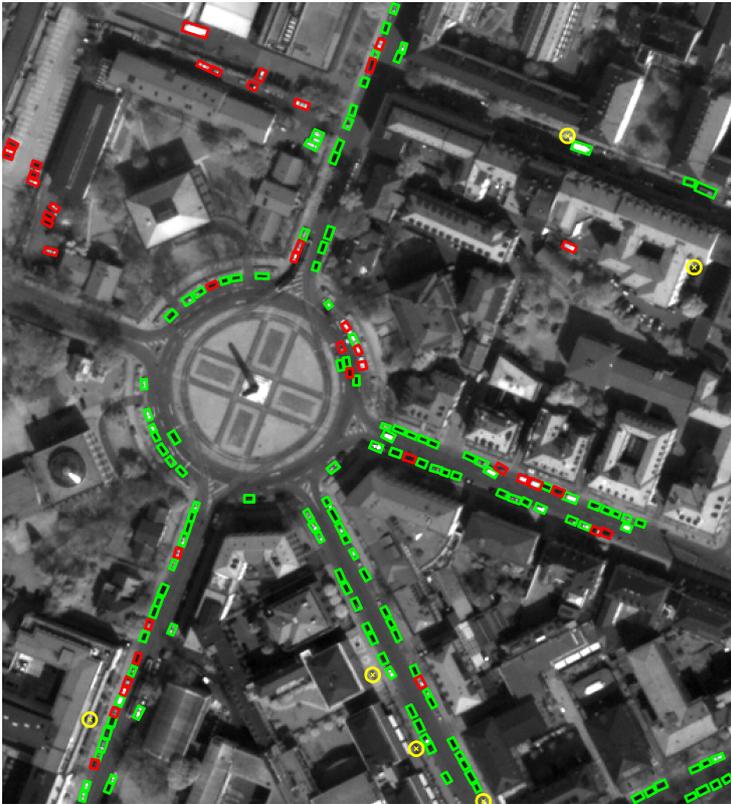
6.2.2 Ergebnisse der Bewegungsschätzung

Insgesamt wurden 238 Fahrzeuge als isoliert markiert (Tabelle 6.3). Für diese Fahrzeuge wurde eine Schätzung des Bewegungszustandes durchgeführt. Für neun Fahrzeuge konnte eine signifikante Verschiebung zwischen der Position im panchromatischen Bild und in den multispektralen Kanälen festgestellt werden. In den anderen Fällen ist keine ausreichend hohe Übereinstimmung durch das Matchingverfahren berechnet worden oder der ermittelte Abstand war kleiner als der Schwellwert von $1,2\text{ m}$. Abbildung 6.8 zeigt die bewegten Fahrzeuge, deren Versatz und die daraus resultierende Geschwindigkeit. Fahrzeuge ohne signifikanten Versatz sind durch rote Kreise gekennzeichnet. In Abbildung 6.8a ist ein weiteres rotes Fahrzeug mit einem deutlichen Versatz zwischen den Spektralkanälen erkennbar. Dieses wurde nicht detektiert, weshalb keine Bewegungsschätzung für das Fahrzeug durchgeführt werden konnte. Bei dem Fahrzeug in Abbildung 6.8i ist anzunehmen, dass es am Straßenrand parkt. Der berechnete Versatz von $1,5\text{ m}$ wäre für ein parkendes Fahrzeug falsch.

6.2. Ergebnisse auf den Testdatensätzen

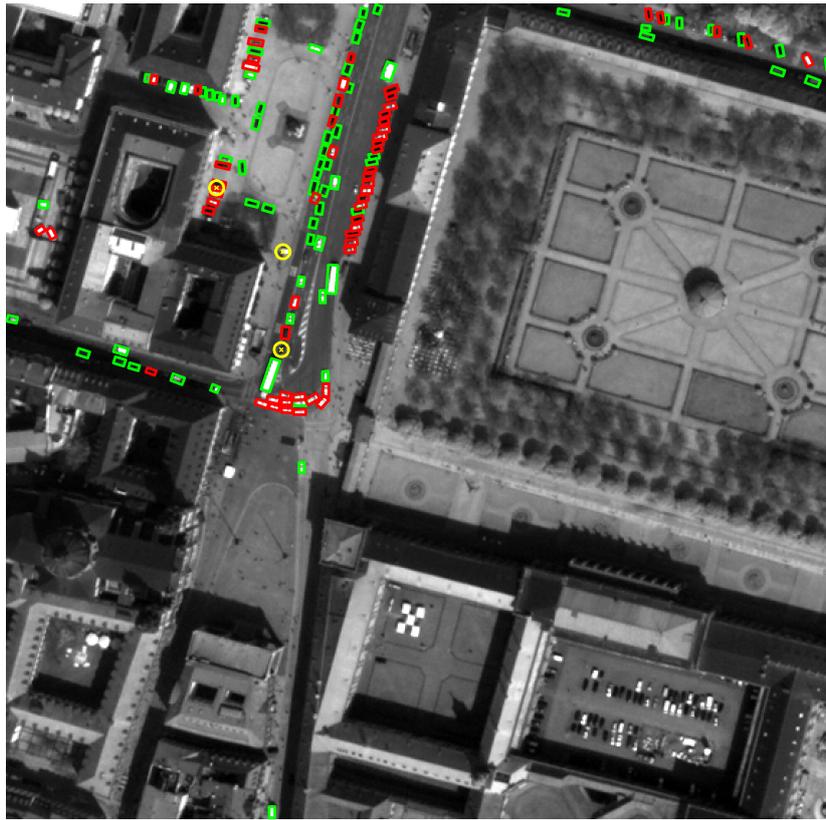


(a)



(b)

Abbildung 6.6: Szenen mit hoher erreichter Qualität. (a) Testgebiet 1, (b) Testgebiet 4



(a)



(b)

Abbildung 6.7: Szenen mit niedriger erreichter Qualität. (a) Testgebiet 6 , (b) Testgebiet 7

6.2. Ergebnisse auf den Testdatensätzen



Abbildung 6.8: Ergebnisse der Bewegungsschätzung. Der grüne Pfeil verläuft von der Fahrzeugposition im panchromatischen Kanal zur Position in den multispektralen Bändern. Die Länge des Pfeils ist proportional zur Geschwindigkeit. Isolierte Fahrzeuge ohne ermittelten Versatz sind durch rote Kreise gekennzeichnet.

6.2. Ergebnisse auf den Testdatensätzen

7 Diskussion

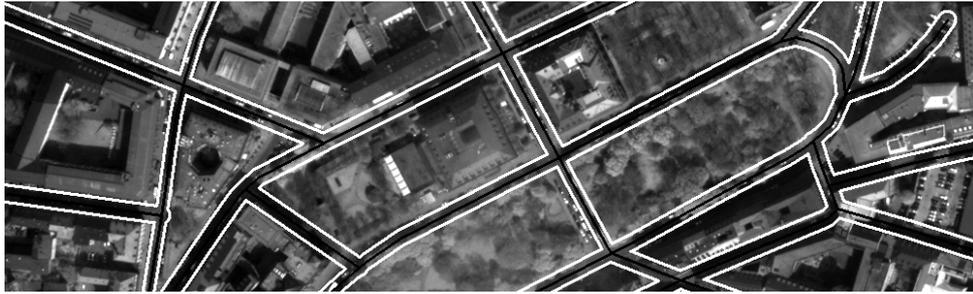
7.1 Ungenauigkeiten der verwendeten Geoinformationen

Nach Durchführung der in Abschnitt 4.2 beschriebene Transformation konnten verbleibende Klaffungen zwischen Bild- und Kartenmaterial beobachtet werden, welche auf die unzureichende Lagegenauigkeit der QuickBird-Daten zurückzuführen sind. Die Positionsunterschiede sind größtenteils parallel zu den Koordinatenachsen. Eine manuelle Überprüfung ergab eine Verschiebung des QuickBird Bildes von ungefähr $-3,0\text{ m}$ in vertikaler und $1,2\text{ m}$ in horizontaler Richtung gegenüber den ATKIS Daten.

Abbildung 7.1 zeigt Ergebnisse nach der ATKIS Transformation und einmaliger (manueller) Lagekorrektur der QuickBird Szene. Abbildung 7.1a verdeutlicht die meist hohe geometrische Genauigkeit der verwendeten Straßeninformationen. Eine Hauptquelle für Fehler im Rahmen der Fahrzeugdetektion ist in Abbildung 7.1b erkennbar. Da die ATKIS Daten auf Basis der Topographischen Karte 25 erfasst werden (diese liegt im Maßstab 1:25.000 vor), treten bei komplexen Kreuzungen starke Generalisierungsfehler auf. In Abbildung 7.1b verlaufen einzelne Fahrbahnspuren durch anliegende Häuserreihen. Dies kann nicht nur zu falschen sondern auch fehlenden Detektionen führen, da die Versetzung der Fahrbahnspuren zu einer unvollständigen Abdeckung der Straße führt. Die Behebung dieser Fehlerquelle ist durch Korrektur der ATKIS Daten oder durch Verwendung alternativer Straßeninformationssysteme - z.B. NAVTEQ, TeleAtlas oder OpenStreetMap - möglich. Alternativ wurde in dieser Arbeit die Breite aller Straßensegmente um 15% vergrößert. Dadurch treten zwar weniger fehlende Detektionen auf, allerdings ist dieser Vorgehensweise ein Teil der falschen Detektionen zu zuschreiben. In Abbildung 7.2 sind einige durch die Vergrößerung der Straße und durch Ungenauigkeiten der Geoinformation verursachte Falschdetektionen durch gelbe Kreise markiert.

7.2 AdaBoost Verfahren und Haar-Merkmale

Im Rahmen des AdaBoost-Trainings werden bestimmte Merkmale häufiger ausgewählt, wenn diese besonders gut die Blob-hafte Struktur von Fahrzeugen erfassen. In Tabelle 7.1 sind die Häufigkeiten bei der Auswahl der verwendeten Haar-Merkmale aufgeführt, welche in unterschiedlicher Größe und Position innerhalb des Suchfensters auftraten. Es ist besonders auffällig, dass die Merkmale der Form 3a und 3b in über 40% der Fälle selektiert wurden. Untersucht man das durch den Boosting Algorithmus zuerst gewählte Merkmal 3a (in der Größe $9\text{ mal }9$, zentriert auf die Spalte 6 und Zeile 7), wird deutlich, wie



(a)



(b)

Abbildung 7.1: Genauigkeit von ATKIS Daten nach Transformation. (a) Hohe Genauigkeit nach Transformation, (b) Generalisierungsfehler in ATKIS Daten

gut bereits mit diesem einen Merkmal Fahrzeuge vom Hintergrund getrennt werden können. Für alle Bildpunkte des Ausschnitts 7.3a wurde dieses Merkmal berechnet. Das Ergebnis der Berechnung ist in Abbildung 7.3b dargestellt. Anschließend wurden zwei manuell gewählte Schwellwerte auf die Merkmalswerte angewandt, wodurch das Bild in drei verschiedene Regionen segmentiert wird. In Abbildung 7.3c beinhalten grüne Regionen alle dunklen Fahrzeugen, während innerhalb der roten Regionen alle hellen Fahrzeuge liegen. Es werden zwar auch Regionen erfasst, welche keine Fahrzeuge sind, insgesamt kann aber auf Basis dieses einen Merkmals, mehr als 85 % des Bildes als Hintergrund erkannt werden.

Neben der beschriebenen Auswahl der markantesten Merkmale ist die deutliche Merkmalsreduzierung ein großer Vorteil des gewählten AdaBoost-Verfahrens. Das Training durchlief bei einer Knotenzahl von vier und einer Größe des Suchfensters von 13 mal 13 Bildpunkte insgesamt 300 Iterationen. Von den ursprünglich zur Verfügung stehenden 20243 Merkmalen wurden nur 1031 Merkmale selektiert. Diese Reduzierung um 95 % wirkt sich während der Online-Phase linear auf die zur Klassifikation benötigte Berechnungszeit aus.

7.3 Parameter zur Erfassung von Fahrzeugreihen

Die Extraktion von Fahrzeugreihen benötigt zehn unterschiedliche Parameter (Abschnitt 6.1.2). Aufgrund der möglichen Parameterkombinationen im Falle einer manuellen Initialisierung, wurden alle Parameter aus den Trainingsdaten für die Klassifikation isolierter Fahrzeuge abgeleitet (siehe Abschnitt 6.1.2).

7.3. Parameter zur Erfassung von Fahrzeugreihen

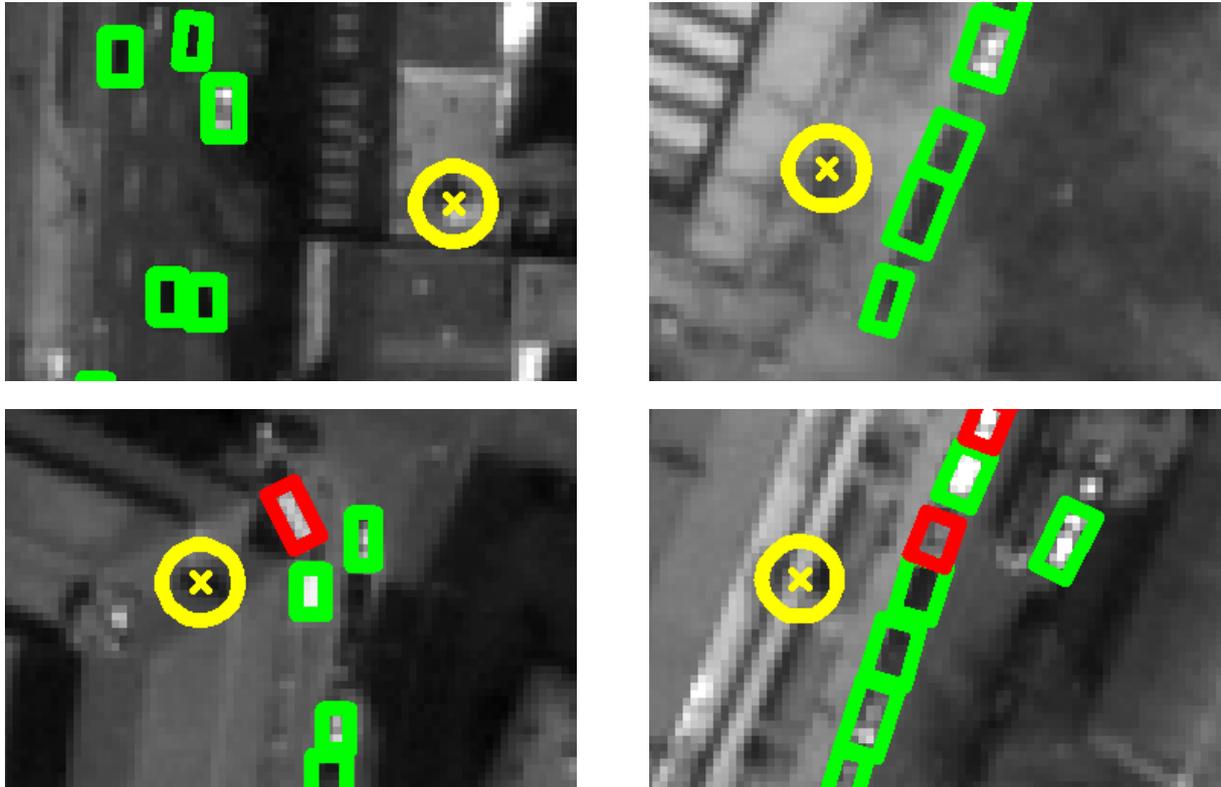


Abbildung 7.2: Fehler durch Ungenauigkeit der Geoinformation und Verbreiterung der Straßensegmente. Falsche Detektionen durch gelbe Kreise markiert, richtige Detektionen durch grüne und fehlende Detektionen durch rote Umrandungen markiert gekennzeichnet

Standardmerkmale		rotierte Merkmale	
1a	3,7 %	1c	11,5 %
1b	3,7 %	1d	10,7 %
2a	3,4 %	2e	2,4 %
2b	2,7 %	2f	1,8 %
2c	3,5 %	2g	1,9 %
2d	1,4 %	2h	0,8 %
3a	22,1 %	3b	19,3 %
4	11,1 %		

Tabelle 7.1: Prozentualer Anteil verschiedener Merkmale am Gesamtklassifikator

7.3. Parameter zur Erfassung von Fahrzeugreihen

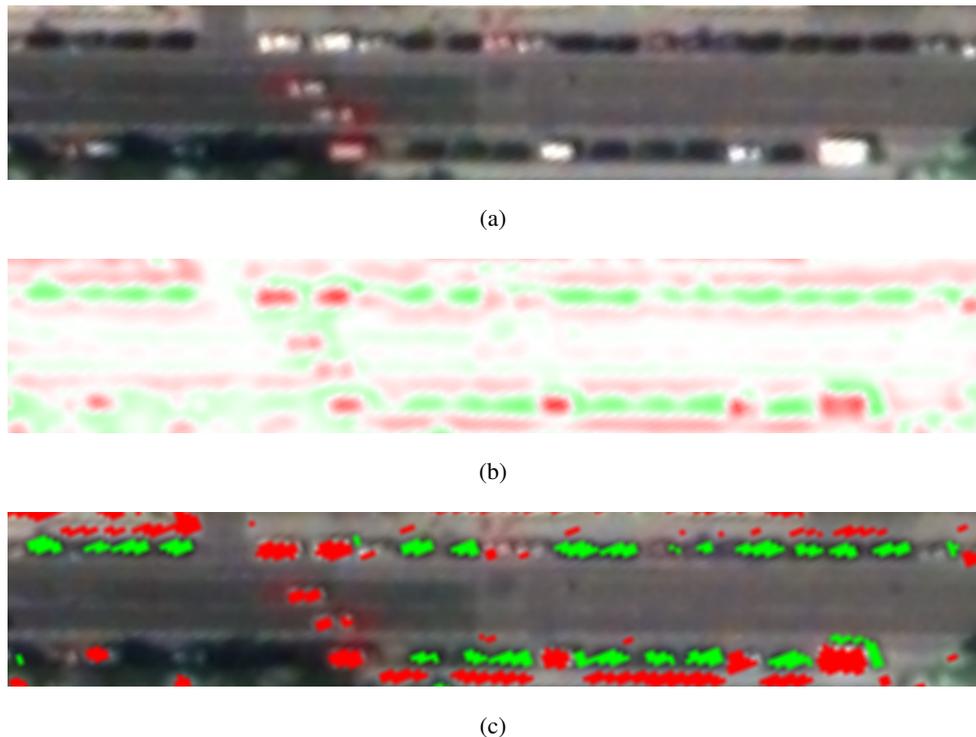


Abbildung 7.3: Anwendung des ersten gewählten Haar-Merkmals. (a) Eingabebild, (b) mit Haarmerkmal gefiltertes Bild, (c) Segmentierung des gefilterten Bild durch zwei Schwellwerte

Abbildung 7.4 zeigt repräsentative Beispiele der Linienextraktion bevor diese nach den Kriterien Breite, Länge und Übereinstimmung zur Straßenrichtung gefiltert werden. Es treten viele Linien auf, welche nicht Fahrzeugreihen zugeordnet werden können. Es werden aber alle Fahrzeugreihen durch die extrahierten Linien abgedeckt und stehen so für weitere Verarbeitungsschritte zur Verfügung. Der zur Extraktion verwendete Kontrast und die Breite sind daher günstig gewählt. Die Anzahl der verbleibenden Linien nach der Filterung ist stark abhängig vom gewählten Merkmal.

So erfüllt die Mehrheit der Linien das Kriterium der maximalen Linienbreite (siehe Tabelle 7.2). Dies liegt auch am gewählten Verfahren, welches die Breite der gesuchten Linien bereits zu deren Extraktion nutzt. Die rot markierten Linien in Abbildung 7.5 besitzen eine zu große Breite und werden nicht weiter berücksichtigt. Grüne Linien erfüllen das Breitenkriterium. Aus der Abbildung ist erkennbar, dass mit Ausnahme des Busses/Lkws in Abbildung 7.5d alle Fahrzeugreihen weiterhin durch (grüne) Linien erfasst sind.

Bei Überprüfung der Länge der extrahierten Linien werden für helle Linien über 80 % und bei dunklen Linien annähernd 75 % verworfen (siehe Tabelle 7.2). Hiervon sind auch Fahrzeuge in Reihen betroffen, wenn sie eine andere Farbe als die umgebende Fahrzeugreihe haben. Die Linienlänge der weißen Fahrzeuge in den Abbildungen 7.6a und 7.6b ist daher zu kurz. Allerdings sind diese Fahrzeuge deutlich als Einzelfahrzeuge erkennbar und können so durch die Klassifikation erfasst werden (siehe Abbildung 7.9a und 7.9b). In Abbildung 7.6d werden auch Fahrzeugreihen verworfen, obwohl diese eine annähernd einheitliche Farbe besitzen. Diese Reihen sind aus mehreren kleinen Linien zusammengesetzt,

7.3. Parameter zur Erfassung von Fahrzeugreihen

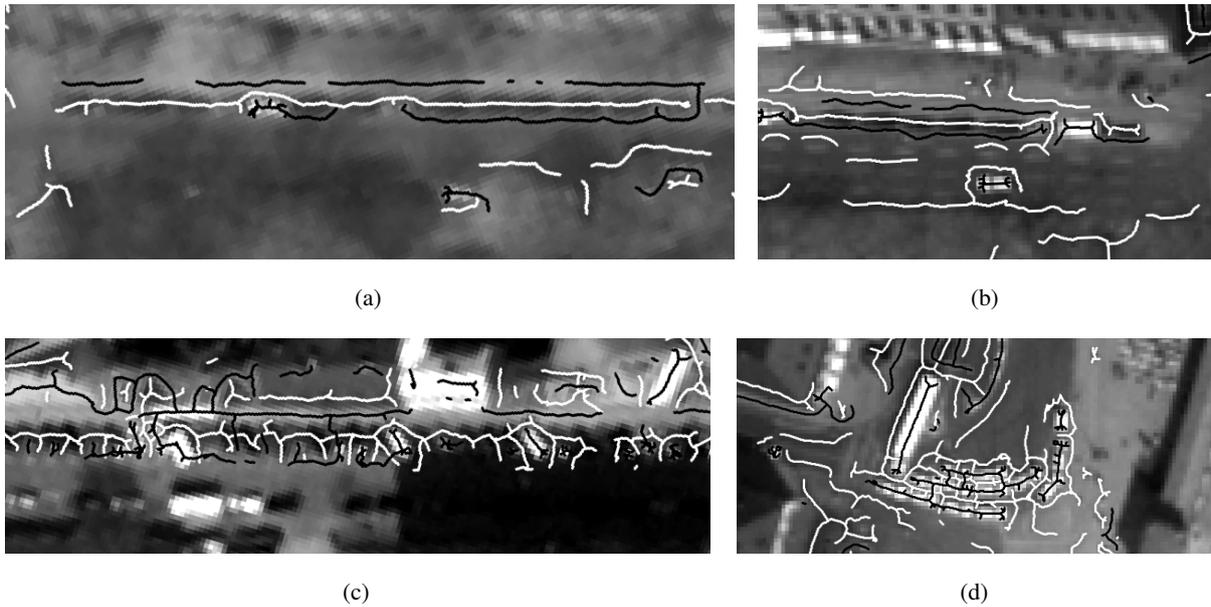


Abbildung 7.4: Ergebnisse der Linienextraktion. Helle Linien sind schwarz und dunkle Linien sind weiß gekennzeichnet

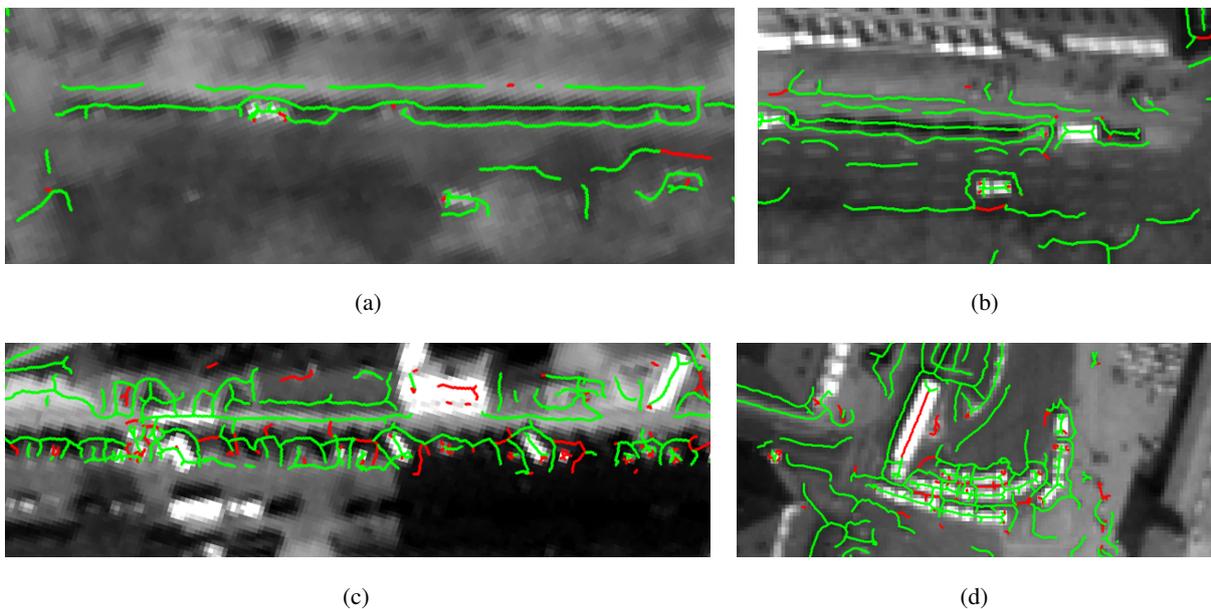


Abbildung 7.5: Filterung der Linien bezüglich der Breite. Grüne Linien erfüllen das Breitenkriterium, rote Linien haben eine zu große Breite

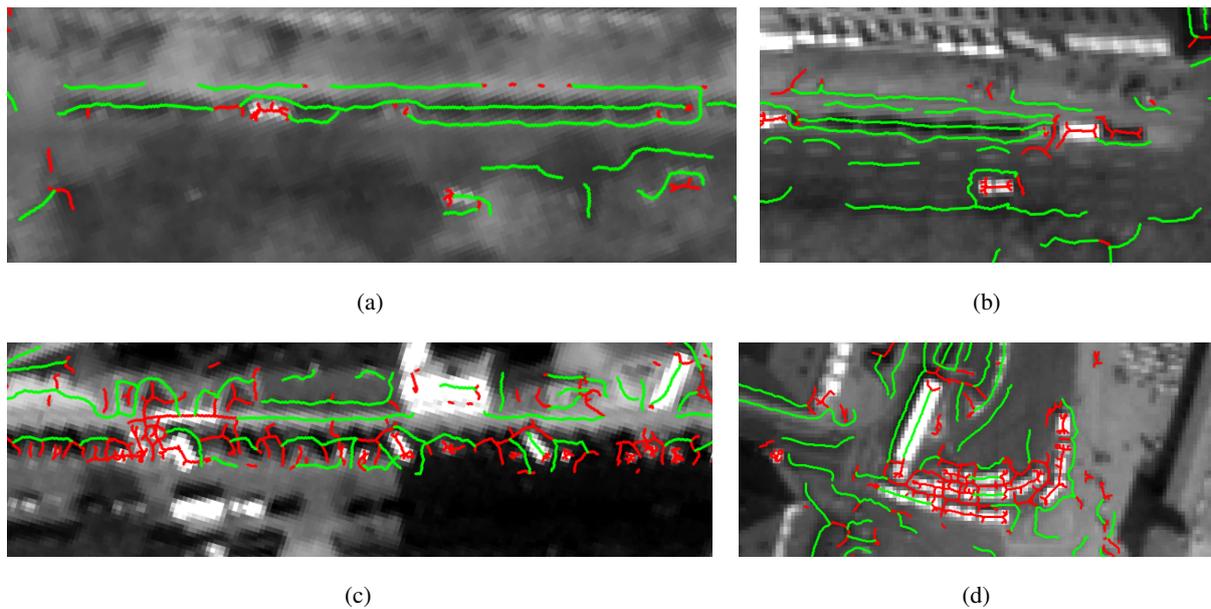


Abbildung 7.6: Filterung der Linien bezüglich der Länge. Grüne Linien erfüllen das Längenkriterium, rote Linien haben eine zu geringe Länge

welche einzeln betrachtet eine zu geringe Länge aufweisen. Der Schwellwert für die Länge von Linien ist in diesen Fällen zu hoch angesetzt. Eine Verringerung erscheint zwar sinnvoll, allerdings würde es sich bei den verbleibenden Linien dann nicht mehr um „Reihen“ handeln. Die Fahrzeuge dieser Abbildung können wegen ihres geringen Abstandes zu einander auch nicht als isolierte Fahrzeuge vollständig detektiert werden (siehe Abbildung 7.9d).

Die fehlende Übereinstimmung zwischen Linien- und Straßenrichtung führt zum größten Ausschluss von Linien (siehe Tabelle 7.2). Besonders offensichtlich wird das in den Abbildungen 7.7c und 7.7d. In Abbildung 7.7c widersprechen alle quer zur Straßenrichtung geparkten Fahrzeuge dem Fahrzeugreihenmodell, welches von parallel zur Fahrtrichtung vorkommenden Fahrzeugen ausgeht. Die Detektion dieser Fahrzeuge ist auch durch die Klassifikation nicht möglich (siehe Abbildung 7.9c), da die Fahrzeuge zu dicht nebeneinander stehen. In Abbildung 7.7d wird der Großteil der Linien verworfen, da diese starke Krümmungen aufweisen und so der Mittelwert der Richtungen aller Teilstücke einer Linie nicht mit der Richtung der Straße übereinstimmt. In beiden Fällen ist der Ausschluss von Fahrzeugreihen, nicht auf eine zu gering gewählte Richtungsdivergenz zurückzuführen, sondern auf die unzureichende Modellierung solcher Fahrzeugkonfigurationen.

Erfüllt eine Linie alle Kriterien, wird diese zur Fusion der hellen und dunklen Linien verwendet. Abbildung 7.8 zeigt Ergebnisse der vollständigen Filterung. In den Abbildungen 7.8a und 7.8b sind die vorhandenen Fahrzeugreihen sehr gut durch Linien erfasst. Im Gegensatz hierzu werden die Fahrzeugreihen in den Abbildungen 7.8c und 7.8d aus den zuvor erläuterten Gründen nur unzureichend oder nicht abdeckt.

7.3. Parameter zur Erfassung von Fahrzeugreihen

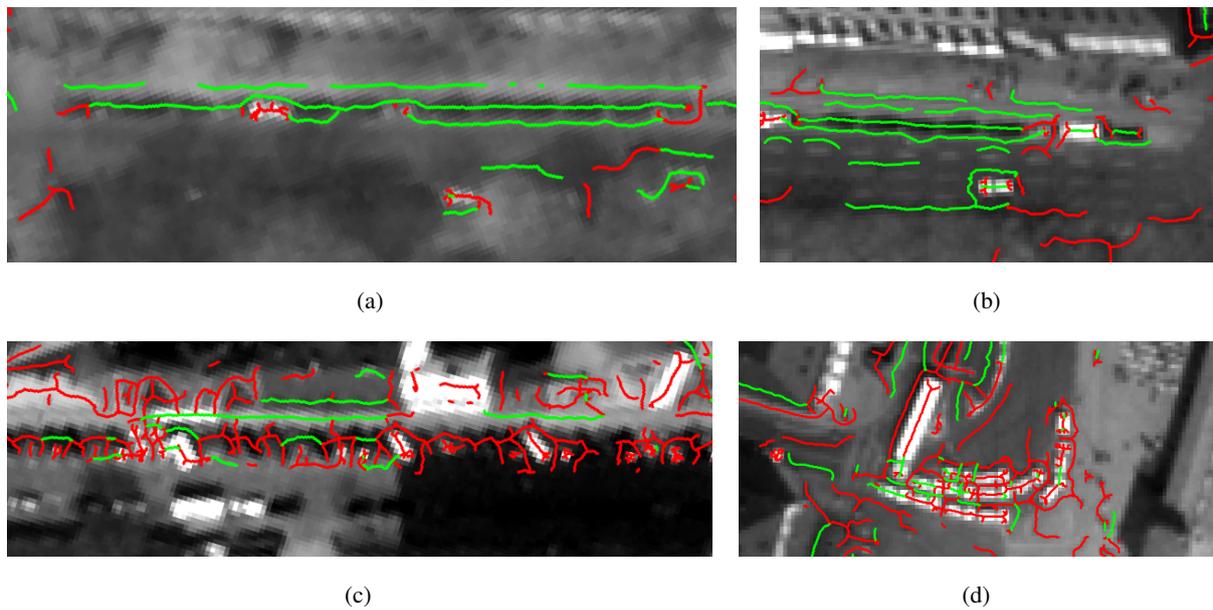


Abbildung 7.7: Filterung der Linien bezüglich der Richtung. Grüne Linien besitzen eine zur Straßenrichtung korrespondierende Orientierung, rote Linien erfüllen das Richtungskriterium nicht

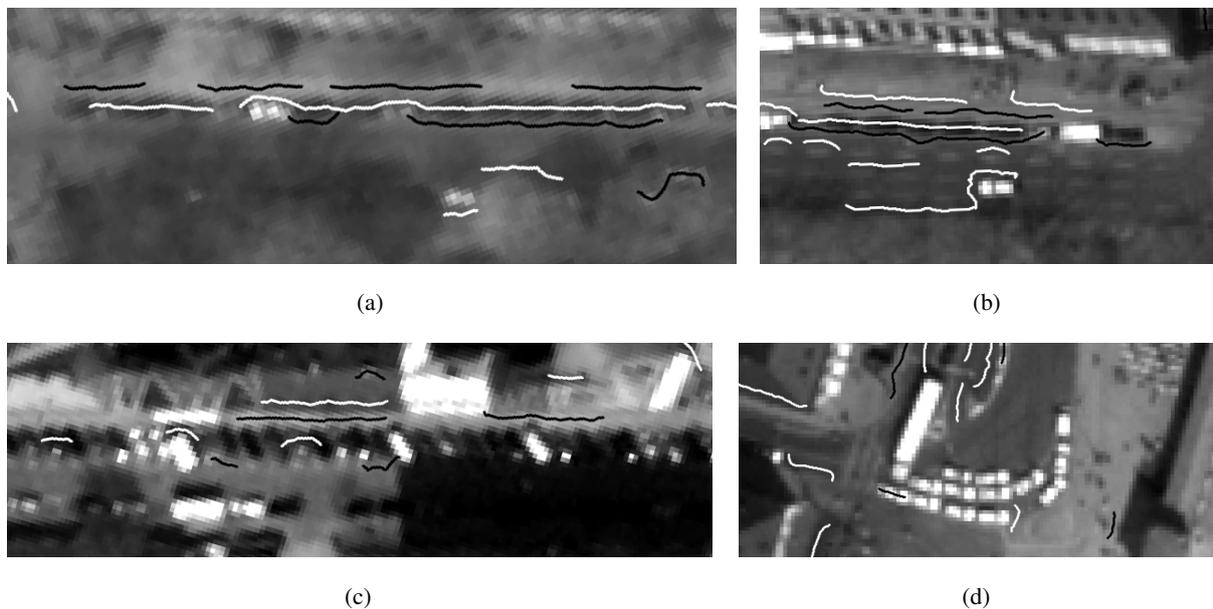


Abbildung 7.8: Ergebnis der gesamten Filterung. Helle Linien sind schwarz und dunkle Linien sind weiß gekennzeichnet

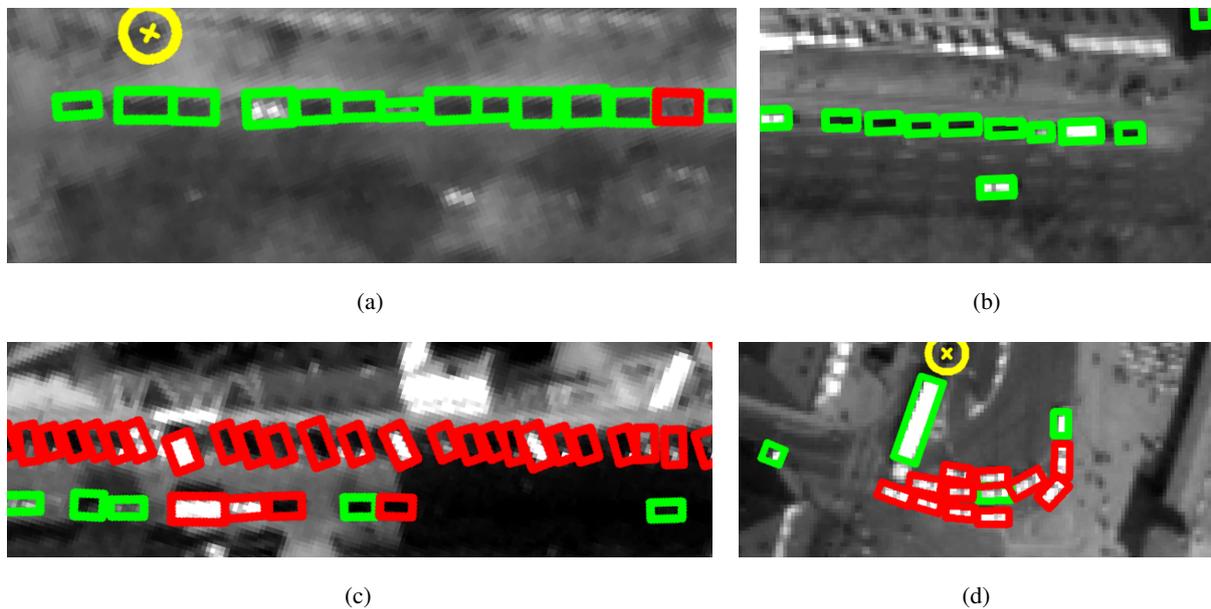


Abbildung 7.9: Ergebnisse der Detektion. Grün umrandete Fahrzeuge sind richtige Detektionen, rot umrandete Fahrzeuge sind fehlende Detektionen, gelbe Markierungen zeigen falsche Detektionen

	ohne Filterung	Filterung nach Breite	Filterung nach Länge	Filterung nach Richtung	alle Filterungen
helle Linien	8312	69,2 %	17,9 %	19,6 %	6,7 %
dunkle Linien	9533	76,8 %	26,3 %	19,8 %	9,7 %

Tabelle 7.2: Verbleibender prozentualer Anteil der gefilterten Linien ausgehend von der Gesamtanzahl der Linien ohne Filterung

7.4 Ergebnisse der Fahrzeugerkennung

Die erreichte Qualität der Fahrzeugerkennung ist für die unterschiedlichen Szenen sehr heterogen. Das erste Testgebiet sticht mit einer Vollständigkeit von annähernd 90 % positiv hervor, während im siebten Testgebiet nicht einmal die Hälfte aller darin enthaltenen Fahrzeuge erkannt wurden. In diesem Testgebiet treten besonders häufig schräg zur Fahrbahn parkende Fahrzeuge auf. Abbildung 7.9c zeigt das Ergebnis eines Ausschnitts dieses Testgebiets. Auch im sechsten Testgebiet sind vermehrt schräg parkende Fahrzeuge präsent, wie der Teilausschnitt in Abbildung 7.10 zeigt. Auch in diesem Fall werden nur etwas mehr als die Hälfte der Fahrzeuge erkannt. Das entwickelte Modell für gruppierte Fahrzeuge zeigt hier seine größte Schwäche, da schräg zur Fahrbahn stehende Fahrzeuge nicht berücksichtigt sind.

Betrachtet man die getrennte Evaluierung der Vollständigkeit für Fahrzeugreihen, isolierte Fahrzeuge und beidseitig detektierte Fahrzeuge, wird deutlich, dass die Anzahl beidseitig detektierter Fahrzeuge meist den höchsten Wert annimmt. Die Klassifikation und die Extraktion von Fahrzeugen aus Reihen liefern für die Mehrzahl der Fahrzeuge gleiche Ergebnisse. Das verwendete zweistufige Klassifikationsverfahren ist also in Lage, auch einen Großteil der gruppierten Fahrzeuge richtig zu erkennen. Das entwickelte Fahrzeugreihenmodell kann in Szenen mit vielen homogenen Fahrzeugreihen eine höhere Vollständigkeit als die Klassifikation erreichen. Dies trifft für drei der sieben Testszenen zu. Ein besonders gutes Beispiel ist in Abbildung 7.11 zu sehen, welches einen Teilausschnitt des ersten Testgebiets zeigt. Bei den dort dargestellten Fahrzeugen handelt es sich meist um am Fahrbahnrand parkende Fahrzeuge. Die Verwendung eines Reihenmodells ist in diesen Fällen von großem Vorteil und trägt maßgeblich zur gesamt erreichten Vollständigkeit bei. Die erzielte Zuverlässigkeit der Erkennung von Fahrzeugen aus Reihen kann in allen Szenen überzeugen. Der erreichte Mittelwert der Zuverlässigkeit von 98,9 % lässt kaum Möglichkeiten zur Verbesserung der robusten Parameterschätzung zu, da dieses Ergebnis fast dem optimalen Wert entspricht.

Die Erkennung isolierter Fahrzeuge ist bei frei fließendem Verkehr deutlich überlegen. Aus diesem Grund ist die Vollständigkeit für isolierte Fahrzeuge im dritten Testgebiet doppelt so hoch wie Vollständigkeit der exklusiven Erfassung aus Fahrzeugreihen. Abbildung 7.12 zeigt einen Teil dieses Testgebiets. Obwohl Fahrzeuge nur als Blobs sichtbar sind, ist die Zuverlässigkeit des Klassifikationsansatzes überzeugend. Diese liegt zwar unterhalb von 90 %, die falschen Detektionen treten aber hauptsächlich in falsch ausgewiesenen Verkehrsflächen auf.

Insgesamt betrachtet wird aus Tabelle 6.3 deutlich, dass durchschnittlich etwas mehr Fahrzeuge durch die Klassifikation isolierter Fahrzeuge erkannt werden können. Allerdings ist die Zuverlässigkeit von Fahrzeugen aus Reihen in allen Fällen höher als die Zuverlässigkeit isolierter Fahrzeuge. Falsche Detektionen treten vermehrt neben den Fahrbahnen auf und werden häufig durch Fahrzeug-ähnliche Strukturen auf Dächern hervorgerufen. Abbildung 7.13 zeigt einen Ausschnitt des dritten Testgebietes, in welchem die höchste Anzahl falscher Detektionen auftrat. Die dort abgebildeten Fehldetektionen liegen größtenteils auf Dächern und werden durch Lageungenauigkeiten der verwendeten Geoinformation oder die deshalb durchgeführte Verbreiterung der Straßen verursacht (siehe Abschnitt 7.1).

Aus der Diskussion über die erreichte Qualität der Detektion können folgende Schlussfolgerungen

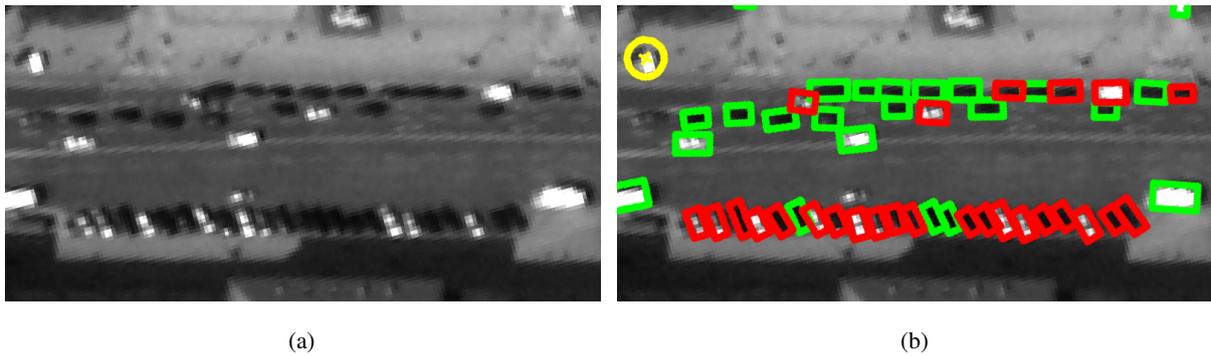


Abbildung 7.10: Ergebnisse bei schräg parkenden Fahrzeugen. (a) Teilausschnitt des Testgebiets 6, (b) Ergebnisse des Teilausschnitts, grün umrandete Fahrzeuge wurden richtig erkannt, rot umrandete Fahrzeuge wurden nicht erkannt, gelbe Kreise markieren falsche Detektionen

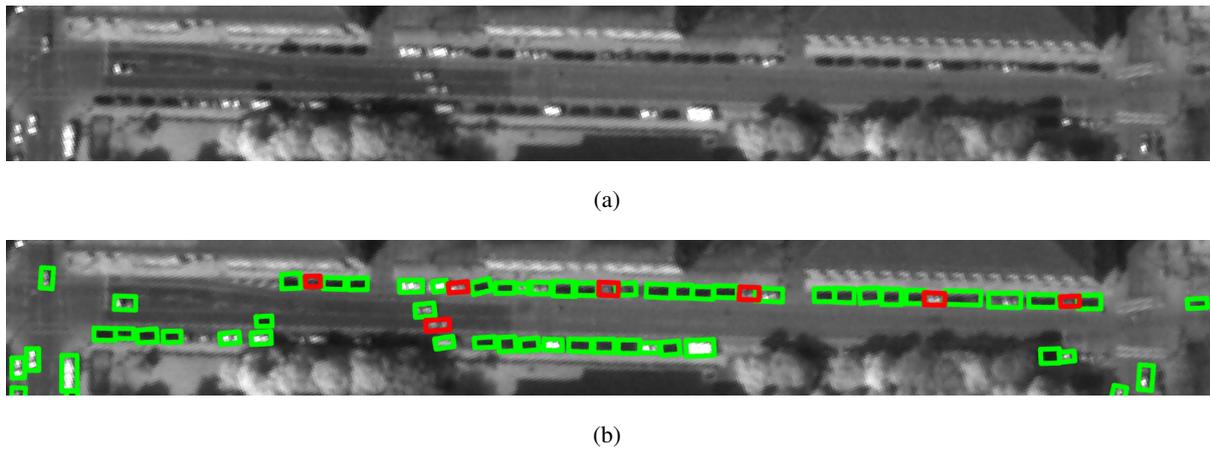


Abbildung 7.11: Ergebnisse bei parallel parkenden Fahrzeugen. (a) Teilausschnitt des Testgebiets 1, (b) Ergebnisse der Teilausschnittees, grün umrandete Fahrzeuge wurden richtig erkannt, rot umrandete Fahrzeuge wurden nicht erkannt



Abbildung 7.12: Ergebnisse bei fließendem Verkehr. a) Teilausschnitt des Testgebiets 3, (b) Ergebnisse der Teilausschnittees, grün umrandete Fahrzeuge wurden richtig erkannt, rot umrandete Fahrzeuge wurden nicht erkannt, gelbe Kreise markieren falsche Detektionen

7.5. Schätzung des Bewegungszustandes

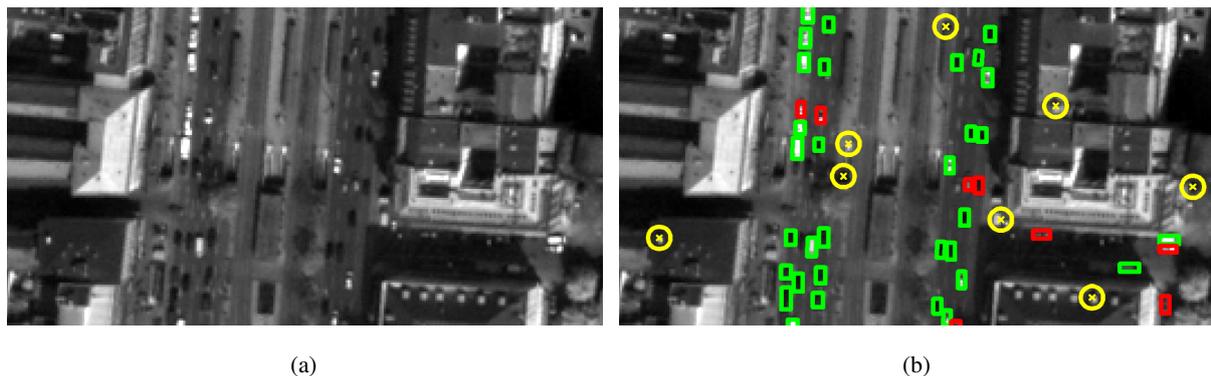


Abbildung 7.13: Ergebnisse bei fehlerhaften Geoinformationen. a) Teilausschnitt des Testgebiets 3, (b) Ergebnisse der Teilausschnittes, grün umrandete Fahrzeuge wurden richtig erkannt, rot umrandete Fahrzeuge wurden nicht erkannt, gelbe Kreise markieren falsche Detektionen

gezogen werden:

- Erkannte Fahrzeuge aus Reihen und erkannte isolierte Fahrzeuge tragen zu annähernd gleichen Teilen zur erreichten Vollständigkeit bei. Größere Differenzen treten bei hoher Anzahl parallel zur Fahrbahn parkender Fahrzeuge bzw. vermehrt fließendem Verkehr auf.
- Beide Verfahren zeigen eine hohe Zuverlässigkeit. Die Mehrzahl der auftretenden Falschdetektionen wird durch Ungenauigkeiten der verwendeten Geoinformationen verursacht.
- Die Qualität der Ergebnisse variiert sehr stark und ist von der dargestellten Szene abhängig. Hierdurch wird eine Abschätzung der erreichbaren Qualität für andere Bilddaten erschwert.

7.5 Schätzung des Bewegungszustandes

Die Schätzung des Bewegungszustandes ist für einen kleinen Teil isolierter Fahrzeuge möglich. Bei erfolgreicher Durchführung wird eine plausible Schätzung für die Geschwindigkeit eines Fahrzeuges erhalten. Die in Abbildung 6.8 dargestellten Geschwindigkeiten können für Stadtgebiete als typisch angesehen werden. Zudem wurde für keines der 238 untersuchten Fahrzeuge eine unverhältnismäßig hohe Geschwindigkeit ermittelt.

Die Ergebnisse der Bewegungsschätzung zeigen aber einen kritischen Nachteil. Für ausschließlich 4 % der untersuchten Fahrzeuge konnte eine Geschwindigkeit berechnet werden, da das Matchingverfahren meist keine Übereinstimmung zwischen Fahrzeugen im panchromatischen Kanal und den multispektralen Kanälen gefunden hat. In den Abbildungen 6.8d und 6.8e sind besonders viele isolierte Fahrzeuge markiert, für die kein Versatz berechnet werden konnte. Meistens handelt es sich um sehr dunkle Fahrzeuge ohne „grelle“ Farbe. Vor allem rot lackierte Fahrzeuge haben eine hohe Farbintensität und sind daher auch gut in den multispektralen Kanälen auffindbar. Liegen keine hohen Intensitätswerte vor kann das entwickelte Verfahren keinen Versatz bestimmen.

7.5. Schätzung des Bewegungszustandes

Aufgrund der geringen Anzahl ermittelter Geschwindigkeiten ist die Ableitung flächendeckender Geschwindigkeitsprofile nicht möglich.

8 Zusammenfassung und Ausblick

8.1 Zusammenfassung

In Satellitenbildern zeigen Fahrzeuge kaum signifikante Strukturen, sondern sind nur als Blobs erkennbar. Treten Fahrzeuge in Reihen gruppiert auf, kann zudem visuell keine klare Trennung von Einzelfahrzeugen vorgenommen werden. Die Abgrenzung von Fahrzeugen gegenüber anderen Objekten wird dadurch zusätzlich erschwert.

Zur Einhaltung einer akzeptablen Fehlerrate wurden Geoinformationen über Verkehrsflächen verwendet. Hierdurch kann der bearbeitete Bildbereich des Verfahrens erheblich eingeschränkt werden und eine Vielzahl falscher Detektionen vermieden werden. Die Beschränkung des zu verarbeiteten Bildmaterials führt zudem zu einer deutlichen Verringerung der Berechnungszeit. Liegen die Bilddaten und Geoinformationen in unterschiedlichen Koordinatensystemen vor, ist zunächst die Transformation in ein einheitliches System notwendig. Dies wurde in der Arbeit am Beispiel von ATKIS demonstriert, welches in das globale UTM-System der Satellitenbilddaten überführt wurde. Es wurde außerdem gezeigt, dass die Mehrzahl falscher Detektionen durch Ungenauigkeiten der Geoinformationen verursacht wurden.

Es wurde ein neues zweistufiges Modell zur Detektion von Fahrzeugen entwickelt:

- Erstmalig wurden Fahrzeugreihen in optischen Satellitenbildern modelliert und erfasst. Zur Erkennung von Fahrzeugen aus diesen Reihen wurden robuste Schätzverfahren intensiv untersucht, welche auf die Problemstellung angepasst wurden. Im Vergleich mit anderen Arbeiten zur Fahrzeugdetektion aus Satellitenbildern konnten so, eine hohe Anzahl gruppierter Fahrzeuge erkannt werden. Durch die Verwendung robuster Schätzverfahren traten nur wenige Falschdetektionen aus Fahrzeugreihen auf.
- Isolierte Fahrzeuge wurden durch Methoden der Mustererkennung detektiert. Zunächst wurden alle Bildpunkte durch den Gentle AdaBoost Algorithmus klassifiziert. Das Ergebnisbild wurde anschließend segmentiert und die entstandenen Segmente durch eine Support Vector Machine erneut klassifiziert. In bisherigen Arbeiten zur Fahrzeugerkennung aus Satellitenbildern wurde ausschließlich eine Klassifikation von Pixeln oder eine Klassifikation von Objekten durchgeführt. Durch die entwickelte Vorgehensweise von aufeinander folgenden Klassifikationen, konnte auch die Erkennung isolierter Fahrzeuge mit hoher Zuverlässigkeit durchgeführt werden. Dieses Konzept ist leicht auf andere Problemstellungen übertragbar und wurde bereits erfolgreich in operationellen

Systemen zur Fahrzeugerkennung aus hochaufgelösten Luftbildern angewandt (Rosenbaum et al., 2010).

In der vorliegenden Arbeit wurde erstmalig eine automatische Bewegungsschätzung von Fahrzeugen durchgeführt, welche durch den Versatz bewegter Objekte zwischen einzelnen Spektralkanälen möglich ist. Es wurde gezeigt, dass nur in wenigen Fällen der Versatz bestimmbar ist und daher keine flächendeckenden Geschwindigkeitsprofile erstellt werden können.

Die Initialisierung der entwickelten Verfahren erfordert kein Expertenwissen. Alle benötigten Parameter werden aus manuell erfassten Trainingsdaten abgeleitet.

8.2 Ausblick

Die Erfassung von Verkehrsdaten aus großflächigen Satellitenbilddaten ist ein ehrgeiziges Vorhaben. Daher zeigt auch der in dieser Arbeit entwickelte Ansatz in bestimmten Situationen Schwächen bezüglich der Methodik oder des verwendeten Datenmaterials.

Beschränkung des Suchbereichs

Ungenauigkeiten in den verwendeten Geoinformationen führten zu falschen Fahrzeugdetektionen. Durch die Verwendung anderer Straßeninformationen könnten diese unter Umständen vermieden werden. Mit der OpenStreetMap (<http://www.openstreetmap.org/>) steht hierfür eine umfangreiche kostenlose Alternative zur Verfügung. Es könnte auch eine Verbesserung der verwendeten Geoinformationen durchgeführt werden. Graphen und aktive Konturmodelle zeigen für diese Anwendung großes Potential (Gerke et al., 2004; Butenuth, 2007).

Extraktion isolierter Fahrzeuge

Die verwendeten Merkmale und Klassifikationsverfahren zeigen auf hochaufgelösten Luftbildern eine durchschnittliche Qualität von über 80 % (Rosenbaum et al., 2010). Die Anwendung des entwickelten Verfahrens auf QuickBird mit 60 cm Bodenpixelgröße kann nicht diese hohe Qualität erreichen. Da aber bereits höher aufgelöste Satellitenbilddaten von WorldView-II und GeoEye-I existieren, wäre eine Evaluierung der Fahrzeugdetektion auf diesen Daten interessant. Die Verringerung der Bodenpixelgröße auf 50 cm, erleichtert zumindest visuell die Erkennung von Fahrzeugen deutlich. Für dieses Bildmaterial wird eine deutliche Steigerung der Qualität erwartet, ohne größere Einschränkungen bei der Abdeckung der Bilddaten zu haben.

Auch das Hinzufügen weiterer Merkmale erscheint sinnvoll. In Jin & Davis (2007) wurden beispielsweise speziell auf Fahrzeuge abgestimmte Merkmale berechnet. Die Verwendung solcher Merkmale könnte zu einer signifikanten Steigerung der Detektionsqualität führen.

In dieser Arbeit wurde kein Vergleich verschiedener Klassifikationsverfahren durchgeführt. Vor allem die Verwendung eines gegenüber Rauschen unempfindlicheren Boosting-Algorithmus (siehe Rättsch et al. (2001)) würde eine Verbesserung des entwickelten Ansatzes darstellen.

8.2. Ausblick

Extraktion gruppierter Fahrzeuge

Auch die Extraktion gruppierter Fahrzeuge würde von höher aufgelösten Satellitenbilddaten profitieren. Der wichtigste Punkt für weiterführende Arbeit ist aber die Erweiterung des Modells für Fahrzeugreihen. Hierzu müsste ein Reihenmodell für schräg zur Fahrbahn parkende Autos entwickelt werden. Die fehlende Erkennung solcher Fahrzeuggruppen ist bislang die Hauptursache für die teils niedrige Vollständigkeit.

Zusätzlich könnte die Eignung anderer approximierender Funktionen untersucht werden. Allerdings lässt dies nur eine geringe oder keine Steigerung der Detektionsrate erwarten.

Schätzung des Bewegungszustands

Die Schätzung des Bewegungszustandes sollte auf weiteren Satellitenbilddaten erprobt werden. Da WorldView-II neben dem panchromatischen Kanal über acht zusätzliche Spektralkanäle verfügt, könnte für deutlich mehr Fahrzeuge ein Versatz bestimmbar sein. Auch der Einsatz von Stereobildpaaren birgt großes Potential bei der Erstellung von flächendeckenden Geschwindigkeitsprofilen. Allerdings würde der zeitliche Abstand von bis zu zehn Sekunden zwischen den einzelnen Aufnahmen eine Zuordnung korrespondierender Fahrzeuge erschweren.

Literaturverzeichnis

- AdV (2003) *ATKIS - Objektartenkatalog Basis-DLM*. Arbeitsgemeinschaft der Vermessungsverwaltungen der Länder der Bundesrepublik Deutschland (AdV), 3.2 edition.
- AdV (2009) *Bundeseinheitliche Transformation für ATKIS*. Arbeitsgemeinschaft der Vermessungsverwaltungen der Länder der Bundesrepublik Deutschland, Technischer Bericht.
- Aizerman A, Braverman E, Rozoner L (1964) Theoretical foundations of the potential function method in pattern recognition learning. *Automation and remote control*, 25: 821–837.
- Alba-Flores R (2005) *Evaluation of the use of high-resolution satellite imagery in transportation applications*. Department of Electrical and Computer Engineering, University of Minnesota, Duluth, Technischer Bericht CTS 05-11.
- Alba-Flores R, Kuthadi S, Rios-Gutiérrez F (2004) Detecting and counting vehicles from high resolution satellite imagery. In: Rashid MH (ed) *Circuits, Signals, and Systems*: 300–305.
- Andrews DF, Bickel P, Hampel F, Huber P, Rogers W, Tukey JW (1972) *Robust estimates of location: survey and advances*. Princeton University Press.
- Aurdal L, Eikvil L, Koren H, Hanssen J, Johansen K (2007) *Road traffic snapshot - Final report*. Nork Regnesentral, Technischer Bericht.
- Ballard DH (1981) Generalizing the Hough transform to detect arbitrary shapes. *Pattern Recognition*, 13 (2): 111–122.
- Bartlett MS, Littlewort G, Fasel I, Movellan JR (2003) Real time face detection and facial expression recognition: development and applications to human computer interaction. *Workshop on Conference on Computer Vision and Pattern Recognition*, 05: 53.
- Beaton AE, Tukey JW (1974) The fitting of power series, meaning polynomials, illustrated on band-spectroscopic data. *Technometrics*, 16 (2): 147–185.
- Bellman R (1961) *Adaptive Control Processes*. Princeton University Press.
- Betke M, Haritaoglu E, Davis LS (2000) Real-time multiple vehicle detection and tracking from a moving vehicle. *Machine Vision and Applications*, 12 (2): 69–83.
- Bishop CM (2006) *Pattern Recognition and Machine Learning*. Springer.

- Bosch A, Munoz X, Marti R (2007) Which is the best way to organize/classify images by content? *Image and Vision Computing*, 25 (6): 778–791.
- Bosserhoff D, Conrath T, König S, Krämer HG, Sauermann D, Schmidt HU, Schulze P (2004) *Strasse und Schiene - Partner zur Sicherung der Mobilität*. Bundesvereinigung der Straßenbau- und Verkehringenieure e.V.
- Boyd S, Vandenberghe L (2009) *Convex Optimization*. Cambridge University Press.
- Breiman L, Friedman J, Olshen RA, Stone CJ (1984) *Classification and Regression Trees*. Chapman & Hall.
- Bundesministerium für Verkehr, Bau- und Stadtentwicklung (2006) Leitfaden Verkehrstelematik.
- Burges CJ (1998) A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, 2: 121–167.
- Burlina P, Chellappa R (1997) A spectral attentional mechanism tuned to object configurations. *IEEE Transactions on Image Processing*, 6 (8): 1117–1128.
- Burlina P, Parameswaran V, Chellappa R (1997) Sensitivity analysis and learning strategies for context-based vehicle detection algorithms. In: *Proc. DARPA Image Understanding Workshop*: 577–584.
- Butenuth M (2007) Segmentation of imagery using network snakes. *Photogrammetrie Fernerkundung Geoinformation*, 2007 (1): 7.
- Campbell G (2007) *Panorama of Transport*. Office for Official Publications of the European Communities.
- Canny J (1986) A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8 (6): 679–698.
- Cerutti-Maori D, Klare J, Brenner A, Ender J (2008) Wide-area traffic monitoring with the SAR/GMTI system PAMIR. *IEEE Transactions on Geoscience and Remote Sensing*, 46 (10): 3019–3030.
- Cortes C, Vapnik V (1995) Support-vector networks. *Machine Learning*, 20: 273–297.
- Cristianini N, Shawe-Taylor J (2000) *An introduction to support vector machines and other kernel-based learning methods*. Cambridge University Press.
- Cucchiara R, Piccardi M, Mello P (2000) Image analysis and rule-based reasoning for a traffic monitoring system. *IEEE Transactions on Intelligent Transportation Systems*, 1 (2): 119–130.
- Daschiel HA (2004) *Advanced methods for image information mining system: evaluation and enhancement of user relevance*. Dissertation, Technische Universität Berlin, Fakultät IV - Elektrotechnik und Informatik.
- de Sá JM (2001) *Pattern recognition: concepts, methods and applications*. Springer.
- Dennis JE, Welsch RE (1978) Techniques for nonlinear least squares and robust regression. *Communications in Statistics - Simulation and Computation*, 7 (4): 345–359.
- Deutsches Institut für Wirtschaftsforschung (2002) *Verkehr in Zahlen 2002/2003*. Deutscher Verkehrs-Verlag.
- Dietterich TG (2000) An experimental comparison of three methods for constructing ensembles of decision trees: bagging, boosting, and randomization. *Machine Learning*, 40: 139–157.

Literaturverzeichnis

DigitalGlobe (2009) *QuickBird Imagery Product - Product Guide*. DigitalGlobe, Inc., revision 5.0 edition.

Dreschler L, Nagel HH (1981) Volumetric model and 3D-trajectory of a moving car derived from monocular TV-frame sequences of a street scene. In: *IJCAI'81: Proc. 7th international joint conference on Artificial intelligence*: 692–697.

Duda RO, Hart PE, Stork DG (2001) *Pattern Classification*. Wiley & Sons.

Eikvil L, Aurdal L, Koren H (2009) Classification-based vehicle detection in high-resolution satellite images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 64 (1): 65 – 72.

Ender J (1999) Space-time processing for multichannel synthetic aperture radar. *IEEE Electronics and Communication Engineering*, 11 (1): 29–38.

Etaya M, Nakano R, Shimoda H, Sakata T (2005) Detection of ocean wave movements after the Northern Sumatra earthquake using SPOT images. In: *Proc. IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*. 2: 1420–1423.

Etaya M, Sakata T, Shimoda H, Matsumae Y (2004) An experiment on detecting moving objects using a single scene of QuickBird data. *Journal of the Remote Sensing Society of Japan*, 24 (4): 357–366.

Eurimage (2010) QuickBird (F.A.Q.). letzter Zugriff: 10.01.2010.

Europäische Kommission (2001) WEISSBUCH - Die europäische Verkehrspolitik bis 2010: Weichenstellung für die Zukunft. Amt für amtliche Veröffentlichungen der Europäischen Gemeinschaften.

Fair R (1974) On the robust estimation of econometric models. *Annals of Economic and Social Measurement*, 3 (4): 667–678.

Förstner W, Gülch E (1987) A fast operator for detection and precise location of distinct points, corners and centers of circular features. In: *Proc. ISPRS Intercommision Workshop*: 281–305.

Freund Y, Schapire RE (1995) A decision-theoretic generalization of on-line learning and an application to boosting. In: Vitányi P (ed) *Computational Learning Theory*, volume 904 of *Lecture Notes in Computer Science* (pp. 23–37). Springer.

Freund Y, Schapire RE (1997) A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55 (1): 119–139.

Freund Y, Schapire RE (1999) A short introduction to boosting. *Journal of Japanese Society for Artificial Intelligence*, 14 (5): 771–780.

Friedman J, Hastie T, Tibshirani R (2000) Additive logistic regression: a statistical view of boosting. *Annals of Statistics*, 28 (28): 337–407.

Fukunaga K (1990) *Introduction to Statistical Pattern Recognition*. Academic Press.

GeoEye, Inc. (2010) *2010 Annual Report*. GeoEye, Inc., Technischer Bericht.

- Gerhardinger A, Ehrlich D, Pesaresi M (2005) Vehicle detection from very high resolution satellite imagery. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 36 (3/W24): 83–88.
- Gerke M, Butenuth M, Heipke C, Willrich F (2004) Graph-supported verification of road databases. *ISPRS Journal of Photogrammetry and Remote Sensing*, 58 (3-4): 152–165.
- Gierull CH (2004) Statistical analysis of multilook SAR interferograms for CFAR detection of ground moving targets. *IEEE Transactions on Geoscience and Remote Sensing*, 42: 691–701.
- Grabner H, Nguyen TT, Gruber B, Bischof H (2008) On-line boosting-based car detection from aerial images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 63 (3): 382–396.
- Gupta R, Hartley RI (1997) Linear pushbroom cameras. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19 (9): 963–975.
- Haag M, Nagel HH (2004) Combination of edge element and optical flow estimates for 3D-model-based vehicle tracking in traffic image sequences. *International Journal of Computer Vision*, 35 (3): 295–319.
- Haar A (1910) Zur Theorie der orthogonalen Funktionensysteme. *Mathematische Annalen*, 69: 331–371.
- Hampel F (1980) Robuste Schätzungen: Ein anwendungsorientierter Überblick. *Biometrical Journal*, 22 (1): 3–21.
- Hampel F, Ronchetti E, Rousseeuw P, Stahel W (1986) *Robust Statistics*. Wiley Series in Probability and Mathematical Statistics. John Wiley and Sons.
- Haralick RM, Shanmugam K, Dinstein I (1973) Textural features for image classification. *IEEE Transactions on Systems, Man and Cybernetics*, 3 (6): 610–621.
- Hinich M, Talwar P (1975) A simple method for robust regression. *Journal of the American Statistical Association*, 70 (349): 113–119.
- Hinz S (2003a) *Automatische Extraktion urbaner Straßennetze aus Luftbildern*. Dissertation, Fakultät für Bauingenieur- und Vermessungswesen der Technischen Universität München.
- Hinz S (2003b) Integrating local and global features for vehicle detection in high resolution aerial imagery. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 34 (3/W8): 119–124.
- Hinz S (2004) Detection of vehicles and vehicle queues in high resolution aerial images. *Photogrammetrie - Fernerkundung - Geoinformation*, 3: 201–213.
- Hinz S (2008) *Bewegobjekterkennung und -charakterisierung in Fernerkundungsbilddaten mit Fokus auf Verkehrsanwendungen*. Habilitation, Technische Universität München.
- Hinz S, Bamler R, Stilla U (2006) Theme Issue: Airborne and spaceborne traffic monitoring. *ISPRS Journal of Photogrammetry and Remote Sensing*, 61 (3-4): 135–136.
- Hinz S, Leitloff J, Stilla U (2005) Context-supported vehicle detection in optical satellite images of urban areas. In: *Proc. Geoscience and Remote Sensing Symposium (IGARSS 05)*. 4: 2937 – 2941.

Literaturverzeichnis

- Hinz S, Stilla U (2006) Car detection in aerial thermal images by local and global evidence accumulation. *Pattern Recognition Letters*, 27: 308–315.
- Holland PW, Welsch RE (1977) Robust regression using iteratively reweighted least-squares. *Communications in Statistics - Theory and Methods*, 6 (9): 813–827.
- Hsu C, Lin C (2002) A comparison of methods for multiclass support vector machines. *IEEE Transactions on Neural Networks*, 13 (2): 415–425.
- Huber P (1964) Robust estimation of a location parameter. *Annals of Mathematical Statistics*, 35 (1): 73–101.
- Huber P (1977) Robust methods of estimation of regression coefficients. *Statistics*, 8 (1): 41–53.
- Jacobsen K (2005) High resolution satellite imaging systems - overview. *Photogrammetrie - Fernerkundung - Geoinformation*, 6: 487–496.
- Jacobsen K (2011) Recent developments of digital cameras and space imagery. In: *Proc. GIS Ostrava*
- Jäger R, Müller T, Saler H, Schwäble R (2005) *Klassische und robuste Ausgleichungsverfahren*. Herbert Wichmann Verlag.
- Jain AK, Duin RP, Mao J (2000) Statistical pattern recognition: a review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22 (1): 4–37.
- Jensen JR, Cowen DC (1999) Remote sensing of urban/suburban infrastructure and socio-economic attributes. *Photogrammetric Engineering & Remote Sensing*, 65 (5): 611–622.
- Jin X, Davis CH (2007) Vehicle detection from high-resolution satellite imagery using morphological shared-weight neural networks. *Image and Vision Computing*, 25 (9): 1422–1431.
- Jolliffe I (2002) *Principal Component Analysis*. Springer.
- Junkins D, Farley S (1995) *National Transformation Version 2*. Geodetic Survey Division - Geomatics Canada, User's guide.
- Kearns M (1988) Thoughts on Hypothesis Boosting. nicht veröffentlichtes Manuskript.
- Kirchhof M, Stilla U (2006) Detection of moving objects in airborne thermal videos. *ISPRS Journal of Photogrammetry and Remote Sensing*, 61 (3-4): 187–196.
- Larsen SO, Koren H, Solberg R (2009) Traffic monitoring using very high resolution satellite imagery. *Photogrammetric Engineering & Remote Sensing*, 75 (7): 859–869.
- Leitloff J, Hinz S, Stilla U (2005a) Automatic vehicle detection in space images supported by digital map data. *International Archives of Photogrammetry, Remote Sensing and Spatial Geoinformation Sciences*, 36 (3/W24): 75–80.
- Leitloff J, Hinz S, Stilla U (2005b) Vehicle queue detection in complex urban areas by extraction and analysis of linear features. *International Archives of Photogrammetry, Remote Sensing and Spatial Geoinformation Sciences*, 36 (1/W3): on CD.

- Leitloff J, Hinz S, Stilla U (2005c) Vehicle queue detection in satellite images of urban areas. *International Archives of Photogrammetry, Remote Sensing and Spatial Geoinformation Sciences*, 36 (8/W27): on CD.
- Leitloff J, Hinz S, Stilla U (2006a) Automatic vehicle detection in satellite images. *International Archives of Photogrammetry, Remote Sensing and Spatial Geoinformation Sciences*, 36 (3): 221–227.
- Leitloff J, Hinz S, Stilla U (2006b) Detection of Vehicle Queues in Quickbird Imagery of City Areas. *Photogrammetrie - Fernerkundung - Geoinformation*, 4: 315–325.
- Leitloff J, Hinz S, Stilla U (2007a) Estimation of traffic parameters in urban areas from satellite images. In: *Urban Remote Sensing Joint Event, 2007*: 1–6.
- Leitloff J, Hinz S, Stilla U (2007b) Inferring traffic activity from optical satellite images. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 36 (3/W49 B): 89–94.
- Leitloff J, Hinz S, Stilla U (2010) Vehicle detection in very high resolution satellite images of city areas. *IEEE Transactions on Geoscience and Remote Sensing*, PP (99): 1–12.
- Lienhart R, Kuranov A, Pisarevsky V (2003) Empirical analysis of detection cascades of boosted classifiers for rapid object detection. *Pattern Recognition*, 2781: 297–304.
- Lienhart R, Maydt J (2002) An extended set of Haar-like features for rapid object detection. In: *Proc. International Conference on Image Processing 2002*. 1: 900–903.
- Lindeberg T (1998) Feature detection with automatic scale selection. *International Journal of Computer Vision*, 30: 79–116.
- Liu G, Gong L, Haralick R (1999) *Vehicle detection in aerial imagery and performance evaluation*. Department of Electrical Engineering, University of Washington, Technischer Bericht.
- Marquardt DW (1963) An algorithm for least-squares estimation of nonlinear parameters. *Journal of the Society for Industrial and Applied Mathematics*, 11: 431–441.
- Mayer H (1998) *Automatische Objektextraktion aus digitalen Luftbildern*. Habilitation, Technische Universität München.
- McCord M, Goel P, Jiang Z, Coifman B, Yang Y, Merry CJ (2002) Improving AADT and VDT estimation with high-resolution satellite imagery. In: *Proc. Pecora 15/Land Satellite Information IV Symposium* on CD.
- McCord M, Merry C, Goel P (1998) Incorporating satellite imagery in traffic monitoring programs. In: *Proc. North American Traffic Monitoring Exhibition and Conference* on CD.
- McCord MR, Merry CJ (2000) *Traffic monitoring using satellite and ground data: preparation for feasibility tests and an operational system*. The Ohio State University, Department of Civil Engineering, Columbus, Technischer Bericht.
- McCord MR, Merry CJ, Bossler JD (1995) *The feasibility of traffic data collection using satellite imagery*. The Ohio State University, Department of Civil Engineering, Columbus, Technischer Bericht.
- Meyer F, Hinz S, Laika A, Weihing D, Bamler R (2006) Performance analysis of the TerraSAR-X traffic monitoring concept. *ISPRS Journal of Photogrammetry and Remote Sensing*, 61 (3-4): 225–242.

Literaturverzeichnis

- Michaelsen E, Stilla U (2001) Estimating urban activity on high-resolution thermal image sequences aided by large scale vector maps. In: *Proc. IEEE/ISPRS Joint Workshop on Remote Sensing and Data Fusion over Urban Areas: 25–29*.
- Moon H, Chellappa R, Rosenfeld A (2002) Performance analysis of a simple vehicle detection algorithm. *Image and Vision Computing*, 20 (1): 1–13.
- Nagao M (1984) Control strategies in pattern analysis. *Pattern Recognition*, 17 (1): 45–56.
- Neumann B (2003) Bildverstehen - Ein Überblick. In: Görz G, Rollinger CR, Schneeberger J (eds) *Handbuch der Künstlichen Intelligenz* (pp. 815–841). Oldenbourg. Kapitel 21.
- Niemann H (2003) *Klassifikation von Mustern*. 2. überarbeitete Auflage.
- Otsu N (1979) A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man and Cybernetics*, 9 (1): 62–66.
- Papageorgiou C, Oren M, Poggio T (1998) A general framework for object detection. In: *Proc. Sixth International Conference on Computer Vision: 555–562*.
- Papageorgiou C, Poggio T (2000) A trainable system for object detection. *International Journal of Computer Vision*, 38 (1): 15–33.
- Pesaresi M, Gutjahr K, Pagot E (2007) Moving targets velocity and direction estimation by using a single optical VHR satellite imagery. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 36 (3/W49 B): 125–130.
- Pesaresi M, Gutjahr KH, Pagot E (2008) Estimating the velocity and direction of moving targets using a single optical VHR satellite sensor image. *International Journal of Remote Sensing*, 29 (4): 1221–1228.
- Platt J (1998) *Sequential Minimal Optimization: A fast algorithm for training support vector machines*. Microsoft Research, Technical report.
- Rajagopalan A, Chellappa R (2001) Higher-order-statistics-based detection of vehicles in still images. *Journal of the Optical Society of America*, 18 (12): 3037–3048.
- Rätsch G, Onoda T, Müller KR (2001) Soft margins for AdaBoost. *Machine Learning*, 42: 287–320.
- Rosenbaum D, Leitloff J, Kurz F, Meynberg O, Reize T (2010) Real-time image processing for road traffic data extraction from aerial images. In: *Proc. ISPRS Technical Commission VII Symposium - 100 years ISPRS: 469–474*.
- Rosenblatt F (1958) The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65 (6): 386–408.
- Ruskoné R, Guiges L, Airault S, Jamet O (1996) Vehicle detection on aerial images: A structural approach. In: Kropatsch G (ed) *Proc. 13th International Conference on Pattern Recognition*. IEEE Computer Society Press, 3: 900–903.
- Schapire RE, Singer Y (1999) Improved boosting algorithms using confidence-rated predictions. *Machine Learning*, 37 (3): 297–336.

- Schischmanow A (2005) *Analyse des Straßenverkehrs mit verteilten opto-elektronischen Sensoren*. Dissertation, Mathematisch-Naturwissenschaftliche Fakultät II der Humboldt-Universität zu Berlin.
- Schölkopf B, Platt J, Shawe-Taylor J, Smola A, Williamson R (2001) Estimating the support of a high-dimensional distribution. *Neural computation*, 13 (7): 1443–1471.
- Schölkopf B, Smola AJ (2001) *Learning with kernels: Support vector machines, regularization, optimization and beyond*. MIT Press.
- Schölkopf B, Smola AJ, Williamson RC, Bartlett PL (2000) New support vector algorithms. *Neural Computation*, 12: 1207–1245.
- Selfridge O (1955) Pattern recognition and modern computers. In: *Proc. Western Joint Computer Conference*: 91–93.
- Sharma G (2002) Vehicle detection and classification in 1-m resolution imagery. Master's thesis, Graduate School, The Ohio State University.
- Sharma G, Merry CJ, Goel P, McCord M (2006) Vehicle detection in 1-m resolution satellite and airborne imagery. *International Journal of Remote Sensing*, 27 (4): 779 – 797.
- Smola A (1998) *Learning with kernels*. Dissertation, Fachbereich 13 - Informatik der Technischen Universität Berlin.
- Spiegel M (2000) Automatische Erkennung von Fahrzeugen aus digitalen Luftbildern mittels Neuronaler Netze. Diplomarbeit, Lehrstuhl für Photogrammetrie und Fernerkundung, Technische Universität München.
- Statistisches Bundesamt (2006) *Im Blickpunkt - Verkehr in Deutschland 2006*. Wiesbaden.
- Steger C (1998) An unbiased detector of curvilinear structures. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20 (2): 113–125.
- Steger C (2001) Similarity measures for occlusion, clutter, and illumination invariant object recognition. In: Radig B, Florczyk S (eds) *Pattern Recognition*. Springer-Verlag, 2191: 148–154.
- Steger C, Ulrich M, Wiedemann C (2008) *Machine vision algorithms and applications*. WILEY-VCH Verlag GmbH & Co. KGaA.
- Suetens P, Fua P, Hanson AJ (1992) Computational strategies for object recognition. *ACM Computing Surveys*, 24 (1): 6–61.
- Sullivan GD, Baker KD, Worrall AD, Attwood CI, Remagnino PM (1997) Model-based vehicle detection and classification using orthographic approximations. *Image and Vision Computing*, 15 (8): 649 – 654. British Machine Vision Conference.
- Theodoridis S, Koutroumbas K (2006) *Pattern Recognition*. Academic Press.
- Toth CK, Grejner-Brzezinska D (2006) Extracting dynamic spatial data from airborne imaging sensors to support traffic flow estimation. *ISPRS Journal of Photogrammetry and Remote Sensing*, 61 (3-4): 137–148.

Literaturverzeichnis

- Toth CK, Grejner-Brzezinska D, Merry CJ (2003) Supporting traffic flow management with high-definition image. In: *Proc. Workshop on High Resolution Mapping from Space* on CD.
- Valiant LG (1984) A theory of the learnable. *Communications of the ACM*, 27 (11): 1134–1142.
- Vapnik V (1982) *Estimation of dependences based on empirical data*. Springer.
- Viola P, Jones MJ (2001) Rapid object detection using a boosted cascade of simple features. *Proc. CVPR*, 1: 511–518.
- Viola P, Jones MJ (2004) Robust real-time face detection. *International Journal of Computer Vision*, 57 (2): 137–154.
- Xiong Z, Zhang Y (2008) An initial study on vehicle information extraction from single pass QuickBird satellite imagery. *Photogrammetric Engineering & Remote Sensing*, 74 (11): 1401–1411.
- Yao W, Hinz S, Stilla U (2008a) Automatic analysis of traffic scenario from airborne thermal infrared video. *International Archives of Photogrammetry, Remote Sensing and Spatial Geoinformation Sciences*, 37 (B3 A): 223–228.
- Yao W, Hinz S, Stilla U (2008b) Automatic vehicle extraction from airborne LiDAR data of urban areas using morphological reconstruction. In: *Proc. 5th IAPR Workshop on Pattern Recognition in Remote Sensing* on CD.
- Zhang Y, Xiong Z (2006) Moving vehicle detection using a single set of QuickBird imagery - An initial study. *International Archives of Photogrammetry, Remote Sensing and Spatial Geoinformation Sciences*, 36 (7): 397–402.
- Zhao T, Nevatia R (2003) Car detection in low resolution aerial images. *Image and Vision Computing Journal*, 21 (8): 693–703.
- Zheng H, Hu X (2008) A vehicle detection approach from satellite imagery using antibody networks. *International Journal of Intelligent Engineering and Systems*, 1 (1): 9–16.
- Zheng H, Li L (2007) An artificial immune approach for vehicle detection from high resolution space imagery. *International Journal of Computer Science and Network Security*, 7 (2): 67–72.
- Zheng H, Pan L, Li L (2006) A morphological neural network approach for vehicle detection from high resolution satellite imagery. In: King I, Wang J, Chan L, Wang DL (eds) *Proc. International Conference on Neural Information Processing*. Springer, 4233: 99–106.

Dank

Zunächst möchte ich meinem Doktorvater Prof. Dr.-Ing. Uwe Stilla danken. Er gab mir die Möglichkeit zum eigenständigen, wissenschaftlichen Arbeiten am Fachgebiet Photogrammetrie und Fernerkundung der TUM. Durch die konstruktiven Gespräche, vor allem in der Endphase der Promotion, hat er maßgeblich zur Gestalt und dem Gelingen dieser Arbeit beigetragen.

Dem Zweitgutachter Prof. Dr.-Ing. habil. Stefan Hinz danke ich für die Einführung in das spannende Gebiet der Bildverarbeitung zu seiner Zeit als Mitarbeiter an der TUM. Trotz seiner Berufung an das KIT fand ich in ihm stets einen kompetenten und freundschaftlichen Mentor.

Allen ehemaligen Kollegen der TU München und des DLR Oberpfaffenhofen danke ich für die angenehme Arbeitsatmosphäre. Mein besonderer Dank gilt Andreas Laika, Karin Hedman und Diana Weihing für die schöne gemeinsame Zeit an der TUM; Prof. Dr.-Ing. Peter Reinartz für die Möglichkeit zur Fertigstellung der Arbeit vor Verlassen des DLR und sein stets entgegengebrachtes Verständnis für kleine und große (persönliche) Krisen; meinem Zimmerkollegen Dominik Rosenbaum für die freundschaftliche Zusammenarbeit und die Übernahme einiger meiner Projektaufgaben in der heißen Phase vor Abgabe der Dissertation.

Zu tiefstem Dank bin ich meiner Frau Claudia verpflichtet. Ohne ihre Motivation und Unterstützung, während der langen Zeit bis zur Fertigstellung der Arbeit, gäbe es diese Dissertation nicht. Sie und unsere beiden Kinder haben mir immer wieder verdeutlicht, was wirklich wichtig ist und sorgten damit für den notwendigen Ausgleich zur akademischen Arbeit.

Lebenslauf

Name Jens Leitloff
Geburtstag und -ort 06.04.1978 in Beckendorf-Neindorf
Familienstand verheiratet, 2 Kinder
Wohnort Huttenstraße 78, 76646 Bruchsal



Ausbildung/Tätigkeit

1984 – 1991 Polytechnische Oberschule in Hadmersleben
1991 – 1996 Gymnasium in Oschersleben (Abschluss: Allgemeine Hochschulreife)
1996 – 1997 Grundwehrdienst in der Erzgebirgs-Kaserne, Marienberg
1997 – 2003 Studium des Vermessungswesens an der Technischen Universität Berlin
Abschluss: Diplom-Ingenieur für Vermessungswesen
2000 – 2002 Wissenschaftliche Hilfskraft an der Technischen Universität Berlin
(Fachgebiet Ingenieursvermessung und Ausgleichsrechnung)
2002 – 2003 Wissenschaftliche Hilfskraft am Geoforschungszentrum Potsdam
(Sektion 1.3 - Erdsystemmodellierung)
2003 – 2004 Auslandsaufenthalt (überwiegend in Australien)
2004 – 2008 Wissenschaftlicher Mitarbeiter an der Technischen Universität München
(Fachgebiet Photogrammetrie und Fernerkundung, Institut für Photogrammetrie und Kartographie)
2008 – 2011 Wissenschaftlicher Mitarbeiter am Deutschen Zentrum für Luft- und Raumfahrt e.V., Oberpfaffenhofen (Abteilung Photogrammetrie und Bildanalyse, Institut für Methodik der Fernerkundung)
seit 2011 Akademischer Mitarbeiter am Karlsruher Institut für Technologie
(Institut für Photogrammetrie und Fernerkundung)