

Fusion von Farbbildern und Laserdaten zur Terrainklassifikation

Marcel Häselich, Jessica Orth, Jens Hedrich, Dietrich Paulus

Universität Koblenz-Landau, Institut für Computervisualitik,
Universitätsstr. 1, 56070 Koblenz
eMail: vorname.nachnahme@uni-koblenz.de,
URL: <http://www.uni-koblenz.de/agas>

Zusammenfassung. Zur Navigation und Pfadplanung autonomer Roboter in unstrukturiertem Gelände ist es notwendig, befahrbare Flächen von unbefahrbaren Bereichen zu unterscheiden. In solchen Szenarien wird in der Robotik oft ein 3D Laserscanner als primärer Sensor eingesetzt, der kontinuierlich rund um das Fahrzeug Entfernungswerte in 3D aufnimmt. Die Analyse der Entfernungsmessungen ermöglicht es im Straßenverkehr, glatte Flächen von unebenen Flächen zu unterscheiden; diese Unterscheidung ist in unstrukturiertem Außengelände weder möglich noch ausreichend. Hier können weitere Sensordaten zur Interpretation herangezogen werden. Beispielsweise kann Farbinformation genutzt werden, um im Gelände die Sicherheit der Terrainklassifikation zu erhöhen. Nach einer Kalibrierung der Farbkameras und des 3D Laser werden die Sensordaten in ein gemeinsames Koordinatensystem fusioniert. Zur Terrainklassifikation werden hier Markov-Zufallsfelder eingesetzt, die Kontextinformationen berücksichtigen. Experimentelle Ergebnisse belegen, dass Farbe die Fehlklassifikationsrate verringern kann, dabei aber derzeit zu wesentlich höheren Laufzeiten führt.

1 Einleitung

Eine aktuelle Herausforderung in der autonomen Robotik ist eine hinreichend genaue Analyse und Interpretation der Beschaffenheit von unstrukturierten Umgebungen. Dies ist eine Grundvoraussetzung um Hindernissen ausweichen zu können und die Beschaffenheit der Terrainoberfläche bei der Navigation zu berücksichtigen. Moderne 3D Laser-Entfernungsmessgeräte (LRFs) bilden die Umgebung in Form von punktuellen Entfernungsmessungen detailliert ab. Diese Entfernungsmessungen reichen jedoch nicht aus, um die Festigkeit von möglichen Hindernissen abzuschätzen. Um elastische Hindernisse, wie z. B. hohes Gras, von festen Hindernissen unterscheiden zu können sind Farb- und Texturinformationen aus RGB-Kameras eine ideale Ergänzung. Weitere Einflussfaktoren, welche die Güte der Klassifikation auf LRF Daten beeinflussen, sind Sensorrauschen, variierende Datendichte, Eigenbewegung bzw. Erschütterungen des Roboters auf rauem Untergrund. Neben Sensorrauschen und Bewegungsunschärfe von den Farbdaten können Schattenwürfe und unregelmäßig beleuchtete Bereiche die Klassifikation negativ beeinflussen. Im Abschnitt 2 wird der Sensor Aufbau zur Terrainklassifikation vorgestellt. Der in dieser Abhandlung beschriebene Klassifikationsansatz setzt eine Fusion zwischen der Kamera und dem 3D LRF Daten voraus. Hierzu wird im Abschnitt 3 der Kalibrierungsansatz von Unnikrishnan und Hebert [1], welchen wir in unser Roboter-Framework integriert haben, motiviert. Das Ergebnis des Kalibrierverfahrens ist eine Schätzung



Abb. 1: Übersicht der verwendeten Sensoren: Ein 3D Laser-Entfernungsmessgerät und zwei unterschiedliche Farbkameras.

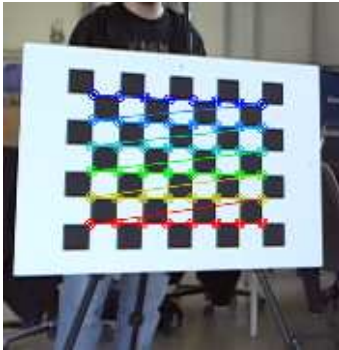
aller intrinsischen Kamera-Parameter sowie die jeweilige Position und Rotation der Kameras in Relation zum LRF. In Abschnitt 4 wird das Verfahren von Guo et al. [2] vorgestellt, welches zum Entfernen von Schatten verwendet wird, um den Einfluss dessen auf die Klassifikationsergebnisse zu minimieren. Der Abschnitt 5 führt in das zugrunde gelegte Klassifikationsverfahren mittels eines Markov-Zufallsfeldes ein. Die gewählte Methode ermöglicht durch Einbeziehen der Informationen aus benachbarten Bereichen eine Kontext-abhängige Klassifikation. In Abschnitt 6 wird die Stand unserer Arbeit zusammengefasst und ein Ausblick auf unsere zukünftigen Forschungsrichtungen gegeben.

2 Hardware

Die Abbildung 1 gibt eine Übersicht über die verwendete Sensorik. Ein LRF des Typs *Velodyne HDL-64E S2* wird gemeinsam mit drei Kameras verwendet. Der Kopf des Velodyne HDL-64E S2 besteht aus 64 Lasern, welche kontinuierlich Entfernungsmessungen der Umgebung liefern; dieser Kopf dreht sich mit einer Frequenz von ca. 15 Hz um die senkrechte Achse. Auf diese Weise generiert dieser LRF einen umfangreichen Datensatz von bis zu 1.8 Millionen Entfernungsmessungen pro Sekunde. Die Datenpunkte einer vollständigen 360° Drehung werden in unserem System in eine Punktwolke konvertiert und dienen als Eingabe für die nachfolgenden Algorithmen. Der vorgestellte Sensoraufbau kann auf einem Roboter montiert (Autonomes Fahren) oder auf einem Automobil (Testfahrten und Aufzeichnungen von Sensordaten) befestigt werden.

3 Kalibrierung

Die Kamera- und LRF-basierte Sensorfusion ist ein aktives Forschungsfeld, in welchem eine Vielzahl unterschiedlicher Ansätze in den vergangenen Jahren publiziert wurden. Diese werden nach der Art der Kamera und der Art des LRFs (2D oder 3D) differenziert. Einen Vergleich von vier verschiedenen Kalibrierverfahren für 3D Entfernungsmessungen und Kamerabilder geben Cobzaş et al. [3]. Diese Techniken umfas-



(a)



(b)

Abb. 2: Darstellung des verwendeten Kalibrieramusters. Abb. (a) zeigt ein Kamerabild mit automatisch detektiertem Schachbrettmuster. Abb. (b) die gleiche Szene aus der Sicht eines Velodyne HDL-64E S2 in Form von punktuellen 3D Laser-Entfernungsmessungen (nach orthographischer Projektion).

sen Punkt und Linien-basierte Rückgewinnung der rigiden Transformationen und bildbasierte Zuordnung der Datensätze. Die Autoren Zhang und Pless [4] präsentieren eine Methode um einen 2D LRF mit einer Kamera manuell aus mehreren Ansichten zu kalibrieren. Der Ansatz registriert die Laser-Entfernungsmessungen von einem planaren Kalibriermuster mit dem Kamerabild des selben Objekts. Unnikrishnan und Hebert [1] erweitern den Ansatz von Zhang und Pless zur Kalibrierung eines 3D LRF mit einer Kamera. Dazu extrahieren sie eine Ebene aus manuell ausgewählten Eckpunkten des Kalibrieramusters und registrieren diese automatisch mit dem erkannten Muster aus dem Kamerabild. Ähnlich zu diesem Ansatz von Unnikrishnan und Hebert arbeitet die Methode von Andreasson et al. [5]. Die Autoren entwerfen ein neues Kalibriermuster welches aus einem Schachbrettmuster besteht das von grauem Klebband umschlossen ist. Auf diese Weise ist das Kalibriermuster detektierbar, da der Klebbandrahmen in den Intensitätswerten des LRFs sichtbar wird. Eine Kalibrierung einer omni-direktionalen Kamera auf einen 3D LRF ohne Kalibriermuster wird von Scaramuzza et al. [6] vorgestellt. Eine manuelle Auswahl der Korrespondenzen in den Laser-Entfernungsmessungen und dem Kamerabild erlaubt die Berechnung der rigiden Transformationen zwischen den ausgewählten Punkten. Der Ansatz von Unnikrishnan und Hebert [1] ermöglicht eine schnelle und praktikable Kalibrierung, bei dem die intrinsischen Kameraparameter und die rigiden Transformationen der Kamera in Relation zum 3D LRF simultan ermittelt werden. Bei diesem Vorgang wird ein Schachbrettmuster manuell in den 3D Laser-Entfernungsmessungen selektiert. Für eine präzise Kalibrierung sind ca. 15 Laser-Entfernungsmessung / Kamerabild Paare, aus unterschiedlichen Entfernungen zum Sensoraufbau notwendig. Aufgrund dessen nimmt dieser Kalibriervorgang einige Zeit in Anspruch. Um diese Zeitspanne zu minimieren ist ein gut erkennbares Kalibriermuster in dem Sensordatenraum von entscheidender Bedeutung. Daher wurden unterschiedliche Experimente durchgeführt und ein Kalibriermuster entwickelt, welches den Kriterien gerecht wird. Eine

schnelle oder sogar vollständig automatisierte Erkennung erfordert ein Kalibrierobjekt, das sowohl in den 3D Laser-Entfernungsmessungen als auch in den Kamerabildern erkennbar ist. Graues Klebeband mit glänzender Oberfläche und matt-schwarze Lacke absorbieren nahezu vollständig die Intensitätswerte in LRF-Daten. Entsprechend konstruierten wir ein Kalibriermuster bestehend aus einer DIN A1 Aluminium Platte mit einem matt-schwarz lackiertem Schachbrettmuster, welches auf einem flexible Kamera-Tripod montiert wird. Der folgende Ansatz unterliegt der Prämisse dass der 3D LRF das Zentrum des Koordinatensystems definiert. Die Rotation R und die Translation t eines Punktes P_c von Kamera- in Laser-Koordinaten ist gegeben als

$$P_l = R P_c + t \quad . \quad (1)$$

Sei p_{lc} ein Tupel bestehend aus einen *Laserscan* (aller während einer 360° Drehung gesammelten 3D Laser-Entfernungsmessungen) und einem korrespondierenden Kamerabild. Für alle p_{lc} müssen die Orientierungen $\theta_{c,i}$, $\theta_{l,i}$ bezüglich Kamera respektive LRF und die Abstände $\delta_{c,i}, \delta_{l,i}$ der Ebenen $\theta_{c,i} \mathbf{x} - \delta_{c,i} = 0$ und $\theta_{l,i} \mathbf{x} - \delta_{l,i} = 0$ gefunden werden. Dazu werden $\theta_{c,i}$ und $\delta_{c,i}$ von der extrinsischen Kalibrierung der Kamera mit der Hilfe von *OpenCV* extrahiert (vgl. Abbildung 2 (a)). Um weiterhin $\theta_{l,i}$ und $\delta_{l,i}$ zu bestimmen muss das Kalibriermuster in den 3D Laserdaten lokalisiert werden. Die schwarzen Rechtecke des Schachbrettmusters sind in den Intensitätswerten detektierbar, da der LRF an diesen Stellen keine Messwert liefert. Anschließend werden alle Entfernungsmessungen auf eine Ebene projiziert und der Tiefenwert in Richtung der zu kalibrierenden Kamera verworfen (Binarisierung). Das Ergebnis ist

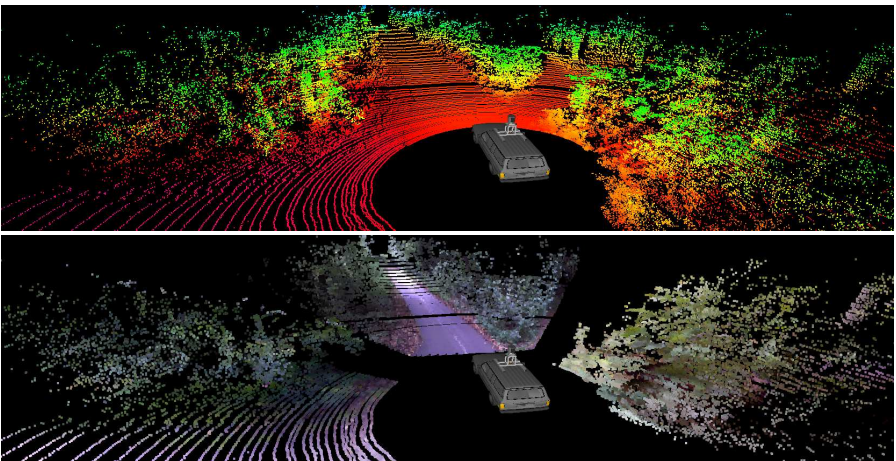


Abb. 3: Beispiel Szene einer unstrukturierten Umgebung. Die 3D Laser-Entfernungsmessungen sind im oberen Bildteil dargestellt wobei die Höhe eines 3D Punktes farblich kodiert ist. Das untere Bild zeigt die gleich Szene nachdem die Laser-Entfernungsmessungen mit den Bilddaten der drei Kameras überlagert worden sind.



Abb. 4: Bild der Frontkamera mit Schattensegmentierung nach [2]. Rote Kanten verbinden Regionen mit der selben Beleuchtungsklasse und blaue Kanten Regionen mit unterschiedlicher Beleuchtung.

ein "Bild" aus dem Blickwinkel der Kamera in dem das Kalibriermuster sichtbar sind (vgl. Abbildung 2 (b)). Nachdem die Eckpunkte des Kalibrierusters manuell ausgewählt wurden kann die Projektion invertiert werden um die zum Kalibriermuster korrespondierenden Messungen zu extrahieren. Daraufhin wird die Transformation bestimmt, welche die Diskrepanz zwischen der Pose des detektierten Kalibrierusters in den Kamera- und Laserdaten minimiert. Dabei folgen wir dem Ansatz von Unnikrishnan und Hebert [1] und optimieren:

$$\underset{\mathbf{R}, \mathbf{t}}{\operatorname{argmin}} \sum_{i=1}^n \sum_{j=1}^{m(i)} \|\boldsymbol{\theta}_{c,i}^T(\mathbf{R}\mathbf{x}_{i,j} + \mathbf{t}) - \delta_{c,i}\| \quad . \quad (2)$$

Mindestens 15 Tupel p_c sind für eine präzise Kalibrierung notwendig, wodurch in etwa 10 Minuten für jede Kamera und eine gesamte Kalibrierdauer von etwa 30 Minuten benötigt wird. Ein Beispiel der fusionierten Sensordaten ist in Abbildung 3 dargestellt. Das Bild zeigt einen Laserscan in einem Waldgebiet mit einer Straße. Im unteren Teil des Bildes wurden die Laser-Entfernungsmessungen mit den drei Kamerabildern überlagert.

4 Erkennen und Entfernen von Schatten

Die Verwendung von Farb- und Texturwerten ohne vorherige Vorverarbeitung hat sich als problematisch bei den ersten durchgeführten Experimenten und einer Evaluation herausgestellt. Hierbei zeigte sich dass Einfluss von Licht und Schatten durch das Sonnenlicht signifikant ist. Ohne die Beachtung bzw. Korrektur der Beleuchtung führt die Verwendung der Farbwerte nicht zur gewünschten Verbesserung der Klassifikation. Aus diesem Grund werden in diesem Abschnitt die aktuellen Arbeiten zum Erkennen und Entfernen von Schatten vorgestellt. Zur Erkennung und Entfernung von Schatten verwenden wir das Verfahren von Guo et al. [2]. Vor der Schattenerkennung wird das Bild zunächst mit dem *Mean Shift Algorithmus* nach Pandey et al.

[7] segmentiert. Anschließend werden die Segmente der Klasse Schattenregion oder Normalregion zugeordnet. Diese Klassifizierung der Segmente in Schatten- oder Normalregion und deren Vergleich wird durch eine Kombination von *Single Region Classification* und *Pair-wise Region Relationship Classification* umgesetzt. Dabei wird die *Single Region Classification* zuvor mit manuell gekennzeichneten Schattenregionen trainiert. Dies geschieht mithilfe von Histogrammen für Farbe und Textur, wobei angenommen wird, dass eine Schattenregion insgesamt dunkler und die Textur schwächer ausgeprägt ist, als bei einer Normalregion. Bei der *Pair-wise Region Relationship Classification* werden die Verhältnisse von Intensität und Verteilung der Farb- und Texturhistogramme zweier Regionen zueinander berechnet. Hieraus wird bestimmt, ob zwei Regionen das selbe Material aufweisen und, falls dies der Fall ist, ob es sich bei beiden Regionen um die selbe Beleuchtungsklasse handelt. Die Ergebnisse der Klassifikation werden in einem Relationship-Graphen abgespeichert. In den Knoten des Graphen sind die Segmente und in den Kanten die Relation der Segmente zueinander gegeben, d. h. ob zwei Segmente die gleiche Beleuchtung aufweisen. Alle vom Graphen miteinander verbundenen Segmente wurden von dem Schatten-Erkennungsalgorithmus als Segmente des gleichen Materials identifiziert. Die Abbildung 4 zeigt ein segmentiertes Bild der Frontkamera mitsamt dem Relationship-Graphen. Sowohl das Material als auch die Beleuchtungsklasse vieler Regionen wurden korrekt erkannt. In ebenso vielen Regionen wurde das Material in den anderen



(a)



(b)



(c)



(d)

Abb. 5: Ergebnisse der Schattenentfernung nach [2] mit Bildern der Roboter-Kamera. Links jeweils die Originalaufnahmen, rechts die Korrektur.

Region nicht wiedergefunden, dies tritt auch bei benachbarten Segmenten des selben Materials auf. Beispielsweise wurde das Segment des Schlagschattens von dem

Auto in der unteren linken Ecke des Bildes, nicht als das selbe Material erkannt, wie die der anderen Schatten. Auch wurde das Schattensegment in der unteren rechten Ecke der Straße nicht erkannt, obwohl es an einige benachbarte Segmente des selben Materials angrenzt, die richtig eingeordnet wurden. Kleinere Bereiche der Schatten, die von Ästen auf die Straße geworfen werden, wurden ebenfalls nicht erkannt. Die Abbildung 5 zeigt Beispielbilder des Schattenentfernungs-Verfahren, links die Originalbilder und rechts die Ergebnisbilder der Schattenentfernung. Bei dem Bildpaar (a)-(b) liefert die Schattenentfernung gute Ergebnisse; die Schlagschatten vom hohen Gras wurden vollständig entfernt. Im Bildpaar (c)-(d) war die Entfernung partiell erfolgreich. Ein Teil des Schattens auf dem Asphalt wurde entfernt, da die Schattenkante allerdings unscharf war, wurde der Schatten in zu viele Segmente unterteilt, die nicht alle als Schatten erkannt wurden. An den unscharfen Schattenkanten wurde der Schatten teilweise nicht entfernt. Gleichzeitig wurden die Ränder der als Schatten erkannten Segmente, die heller waren als die restliche Fläche der Schattensegmente zu stark aufgehellt. Während der Fahrt konnten die übrig gebliebenen Schattenflecken in der Textur, besonders in Verbindung mit den hellen Flecken, die Hinderniserkennung im Vergleich zu den Originalbildern sogar verschlechtern. Bei Testaufnahmen, die bei bewölktem Himmel aufgenommen wurden, und dadurch keine Schatten enthielten, wurden die Eingabedaten durch die Schattenentfernung stark verschlechtert, da an Stellen im Bild Schatten erkannt und aufgehellt wurden, wo keiner war. Es zeigte sich, dass die Segmentierung und Schattenerkennung nur auf den Bildern der Frontkamera ausgeführt werden konnte, da die Bilder der seitlich angebrachten Kameras einerseits eine zu niedrige Auflösung hatten, und andererseits die Bewegungsunschärfe der Seitenkameras zu groß ist, sodass die Segmentierung schlechter war und die Schattenerkennung keine befriedigenden Ergebnisse lieferte.

5 Terrainklassifikation

Im Folgenden wird der Stand der Technik in den beiden Bereichen Bildsegmentierung und Terrainklassifikation dargelegt. Die Bildsegmentierung ist ein grundlegender Vorverarbeitungsschritt in vielen Anwendungen der Bildverarbeitung, häufig werden hierzu Verfahren die auf Markov-Zufallsfelder beruhen verwendet. Das Problem der hohen Berechnungskomplexität wird von den Autoren Szirányi et al. [8] aufgezeigt. Hierzu wird eine vollständig parallelisierte Markov-Zufallsfeld-Architektur zur Bildsegmentierung eingeführt. Dabei konnte gezeigt werden, dass der Markov-Ansatz mit einer voll parallelisierten Zellnetzwerkarchitektur implementiert werden konnte. Meas-Yedid et al. [9] verwenden einen Markov-Zufallsfeld Cluster-Ansatz zur Farbsegmentierung welcher auf Farb- und Rauminformationen basiert. Dieser Ansatz zeigt mit den durchgeführten Experimenten empirisch die Robustheit gegenüber Rauschen, Farbänderungen, Beleuchtungsänderungen und Unschärfe. Vandapel et al. [10] segmentieren 3D Entfernungsdaten und klassifizieren die Segmente in drei unterschiedliche Klassen: Oberflächen, unstrukturiert oder Kabel. Der Ansatz verwendet einen stationären 3D LRF sowie einen rotierenden 2D LRF auf einer mobilen Plattform. Eine Stereo-Farbkamera und ein 2D LRF werden von Manduchi et al. [11] zur Hinderniserkennung eingesetzt. Das vorgestellte System führt eine Farb-basierte Klassifikation durch und analysiert die Laser Entfernungsmessungen um zwischen Rasen und

Hindernissen zu unterscheiden. Eine Stereokamera, eine Infrarotkamera und zwei 2D LRFs werden von Wellington et al. [12] verwendet. Mehrere Markov-Zufallsfelder interagieren miteinander durch ein Hidden Semi-Markov Modell. Die Ergebnisse belegen dass die Verwendung von Nachbarschafts-Informationen die Hindernis-Klassifikation signifikant verbessern konnte und lieferte die Idee unsere Kameras mit dem 3D LRF für die Terrainklassifikation zu fusionieren.

5.1 Markov-Zufallsfeld Nachbarschaft

Markov-Zufallsfelder können als zweidimensionale Erweiterung von Markov-Ketten verstanden werden. Dies bedeutet, dass die Orte S zueinander in Beziehung stehen über ein Nachbarschaftssystem, welches definiert ist als

$$\mathcal{N} = \{\mathcal{N}_i \mid \forall i \in S, \mathcal{N}_i \subset S\} \quad , \quad (3)$$

wobei \mathcal{N}_i die Menge der Orte ist, welche als Nachbarn von i gelten. Kein Ort ist Teil seiner eigenen Nachbarschaft und die Nachbarschaft gilt in beide Richtungen gemäß

$$i \notin \mathcal{N}_i \quad , \quad (4)$$

$$i \in \mathcal{N}_{i'} \Leftrightarrow i' \in \mathcal{N}_i \quad . \quad (5)$$

Benachbarte Zustände können unterschiedliche Abstände zueinander haben und bilden unterschiedliche Nachbarschafts-Cliquen. In dieser Arbeit wurden die 8 direkt angrenzenden Nachbarn eines Zustand in Gitter-Repräsentation verwendet. Dies liefert, mit Ausnahme von Zustände am Rande der Datenrepräsentation, die Nachbarschaft

$$\mathcal{N}_{i,j} = \{(i-1, j-1), (i-1, j), (i-1, j+1), (i, j-1), (i, j+1), (i+1, j-1), (i+1, j), (i+1, j+1)\} \quad , \quad (6)$$

für die Orte $S = \{(i, j) \mid 1 \leq i, j \leq n\}$ welche zu den Terrainzellen eines $n \times n$ Terrain-Gitters gehören. Die Annahme das jeder Zustand genau einen Wert annimmt liefert die Konfiguration ω eines Zufallsfeldes. Ein Zufallsfeld ist ein Markov-Zufallsfeld, wenn die Wahrscheinlichkeit einer Konfiguration einen positiven Wert annimmt als

$$P(\omega) > 0 \quad , \quad (7)$$

und wenn eine zufälliger Zustand ausschließlich von seinen Nachbarn beeinflusst wird.

$$P(\omega_{i,j} \mid \omega_{-\omega_{i,j}}) = P(\omega_{i,j} \mid \omega_\lambda) \quad , \quad (8)$$

wobei $\omega_{-\omega_{i,j}}$ die Menge ω ist ohne $\omega_{i,j}$ und ω_λ die Menge der Zustände der Nachbarschaft. Um die Berechnungen zu vereinfachen werden Markov-Zufallsfelder oft als Gibbs-Zufallsfelder modelliert. Markov-Zufallsfelder und Gibbs-Zufallsfelder können als äquivalent betrachtet werden, was durch das *Hammersley-Clifford Theorem* bewiesen wurde. In einem Gibbs-Zufallsfeld korrespondieren die zufälligen Zustände mit der Gibbs-Verteilung, welche definiert ist als

$$P(\omega) = \frac{1}{Z} \exp\left(-\frac{1}{T}U(\omega)\right) \quad (9)$$

wobei T ein Temperaturparameter ist und $U(\cdot)$ eine Funktion welche die Energie E einer Konfiguration berechnet. Die Normalisierungskonstante Z ist definiert als

$$Z = \sum_{\omega} \exp\left(-\frac{1}{T}U(\omega)\right) \quad (10)$$

und berechnet die Summe über alle mögliche Konfigurationen (den Konfigurationsraum). Die Modellierung von $U(\cdot)$ richtet sich dabei nach dem jeweiligen Anwendungsfall. $P(\omega)$ kann maximiert werden indem eine Konfiguration ω gefunden wird die $U(\omega)$ minimiert. Zu diesem Zweck existieren mehrere Stichprobeverfahren da es zu zeitaufwändig wäre alle möglichen Konfigurationen zu vergleichen. Markov-Zufallsfelder für Klassifikationsprobleme werden von Theodoridis und Koutroumbas [13] erörtert und mehrere Möglichkeiten für die Modellierung von Markov-Zufallsfelder werden von Li [14] diskutiert.

5.2 Merkmale

Merkmale sind essentiell für die Klassifikation. Wir definieren eine Vielzahl von Merkmalen welche die unterschiedlichen Terrainklassen unseres Ansatzes optimal repräsentieren. Dabei können die Merkmale in zwei Kategorien unterteilt werden. Bei der ersten Kategorie handelt es sich um Merkmale die aus dem 3D Laser-Entfernungsmessungen gewonnen werden können. Die Merkmale der zweiten Kategorie zeichnen sich dadurch aus, dass sie aus den Kamerabildern extrahiert werden können. Unser erstes Merkmal ist die *Rauheit* f_r einer Terrainzelle. Wie von Neuhaus et al. [15] vorgeschlagen lässt sich die Rauheit durch iterieren über alle Kreise die darin enthaltenen Entfernungsmessungen unseres LRFs berechnen. Ein Kreis entsteht dabei durch die Rotation eines der 64 Laserscanner im Sensorkopf des Velodyne HDL-64E S2. Innerhalb jedes dieser Kreise werden die Messungen Hoch-Pass gefiltert. Das Ergebnis des gefilterten Signal wird dann den jeweiligen Terrainzellen räumlich zugeordnet und als Maßstab der lokalen Rauheit interpretiert. Die Rauheits-Information eignet sich besonders für die Unterscheidung von Gras- und Ackerflächen gegenüber Straßen oder Wegen. Das zweite Laser-basierte Merkmal ist die *Höhendifferenz* f_h und ermöglicht eine Unterscheidung von Hindernissen und befahrbaren Regionen. Dieses Merkmal wurde von Happold et al. [16] vorgeschlagen und repräsentiert den Abstand zwischen der höchsten und der niedrigsten Laser-Entfernungsmessung in einer Terrainzelle. Die Verwendung dieses Merkmals geschieht unter unserer Annahme, dass das Hauptkriterium eines Hindernisses seine Höhe ist. Unter der Annahme, dass die Textur unterschiedlicher Terrain-Arten variiert, werden mehrere Bild-basierte Merkmale verwendet. Dabei verwenden wir die Texturmerkmale *Zweites Winkel-Moment* (engl. angular second moment) f_{asm} , die *Varianz* f_v und das *inverse Differenz-Moment* (engl. inverse difference moment) f_{dim} . Diese drei Merkmale stammen aus den 14 von Haralick et al. [17] vorgeschlagenen Texturmerkmalen und zeichnen sich durch ihre schnelle Berechenbarkeit aus. Ein weiteres schnell zu berechnendes Merkmal stellt die von Knauer et al. [18] vorgeschlagene Textur-Homogenität f_{th} dar. Diese wird unter der Annahme verwendet, dass raues Terrain eine inhomogenere Textur aufweist, als die Textur von flachen bzw. glattem Terrain. Darüber hinaus

wird ebenfalls die beobachtete Farbe f_c einer Terrainzelle für den Klassifikationsprozess verwendet, wobei das schattenkorrigierte Bild verwendet wird. Die Kombination aller Merkmale liefert den Merkmalsvektor $\mathbf{f} = (f_r, f_h, f_{sam}, f_v, f_{idm}, f_{th}, f_c)$, welcher pro Terrainzelle für den Klassifikationsprozess verwendet wird.

5.3 Markov-Modell des Terrains

Befahrbare Terrainzellen können je nach ihrer Oberflächenbeschaffenheit entweder schwer oder einfach für einen Roboter zu befahren sein. Aus diesem Grund verwenden wir die beiden Klassen *Straße* und *rau* um zwischen den Oberflächentypen unterscheiden zu können. Daher gibt es insgesamt vier Klassen *Unbekannt*, *Straße*, *rau*, *Hindernis*. Das Markov-Modell des Terrains sollte sowohl die Kontextinformationen der benachbarten Terrainzellen als auch die beobachteten Merkmale berücksichtigen. Daher verwenden wir ein Markov-Modell welches bereits in der Bildsegmentierung erfolgreich eingesetzt wird und von Deng et al. [19] vorgestellt wird. Dieses Modell ist in der Lage beide Anforderungen zu erfüllen und kann auf das Terrainklassifikations-Problem angewendet werden indem Terrainzellen als Zufalls-Zustände im Markov-Zufallsfeld betrachtet werden. Dabei wird die Energie E eines Gibbs-Zufallsfelder stückweise berechnet durch das Aufspalten in die Nachbarschaftskomponente $E_{\mathcal{N}}$ für benachbarte Terrainzellen und die Merkmalskomponente E_f , welche die Merkmale der jeweiligen Terrainzelle beschreibt. Die Nachbarschaftskomponente $E_{\mathcal{N}_{i,j}}$ einer Zelle $C_{i,j}$ an der Position (i, j) im Terraingitter ist definiert als

$$E_{\mathcal{N}_{i,j}} = \sum_{\lambda \in \mathcal{N}_{i,j}} \beta \cdot \delta(c_{i,j}, c_\lambda) \quad (11)$$

wobei $c_{i,j}$ die Terrainklasse ist, welche der Zelle $C_{i,j}$ zugewiesen ist und c_λ die Terrainklasse ist, welche Teil der Nachbarschaft von $C_{i,j}$ ist. Die Funktion $\delta(\cdot)$ liefert einen Rückgabewert von -1 für $c_{i,j} \neq c_\lambda$ und $+1$ für $c_{i,j} = c_\lambda$. β kann zur Gewichtung des Einflusses der Nachbarschaftskomponente in Relation zum totalen Abstand der Terrainzellen zueinander verwendet werden. In unserer Arbeit haben wir die Interaktion der Terrainzellen erfolgreich mit einer 8er-Nachbarschaft und einem fixen β durchführen können. Die korrespondierende Merkmalskomponente $E_{f_{i,j}}$ einer Zelle $C_{i,j}$ beruht auf der Annahme, dass die Zellmerkmale einer Gaußverteilung unterliegen. Als Energie ist ihre Berechnung definiert als

$$E_{f_{i,j}} = \sum_k \left(\frac{(f_{i,jk} - \mu_{i,jk})^2}{2\sigma_{i,jk}^2} + \log(\sqrt{2\pi}\sigma_{i,jk}) \right) , \quad (12)$$

wobei $f_{i,jk}$ dem k -ten Merkmal von $C_{i,j}$ und $\mu_{i,jk}$ und $\sigma_{i,jk}$ der Mittelwert bzw. die Standardabweichung des k -ten Merkmals der Klasse $c_{i,j}$ entsprechen, welches der Terrainzelle $C_{i,j}$ zugewiesen wurde. Die Kombination dieser beiden Energiekomponenten liefert die vollständige Energie $E_{i,j}$ einer Terrainzelle $C_{i,j}$, welche berechnet wird als

$$E_{i,j} = E_{\mathcal{N}_{i,j}} + \alpha \cdot E_{f_{i,j}} , \quad (13)$$

wobei α ein Gewichtungsfaktor ist, um den Einfluss der jeweiligen Energie-Typen zu steuern. Zur Klassifikation muss die Summer aller berechneten Energien $E_{i,j}$ minimiert werden. Diese Energien können minimiert werden indem eine Klasse für jede

Terrainzelle gefunden wird, die optimal zu dem Merkmal und den Klassen der benachbarten Terrainzellen passt. Für diesen finalen Schritt verwenden wir den Gibbs Sampler von Geman und Geman [20].

6 Zusammenfassung

Die Terrainklassifikation von fusionierten Farbkamerabildern und 3D Laser Entfernungsmessungen mit Markov-Zufallsfelder ist in der Lage, die Umgebung eines autonomen, mobilen Roboters in Echtzeit zu klassifizieren. Die Laufzeit des Ansatzes skaliert dabei mit der Auflösung und Reichweite des Terraingitters, sodass je nach verwendeter Hardware Terraingitter von bis zu $200 \times 200 \text{ m}^2$ in Echtzeit verarbeitet werden können. Als Herausforderung hat sich die direkte Verwendung von Farb- und Texturwerten und Merkmalen im Außengelände herausgestellt. Bisherige Ergebnisse zeigen, dass Farbe keinen signifikante Verbesserung in der Terrainklassifikation liefert [21]. Hier wird aktuell an einer Erweiterung gearbeitet, indem ein optimales Vorgehen zur Schattenentfernung speziell für unsere Anwendung gesucht wird. Hierzu müssen die Schattenverläufe für die Seitenkameras getrennt von der Frontkamera trainiert werden. Der Algorithmus für die Schattenerkennung sollte besonders robust auf Bildern arbeiten, die durch die Eigenbewegung des Fahrzeuges unscharf sind. Das Gesamtergebnis könnte außerdem verbessert werden, indem vor der Schattenentfernung untersucht wird, ob sich auf dem Bild überhaupt Schatten befinden, da die Schattenentfernung die Qualität der Bilder verschlechtern kann, falls keine Schatten vorhanden sind.

Danksagung

Diese Arbeit wurde von der Deutschen Forschungsgemeinschaft (DFG) unter Forschungsantrag PA 599/11-1 finanziert.

Literatur

- [1] UNNIKRISHNAN, R. ; HEBERT, M.: Fast Extrinsic Calibration of a Laser Rangefinder to a Camera / Robotics Institute. Pittsburgh, USA, 2005. – Forschungsbericht
- [2] GUO, R. ; DAI, Q. ; HOIEM, D.: Single-Image Shadow Detection and Removal using Paired Regions. In: *Proc. of the IEEE Int. Conf. on Computer Vision and Pattern Recognition*. Colorado Springs, USA, 2011, S. 2033–2040
- [3] COBZAŞ, D. ; ZHANG, H. ; JÄGERSAND, M.: A Comparative Analysis of Geometric and Image-Based Volumetric and Intensity Data Registration Algorithms. In: *Proc. of the IEEE Int. Conf. on Intelligent Robots and Systems*. Washington, USA, 2002, S. 2506–2511
- [4] ZHANG, Q. ; PLESS, R.: Extrinsic Calibration of a Camera and Laser Range Finder. In: *Proc. of the IEEE Int. Conf. on Intelligent Robots and Systems*. Sendai, Japan, 2004, S. 2301–2306
- [5] In: ANDREASSON, H. ; TRIEBEL, R. ; LILIENTHAL, A.: *Non-iterative Vision-based Interpolation of 3D Laser Scans*. Springer, Germany, 2007, S. 83–90
- [6] SCARAMUZZA, D. ; HARATI, A. ; SIEGWART, R.: Extrinsic Self Calibration of a Camera and a 3D Laser Range Finder from Natural Scenes. In: *Proc. of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, 2007, S. 4164–4169

- [7] COMANICIU, D. ; MEER, P.: Mean shift: A robust approach toward feature space analysis. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24 (2002), S. 619–619
- [8] SZIRÁNYI, T. ; ZERUBIA, Josiane ; CZÚNI, László ; GELDREICCH, D. ; KATO, Z.: Image Segmentation Using Markov Random Field Model in Fully Parallel Cellular Network Architectures. In: *Real-Time Imaging* 6 (2000), Nr. 3, S. 195–211
- [9] MEAS-YEDID, V. ; TILIE, S. ; OLIVO-MARIN, J.-C.: Color Image Segmentation Based on Markov Random Field Clustering for Histological Image Analysis. In: *Proc. of the 16th Int. Conf. on Pattern Recognition*. Quebec City, Canada, 2002, S. 796–799
- [10] VANDAPEL, N. ; HUBER, D. ; KAPURIA, A. ; HEBERT, M.: Natural Terrain Classification Using 3-D Ladar Data. In: *Proc. of the IEEE Int. Conf. on Robotics and Automation*. New Orleans, USA, 2004, S. 5117–5122
- [11] MANDUCHI, R. ; CASTANO, A. ; TALUKDER, A. ; MATTHIES, L.: Obstacle Detection and Terrain Classification for Autonomous Off-road Navigation. In: *Autonomous Robots* 18 (2004), Nr. 1, S. 81–102
- [12] WELLINGTON, C. ; COURVILLE, A. ; STENTZ, A.: Interacting Markov Random Fields for Simultaneous Terrain Modeling and Obstacle Detection. In: *Proc. of Robotics Science and Systems*. Cambridge, USA, 2005, S. 1–8
- [13] THEODORIDIS, S. ; KOUTROUMBAS, K.: *Pattern Recognition*. Academic Press, 2009. – ISBN 978–0–123–69531–4
- [14] LI, S.: *Markov Random Field Modeling in Computer Vision*. Springer, 2009. – ISBN 978–4–431–70145–3
- [15] NEUHAUS, F. ; DILLENBERGER, D. ; PELLEZ, J. ; PAULUS, D.: Terrain Drivability Analysis in 3D Laser Range Data for Autonomous Robot Navigation in Unstructured Environments. In: *Proc. of the IEEE Int. Conf. on Emerging Technologies and Factory Automation*. Catalonia, Spain, 2009, S. 1686–1689
- [16] HAPPOLD, M. ; OLLIS, M. ; JOHNSON, N.: Enhancing Supervised Terrain Classification with Predictive Unsupervised Learning. In: *Proc. of Robotics: Science and Systems*. Philadelphia, USA, 2006
- [17] HARALICK, R. ; SHANMUGAM, K. ; DINSTEN, I.: Textural Features for Image Classification. In: *IEEE Transactions on Systems, Man, and Cybernetics* 3 (1973), Nr. 6, S. 610–621
- [18] KNAUER, U. ; MEFFERT, B.: Fast Computation of Region Homogeneity with Application in a Surveillance Task. In: *Proc. of ISPRS Commission V Mid-Term Symposium Close Range Image Measurement Techniques*. Newcastle, UK, 2010, S. 337–342
- [19] DENG, H. ; CLAUSI, D. A.: Unsupervised Image Segmentation Using a Simple MRF Model with a New Implementation Scheme. In: *Pattern Recognition* 37 (2004), Nr. 12, S. 2323–2335
- [20] GEMAN, S. ; GEMAN, D.: Stochastic Relaxation, Gibbs Distribution and Bayesian Restoration of Images. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 6 (1984), Nr. 6, S. 721–741
- [21] HÄSELICH, M. ; ARENDS, M. ; LANG, D. ; PAULUS, D.: Terrain Classification with Markov Random Fields on fused Camera and 3D Laser Range Data. In: *Proc. of the 5th Euro. Conf. on Mobile Robotics*. Örebro, Schweden, 2011, S. 153–158