Semantische 3D Modellierung aus Bildern mit geometrischem Vorwissen[1]

Christian Häne[2]

Abstract: 3D Rekonstruktion aus Bildern und semantische Segmentierung von Farbbildern wurden traditionellerweise als zwei separate Probleme betrachtet. Meine Dissertation führt Algorithmen ein, die es erlauben die 3D Rekonstruktion und die semantische Segmentierung in einer einzelnen konvexen Optimierung zu berechnen. Durch eine starke Kopplung zwischen Geometrie und Semantik kann dadurch eine bessere Rekonstruktion erreicht werden.

1 Einleitung

Vollautomatische digitale 3D Modellierung von realen Umgebungen und Objekten aus einem Satz von Eingabebildern wurde in den letzten zwei Jahrzehnten ausgiebig untersucht. Das Ziel ist es, eine digitale Kopie von realen Umgebungen und Objekten zu erstellen. Die Anwendungsmöglichkeiten dieser Technologie sind vielseitig. Während für einige der Anwendungen das digitale 3D Modell das schlussendliche Ziel ist, wird für andere Anwendungen die digitale 3D Rekonstruktion als Vorbereitungsschritt verwendet. Digitale 3D Modelle können zum Beispiel als digitales Archiv für Kunstgegenstände oder archäologische Stätten dienen. Das *Digital Michelangelo* Projekt [Le00] hat Skulpturen von Michelangelo mittels Lasern vermessen und digital in 3D rekonstruiert. Das *MURALE* Projekt [Va02] hat Technologien entwickelt um archäologische Stätten digital einzuscannen und zu visualisieren. Ein anderer Anwendungsfall sind Kinofilme. Das System, welches in [DTM96] präsentiert wurde, ist später verwendet worden um einige Spezialeffekte für den Film *The Matrix* zu erstellen. Digitale Rekonstruktion von Umgebungen spielt auch eine wichtige Rolle in der Robotik. Ein Roboter möchte über seine Umgebung informiert sein, um Pfadplanung durchzuführen und Hindernisse zu erkennen. Für diese Anwendungen ist die 3D Rekonstruktion meist ein vorbereitender Schritt. Sobald ein 3D Modell der Umgebung verfügbar ist, wird das 3D Modell weiter analysiert, um zu verstehen wo der Roboter sich frei bewegen kann, und wo sich die Hindernisse befinden. Ein Beispiel einer solchen Anwendung für selbstfahrende Fahrzeuge ist [BFP09]. Ein weiteres Gebiet in dem die Semantik der Geometrie eine wichtige Bedeutung hat, ist im Computer Aided Design (CAD). Angenommen ein Architekt möchte ein existierendes Gebäude digital erfassen als Ausgangspunkt für einen Umbau. Ein System das automatisch zwischen semantischen Klassen wie Gebäude, Boden oder Vegetation unterscheidet, kann sehr hilfreich sein da es die Bearbeitungszeit gegenüber einer manuellen Verarbeitung stark verringern kann.

[1] Englischer Titel der Dissertation: "Semantic 3D Modeling from Images with Geometric Priors"

[2] Department of Electrical Engineering and Computer Sciences, UC Berkeley, chaene@eecs.berkeley.edu

Das Hauptziel meiner Dissertation [Hä16] war es, Algorithmen zu entwickeln, welche die Geometrie und die semantische Segmentierung in einem gemeinsamen Vorgang berechnen. Frühere Systeme haben die zwei Probleme separat gelöst. Durch die Optimierung von Geometrie und Segmentierung mittels einer gemeinsamen Optimierung können bessere Lösungen erzielt werden. Eine Gebäudefassade ist zum Beispiel viel wahrscheinlicher vertikal als horizontal. Solche Abhängigkeiten können in einer gemeinsamen Formulierung ausgenützt werden und bessere Resultate erzielt werden, als wenn beide Probleme separat gelöst werden.

Um eine digitale 3D Rekonstruktion zu erstellen wird ein Sensor benötigt, der die reale 3D Geometrie vermessen kann. Die Sensoren können in zei Gruppen eingeteilt werden: aktive und passive Sensoren. Unter den aktiven Sensoren sind die beliebtesten Sensoren Laser Scanner, Structured Light Scanner oder Time-of-Flight Kameras. Passive Sensoren wie zum Beispiel Fotokameras können auch als 3D Sensor verwendet werden. Dafür werden mehrere Bilder aufgenommen. Durch die Erstellung von Korrespondenzen zwischen den Bildern wird die Position von Punkten im dreidimensionalen Raum mittels Triangulation bestimmt. Um mehrere Messungen, die vom selben Blickwinkel aufgenommen wurden zu speichern, werden Bilder verwendet, in denen jedes Pixel einer Tiefenmessung entspricht, genannt Tiefenkarte. Kameras haben den Vorteil gegenüber aktiven Systemen, dass sie bei vielen Wetterbedingungen verwendet werden können, sehr wenig Instandhaltung benötigen, günstig sind und in vielen Systemen bereits verbaut sind. Zusätzlich enthalten die Bilddaten wichtige Informationen für die semantische Segmentierung. Die Experimente, die in meiner Arbeit gemacht wurden, basieren ausschliesslich auf Tiefenkarten die aus Farbbildern berechnet wurden. Viele der vorgeschlagenen Algorithmen sind jedoch nicht auf diesen Sensortyp limitiert. Alle Sensoren haben gemeinsam, dass für ein komplettes und konsistentes 3D Modell mehrere Tiefenkarten in eine gemeinsame Repräsentation fusioniert werden. Um Messungenauigkeiten und Rauschen zu entfernen wird eine Optimierung verwendet, die glatte Lösungen bevorzugt.

Um semantische Information aus Bildern zu extrahieren wird zuerst eine Merkmalbeschreibung eines Pixels und seines Kontexts aus den Eingabebildern extrahiert. Solche Merkmale versuchen die Textur des Bildes, vorherrschende Gradienten, Farbe und viele weitere Informationen, die aus Farbbildern extrahiert werden können, zu beschreiben. Mittels manuell segmentierten Referenzbildern und maschinellem Lernen wird ein Klassifikator gelernt, der zwischen den semantischen Klassen unterscheiden kann. Die Ausgabe solcher Systeme ist eine pro Superpixel oder pro Pixel Wahrscheinlichkeit für jede der semantischen Klassen. Meistens sind diese rohen Klassifikatorausgaben verrauscht und folgen den Objektkanten nicht. Deshalb wird häufig eine globale Optimierung in einem Markov Random Field (MRF) oder eine konvexe Optimierung verwendet.

Wie bereits oben erwähnt wurde die Optimierung von 3D Geometrie und semantische Segmentierung von Bildern traditionellerweise als zwei separate Probleme betrachtet und deshalb in einer individuellen Optimierung berechnet. Meine Dissertation hat Methoden eingeführt, die es erlauben, die semantische und geometrische Information in einem gemeinsamen Fusionierungsprozess zu bestimmen. Die semantische Segmentierung von 3D Geometrie kann je nach Anwendungsfall verschiedene Granularität aufweisen. Um ein

selbstfahrendes Fahrzeug zu bewegen, kann es ausreichend sein, zwischen fahrbarer Geometrie und Hindernissen zu unterscheiden. Für Rekonstruktionen von Gebäuden, zum Beispiel für Katastermodelle oder Gebäudeverwaltung werden Klassen wie Boden, Gebäude und Vegetation verwendet. Obwohl das immer noch relativ wenige Klassen sind, gibt es bereits Wechselbeziehungen zwischen Geometrie und Semantik. Zum Beispiel hat die Klasse Boden eine starke Präferenz horizontal zu sein und horizontale Geometrie hat eine starke Bevorzugung Boden zu sein. Wenn man auch Übergänge zwischen zwei semantischen Klassen, die nicht direkt in den Eingabebildern gesehen werden können in Betracht zieht, wie zum Beispiel den Übergang zwischen Boden und Gebäude, kann man auch geometrisches Vorwissen über solche Übergäge ausnützen. Um Objekte zu rekonstruieren, für welche Instanzen aus einer spezifischen Objektklasse eine ähnliche Geometrie haben, kann man auch objektklassenspezifisches Vorwissen ausnützen. Diese Formulierungen rekonstruieren nicht nur das Objekt, sondern auch die unmittelbare Umgebung und geben eine Segmentierung zwischen Objekt und Umgebung aus.

Meine Dissertation ist in drei Teile gegliedert. Für diese Kurzfassung werde ich im Folgenden die drei Teile vorstellen und deren wichtigsten Ergebnisse präsentieren.

2 Echtzeitrekonstruktion mit Fischaugenbildern

Der erste Teil meiner Arbeit stellt ein Echtzeitsystem zur Hinderniserkennung für selbstfahrende Fahrzeuge vor. Das System verwendet vier Fischaugenkameras die so auf dem Fahrzeug angebracht sind, dass der ganze Raum rund um das Fahrzeug von Kameras gesehen werden kann. Ähnliche Kamerasysteme sind in einigen Serienfahrzeugen bereits eingebaut. Die vorgeschlagene Methode, ermöglicht es, Tiefenkarten direkt auf Fischaugenbildern zu berechnen (siehe Abb. 1). Frühere Verfahren haben dazu die Fischaugenbilder zuerst in normale Lochbildkamerabilder oder andere geeignete Projektionsmodelle umgerechnet, dabei war häufig eine binokulare Kameranordnung notwendig. Das vorgeschlagene System verwendet mehrere, zeitlich nacheinander aufgenommene Bilder von derselben Kamera. Um eine 3D Rekonstruktion zu ermöglichen muss sich das Fahrzeug dabei bewegen. Der Vorteil, die Tiefenkarten direkt auf den Fischaugenbildern zu berechnen liegt darin, dass dadurch das gesamte Sichtfeld der Fischaugenkamera verwendet werden kann. Des Weiteren zeigen die Experimente, dass die Tiefenmessungen durch das Verwenden des originalen Projektionsmodells auch auf dem Sichtbereich den beide Versionen (Fischaugenbild und ein in ein Lochbildkamerabild umgerechnetes Bild) abdecken, weniger Ausreisser enthalten. Die Berechnungen werden auf einer Grafikkarte in Echtzeit durchzuführt. Die Kameraposen, die für die Berechnung benötigt werden, können direkt aus der Odometrie vom Fahrzeug extrahiert werden.

Die Daten, die in den Tiefenkarten enthalten sind, sind für Navigationsaufgaben von selbstfahrenden Fahrzeugen in der Form ungeeignet, da sie nur die Geometrie darstellen und keine semantische Information enthalten. Deshalb verwendet das System elementare semantische Information, nämlich ob eine gewisse Geometrie ein Hindernis darstellt oder nicht. Es wird für jeden 3D Punkt ausgewertet, ob er zum Boden gehört oder vom Boden hervorsteht. Gleichzeitig werden mehrere 3D Punkten zusammengefasst und nur jeweils

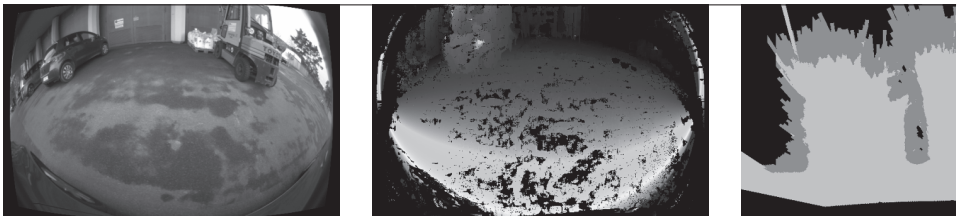

Abb. 1: Beispielresultate des vorgeschlagenen Rekonstruktionssystems für selbstfahrende Fahrzeuge. (links) Beispiel Eingabebild, (mitte) Beispiel Tiefenkarte (rot nah, blau fern), (rechts) Beispiel Hinderniskarte (grün fahrbar, rot Hindernis, schwarz unbeobachtet).

das Hindernis, welches auf der 2D Bodenebene in einer gewissen Richtung am nächsten zur Kamera ist, ausgegeben. Durch das Zusammenfassen mehrerer Punkte wird die Information komprimiert und gleichzeitig werden Ausreisser gefiltert und die Messungenauigkeit über mehrere Messungen gemittelt. Diese Hindernisinformationen, die für jede Kamera separat berechnet werden und mit einer Wiederholrate von 12.5Hz (Bildwiederholrate der verwendeten Kameras) über die Zeit ausgegeben werden, werden in einem Fusionierungsschritt in eine 2D Hinderniskarte eingetragen (siehe Abb. 1). Diese Hinderniskarte kann dann für Navigationsaufgaben, wie zum Beispiel Pfadplanung, verwendet werden. Gegenüber früheren Systemen hat das vorgeschlagene System den Vorteil, dass es keine spezielle binokulare Anordnung der Kameras benötigt, das ganze Sichtfeld der Fischaugenkameras verwendet wird und es die Kameraposen direkt von der Fahrzeugodometrie beziehen kann. Das System wird im selbstfahrenden Auto vom V-Charge Forschungsprojekt [Sc16] verwendet.

3 Volumetrische semantische 3D Rekonstruktion

Im zweiten Teil meiner Arbeit stelle ich ein System vor, welches gemeinsame volumetrische 3D Rekonstruktion und semantische Segmentierung ermöglicht und es erlaubt, klassenabhängiges Vorwissen über die Geometrie zu verwenden. Die klassische volumetrische Tiefenkartenfusionierung formuliert das Problem als Segmentierung eines Volumens in freier Raum und besetzter Raum. Die Oberfläche kann dann als Übergang vom freien in den besetzten Raum extrahiert werden. Um semantische Information in die Formulierung einzuführen, wird die 2-Klassen Segmentierung in eine semantische Segmentierung erweitert, in welcher jedem Volumenelement eine semantische besetzte Klasse oder freier Raum zugewiesen wird. Alle Rekonstruktionsmethoden die in diesem Teil vorgestellt werden haben gemeinsam, dass die semantische Segmentierung des Volumens und somit die finale Rekonstruktion, mittels eines einzelnen konvexen Minimierungsproblems gelöst wird.

3.1 Mathematische Segmentierungsformulierung

In der Literatur wurden zwei verschiedene Möglichkeiten vorgeschlagen um Segmentierungsprobleme zu formulieren. Eine Möglichkeit ist es, den Raum von Beginn an als diskreten Raum zu betrachten. Die Formulierung ist dann ein probabilistisches graphisches

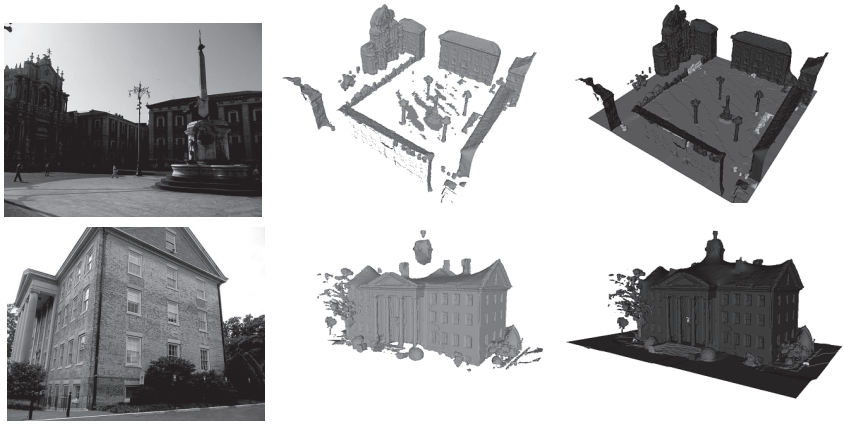

Abb. 2: Beispielresultate, (links) Beispieleingabebild, (mitte) Standardrekonstruktion der Geometrie, (rechts) Resultate des gemeinsamen Rekonstruktions- und Segmentierungssystems.

Modell, wobei die Elemente des Raums als Knoten im Graph repräsentiert werden, die mit Kanten verbunden sind, welche verwendet werden um die Segmentierung zu regularisieren. Die zweite Möglichkeit ist es, den Raum im Kontinuum darzustellen und das Segmentierungsproblem direkt darin zu formulieren. In dieser Art von Formulierung wurden konvexe Relaxationen vorgeschlagen, die es erlauben global optimale Lösungen für das relaxierte Problem zu finden. Der Raum wird nur am Schluss für die numerische Optimierung diskretisiert. Die Formulierungen im Kontinuum haben den Vorteil, dass die Lösungen wesentlich weniger Artefakte vom diskreten Gitter aufweisen. Sie haben jedoch den Nachteil, dass sie nur Regularisierungskosten, die eine Metrik über die Klassen bilden, erlauben. Das heisst, die Dreiecksungleichung muss erfüllt sein, was bedeutet dass die Summe der Kosten für die Sprünge von Klasse A zu Klasse B und von Klasse B zu Klasse C grösser gleich einem direkten Sprung von Klasse A zu Klasse C sein muss. Diese Einschränkung existiert bei den diskreten graphischen Modellen nicht. Die vorgeschlagene Formulierung wird durch die Analyse dieser zwei Formulierungen hergeleitet. Sie hat keine Einschränkung auf metrische Regularisierungskosten und hat trotzdem die Vorteile bezüglich der Gitterartefakte die Formulierungen im Kontinuum haben. Des Weiteren ermöglicht die Formulierung die Verwendung von Regularisierungskosten, die von der Richtung des Übergangs zwischen den Klassen abhängt.

3.2 Gemeinsame 3D Rekonstruktion und semantische Segmentierung

Um eine gemeinsame 3D Rekonstruktion und semantische Segmentierung zu erreichen, wird die oben eingeführte mathematische Formulierung für semantische Segmentierung verwendet. Die Methode wird im Volumen formuliert. Jedes Volumenelement bekommt eine Klasse zugewiesen. Um die Segmentierung durchzuführen werden die Datenkosten und die Regularisierungskosten benötigt. Die Datenkosten beschreiben, wie wahrscheinlich es ist, dass ein gewisses Volumenelement eine gewisse Klasse zugewiesen bekommt. Die Re-

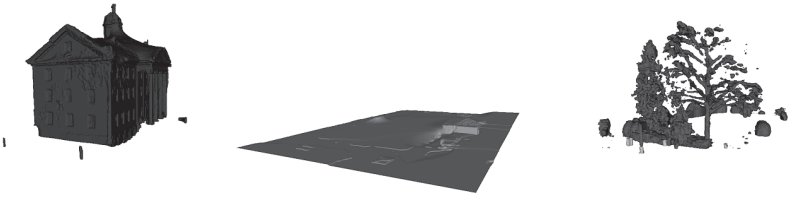

Abb. 3: Individuelle semantische Segmente. Die Ausgabe des Systems lässt sich direkt in die einzelnen semantischen Segmente zerlegen. Auch versteckte Oberflächen wie der Übergang zwischen Boden und Gebäude werden vom vorgeschlagenen Algorithmus abgeschätzt.

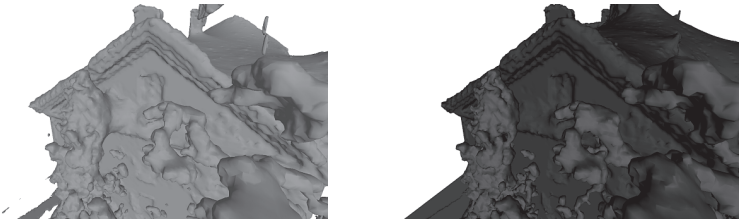

Abb. 4: Detailvergleich, (links) in der geometrischen Rekonstruktion wird der Baum fälschlicherweise mit dem Gebäude verbunden, (rechts) dank den semantischen Klassen rekonstruiert das System die Gebäudefassade korrekt.

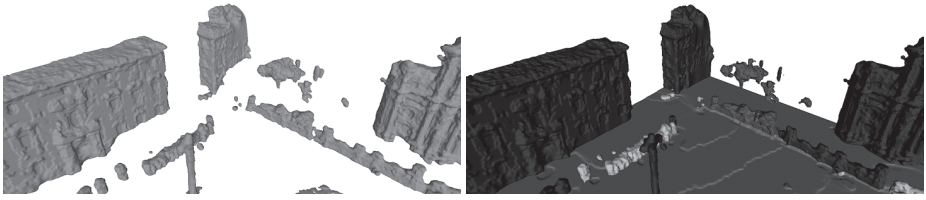

Abb. 5: Detailvergleich, (links) in der geometrischen Rekonstruktion fehlt der Boden und die Gebäude sind im unteren Teil nicht komplett, (rechts) der vorgeschlagene Algorithmus rekonstruiert den Boden und die Gebäude stehen korrekt auf dem Boden.

gularisierungskosten beschreiben wie wahrscheinlich ein Übergang zwischen zwei semantischen Klassen in eine bestimmte Richtung ist oder anders gesagt, wie stark ein Übergang bestraft wird. Die Datenkosten werden aus den Tiefenkarten und pro Pixel semantischen Klassenwahrscheinlichkeiten des semantischen Klassifikators berechnet. Die Regularisierungskosten werden von einem Katastermodell gelernt. Das Katastermodell enthält die Oberflächen zwischen semantischen Klassen, was es ermöglicht die Wahrscheinlichkeiten eines Übergangs zwischen zwei bestimmten semantischen Klassen mit einer bestimmten Richtung zu bestimmen. Um den konvexen Regularisierungsterm daraus zu bestimmen, werden Wulff-Formen verwendet. Mathematisch sind Wulff-Formen konvexe Mengen. In der vorgeschlagenen Methode werden sie verwendet um die Regularisierungskosten durch das Auswählen einer geeigneten konvexen Form zu definieren.

Der Algorithmus wurde auf Datensätzen aus Bildern von Gebäuden ausgewertet. Beispiele aus Datensätzen von ca. 120 Bildern sind in Abb. 2 dargestellt. Die Ausgabe des Systems kann direkt in die einzelnen semantischen Segmente zerlegt werden (siehe Abb. 3). Dank dem geometrischen Vorwissen, welches im Regularisierungsterm eingebunden ist, werden auch verdeckte oder für traditionelle geometrische Rekonstruktionen nicht genügend gesehene Oberflächen rekonstruiert (siehe Abb. 4, 5).

3.3 Objektrekonstruktion mit kategoriespezifischem Vorwissen

Objekte die transparente, reflektierende oder sonst für die 3D Rekonstruktion schwierige Oberflächen aufweisen stellen ein besonderes Problem für die 3D Rekonstruktion dar. Um trotzdem eine überzeugende Rekonstruktion zu erhalten, wird in der Literatur vorgeschlagen, die Ähnlichkeit der Objektformen innerhalb einer Objektklasse auszunutzen. Meine Arbeit schlägt eine neuartige Formulierung vor, die auf den Oberflächenrichtungen basiert und direkt in die oben beschriebene Multiklassenrekonstruktion eingebunden werden kann. Die zugrunde liegende Beobachtung ist, dass die Oberflächenrichtungen an korrespondierenden Stellen von Objekten der gleichen Klasse ähnlich sind. Zum Beispiel ist bei einem Auto das Dach nahe der horizontalen Richtung und die Windschutzscheibe ist ungefähr im 45° Winkel. Um während dem Rekonstruktionsvorgang Formen der Klasse des Objekts gegenüber beliebigen Formen zu bevorzugen, wird ein Regularisierungsterm verwendet, der die Oberflächenrichtungen, die in dieser Objektklasse lokal vorkommen, bevorzugt. Gegenüber den Rekonstruktionen von Gebäuden von oben ist nun der Regularisierungsterm zusätzlich positionsabhängig. Das heisst, für jedes Volumenelement vom Rekonstruktionsvolumen wird eine separate Wulff-Form aus Trainingsdaten extrahiert. Um diese Prozedur effizient durchzuführen werden die Wulff-Formen als Schnittmenge von Halbräumen parametrisiert. Für diese Formulierung werden als Datenkosten ausschliesslich Tiefenkarten verwendet, die aus den Eingabebildern berechnet wurden. Es wird keine bildbasierte semantische Segmentierung benötigt. Die Segmentierung wird alleine dadurch bestimmt, dass die Umgebung die nicht zur gewählten Objektklasse gehört, nicht mit dem kategoriespezifischen Vorwissen übereinstimmt. Einige Resultate sind in Abb. 6 dargestellt. Mit der vorgeschlagenen Methode wird für die Klasse Auto der Boden korrekt unter dem Auto rekonstruiert, das Auto berührt den Boden nur an den Rädern. Dies ist möglich, da das vorgeschlagene Modell von den Trainingsdaten gelernt hat, dass ein Übergang zwischen Auto und Boden nur an den Rädern vorkommt.

Die oben vorgestellte oberflächenrichtungsbasierte Formulierung hat den Nachteil, dass für jedes Volumenelement eine Wulff-Form gespeichert werden muss. Zusätzlich benötigt die Optimierung als Eingabe die Orientierung und Position des zu rekonstruierenden Objekts. Für Formulierungen wo der Regularisierungsterm nicht positionsabhängig ist, und deshalb Übergänge im Ganzen Rekonstruktionsvolumen gleich bestraft werden, ist eine Ausrichtung bezüglich der Orientierung, welche in der Praxis einfacher berechnet werden kann, genügend. Die Alternative ist eine Rekonstruktionsmethode für Objekte, welche keine positionsabhängigen Regularisierungsterme verwendet. Um trotzdem interessante Objektformen zu beschreiben, teilt die vorgeschlagene Methode die Objekte in Segmente auf. Für die Klasse Tisch sind das zum Beispiel die Tischplatte und die Tischbeine. Gleichzeitig

Abb. 6: Ergebnisse für die Rekonstruktion von Autos und Flaschen. (von links nach rechts pro Beispiel), Beispiel Eingabebild, Rekonstruktion ohne kategoriespezifisches Vorwissen, Rekonstruktion mit dem oberflächenrichtungsbasierten kategoriespezifischen Vorwissen.

zur Rekonstruktion erhält man dadurch auch eine Segmentierung des Objekts. Um nun den Regularisierungsterm zu definieren, muss nur eine geeignete Wulff-Form, also eine konvexe Form, für jeden Übergang definiert werden. Für das Beispiel Tisch wird der Übergang Tischplatte zu freier Raum so gewählt, oder von Trainingsdaten gelernt, dass horizontale Flächen wesentlich weniger stark bestraft werden als vertikale Flächen. Der Übergang von Tischbein zu freier Raum soll hauptsächlich vertikal sein. Des Weiteren soll das Tischbein oben einen Übergang zur Tischplatte und unten zum Boden haben. Einige Beispielergebnisse des Algorithmus sind in Abb. 7 dargestellt.

4 Tiefenkartenregularisierung mit geometrischem Vorwissen

Im dritten Teil meiner Dissertation werden Methoden untersucht, deren Ausgabe eine Tiefenkarte ist. Eine erste Methode ist dazu gedacht, Umgebungen zu rekonstruieren, die hauptsächlich aus planaren Flächen bestehen, also von Menschenhand geschaffene Umgebungen wie zum Beispiel Städte. Der Algorithmus rendert zuerst mehrere Tiefenkarten die einen ähnlichen Sichtpunkt haben, in eine einzelne Ansicht. Der eingeführte Regularisierungsterm bestraft Flächen die nicht planar sind. Dazu wird ein Patchverzeichnis verwendet. Die Optimierung bestimmt daher pro Pixel die Koeffizienten der Verzeichniselemente. Durch die Verwendung eines Totalvarationsterms als Regularisierung der Koeffizienten wird eine stückweise planare Rekonstruktion bevorzugt. Ein Beispiel ist in Abb. 8 dargestellt.

Ein zweiter Algorithmus verwendet einen Klassifikator, der von einem einzelnen Farbbild Wahrscheinlichkeiten für Oberflächenrichtungen vorhersagen kann. Diese zusätzlichen Informationen über die Geometrie helfen an Stellen, wo eine direkte Berechnung der Tiefe nicht möglich ist, zum Beispiel wenn die Bilder nicht genügend Textur aufweisen. Im vor-

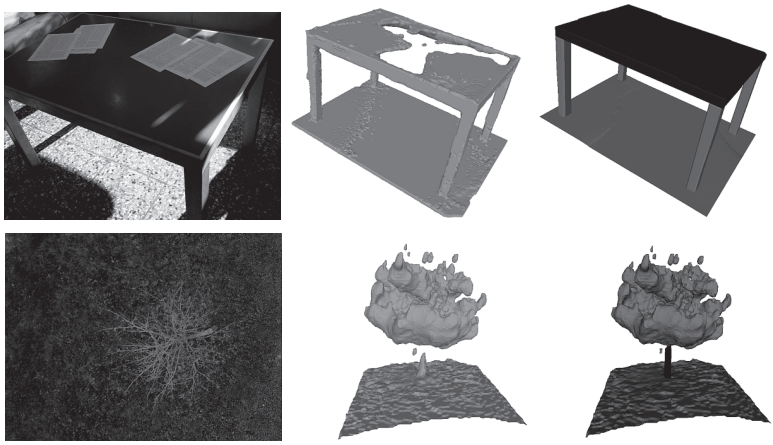

Abb. 7: Ergebnisse für die Rekonstruktion von Tischen und Bäumen, (links) Beispiel Eingabebild, (mitte) Rekonstruktion ohne kategoriespezifisches Vorwissen, (rechts) Rekonstruktion mit dem segmentbasierten kategoriespezifischen Vorwissen.

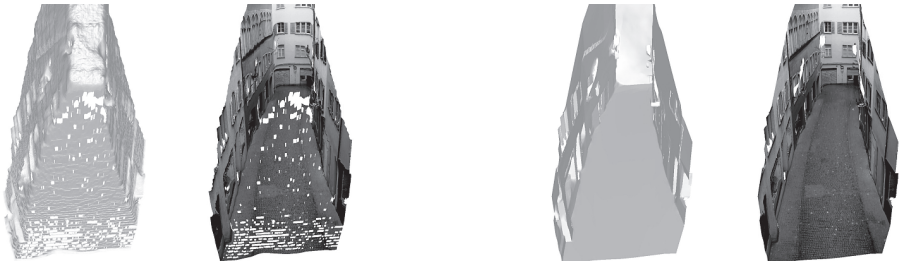

Abb. 8: Stückweise planare Rekonstruktion, (links) Rekonstruktion mit standard Regularisierungsterm, (rechts) Rekonstruktion mit dem vorgeschlagenen Regularisierungsterm der stückweise planare Flächen bevorzugt.

geschlagenen Algorithmus werden die Wahrscheinlichkeiten der Oberflächenrichtungen, in ähnlicher Weise wie oben für die volumetrischen Rekonstruktionen, in Wulff-Formen dargestellt und als Regularisierungsterm verwendet. Im Extremfall kann man damit Rekonstruktionen aus nur einem Farbbild berechnen, wenn auch ein von Trainingsdaten gelernter Klassifikator für die Abschätzung der Tiefe verwendet wird (siehe Abb. 9).

5 Schlussfolgerung

Meine Arbeit hat Algorithmen eingeführt, die gemeinsame semantische Segmentierung und 3D Rekonstruktion ermöglichen. Diese zwei Probleme wurden in früheren Systemen unabhängig gelöst. Durch eine Kopplung der Geometrie und der Semantik mittels Oberflächenrichtungen erreichen die eingeführten Algorithmen eine bessere Rekonstruktion, als wenn nur die Geometrie rekonstruiert wird. Zusätzlich erhält man eine konsistente semantische Segmentierung der Geometrie als Ausgabe vom Algorithmus.

Abb. 9: Rekonstruktion von einem einzelnen Farbbild, (links) Eingabebild, (mitte, rechts) berechnete Punktwolke.

Literaturverzeichnis

[BFP09] Badino, Hernán; Franke, Uwe; Pfeiffer, David: The stixel world-a compact medium level representation of the 3d-world. In: Symposium of the German Association for Pattern Recognition (DAGM). 2009.

[DTM96] Debevec, Paul E; Taylor, Camillo J; Malik, Jitendra: Modeling and rendering architecture from photographs: A hybrid geometry-and image-based approach. In: International Conference on Computer graphics and interactive techniques (SIGGRAPH). 1996.

[Hä16] Häne, Christian: Semantic 3D Modeling from Images with Geometric Priors. Dissertation, ETH Zürich, 2016.

[Le00] Levoy, Marc; Pulli, Kari; Curless, Brian; Rusinkiewicz, Szymon; Koller, David; Pereira, Lucas; Ginzton, Matt; Anderson, Sean; Davis, James; Ginsberg, Jeremy; Shade, Jonathan; Fulk, Duane: The digital Michelangelo Project: 3D Scanning of Large Statues. In: International Conference on Computer graphics and interactive techniques (SIGGRAPH). 2000.

[Sc16] Schwesinger, Ulrich; Bürki, Mathias; Timpner, Julian; Rottmann, Stephan; Wolf, Lars; Paz, Lina Maria; Grimmett, Hugo; Posner, Ingmar; Newman, Paul; Häne, Christian; Heng, Lionel; Lee, Gim Hee; Sattler, Torsten; Pollefeys, Marc; Allodi, Marco; Valenti, Francesco; Mimura, Keiji; Goebelsmann, Bernd; Derendarz, Wojciech; Mühlfellner, Peter; Wonneberger, Stefan; Waldmann, Rene; Grysczyk, Sebastian; Last, Carsten; Brüning, Stefan; Horstmann, Sven; Bartholomäus, Marc; Brummer, Clemens; Stellmacher, Martin; Pucks, Fabian; Nicklas, Marcel; Siegwart, Roland: Automated Valet Parking and Charging for e-Mobility Results of the V-Charge Project. In: Intelligent Vehicles Symposium (IV). 2016.

[Va02] Van Gool, Luc; Pollefeys, Marc; Proesmans, Marc; Zalesny, Alexey: The MURALE project: image-based 3d modeling for archaeology. British Archaeological Reports (BAR) International Series, 1075:53–64, 2002.

Christian Häne hat sein Doktorat mit einem Dr. sc. Titel an der ETH Zürich 2016 abgeschlossen. Davor hat er seinen BSc und MSc Titel in Informatik von der ETH Zürich erhalten. Während seines Doktorats hat er ein dreimonatiges Praktikum bei Microsoft Research in Cambridge, UK absolviert. Heute ist er als Postdoctoral Scholar an der University of California, Berkeley tätig. Er hat dafür vom Schweizerischen Nationalfonds ein Mobilitätsstipendium bekommen. Für seine Doktorarbeit hat er die ETH Medaille für ausgezeichnete Doktorarbeiten erhalten.