ORIGINALARBEIT

Elektrotechnik & Informationstechnik (2017) 134/6: 304–311. DOI 10.1007/s00502-017-0515-1

Roboter lernen mit Gegenständen umzugehen: neue Entwicklungen und Chancen

M. Vincze, M. Zillich, J. Prankl

Für den zukünftigen Einsatz von Robotern fordern Experten aus der Industrie Roboter, die sicher sind und dem Benutzer eine klare Rückmeldung geben, was der Roboter macht und warum. Eine der wichtigsten Funktionen, um dies zu erreichen, ist die zuverlässige Erkennung der Umgebung, um den Roboter mit einem besseren Verstehen der Welt auszustatten. In diesem Artikel geben wir eine Übersicht derzeitiger Entwicklungen. Mit der stetig ansteigenden Rechenleistung und neuen bildgebenden Sensoren, wie Stereo- oder Tiefenbildkameras, steigen die Möglichkeiten, Objekte und deren Umgebung immer besser und besser zu erkennen und zu verstehen. So können aus Bilddatenbanken eine große Anzahl an Objekten gelernt und auch wieder erkannt werden. Des Weiteren ist es möglich, Modelle auch aus den 3D CAD Daten von Objekten zu lernen. Dadurch können Klassen von Objekten erkannt werden, ohne vorher einzelne Objekte modellieren zu müssen. Zusätzlich können aus Beispielen die typischen Einrichtungsgegenstände wie Tische, Montageplätze oder Stühle und Kästen gelernt werden. Damit wird es möglich, Robotern ein erstes Verständnis der Umgebung mitzugeben. Dadurch eröffnen sich neue Anwendungen sowohl für Industrieroboter als auch Service–Roboter in der Industrie, im Bereich von Dienstleistungen und auch für zukünftige Roboter zu Hause.

Schlüsselwörter: Objekterkennung; Szenenverstehen; Klassifikation

Robots learn to handle objects: new developments and chances.

Experts predict that future robot applications will require safe and predictable operation: robots will need to be able to explain what they are doing to be trusted. To reach this goal, they will need to perceive their environment and its object to understand better the world and task they have to perform. This article gives an overview of present advances with the focus on options to model, detect, classify, track, grasp and manipulate objects. With the approach of colour and depth (RGB-D) cameras and the approaches in deep learning, robot vision was pushed considerably over the last years. It is possible to model and recognise objects, though prove in industrial settings is yet outstanding. Given a first detection of larger structures such as tables, chairs or assembly places, relations between object and setting can be obtained leading to a first interpretation of the scenes. We highlight present developments and point out future developments towards service and industrial robotics applications.

Keywords: object detection; scene understanding; classification

Eingegangen am 9. August 2017, angenommen am 15. September 2017, online publiziert am 18. September 2017

1. Einleitung

Zwei Richtungen prägen derzeit die Entwicklung in der Robotik. Dank neuer Sicherheitsmerkmale, wie nachgiebiger Arme, Kraftsensoren in Gelenken und berührungsempfindlicher Haut, kommen Industrieroboter hinter den Zäunen hervor. Dadurch entstehen neue Anwendungen in der Kooperation zwischen Roboter und Mensch in der Fertigung. Und dank besserer Methoden basierend auf künstlicher Intelligenz, wie in der Bildverarbeitung, Navigation oder Pfadplanung, ist zu erwarten, dass Roboter in Service-Anwendugen mehr und mehr neben den Menschen arbeiten und helfen werden. Beispiele hiefür sind Roboter in Pflegeheimen zur Unterstützung des Personals, Roboter in Geschäften beim Empfang, und auch Roboter zu Hause die über die Staubsaugroboter hinaus vielfältige Tätigkeiten ausführen werden wie zum Bespiel Dinge vom Boden aufzuheben [3].

Für diese beiden Arten von zukünftigen Roboteranwendungen fordern Experten, dass die Roboter sicher sind und dem Benutzer eine klare Rückmeldung geben [26]. Letzteres ist vor allem wichtig, um Vertrauen in die neuen Roboter zu bekommen. Nur so wird es gelingen, dass sich Roboter nahtlos in Fertigungsprozesse und eine Zusammenarbeit mit dem Menschen einfügen werden, also zukünftige, *vertrauenswürdige Roboter*.

Eine der wichtigsten Funktionen um Vertrauen zu erreichen ist die zuverlässige Erkennung der Umgebung. Nur so werden Roboter mit einem besseren Verstehen der Welt ausgestattet und können mit dem Menschen arbeiten und den Menschen in seiner Arbeit unterstützen. Im industriellen Umfeld spricht man von einer Kooperation von Mensch und Roboter mit einer klaren Übergabe und Trennung der Aufgaben. Dies heißt aber auch, dass man vertrauenswürdigen Roboter nicht nur vertrauen, sondern auch etwas zutrauen kann. Wozu ein gewisses Verstehen der Umwelt Voraussetzung ist.

Zum Verstehen der Umgebung eines Roboters tragen zuletzt Entwicklungen insbesondere in der Bildverarbeitung bei. Mit der ste-

Vincze, Markus, Technische Universität Wien, Institut für Automatisierungs- und Regelungstechnik, Gußhausstraße 25-25a, 1040 Wien, Österreich (E-Mail: markus.vincze@tuwien.ac.at); **Zillich, Michael,** Technische Universität Wien, Institut für Automatisierungs- und Regelungstechnik, Gußhausstraße 25-25a, 1040 Wien, Österreich; **Prankl, Johann,** Technische Universität Wien, Institut für Automatisierungs- und Regelungstechnik, Gußhausstraße 25-25a, 1040 Wien, Österreich

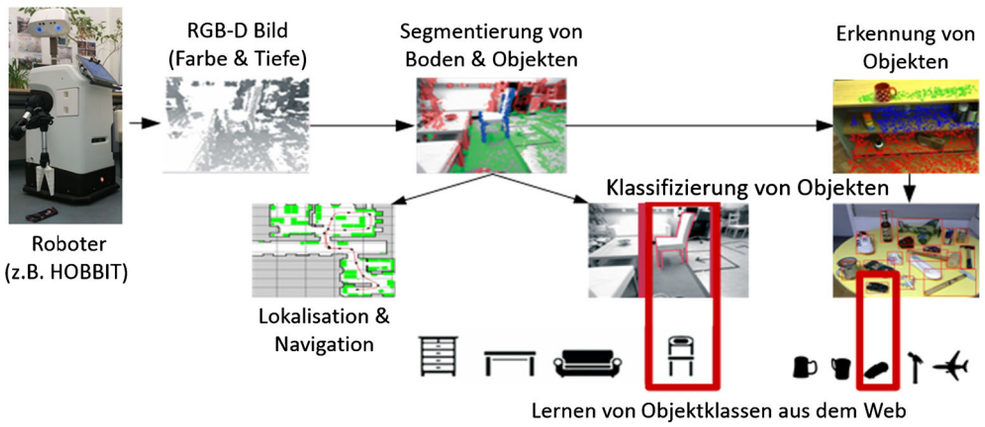

Abb. 1. Der situierte Ansatz zum Verstehen von Szenen: jede Information wie Boden, Grenzen und Wände, horizontale Ebenen, etc., werden verwendet, um Methoden zur Erkennung von Objekten mit Wissen aus der Situation zu verbessern und zuverlässiger zu machen. Ziel des Ansatzes ist es, jeden Teil der Szene semantisch erklären zu können

tig ansteigenden Rechenleistung und neuen bildgebenden Sensoren, wie Stereo- oder Tiefenbildkameras, steigen die Möglichkeiten Objekte und deren Umgebung immer besser und besser zu erkennen und zu verstehen. So können aus Bilddatenbanken eine große Anzahl an Objekten gelernt werden und auch wieder erkannt werden. Dies ermöglicht den Einsatz in Service Anwendungen, wo der Mensch Bilder aufnimmt und Informationen zu den aufgenommen Gegenständen erhält. Diese Entwickelungen treiben aber auch Anwendungen im Sehen für Roboter voran. So können Objekte komplett aus allen Richtungen modelliert werden und diese Modelle für eine Objekterkennung aber auch die Verfolgung der Bewegung des Objektes verwendet werden. Des Weiteren ist es möglich, Modelle auch aus den 3D CAD Daten von Objekten zu lernen. Dadurch können Klassen von Objekten erkannt werden, ohne vorher einzelne Objekte modellieren zu müssen. Zusätzlich können aus Beispielen die typischen Einrichtungsgegenstände wie Tische, Montageeinrichtungen oder Stühle und Kästen gelernt werden. Damit wird es möglich Robotern ein erstes Verständnis der Umgebung mitzugeben.

In diesem Beitrag geben wir einen Einblick in derzeitige Möglichkeiten situationsbedingt Gegenstände zu (1) modellieren und zu erkennen, (2) zu verfolgen, und (3) zu greifen, basierend auf visuellen Informationen. Durch konsequentes Nutzen bereits bekannter Informationen wird ein Bild der Situation aufgebaut, siehe Abb. 1. Dieses kontextuelle Wissen ermöglicht ein besseres Verstehen der Umgebung.

Dieser Überblick soll dazu beitragen, besser zu verstehen, welche Fähigkeiten bereits möglich sind und welche Entwicklungen in naher Zukunft zu erwarten sind. Abschnitt 2 geht kurz auf den situierten Ansatz ein, Abschn. 3 beschäftigt sich mit der Modellierung und Erkennung von Objekten, und Abschn. 4 mit der visuellen Erkennung für das Greifen von Objekten.

2. Situierte Wahrnehmung: Sehen aus der Sicht von Robotern

Bevor ein näherer Blick in Methoden der Objekterkennung, Verfolgung oder des Greifens erfolgt, ist es hilfreich die Besonderheiten visueller Fähigkeiten von Robotern hervorzuheben. Derzeit wird viel über Neuronale Netze und Deep Learning gesprochen (z. B. [8, 12]). Diese Methoden setzen große Datenbanken von Bildern voraus. Die Bilder wurden von Menschen aufgenommen, beinhalten daher eine überlegte Auswahl des Bildausschnittes, und sind meist Farbbilder. Tiefen- oder 3D Information ist nicht vorhanden. Erst neuere kleine Datenbanken umfassen auch Bilder mit Tiefeninformation. Große Fortschritte wurden erzielt in der Erkennung von Objekten die in sehr vielen (mehrere tausend) Bildern gelernt wurden, z. B. [6, 12].

Da Roboter jedoch selbst Bildausschnitte wählen müssen und über keinen Sinn für "gute" Bilder verfügen, sind die Erkennungsraten jedoch meist sehr viel geringer [27]. Bilder wurden bei nicht optimalen Beleuchtungsverhältnissen aufgenommen, Bildausschnitte schneiden Objekte an, die Situation in der sich Objekte befinden ist nicht immer ersichtlich, oder Gegenstände wurden von speziellen Ansichten oder mit Verdeckung aufgenommen.

In der Robotik hat sich daher ein Ansatz bewährt, der das Wissen über die Situation des Roboters aktiv verwendet [11, 28]. Dieser situierte Ansatz verwendet schrittweise alle Information, die dem Robotersystem zu Verfügung steht, um durch dieses kontextuelle Wissen ein besseres Verstehen der Umgebung zu erlangen. Abb. 1 macht dies deutlich.

Jede Bewegung des Roboters erfordert als ersten Schritt die Erkennung von freier Fläche. Die freie Fläche wird für die Navigation verwendet, aber fließt auch direkt in die Lokalisierung ein. Damit ist aber auch die Lage des Roboters und der visuellen Sensoren bestimmt, und es können zum Boden parallele Flächen gefunden werden. Denn auf diesen horizontalen Flächen werden weitere Objekte zu finden sein, die für die Aufgaben des Roboters relevant sind. Und so lassen sich die nächsten Schritte deutlich verbessern, indem die größeren Strukturen in der Umgebung des Roboters wie Wände, Tische, Stühle, und Ablageflächen als Unterstützung für die Erkennung von Gegenständen herangezogen werden können [31].

Ein Beispiel zur Erkennung dieser Strukturen zeigt Abb. 2. Dies ist ein erster Schritt, sodass Roboter nicht nur wissen in welchem Raum sie sind, sondern auch die großen Strukturen erkennen können um in weiterer Folge darauf verweisen zu können. Zum Beispiel zum gezielten Erkennen von Objekten.

3. Modellierung, Erkennung und Verfolgung von Objekten

Eine wichtige Aufgabe von Robotern sowohl in Industrie als auch Service Anwendungen ist das Erkennen und Verfolgen von Objekten. Jede Handhabung eines Objektes benötigt die Erkennung des Gegenstandes und die Bestimmung der Position und Orientierung (Pose). Entsprechende Methoden können eingeteilt werden in die Erkennung von Objekten anhand ihrer Textur (oder dem Aussehen) im Farbbild oder ihrer Form, meist im Tiefenbild.

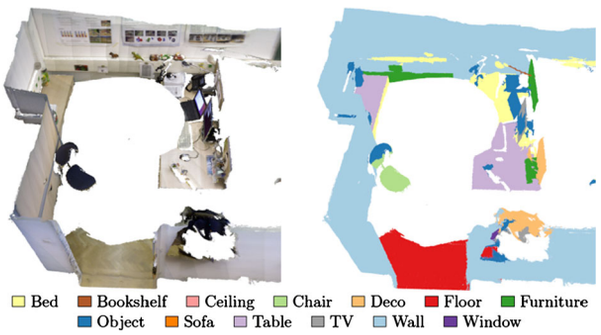

Abb. 2. Erkennung der großen Strukturen als erster Schritt zu einem Verstehen der Umwelt

3.1 Objekterkennung in Farbbildern anhand der Textur

Texturierte Gegenstände werden durch lokale markante Merkmale (sogenannte Interest Points) erkannt [13]. Aufbauend auf diesen Arbeiten gelingt es derzeit, Gegenstände aus mehreren Ansichten zu modellieren und wiederzuerkennen. Frühere Methoden basieren auf der Verwendung regelmäßiger Grundkörper wie Quader oder Zylinder [17], siehe Abb. 3.

Neuere Methoden erweitern die Modellierung auf Objekte beliebiger Geometrie. Dazu gibt es Ansätze, die die 3D Geometrie rekonstruieren und damit auch für weitere Anwendungen, wie z. B. das Greifen zugänglich machen [23] oder Ansätze, die mittels neuronalen Netzwerken jede Ansicht des Objekts lernen [14].

Abbildung 4 zeigt ein Beispiel in dem nicht nur die Objekte erkannt werden, sondern direkt die Pose bestimmt wird um die Gegenstände mit dem Roboter greifen zu können [19]. Ein weiterer Vorteil dieser Methode ist, dass beliebige Modelle durch drehen vor der Kamera leicht gelernt werden können und dass gelernte Modell direkt mit Methoden zur Erkennung des Objektes und dem Verfolgen von Objekten, beides mit 3D Pose, verknüpft sind (RTMT – Recognition, Tracking and Modelling of Objects Toolbox, open source verfügbar unter http://www.acin.tuwien.ac.at/?id=450).

3.2 Objekterkennung in Tiefenbildern anhand der Form

Sowohl der Einsatz immer besserer Farbkameras als auch die Verfügbarkeit von Tiefenbildern beschleunigten die Entwicklungen in der Bildverarbeitung in den letzten Jahren. Beispiele für Sensoren, die Videos mit Tiefenbildern liefern, sind Sensoren wie die Kinect (Microsoft, Asus, neue Varianten auch von Orbbec), RealSense (Intel), Astra Pro (Orbbec) oder PMD (Infineon). Während davor übliche Stereokamerasysteme große Rechenleistung benötigen, liefern diese Tiefenbildkameras direkt ein 2.5D Bild und die Rechenleistung ist für die Aufgaben der Bildverarbeitung verfügbar.

Erste Arbeiten mit Tiefenbildern erforschten Merkmale, um das klassische Problem in der Robotik, die Erkennung von Objekten zu ermöglichen. Beispiele für erste 3D Merkmale sind FPFH [20] und VFH [21]. Beide beschreiben die Ansichten eines Objektes, da sich rasch herausstellte, dass herkömmliche Ansätze wie spin images [7]

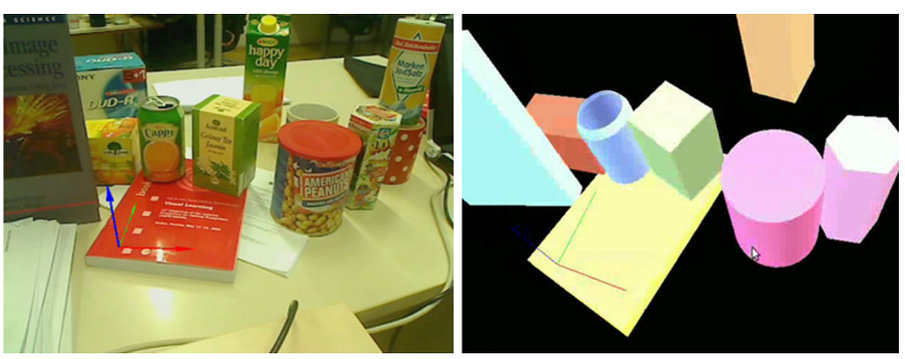

Abb. 3. Objekterkennung mittels Textur in Farbbildern: Erkennung der Objekte und Modellierung mit Grundkörpern. *Links:* typische Szenen mit Objekten und Verdeckungen. *Rechts:* Erkannte Objekte in 3D-Ansicht

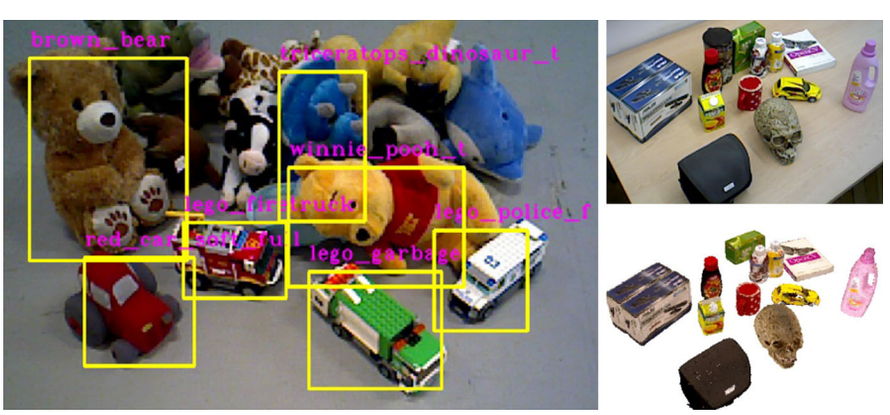

Abb. 4. Objekterkennung mittels Textur und Form in RGBD Bildern: *Links:* umschreibende Rechtecke – die Erkennung erfolgt in 3D. *Rechts:* Szene und darunter die Ansicht in 3D. In beiden Beispielen wird die 3D-Pose der Objekte erkannt und kann für das Greifen der Objekte verwendet werden

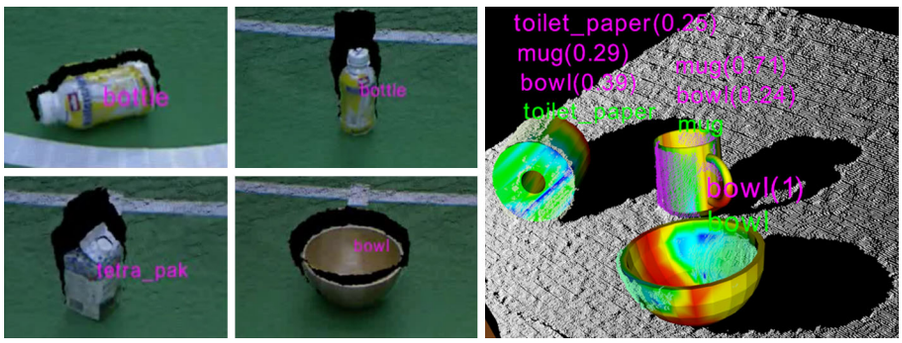

Abb. 5. Objekterkennung anhand der Form von Objekten: aus CAD-Modellen gelernte Objektklassen können mittels ihrer Form erkannt werden. *Links*: vier Beispiele für das Erkennen von Objekten in RGBD-Bildern, die das System nie gesehen hat, aber aufgrund der Form den richtigen Klassen zuordnen kann. *Rechts*: Wahrscheinlichkeiten und korrekt eingepasstes und verifiziertes Objektmodell. Die Position und Orientierung des Gegenstandes wird vollständig bestimmt.

Abb. 6. Objekterkennung mit starken Verdeckungen: mit Hilfe der genannten Methoden zum Erkennen verschiedener Objekte mittels Farb- und Tiefenbilder werden viele Hypothesen kreiert (*Mitte*), um dann in einer Optimierung die Daten der Szenen zu erklären (*rechts*). Damit könen auch Szenen mit vielen sich verdeckenden Objekten interpretiert werden (*links*) [2]

wenig tauglich für Anwendungen in der Robotik sind und eher rein theoretische Beiträge liefern.

Heute ist die beste Referenz für Methoden um Tiefenbilder zu bearbeiten die Point Cloud Library (PCL) [22]. Für die Modellierung einzelner Objekte ist die Form nicht so aussagekräftig wie die Textur, siehe oben. Die Form ist jedoch besonders geeignet, um Objektklassifizierungen zu ermöglichen. Da das Lernen einer Klasse aus vielen Beispielen sehr zeitaufwendig ist, stellt [30] eine Methode vor, die direkt aus 3D Modellen eine Beschreibung zur Wiedererkennung des Objekts lernt. Durch das automatische Kreieren von Ansichten mit einem spezifischen Sensormodell werden sehr gute Klassifiziererfolge erzielt. Insbesondere, da ja die in Abb. 5 gezeigten Objekte nie vorher gesehen wurden, es wurde ja nur von Modellen aus dem Internet gelernt. Ebenso ist das Lernen neuer Klassen rasch möglich, ohne dem System viele Beispielobjekte zeigen zu müssen [29].

So wie auch in Farbbildern sind die meisten Ansätze ausgelegt um einzeln stehende Objekte zu erkennen. SIFT erweitert dies bei gut sichtbarer Textur auf Teile von Objekten. In Tiefenbildern ist Verdeckung ebenso zu beachten und abhängig von der Markanz der Form des noch sichtbaren Teiles des Gegenstandes. Da in 3D die Tischplatte leichter als in Farbbildern segmentierbar ist, sind hier 3D Ansätze im Vorteil. Aber auch dann bedarf es spezieller Ansätze, um Szenen mit vielen Objekten zuverlässig zu interpretieren. Ein Beispiel ist die Arbeit in [2], die einige der üblichen Datenbanken mit Szenen mit vielen Objekten und Verdeckungen in 3D komplett gelöst hat. Abbildung 6 zeigt ein Beispiel.

Vor allem wenn gute Tiefendaten wie aus Lasersystemen vorhanden sind, sind auch große Verdeckungen möglich. Bei der derzeit noch eingeschränkten VGA Auflösung der gängigen Tiefenbildkameras gelingen komplettes verstehen von Szenen mit ausreichend großen Objekten. Die laufende Verbesserung der Sensoren wird auch hier mehr und mehr Anwendungen mit immer kleineren Teilen ermöglichen.

3.3 Verfolgung von Objekten und Posebestimmung

Auch bei der Verfolgung von Objekten kann man unterscheiden welche Merkmale benützt werden. Für Gegenstände mit Textur ist meist das Aussehen ausreichend für eine Verfolgung der Objektbewegung. Zusätzlich können Methoden aber auch die Form oder Informationen aus Tiefenbildern verwenden, siehe oben oder auch [1].

Wichtig in der Anwendung von Methoden zur Objektverfolgung ist die Initialisierung, die meist auf einer Objekterkennung basiert. Deshalb sind Methoden zur Modellierung von Objekten wie BLORT und RTM (siehe oben) wertvoll, um beides direkt zu bewerkstelligen. Fortschritte in der Objekterkennung basieren einerseits auf einer Adaption von Bild zu Bild [24], andererseits verwenden die noch robusteren Ansätze ein 3D Modell des Gegenstandes um daraus die Pose abzuleiten, z. B. [18] oder [15].

Besonders interessant ist die Verknüpfung von Objektverfolgung mit einer Vorhersage, was das Objekt machen wird. Beispiele sind Rollen oder Schieben von Objekten [16], oder sogar ein Schieben und Umwerfen [10]. Während im ersten Ansatz aus den Beobachtungen Objektaktionen und -Folgen gelernt werden können, verwendet die Vorhersage im zweiten Ansatz eine physikalische Simulation um abzuleiten, was als Folge der Bewegung mit dem Objekt passieren wird. Abbildung 7 zeigt eine Sequenz, in der trotz deutlicher Verdeckung die richtige Aktion, das Umfallen des Objektes, erkannt und verfolgt wird. Diese Kombination aus Simulation und Abgleich mit dem Gesehenen verspricht immer bessere Methoden zu entwickeln, die nicht nur Objekte erkennen sondern die Szenen verstehen und die Vorgänge deuten können.

Abb. 7. Verfolgen der Pose eines Objektes mit Verdeckung und Vorhersage der wahrscheinlichen Position nach der Bewegung mit dem Roboter. Mit Hilfe einer Simulation werden physikalisch wahrscheinliche Stellungen vorausgesagt, um Hinweise zu geben, dass die Schachtel umgefallen sein könnte. Verfolgen der Bewegung mit dieser Methode ist erfolgreich, siehe punktierte Linien im letzten Bild. *Dunkle/rote Linien*: Verfolgung nur mit Konturen. *Helle/blaue Linien*: Erkennung der verdeckten Konturen, um Verfolgen zu verbessern.

4. Posebestimmung und Greifen von Objekten

Sowohl die Methoden zur Erkennung (Abb. 6) als auch zur Verfolgung von Objekten (Abb. 7) beinhalteten bereits Methoden zur Bestimmung der 3D Position und Orientierung (Pose) von Objekten. Die klassische Aufgabe in der Robotik ist das Greifen von Objekten. Dank der visuellen Eingabe und der 6D Pose Bestimmung, ist dies für die so erkannten Objekte relativ leicht möglich. Hier sei auch auf den Artikel von Justus Piater in dieser Ausgabe verwiesen.

Ein Schritt weiter ist das Verstehen, welche Funktion Gegenstände mit bringen. So sind Häferl als Behälter zu gebrauchen, aber nur wenn die Öffnung nach oben deutet. Diese Aktionspotenziale (affordances, [5]) genannten Funktionen, lassen sich auch auf Roboter umsetzen. So können Gegenstände gelernt werden, die als Behälter dienen, also nach oben offene Inhalte präsentieren. Abbildung 8 zeigt, wie nicht nur das Glas erkannt wird, sondern auch die Schachtel als Behälter, und aus der relativen Pose der beiden Objekten die notwendigen Bewegungen für den Roboter zum Ausleeren abgeleitet werden [25].

Es gibt viele Anwendungen, bei denen nicht immer alle oder einige der Objekte erkannt werden können. Die Aufgabe "Alles vom

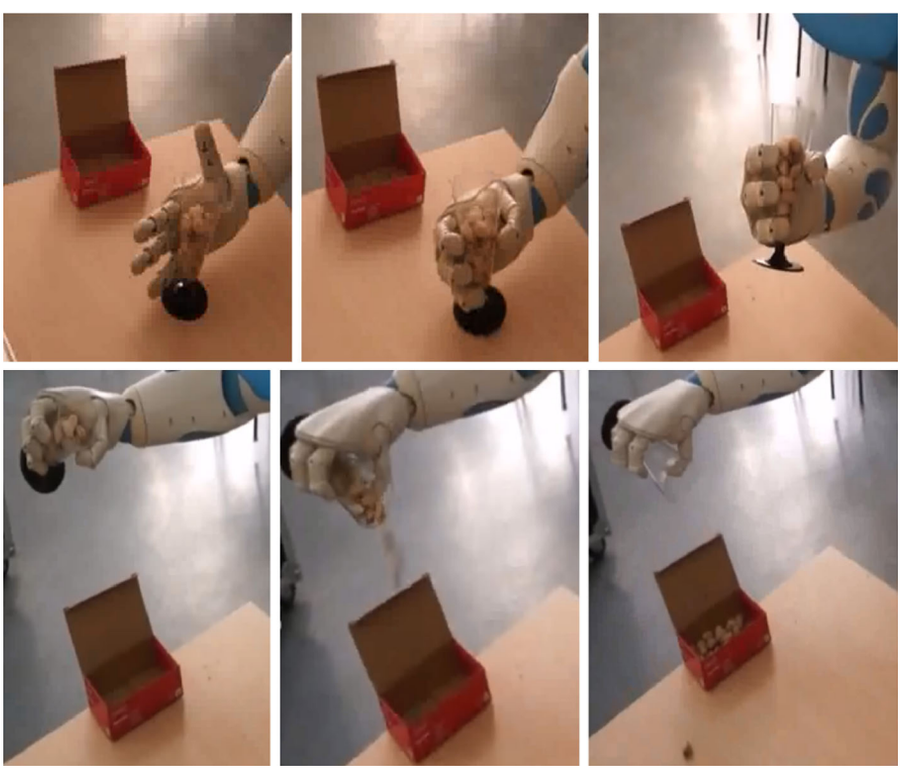

Abb. 8. Ausgießen mit humanoidem Roboter Romeo: Die Aufgabe ist, den Inhalt des Bechers in einen Behälter zu gießen. Die Schachtel wird als Behälter erkannt aufgrund der nach oben offenen 3D-Struktur. Ausgießen erfolgt über dem Behälter. Da der Roboter nicht wasserdicht ist, werden Papierkügelchen verwendet, um die Funktion zu zeigen

Abb. 9. Greifen beliebiger Objekte: *Links und Mitte:* zwei Beispiele zum Greifen unbekannter Objekte aus einem Haufen. Greifen erfolgt aufgrund einer greifbaren Form [4]. *Rechts:* Anwendung für die Unterstützung älterer Personen: der Hobbit-Roboter hebt alles vom Boden auf [3]

Boden oder Tisch wegzuräumen" kann auch bedeuten, dass Gegenstände nicht erkannt werden müssen aber entfernt werden sollten. Ursache kann sein, dass komplexe Szenen mit viel Verdeckung entstehen, Haufen mit Objekten, oder auch einfach neue Objekte vorkommen. Für diese Fälle gibt es Methoden, die aus der Form der Hand ableiten, welche der beobachteten Dinge greifbar sind. Beispiele sind Methoden wie eine Extrapolation von bekannten Greifgeometrien [9] oder ein Lernen von greifbaren 3D Merkmalen [4]. Abbildung 9 gibt ein Beispiel: unbekannte Objekte werden zu einem Haufen zusammengeworfen um alle möglichen Posen und Verdeckungen zuzulassen. Alle diese Haufen können vereinzelt werden. Einzelne Objekte sind dann weitaus einfacher zu Erkennen, oder können im falle neuer Objekte modelliert werden.

Die Abb. 9 zeigt rechts auch ein Bild einer Anwendung dieser Methode beim Einsatz als Hilfsroboter für alte Personen in deren Wohnung. Um die Gefahr von Sturz zu reduzieren, hebt der Roboter alles auf was er am Boden findet. Dabei ist die Erkennung nicht wichtig, da es wichtiger ist alles aufzuheben. So werden am Boden nach herausstehenden Formen gesucht und dann die Technik zum Greifen unbekannter Dinge angewendet [4]. Ähnliche Anwendungen werden bei Robotern, die nicht nur Saugen sondern wirklich alles Aufräumen, bald zum Einsatz kommen.

5. Schlussfolgerungen und Ausblick

Ziel dieses Beitrages ist das Aufzeigen, welche Methoden zur visuellen Wahrnehmung der Umgebung bereits verfügbar sind. Hierbei wurde insbesondere auf Methoden zur Modellierung, Erkennung und zum Greifen und Handhaben von Gegenständen eingegangen. Wie der Überblicksartikel in [11] aufzeigt gibt es neben dieser Aufgabe noch die beiden anderen großen Aufgabengebiete in der Robotik, die Navigation und die Erkennung von Personen.

Der Artikel zeigt, dass es robuste Methoden gibt um Objekte zu modellieren, wobei die Textur der Unterlage verwendet werden kann um auch texturlose Objekte zu modellieren. Diese Methoden sind manchmal, z. B. RTM Toolbox, direkt an Methoden zur Objekterkennung und -verfolgung angebunden.

Da ein vollständiges Modellieren von vielen Objekten sehr aufwendig ist, gibt es auch die Möglichkeit aus Datenbanken mit Bildern oder über 3D CAD-Modelle Funktionen zu lernen, um Klassen von Objekten zu Erkennen. Wir sehen in dieser Entwicklung einen deutlichen Schritt vorwärts, da ich dem System somit rasch die gewünschten Zielobjekte Einlernen kann. Es wird möglich, nähere Informationen über die Objekte zu speichern, so wie deren Funktion und die Verbindung mit geometrischen Merkmalen wie in Abb. 8 gezeigt. Somit können neue Einsätze in Industrie und Service entstehen, bei denen es darauf ankommt, dass der Roboter versteht was er tut und so auch Fehler erkennt und gegebenenfalls korrigieren kann.

Während dies erste erfolgversprechende Schritte sind, bleiben noch einige offene Fragen, um Vertrauenswürdige Roboter zu bauen. Noch gelingt es nicht, Robotern beizubringen, was falsch läuft, wenn daneben geschüttet wird. Es wird eher eine Funktion gelernt, die aus der Geometrie abgeleitet wird als eine Relation zwischen geometrischen Bezugspunkten. Auch Erkennen Roboter noch nicht spezifische Teile an einem Objekt, zum Beispiel, den Ausgießpunkt einer Kanne. Da neuronale Netzwerke (CNN) eher direkte Funktionen lernen, bleibt zu erarbeiten, ob modellbasierte Methoden eher die Kluft zwischen einer formalen Beschreibung und einer Aufgaben überbrücken können. Ziel ist es Funktionen zur Verfügung zu stellen, um Robotern rasch eine Aufgabe einzulernen und dem Roboter ein immer besseres Verstehen er Umwelt mitzugeben, sodass der Roboter selbständig eine Aufgaben übernehmen kann. In vielen Interviews mit Firmen wurde genau dies gefordert, man sollte eine Aufgabe einem Roboter beibringen können wie einem Praktikanten: vorzeigen, üben, verbessern und dann zuverlässig ausführen [26]. Denn die Fakten für die fertigende Industrie sind klar, die Firmen, die automatisieren, bleiben erfolgreich.

Für den Bereich der Service Roboter ergibt sich noch weitaus größeres Potential. Die Vision jeden Haushalt mit einem Roboter zum Aufräumen zu versehen beinhaltet ein Marktpotential wie die Autoindustrie. Mit ersten Lösungen wie zum Modellieren, Erkennen von Objektklassen und dem Aufheben von Dingen, wurden erste Schritte in diese Richtung aufgezeigt. Es gilt die Mittel bereitzustellen, um Europa nicht nur als Lieferanten von Methoden zu positionieren, sondern direkt die Fertigung der nächsten Generation von Service Robotern für Anwendungen zu Hause anzugehen.

Danksagung

Open access funding provided by TU Wien (TUW). This work was partially supported by the European Community's Seventh Framework Programme under grant agreement No. 610532 SQUIRREL, through the Horizon 2020 Programme, grant agreement No. 519625, project Flobot, and funding from the Austrian Science Foundation under grant ALOOF No. I1856-N30.

Literatur

1. Akkaladevi, S., Ankerl, M., Heindl, C., Pichler, A. (2016): Tracking multiple rigid symmetric and non-symmetric objects in real-time using depth data. In 2016 IEEE international conference on robotics and automation ICRA (S. 5644–5649). doi:10.1109/ICRA.2016.7487784.
2. Aldoma, A., Tombari, F., Stefano, L. D., Vincze, M. (2016): A global hypothesis verification framework for 3d object recognition in clutter. IEEE Trans. Pattern Anal. Mach. Intell., 38(7), 1383–1396.
3. Fischinger, D., Einramhof, P., Papoutsakis, K., Wohlkinger, W., Mayer, P., Panek, P., Hofmann, S., Körtner, T., Weiss, A., Argyros, A., Vincze, M. (2016): Hobbit, a care robot supporting independent living at home: first prototype and lessons learned. Robot. Auton. Syst., 75, 60–78.
4. Fischinger, D., Weiss, A., Vincze, M. (2015): Learning grasps with topographic features. Int. J. Robot. Res., 3, 3.
5. Gibson, J. J. (1979): The ecological approach to visual perception. London: Routledge.
6. He, K., Zhang, X., Ren, S., Sun, J. (2015): Deep residual learning for image recognition. CoRR. http://arxiv.org/abs/1512.03385. 1512.03385.
7. Johnson, A. E., Hebert, M. (1999): Using spin images for efficient object recognition in cluttered 3d scenes. IEEE Trans. Pattern Anal. Mach. Intell., 21(5), 433–449. doi:10.1109/34.765655.
8. Karpathy, A., Joulin, A., Li, F. (2014): Deep fragment embeddings for bidirectional image sentence mapping. CoRR. http://arxiv.org/abs/1406.5679. 1406.5679.
9. Kopicki, M., Detry, R., Adjigble, M., Stolkin, R., Leonardis, A., Wyatt, J. L. (2016): One-shot learning and generation of dexterous grasps for novel objects. Int. J. Robot. Res., 35(8), 959–976. doi:10.1177/0278364915594244.
10. Kopicki, M., Zurek, S., Stolkin, R., Mörwald, T., Wyatt, J. (2011): Learning to predict how rigid objects behave under simple manipulation. In Proc. of the ICRA 11 int. conference on robotics and automation.
11. Kragic, D., Vincze, M. (2010): Vision for robotics. Found. Trends Robot., 1(1), 1–78.
12. Krizhevsky, A., Sutskever, I., Hinton, G. E. (2012): Imagenet classification with deep convolutional neural networks. In F. Pereira, C. J. C. Burges, L. Bottou, K. Q. Weinberger (Hrsg.), Advances in neural information processing systems (Bd. 25, S. 1097–1105). Red Hook: Curran Associates, Inc. http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf.
13. Lowe, D. G. (2004): Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vis., 60(2), 91–110. doi:10.1023/B:VISI.0000029664.99615.94.
14. Mitash, C., Bekris, K. E., Boularias, A. (2017): A self-supervised learning system for object detection using physics simulation and multi-view pose estimation. CoRR. http://arxiv.org/abs/1703.03347. 1703.03347.
15. Moerwald, T., Prankl, J., Zillich, M., Vincze, M. (2013): Advances in real-time object tracking – extensions for robust object tracking with a Monte Carlo particle filter. J. Real-Time Image Process.
16. Montesano, L., Lopes, M., Bernardino, A., Santos-Victor, J. (2008): Learning object affordances: from sensory–motor coordination to imitation. IEEE Trans. Robot., 24(1), 15–26. doi:10.1109/TRO.2007.914848.
17. Mörwald, T., Prankl, J., Richtsfeld, A., Zillich, M., Vincze, M. (2010): BLORT – the blocks world robotic vision toolbox best practice in 3d perception and modeling for mobile manipulation. In Proc. ICRA.
18. Pauwels, K., Rubio, L., Diaz, J., Ros, E. (2013): Real-time model-based rigid object pose estimation and tracking combining dense and sparse visual cues. In 2013 IEEE conference on computer vision and pattern recognition (S. 2347–2354). doi:10.1109/CVPR.2013.304.
19. Prankl, J., Buchaca, A. A., Svejda, A., Vincze, M. (2015): RGB-D object modelling for object recognition and tracking. In IEEE/JRS International Conference on Intelligent Robots and Systems, IROS.
20. Rusu, R. B., Blodow, N., Beetz, M. (2009): Fast point feature histograms (FPFH) for 3D registration. In Proceedings of the IEEE international conference on robotics and automation ICRA, Kobe, Japan (S. 12–17).
21. Rusu, R. B., Bradski, G., Thibaux, R., Hsu, J. (2010): Fast 3d recognition and pose using the viewpoint feature histogram. In 2010 IEEE/RSJ international conference on intelligent robots and systems (S. 2155–2162). doi:10.1109/IROS.2010.5651280.
22. Rusu, R. B., Cousins, S. (2011): 3d is here: point cloud library (PCL). In 2011 IEEE international conference on robotics and automation (S. 1–4). doi:10.1109/ICRA.2011.5980567.
23. Sturm, J., Bylow, E., Kahl, F., Cremers, D. (2013): CopyMe3D: scanning and printing persons in 3D (S. 405–414). Berlin: Springer.
24. Tan, D. J., Tombari, F., Ilic, S., Navab, N. (2015): A versatile learning-based 3d temporal tracker: scalable, robust, online. In 2015 IEEE international conference on computer vision, ICCV (S. 693–701). doi:10.1109/ICCV.2015.86.
25. Varadarajan, K. M., Vincze, M. (2013): Parallel deep learning with suggestive activation for object category recognition. In Proceedings of the 9th international conference on computer vision systems, ICVS 2013.
26. Vernon, D., Vincze, M. (2016): Industrial priorities for cognitive robotics. In EUCognition meeting on cognitive robot architectures.
27. Vincze, M., Bajones, M., Suchi, M., Wolf, D., Weiss, A., Fischinger, D., de la Puente, P. (2016): Learning and detecting objects with a mobile robot to assist older adults in their homes. In Fourth international workshop on assistive computer vision and robotics – ACVR, ECCV workshops. LNCS (Bd. 9914).
28. Vincze, M., Wachsmuth, S., Sagerer, G. (2014): Perception and computer vision. In K. Frankish, W. M. Ramsey (Hrsg.), The Cambridge handbook of artificial intelligence (S. 168–246). Cambridge: Cambridge University Press.
29. Wohlkinger, W., Aldoma, A., Rusu, R. B., Vincze, M. (2012): 3dnet: large-scale object class recognition from cad models. In Proc. of the IEEE international conference on robotics and automation, ICRA (Bd. 2012).
30. Wohlkinger, W., Vincze, M. (2011): Ensemble of shape functions for 3d object classification. In Proc. of the 2011 IEEE international conference on robotics and biomemetics, ROBIO.
31. Wolf, D., Prankl, J., Vincze, M. (2016): Enhancing semantic segmentation for robotics: the power of 3d entangled forests. IEEE Robotics and Automation Letters, 1(1), 49–56.

Autoren

Markus Vincze
ist Leiter der Abteilung "Sehen für Roboter" am Institut für Automatisierungs- und Regelungstechnik (ACIN) der Technischen Universität Wien. Er studierte an der Technischen Universität Wien Maschinenbau und schloss 1988 das Studium mit dem Titel Dipl.-Ing. ab. Es folgte der Master of Science in Maschinenbau am Polytechnischen Institut Rensselear, Troy, New York, 1990. Im Jahr 1993 promovierte er an der TU Wien und war als Forschungsassistent am Institut für Automatisierungs- und Regelungstechnik tätig. Mit Unterstützung durch ein APART-Stipendium (bis 1998) der Österreichischen Akademie der Wissenschaften arbeitete er 1995 bis 1996 bei HelpMate Robotics Inc. sowie an der Yale University, New Haven, Connecticut, USA, in der Gruppe von Prof. G. D. Hager. Seit Februar 1998 ist er als Universitätsassistent am ACIN der TU Wien beschäftigt. Im Jänner 2004 legte er die Habilitation für "Robotertechnik" ab. Er leitete die EU-Projekte RobVision, ActIPret, robots@home, HOBBIT sowie ER4STEM sowie das FWF Forschungsnetzwerk "Kognitives Sehen". Er hat bislang 40 Journalartikel und über 300 begutachtete Beiträge im Bereich Objekte modellieren, erkennen, verfolgen und greifen mit dem Ziel, Robotern das Sehen beizubringen, veröffentlicht.

Michael Zillich
studierte Mechatronik an der Johannes Kepler Universität Linz und promovierte 2007 an der Technischen Universität Wien. Im Jahr 2000 verbrachte er sechs Monate als Gastforscher an der Königlich Technischen Hochschule in Stockholm, Schweden, und war 2006 bis 2008 Postdoc an der Universität von Birmingham in Großbritannien. Michael Zillich ist (Ko-)Autor von über 90 Publikationen, Guest Editor eines internationalen Journals und ist regelmäßig als

Reviewer und Associate Editor für internationale Konferenzen sowie internationale Journale tätig. Sein Forschungsinteresse liegt in Bildverarbeitung (Objekt-Segmentierung, -Erkennung und -Verfolgung) sowie taktiler Sensorik, mit einem Fokus auf integrierte kognitive Systeme und Robotik. Michael Zillich war an nationalen Projekten (FWF NFN Cognitive Vision, FWF InSitu – Leitung) sowie an EU-Projekten beteiligt (RobVision, ActIPret, CoSy, CogX, STRANDS), koordiniert zurzeit das EU-Projekt SQUIRREL und fungiert als Principal Investigator im WWTF-Projekt RALLI. 2013 gründete er mit seinem TU-Kollegen Dr. Walter Wohlkinger Blue Danube Robotics, ein Unternehmen, das sich mit taktiler Sicherheits-Sensorik für sichere Mensch-Roboter Kollaboration beschäftigt.

Johann Prankl
studierte an der Technischen Universität Wien Elektrotechnik mit dem Schwerpunkt Automatisierungs- und Regelungstechnik und schloss 2005 mit dem Titel Dipl.-Ing. ab. Darauf folgend arbeitete er als Forschungsassistent am ACIN in der Gruppe Vision for Robotics (V4R) und promovierte 2011 an der TU Wien. Seit Oktober 2011 ist er als Postdoc am ACIN beschäftigt. Der Schwerpunkt seiner wissenschaftlichen Tätigkeit liegt im Erkennen und Modellieren von Objekten, um autonomen Robotern das Interagieren in komplexen Umgebungen zu ermöglichen. Weitere berufliche Tätigkeiten umfassen den Bereich Bildbearbeitung in der Landtechnik als Senior Researcher bei Josephinum Research, Wieselburg, (seit 2013) und den Bereich Datenverarbeitung in der Landwirtschaft als Co-Founder der Farmdok GmbH (seit 2015).