

Bildverarbeitung in der Robotik: Anwendungen, Methoden und Herausforderungen

Jens Kubacki (Fraunhofer IPA, Stuttgart)

In der modernen Robotik können Bildverarbeitungssysteme für zahlreiche Aufgaben eingesetzt werden. In diesem Beitrag werden aktuelle Problemstellungen und Methoden im Kontext moderner Roboterassistenten für die Bildverarbeitung dargestellt. In der Einleitung werden Anwendungsgebiete behandelt. Dann folgt ein Kapitel über moderne Sensoren. Anschließend werden aktuelle Ansätze vorgestellt. Dieser Teil bildet den Kern des Artikels. Zuletzt werden die wesentlichen Inhalte des Beitrags kurz zusammengefasst.

1 Einleitung

Bildverarbeitungssysteme sind aus der modernen Automatisierungstechnik kaum mehr weg zu denken. Für zahlreiche Prozesse in Produktionsanlagen liefern sie nötige Messergebnisse für weitere Fertigungsprozesse oder Qualitätsprüfungen. Neben der traditionellen Automatisierungstechnik werden Bildverarbeitungssysteme aber auch in anderen Bereichen eingesetzt. Beim Fraunhofer IPA wird der Einsatz von Bildverarbeitungssystemen in modernen Roboterassistenten untersucht.

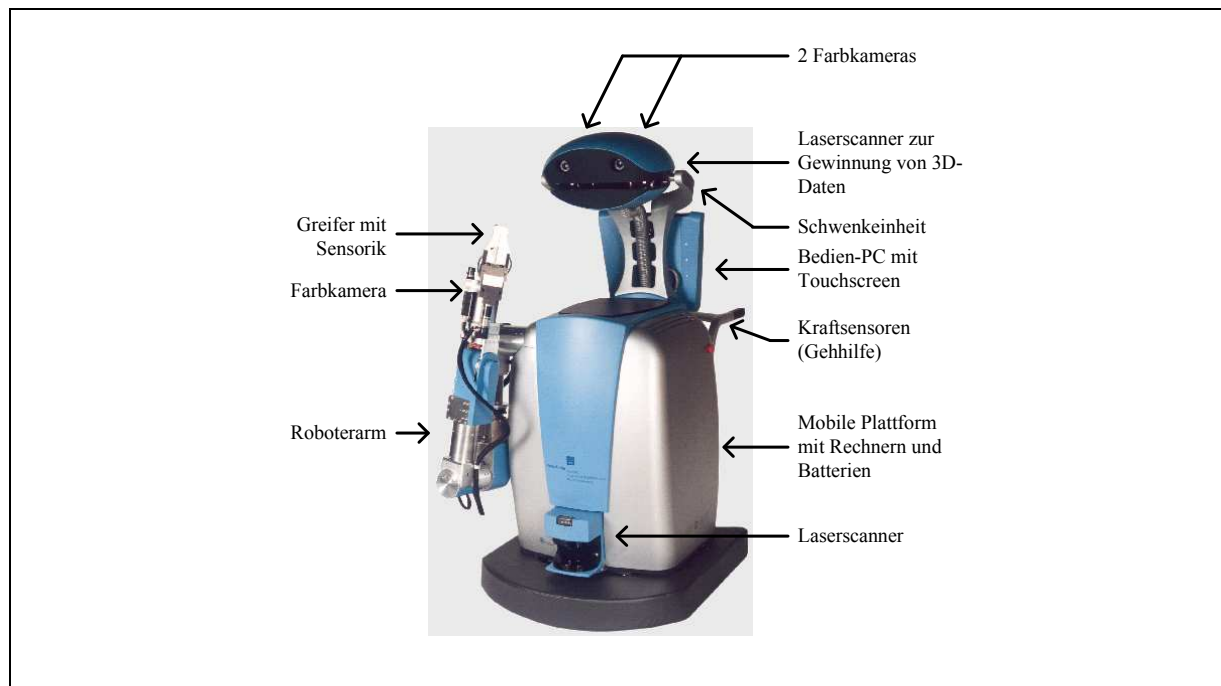


Bild 1-1: Beispiel für einen Roboterassistenten: Der Care-O-bot II (Fraunhofer IPA) mit Beschreibung der Sensor- und Aktorkomponenten

Mit Roboterassistenten sind Roboter gemeint, die über komplexe Sensor- und Aktorsysteme verfügen und mit einer gewissen "technischen" Intelligenz komplizierte Aufgaben ausführen können. Ein Beispiel für einen Roboterassistenten ist in Bild 1-1 abgebildet. Folgende typische Funktionalitäten, die mit Bildverarbeitung im Zusammenhang stehen, werden für moderne Roboterassistenten untersucht und entwickelt:

- **Navigation.** Der Roboter muss, um sich orientieren zu können, eine "Karte" der Einsatzumgebung erstellen und seine Position anhand der Aktuellen Sensordaten bestimmen und ausgleichen. Bildverarbeitungssysteme können eingesetzt werden um die Position des Roboters im Raum zu bestimmen. Dazu müssen die Ortinformationen in der Karte mit visuellen Hinweisen gekoppelt werden.
- **Manipulation.** Die Fähigkeit des Menschen, mit scheinbarer Leichtigkeit Objekte in seiner Umgebung hand zu haben und komplexere Handwerkliche Aufgaben zu lösen ist immer noch eine Herausforderung in der Robotik. Bildverarbeitungssysteme spielen bei der Suche eines Objektes im Arbeitsraum eine zentrale Rolle.
- **Mensch-Maschine-Interaktion.** Bildverarbeitungssysteme werden auch zur Interaktion zwischen Mensch und Maschine eingesetzt. Hier stehen Themen wie Erkennung, Lokalisierung und Identifikation des menschlichen Gegenübers sowie Gestenerkennung im Mittelpunkt.

2 Moderne Sensoren für die Bildverarbeitung

Der technische Erfolg von Bildverarbeitungssystemen gerade beim Einsatz in modernen Roboterassistenten hängt auch vom ersten Glied in der Kette ab: vom Bildsensor. Nicht nur Farbbildsensoren (Farbkameras) werden eingesetzt, sondern auch so genannte Tiefenbildsensoren, die z. B. mit dem Time-of-flight-Prinzip arbeiten. Hier wird die Zeit gemessen, die ein künstlich erzeugter Lichtstrahl vom Sensor zu einem Objekt in der Umgebung benötigt und zurück. Diese Messung geschieht pro Rezeptor auf dem Bildchip und so ist ein ganzes "Tiefenbild" verfügbar. Eine weitere Methode zur Gewinnung von Tiefendaten ist, zwei Farbkameras parallel anzuordnen (Stereobildverarbeitung) und durch die Verschiebung eines Punktes in den beiden Bildern den Abstand zu berechnen. In Bild 2-1 sind verschiedene repräsentative moderne Sensoren für die Bildverarbeitung aufgeführt. Diese stellen aber nur einen kleinen beispielhaften Ausschnitt der am Markt verfügbaren Komponenten dar.



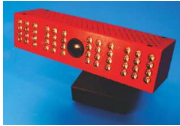
Bild			
Produkt	DFK41F02	Bumblebee	SwissRanger SR2
Hersteller	Imaging Source	Point Grey Research	CSEM
Typ	Farbbildsensor	Stereobildfarbsensor	Tiefenbildsensor
Informationen	www.theimagingsource.com	www.ptgrey.com	www.csem.ch

Bild 2-1: Beispiele für moderne Sensoren in der Bildverarbeitung

3 Ausgewählte Verfahren aus der Bildverarbeitung im Kontext der modernen Robotik

In diesem Kapitel werden einige Verfahren, die für die moderne Robotik interessant sind, vorgestellt. Das Kapitel wurde in drei grobe Bereiche geteilt: Verfahren mit geometrischen Modellen, ansichtbasierte Verfahren und direkte Sensor-Aktor-Kopplungen. Die Bereiche sind allerdings nicht eindeutig voneinander trennbar. Dennoch stellen sie drei methodische Sichtweisen dar, die bei der Einordnung der Vielzahl von Verfahren in der Bildverarbeitung mit Bezug zur modernen Robotik nützlich sind.

1.1 Verfahren mit geometrischen Modellen

Die Geometrie von Objekten steht im Zusammenhang mit deren visuellen Erscheinung. Ist ein geometrisches Modell eines Objektes verfügbar, so kann dieses z. B. zur Schätzung der Lage und Rotation eines Objektes relativ zu einem festgelegten Koordinatensystem dienen. Die Annahmen sind hierbei:

- Die geometrischen Daten sind in Form eines Modells bekannt.
- Die Komponenten sind kalibriert, d. h. die mathematischen Modelle (Gleichungen) von gemessenen Sensorwerten zu realen Koordinaten sind bekannt.

Verfahren, die die Koordinaten eines Objektes in Abhängigkeit von der aktuellen visuellen Erscheinung schätzen lösen in der Regel ein Optimierungsproblem bei dem das Modell in die Aufnahmen der aktuellen Szene "eingepasst" wird.

Eine prinzipielle Beschränkung dieser Verfahren ist, die Annahme, dass die Modelle der Objekte bekannt sind. Diese ist bei Robotern, die in beliebigen Umgebungen immer neue Aufgaben ausführen sollen nicht immer erfüllt. Deshalb beschäftigen sich zahlreiche Wissenschaftler mit der Fragestellung: Wie können sinnvolle Modelle *online*, also während des Betriebes generiert werden?

Eine Perspektive sind hier parametrisierbare Körper. Geometrische Modelle wie z. B. Kugeln, Quader oder andere lassen sich durch Veränderung der Formparameter an bestimmte reale Objekte anpassen, so dass diese angenähert modelliert werden können. Bei komplexeren, "alltäglichen" Objekten kann z. B. über die Vereinigungsmenge ein kombiniertes Modell gefunden werden. Beim Fraunhofer IPA werden u. A. so genannte Superquadriken eingesetzt [1]. Diese können implizit über eine Oberflächengleichung dargestellt werden. Ein einfacher Sonderfall der Superquadriken ist der Superellipsoid. Die implizite Gleichung des Superellipsoiden ist in Bild 3-1 dargestellt.

$$\left(\left(\frac{x}{a_1} \right)^{\frac{2}{\varepsilon_2}} + \left(\frac{y}{a_2} \right)^{\frac{2}{\varepsilon_2}} \right)^{\frac{\varepsilon_2}{\varepsilon_1}} + \left(\frac{z}{a_3} \right)^{\frac{2}{\varepsilon_1}} = 1$$

Bild 3-1: Implizite Darstellung des Superellipsoiden nach [1].

Dabei sind a_1 , a_2 , a_3 , ε_1 und ε_2 die Formparameter und x , y und z die Koordinaten im Raum R^3 . Diese Gleichung kann verwendet werden um festzustellen, ob Punkte in der aufgenommenen Szene auf der Superquadrik liegen. Die Summe der Abstände geht

üblicherweise in die Zielfunktion eines Optimierungsproblems ein. Durch Lösung des Optimierungsproblems ist es möglich, das geometrische Modell eines Objektes aus den Sensordaten z. B. eines Tiefenbildsensors zu schätzen. Das Modell kann später zu Detektierung und Lokalisierung des Objektes in der Szene genutzt werden. Dazu wird dann ein Optimierungsproblem formuliert indem die Formparameter als konstant angenommen werden, dafür aber die Translations- und Rotationsparameter des Koordinatensystems der Superquadrik variabel sind.

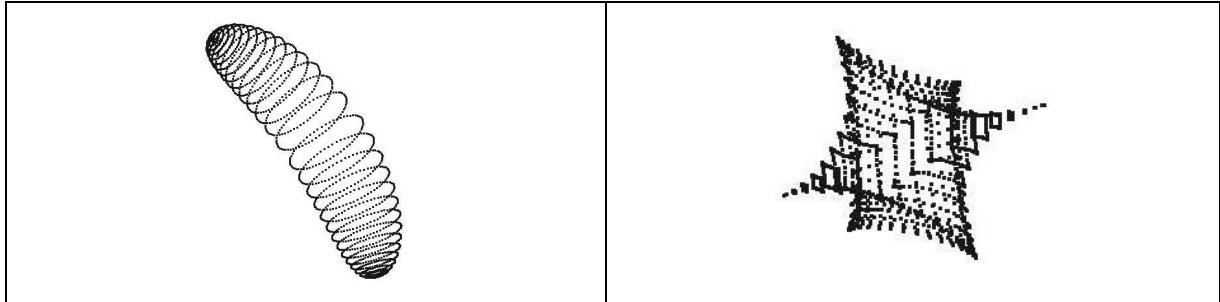


Bild 3-2: Superellipsoiden mit verschiedenen Formparametern. Links: Grundgleichung mit zusätzlicher Krümmung, rechts: konkave Form.

1.2 Ansichtbasierte Verfahren

Die modellbasierte Berechnung des Zustands eines Objektes in der Szene aus aktuellen Sensordaten ist nicht immer mit niedrigem Rechenaufwand lösbar, in vielen Fällen sogar mathematisch kaum darstellbar oder nur mit sehr hohem analytischem Aufwand. Während bei Tiefenbildsensoren der geometrische Zustand indirekt gemessen werden kann, ist es bei Farbbildsensoren nicht so leicht möglich. Deshalb haben sich in der Forschung auch so genannte ansichtbasierte Verfahren einen nicht zu unterschätzenden Stellenwert verschafft. Diese können auch mit Lernalgorithmen realisiert werden. In [2] ist ein Verfahren beschrieben, dass aus der aktuellen Ansicht eines Objektes eine eindeutige Objektkennzeichnung berechnet. Dazu wird ein Klassifizier eingesetzt. Ein Klassifizier hat die Aufgabe aus einem Vektor $x \in R^n$ einen diskreten Klassenbezeichner $y \in \{-1,1\}$ zu bestimmen. Die aktuelle Ansicht des Objektes kann als Vektor von Grau- oder Farbwerten dargestellt werden. In [2] ist der Klassifizier ein Lernalgorithmus, eine so genannte Support-Vektor-Maschine (SVM). In einem Lerndatensatz wird der Klassenbezeichner 1 jeder Beispielansicht eines Objektes zugeordnet, der Bezeichner -1 den Gegenbeispielen. Mehrere SVM's werden kombiniert um 100 Objekte einer Testdatenbank zu unterscheiden. Die Ergebnisse überzeugen durch eine sehr kleine Fehlerrate auch bei künstlich Verrauschten Bildern. Eine SVM für Klassifikation löst in ihrer Grundversion ein Optimierungsproblem welches die Vektoren der Ansichten optimal d. h. mit einer Trennungsebene mit größtem Abstand zum nächsten Lernbeispiel trennt. Die gute Generalisierungsrate dieser Methode wurde in der statistischen Lerntheorie mathematisch bewiesen [3].

Auch Verfahren, die so genannte Merkmalspunkte einsetzen, können zur Detektierung eines Objektes in der Szene genutzt werden. Ein bekanntes Verfahren, das in [4] veröffentlicht wurde sucht Merkmalspunkte im Bild deren Auftreten stabil unter verschiedenen Bildtransformationen ist und denen eine eindeutige Beschreibung in Form eines mehrdimensionalen Beschreibungsvektors zugeordnet werden kann. Ziel ist es, das Objekt in der Szene mit nur wenigen detektierten Merkmalspunkten auszumachen. Ist ein Objekt mit

ausreichend vielen Merkmalspunkten ausgestattet, dann ist es auch bei extremen Verdeckungen noch auffindbar.

1.3 Direkte Sensor-Aktor-Kopplungen

Das Steuerungsparadigma in der Robotik, um zum Beispiel Objekte zu Greifen lehnt sich traditionell an Erkennungsverfahren, die auf geometrischen Modellen beruhen. Eine typische Vorgehensweise zum Greifen ist z. B.:

1. Detektierung und Lageschätzung des Objektes in der Szene
2. Planung einer so genannten Trajektorie, das ist eine durch Raumpunkte zusammengesetzte Wegstrecke für den Roboterarm
3. Ausführung des Greifvorgangs

In der Robotik hat sich gezeigt, dass diese Vorgehensweise nicht die einzig mögliche ist. Reaktive Verfahren bilden die aktuelle Aufnahme des Sensors auf einen kurzfristigen Steuerungsbefehl ab. Solche Verfahren sind auch unter dem Namen verhaltensbasierte Robotik bekannt geworden [5].

Ein verwandtes Beispiel ist das so genannte *Visual Servoing*. Hier werden Steuerungsbefehle z. B. für einen Roboterarm aus Kamerabildern online generiert. Die Anordnung der Systemkomponenten bildet einen Regelkreis, der sich dem Zielzustand, z. B. "Greifposition erreicht" sukzessive und flexibel annähert.

In diesem Zusammenhang soll auch der Optische Fluss genannt werden. Diese Methode berechnet Bewegungsmuster aus Bildsequenzen und kann so ebenfalls für reaktive Steuerungen genutzt werden.

Die erwähnten Verfahren zeichnen sich durch eine hohe Robustheit in Umgebungen mit bewegten Objekten und anderen Unsicherheiten aus. Ohne Unterstützung einer modellbasierten Planungsmethode können die Verfahren aber auch zu Fehlverhalten führen. Neuere biologisch inspirierte Ansätze versuchen reaktive Steuerungsstrategien noch weiter zu generalisieren, in dem adaptive Lernalgorithmen mit verhaltensbasierten Steuerungen kombiniert werden. So ist z. B. in [5] ein interessantes Verfahren beschrieben, das einen Roboter beibringen kann, reaktiv gesteuert auf einen Kasten zu zufahren.

4 Zusammenfassung

In modernen Roboterassistenten können Bildverarbeitungssysteme für die Navigation, Manipulation und im Bereich der Mensch-Maschine-Interaktion eingesetzt werden. Moderne Sensoren liefern nicht nur Farbdaten, auch Tiefenbilder sind verfügbar. Ansätze mit geometrischen Modellen können zur Erkennung von Objekten dienen, wenn das Objektmodell bekannt ist. Wenn diese Anforderung nicht erfüllt ist, müssen Objektmodelle vom Roboter selbst generiert werden können. Bei ansichtbasierten Verfahren wird in der Regel keine geometrische Berechnung vorgenommen. Mit Methoden aus der statistischen Lerntheorie lassen sich Objektansichten so "einlernen", dass Objekte voneinander unterschieden werden können. Neuerer Ansätze widmen sich dem Thema der direkten Sensor-Aktor-Kopplungen und stehen mit der so genannten verhaltensbasierten Robotik im Zusammenhang. Hier werden auch Methoden der Regelungstechnik eingesetzt. Lernverfahren werden eingesetzt um reaktive Steuerungsstrategien online zu lernen.

Die Fülle der Ansätze und die (noch) fehlende Ordnung in dem Themengebiet der Bildverarbeitung für moderne Roboter zeigt, wie jung dieses wissenschaftliche Gebiet noch ist. Ohne Zweifel lässt sich die Bedeutsamkeit visueller Sensoren von deren massivem Einsatz in natürlichen Systemen (als Augen) erahnen. Möglicherweise sind es gerade diese Sensoren und zugehörige Verarbeitungssysteme, die zukünftig so etwas wie visuelles Vorstellungsvermögen auch für Roboter möglich machen und damit die technische Intelligenz erweitern können.

Referenzen

1. Jaklic, A., Leonardis, A., und Solina, F.: *Segmentation and Recovery of Superquadrics*. Kluwer, 2000.
2. Pontil, M. und Verri, A.: *Support vector machines for 3-d object recognition*. IEEE Trans. PAMI, 20:637-646, 1998.
3. Vapnik, V.: *The Nature of Statistical Learning Theory*. Springer, 1995.
4. David G. L.: *Distinctive image features from scale-invariant keypoints*. International Journal of Computer Vision, 60, 2, pp. 91-110, 2004.
5. Nehmzow, U.: *Mobile Robotics: A Practical Introduction*. ISBN: 1-85233-173-9. Springer Verlag, 1999.