Zeitschrift: Bulletin des Schweizerischen Elektrotechnischen Vereins, des Verbandes Schweizerischer Elektrizitätsunternehmen = Bulletin de l'Association suisse des électriciens, de l'Association des entreprises électriques suisses

Herausgeber: Schweizerischer Elektrotechnischer Verein ; Verband Schweizerischer Elektrizitätsunternehmen

Band: 83 (1992)

Heft: 21

Artikel: Nur wer die Umwelt erkennt, kann sie verändern : neue und alte Lösungsansätze der Robot Vision

Autor: Ade, Frank

DOI: https://doi.org/10.5169/seals-902888

Nutzungsbedingungen

Die ETH-Bibliothek ist die Anbieterin der digitalisierten Zeitschriften auf E-Periodica. Sie besitzt keine Urheberrechte an den Zeitschriften und ist nicht verantwortlich für deren Inhalte. Die Rechte liegen in der Regel bei den Herausgebern beziehungsweise den externen Rechteinhabern. Das Veröffentlichen von Bildern in Print- und Online-Publikationen sowie auf Social Media-Kanälen oder Webseiten ist nur mit vorheriger Genehmigung der Rechteinhaber erlaubt. Mehr erfahren

Conditions d'utilisation

L'ETH Library est le fournisseur des revues numérisées. Elle ne détient aucun droit d'auteur sur les revues et n'est pas responsable de leur contenu. En règle générale, les droits sont détenus par les éditeurs ou les détenteurs de droits externes. La reproduction d'images dans des publications imprimées ou en ligne ainsi que sur des canaux de médias sociaux ou des sites web n'est autorisée qu'avec l'accord préalable des détenteurs des droits. En savoir plus

Terms of use

The ETH Library is the provider of the digitised journals. It does not own any copyrights to the journals and is not responsible for their content. The rights usually lie with the publishers or the external rights holders. Publishing images in print and online publications, as well as on social media channels or websites, is only permitted with the prior consent of the rights holders. Find out more

Download PDF: 17.02.2026

ETH-Bibliothek Zürich, E-Periodica, https://www.e-periodica.ch

Nur wer die Umwelt erkennt, kann sie verändern

Neue und alte Lösungsansätze der Robot Vision

Frank Ade

Dieser Beitrag beleuchtet einige aktuelle Fragestellungen der Robot Vision. Die strukturelle Indexierung beschleunigt die Suche nach Objektmodellen in Datenbanken, generische Objektmodelle gestatten die hierarchische Strukturierung von Objektmodellmengen, CAD-basierte Methoden könnten Erleichterung beim Eintrainieren von Objektmodellen bringen und die Multisensor-Fusion verspricht eine Verbesserung der Sicherheit der Objekterkennung. Ein schwieriges Thema betrifft die dreidimensionale Form von Objekten.

Cet article-ci présente quelques problèmes actuels de la vision robotique. L'indexation structurelle accélère la recherche de modèles d'objets dans des banques de données, les modèles génériques permettent la structuration hiérarchique d'ensembles de modèles d'objets, des méthodes à base de CAD promettent un apprentissage de modèles d'objets simplifié et la fusion multisensorielle améliorera la robustesse de la reconnaissance d'objets.

Adresse des Autors
Dr. *Frank Ade*, Institut für Kommunikationstechnik, Fachgruppe Bildwissenschaft, ETH Zentrum, 8092 Zürich.

In Autofabriken ziehen sie Schrauben an, schweissen Bleche zusammen, sprühen Farbe auf Karosserieteile und erledigen noch viele andere Arbeiten. Die fleissigen Arbeiter, von denen die Rede ist, sind Roboter. Sie sind aus der modernen Automatisierungstechnik nicht mehr wegzudenken. Damit ein Roboter die ihm aufgetragenen Arbeiten durchführen kann, muss sein Robotergehirn jederzeit wissen, wo und in welcher Orientierung sich die beteiligten Werkstücke und sein eigener Roboterarm befinden. Es gibt zwei Wege, diese Voraussetzung zu erfüllen: Der heute am häufigsten realisierte Weg besteht in der Schaffung und Aufrechterhaltung einer strengen Ordnung durch eine präzise Zuführung der Teile und eine präzise Steuerung des Roboters. In diesem Fall ist es nicht notwendig, den Roboter mit einem Sehsinn auszustatten. Diese Methode ist jedoch teuer und kann nicht immer realisiert werden. Sie entspricht auch nicht dem Vorgehen des Menschen, der dank seinem Sehsinn und einer adaptiven Motorik eine grosse Toleranz bezüglich der Lage und Orientierung der Werkstücke und Werkzeuge aufweist.

Es ist das Fernziel der Robot Vision, Roboter mit einem künstlichen Sehsinn so auszustatten, dass sie eine ähnlich grosse Toleranz und Robustheit gegenüber der Variabilität der Welt wie der Mensch aufweisen. Dazu kann sie aus einem grossen Arsenal von Methoden schöpfen, das von der Disziplin der Computer Vision im Verlauf der letzten drei Jahrzehnte erarbeitet worden ist. Dieser interessante Weg stösst allerdings heute noch in vielen Anwendungen an eine Kostengrenze, weshalb die Robot Vision im industriellen Umfeld nur bestimmte Nischen zu besetzen vermochte, die sich durch starke Strukturiertheit der Roboter-Arbeitsumgebung auszeichnen.

Hier muss angemerkt werden, dass sich in den letzten zwei Jahren zur Computer Vision äusserst kritische Stimmen zu Wort gemeldet haben. Verschiedene Forscher stellen die Frage, warum der Fortschritt auf dem Gebiet der Computer Vision so langsam sei. «Wir haben uns schuldig gemacht, die experimentellen Aspekte und Anwendungen in unserem Gebiet vernachlässigt zu haben.» Es wird sogar angezweifelt, dass die Computer Vision den Status und den Reifegrad einer respektablen Wissenschaft erreicht hat. Kurz, man beginnt zu entdecken, dass Computer Vision tatsächlich so schwierig ist, wie schon immer behauptet wurde. In überraschender Einstimmigkeit ziehen Warner und Mahner das Fazit, man solle vorerst einmal davon lassen, das Computer Vision-Problem allgemein (für beliebige Szenen) lösen zu wollen und statt dessen erst einmal einen Satz von Zwischenzielen formulieren, um über diese während der nächsten fünf bis zehn Jahre messbare Fortschritte zu erzielen.

Ein Beispiel für realistische Zwischenziele wäre die Lösung sehr spezieller Probleme, bei denen man ein gutes Verständnis der Physik und des Bildentstehungsprozesses hat und wo der Kontext stark eingeschränkt ist. Viele industrielle Anwendungen fallen in diese Kategorie. Es wird auch einem alten Paradigma neuer Nachdruck (und ein neuer Name) gegeben, dem zweckgerichteten Sehen (purposive Vision), welches dem Paradigma der reinen Computer Vision, deren einziges Ziel ist, ein Bild zu verstehen, die in einen Handlungs-Zusammenhang eingebettete zweckgerichtete Vision gegenüberstellt. Es ist natürlich klar, dass die ganze industrielle Com-

puter Vision seit jeher zweckgerichtete Vision war. Eine wichtige Konsequenz dieses Paradigmas ist die Forderung nach Sparsamkeit bei den eingesetzten Mitteln. Wenn der Zweck des Sehens im Vordergrund steht, sollte man nur das erfassen, messen und berechnen, was zur Lösung der gestellten Aufgabe notwendig ist.

Das neue Ansehen der angewandten Computer Vision verheisst Gutes für die Robot Vision. Erfolgreiche Anwendungsgebiete werden sich ausweiten und vertiefen. Da die Schwierigkeit der Probleme, die man anpakken wird, immer mehr zunimmt, werden auch die Algorithmen komplexer und verlangen bei ihrer Implementierung immer leistungsfähigere (und daher teurere) Rechner. Glücklicherweise nimmt das Preis-Leistungs-Verhältnis sowohl von allgemeiner als auch von Hardware, die auf Bildverarbeitung zugeschnitten ist, kräftig ab. Dies bewirkt, dass jetzt viele schon länger bekannte, aufwendige Computer Vision-Algorithmen und -Methoden in die Reichweite industrieller Anwendbarkeit kommen. Beide Faktoren, das günstige intellektuelle Klima gegenüber Anwendungen wie auch der ungebrochene Aufwärtstrend bei der Hardware machen es leicht, der industriellen Robot Vision gute Zukunftsaussichten zu bescheinigen.

Die Computer Vision und mit ihr die Robot Vision haben in der Tat auch schon in den letzten Jahren bedeutende Fortschritte erzielt. Es ist die Absicht dieses Artikels, einige dieser Fortschritte zu skizzieren und zusätzlich die dringendsten Probleme der unmittelbaren Zukunft darzustellen. Der nächste Abschnitt gibt eine gestraffte Zusammenfassung eines implementierten 3D-Robot Vision-Systems im Rahmen des ETH-Polyprojekts COR, der dritte Abschnitt diskutiert einige Entwicklungstendenzen bei der objektmodell-basierten Robot Vision, der vierte untersucht die Möglichkeiten, ohne detaillierte Objektmodelle auszukommen und der fünfte Abschnitt gibt eine Bewertung des Standes der Robot Vision.

Das 3D-Vision-System des Projekts COR

Die Arbeitsgruppe Mechatronik der ETH Zürich, die aus den Instituten für Robotik, Elektronik, Mess- und Regeltechnik, Kommunikationstechnik sowie der Professur für elektrotechnische Entwicklungen und Konstruktionen besteht, führte ein grosses interdisziplinäres Projekt mit dem Namen COR (Cooperating Robot with Visual and Tactile Capabilities) durch. Um die Arbeit an dem Projekt besser fokussieren und überwachen zu können, wurde beschlossen, eine konkrete, herausfordernde Aufgabe zu definieren. Diese bestand darin, ein vollständiges System zum Abräumen von Mensatabletts (Kantinen-Geschirr-Tabletts) zu bauen. Ein Vision-Subsystem hatte die Aufgabe, den Aufbau von Geschirr- und Besteckteilen auf einem Mensatablett soweit zu analysieren, dass es einem Roboter die Information übermitteln konnte, die dieser benötigte, um ein Objekt korrekt zu greifen, wegzunehmen und an einem bestimmten Platz abzustellen. Wenn eine Folge dieser Operationen ein Mensatablett vollständig abgeräumt hatte, wurde das leere Tablett weggeführt und ein neues, gefülltes Tablett nachgeschoben [1]. Das Vision-System beschaffte sich mit Hilfe von drei Kameras die notwendigen visuellen Daten und interpretierte die Szene vollständig mittels eines objektmodell-basierten Verfahrens. Die benutzte Entscheidungsprozedur kann knapp als regelbasierte Evidenz-Akkumulation charakterisiert werden. Das Bild 1 zeigt schematisch die Komponenten-Anordnung des Robot Vision-Systems und Bild 2 ein typisches Bildtripel, das analysiert wurde [2]. Dieses Projekt wurde nach einer Projektdauer von 4 Jahren mit einer Demonstration des vollen Systems an Industrial Handling 92 in Zürich abgeschlossen.

Als Errungenschaften des Projekts in bezug auf das Vision-System können folgende Punkte angesehen werden: Das System löste ein echtes 3D-Problem. Es war fähig, mit einer beträchtlichen Anzahl von Objekten, die sich zudem teilweise verdecken durften, fertig zu werden. Die Objekte durften transparent sein und spiegelnde Oberflächen haben. Dass die Lösung dieses Problems keineswegs trivial ist, kann man bei einer Betrachtung der Bilder 2a–2c, insbesondere der Aufnahme von oben, verstehen.

Das System bewies Robustheit durch die Verwendung redundanter Information. Die Systemintegration warf dank gut strukturierter Software keine grösseren Probleme auf. Während der Industrial Handling 92 arbeitete das Gesamtsystem während einer

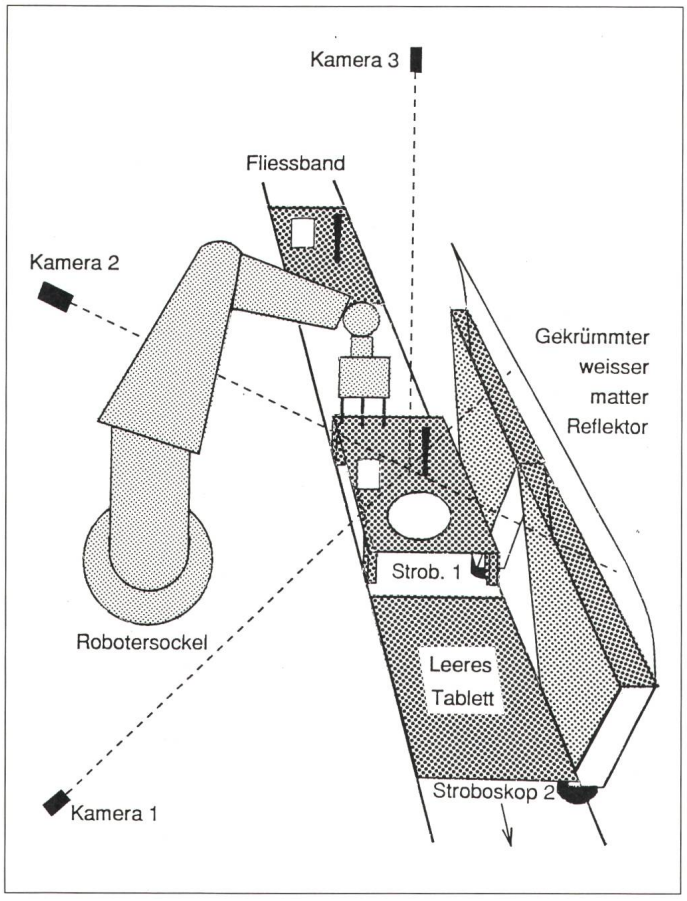


Bild 1
Anordnung des Experimentes

Woche etwa 5 Stunden am Tag und analysierte erfolgreich mehrere hundert Szenen (Mensatabletts).

Das Projekt COR umfasste neben dem oben geschilderten Projekt noch ein zweites Teilprojekt, bei dem es um die Vereinzelung von Paketen aus einem Paketstrom mit Hilfe von Tiefendaten (Range Data) ging [3].

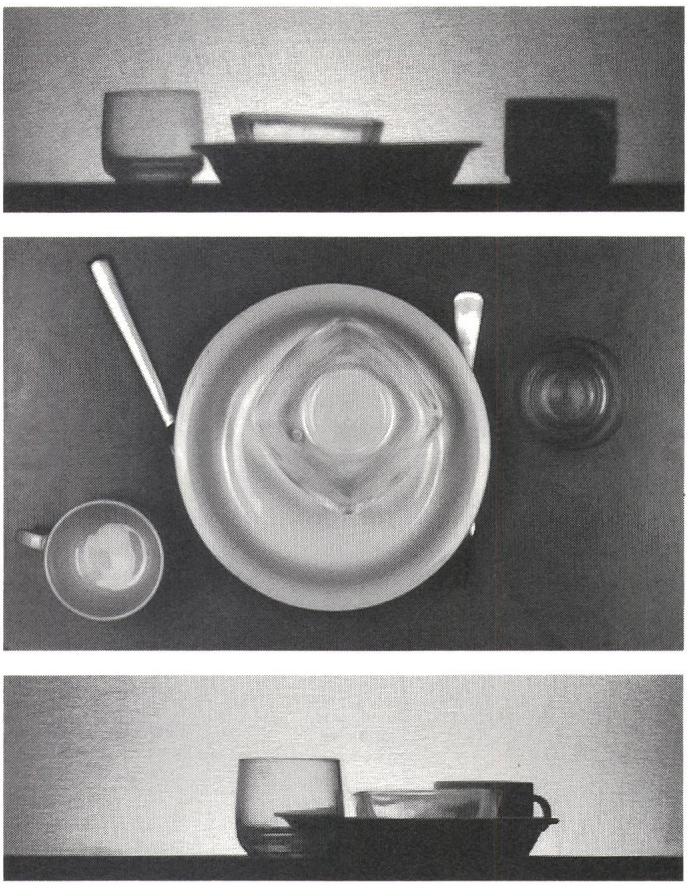


Bild 2
Bilder der drei CCD-Kameras aus verschiedenen Blickwinkeln

Trends beim objektmodellbasierten Vorgehen

Im industriellen Kontext ist eine der häufigsten Aufgaben der Robotik die Manipulation von Werkstücken (Objekten). Der Robot Vision fällt dabei die Unteraufgabe der Erkennung von Objekten und die Bestimmung ihrer Position und Orientierung zu. Die Objekterkennung aufgrund von Bildern geschieht grob gesagt in drei Etappen: Datenakquisition, Rekonstruktion von Objekten (Segmentierung und Gruppierung) und Suche nach dem bestpassenden Objektmodell.

Die Art der zur Verfügung stehenden Daten entscheidet teilweise über die auf sie anzuwendenden Algorithmen. Das Aufkommen schneller, kostengünstiger Tiefendaten (Range Data) hatte und hat einen grossen Einfluss auf das Gebiet. Der Trend zu echt dreidimensionalen Objektdarstellungen und Algorithmen verstärkte sich sehr, da Tiefendaten direkte Informationen über die Form der Oberflächen der Objekte enthalten. Die Datenakquisition mit Hilfe von raumzeitlich kodiertem Licht zur Ableitung eines für die Objekterkennung geeigneten Modells wird in diesem Heft im Beitrag [3] detailliert beschrieben.

Rekonstruktion und Erkennung der Objekte in der Szene sind meist stark miteinander verflochten. Die Gruppierung von Bildteilen zu Objekten ist einfach, wenn man das gesuchte Objekt kennt, und umgekehrt ist die Objekterkennung leicht, wenn die Gruppierung gut ist. In der Phase der Gruppierung werden Heuristiken verwendet, das heisst Faustregeln, die sagen, welche Bildteile zusammengehören könnten. Das menschliche visuelle System ist sehr gut im Gruppieren; diese Leistung wird automatisch und unbewusst vollbracht. Aufgrund des Resultats der Gruppierung werden Objekthypothesen erstellt. Da Heuristiken zwar meistens, aber doch nicht immer, gute Ratschläge geben, muss auch die Möglichkeit vorgesehen sein, Hypothesen zu verwerfen und neue zu erstellen.

Objektmodelle, die in einem bestimmten Kontext in Frage kommen, sind in einer Objektmodell-Datenbank abgespeichert. Mit ihnen werden die in der Szene durch Gruppierung erzeugten Objektfragmente verglichen. Die in der Robot Vision gebräuchlichen Objektmodelle können sehr verschiedenartig sein, sie reichen von einfachen Merkmalsvektoren bis zu komplizierten Graph-Strukturen. Ein deutlicher Trend geht auf die Verwendung von echt dreidimensionalen Modellen. Gleichzeitig wird versucht, die Anzahl der in der Objektmodell-Datenbank abgespeicherten Modelle zu erhöhen.

Die Objektmodelle gelangen in die Datenbank durch einen Prozess, der Einlehren (Teach-in, Training) genannt wird. Dies kann manuell geschehen oder durch Training by Showing. Dabei wird ein bestimmtes Objekt einzeln und unter guten Sichtbarkeitsbedingungen dem System gezeigt. Dieses extrahiert dann die Merkmale, für die man sich beim Entwurf des Systems entschieden hat, und speichert diese strukturiert zusammen mit der Identität des Objekts ab. Neuerdings richten sich grosse Hoffnungen auf die Nutzung von CAD-Modellen. Man untersucht, ob die im CAD-Modell implizit vorhandene Information nutzbar gemacht werden kann. Dies sollte natürlich automatisch geschehen.

Ein weiterer zu beobachtender Trend betrifft die Verwendung von mehreren Datenquellen oder sogar von mehreren Datenarten (Multisensor Fusion). Es wird erwartet, dass dadurch die Robustheit der Vision-Systeme erhöht werden kann. In den folgenden Abschnitten werden nun einzelne dieser Trends etwas genauer unter die Lupe genommen.

Strukturelle Indexierung

Der Trend zu immer grösseren Objektmodell-Datenbanken führt zu einer überproportionalen Zunahme des Rechenaufwands bei der Suche nach dem bestpassenden Objektmodell. Man spricht dann von einer kombinatorischen Explosion. Jede aufgrund der Szenen-Information erstellte Objekthypothese wird mit allen in der Datenbasis abgespeicherten Modellen verglichen. Wenn aber die Datenbank so aufgebaut ist, dass die in der Szene aufgefundenen Merkmale dazu dienen, einen schnellen Zugriff zu Teilen der Datenbank zu erhalten oder um-

gekehrt ausgedrückt, ganze Teile des Suchbaums auszuschliessen, ist die Suche effizienter. Man hat dann gewissermassen eine inhaltsadressierbare Datenbank. Man nennt diese Methode auch strukturelle Indexierung. Sie wurde bereits auf dem Gebiet der 2D-Computer Vision entwickelt und wird jetzt auf den 3D-Fall verallgemeinert. Als strukturelle Merkmale werden Merkmale vor allem geometrischer Natur verwendet, die gegenüber Transformationen wie Translation, Rotation und Skalierung invariant sind. Die strukturellen Merkmale sind meist lokale Gruppierungen noch einfacherer Merkmale wie Löcher, Ecken, Kantensegmente usw., wobei insbesondere deren räumliche Beziehungen notiert werden.

Generische Objektmodelle

Generische Modelle werden verwendet, wenn die Objekte einer Klasse zwar Gemeinsamkeiten aufweisen, aber nicht bis in die letzte Einzelheit übereinstimmen. Sie definieren mit Hilfe von Zwangsbedingungen, die erfüllt sein müssen, Äquivalenzklassen von Objekten. Wenn nur wenige Zwangsbedingungen erfüllt sein müssen, gelangt man zu sehr weit gefassten Klassen; mehr Zwangsbedingungen führen zu sehr genau festgelegten Klassen. Es ist auch möglich, die Zwangsbedingungen eine nach der anderen einzuführen. Auf diese Weise kann man ganze Hierarchien von Klassen (Taxonomien) bilden.

In [3] wurde für das Problem des Vereinzelns von Postpaketen das generische Modell eines Quaders eingeführt. Im Tiefenbild aufgefundene ebene Flächen, sofern sie zum selben Objekt gehören sollen, müssen rechtwinklig zueinander stehen, falls sie aneinander angrenzen, müssen aber parallel sein, wenn sie sich gegenüber stehen. Objekte, die diese Bedingungen erfüllen, qualifizieren sich als Quader, bzw. Pakete. In [4] wird diese Methode auf zwei generische Objekte ausgedehnt (Quader und Zylinder). In [5] wird die Taxonomie von Schrauben eingeführt. Das Minimum an Bedingungen, das ein Objekt erfüllen muss, um sich als Schraube zu qualifizieren, besteht darin, dass ein Kopf da sein muss und ein Schaft mit einem Gewinde. Weitergehende Bedingungen spezifizieren die Steigung und die Art des Gewindes, die Länge des Schafts und des mit Gewinde versehenen Teils, die Form des Kopfs und die Art der Aussparung in ihm. So lassen sich Klassen wie Holzschrauben allgemein oder Metallschrauben M8 mit durchgehendem Schlitz im Schraubenkopf definieren. Es ist offensichtlich, dass generische Modelle eine effizientere Art der Objekterkennung ermöglichen.

Glossar

Computer Vision: Die Computer Vision verwendet Algorithmen, Datenrepräsentationen und Computerarchitekturen, welche auf Prinzipien beruhen, die ganz generell visuellen Fähigkeiten zugrunde liegen.

Robot Vision: Die Robot Vision nutzt die Methoden der Computer Vision im Hinblick auf eine intelligente Verknüpfung von Wahrnehmung und Handlung.

Tiefenbild, Abstandsbild (Range Image): Bildmässige Darstellung des Abstands jedes sichtbaren Szenenpunkts von einem Projektionszentrum.

Heuristik: Regel, die nicht immer, aber meistens gültig ist.

Paradigma: Anerkanntes Denk- oder Vorgehensmuster.

Szene: Dreidimensionale Konstellation von Dingen.

Purposive Vision: Zweckgerichtetes Sehen.

Active Vision: Datenakquisition verläuft szenengesteuert, unter visuellem Feedback.

Generic Model: Durch Bedingungen, die erfüllt sein müssen, wird eine (i.a. weitgefasste) Objektklasse definiert.

Focus of attention: Nur der Teil des Bildes wird analysiert, der interessante Informationen zu liefern verspricht.

Gruppierung: Automatische und unbewusste Fähigkeit des menschlichen visuellen Systems, zu erkennen, welche primären Bildelemente, z.B. Linienelemente, zusammengehören. Auch bei einem Computer Vision-System ist diese Fähigkeit erwünscht.

Multi-Sensor-Fusion: Kombination von Information von verschiedenen Datenquellen. Das Niveau, auf dem das geschieht, sollte möglichst hoch sein.

CAD-basierte Methoden

Die symbolische CAD-Beschreibung eines Objekts enthält implizit einen grossen Teil der Information, die zur Erkennung des Objekts benötigt wird, und zwar hauptsächlich geometrische Information. Es hat sich aber gezeigt, dass die bei der modellbasierten Objekterkennung gebräuchlichen geometrischen Merkmale, obwohl sie in den CAD-Modellen implizit vorhanden sind, erst explizit gemacht und bezüglich ihres Diskriminanzvermögens bewertet werden müssen. Andere Aspekte des visuellen Erscheinungsbildes, wie Farbe, Textur, Reflexionseigenschaften, Schattenwurf usw. müssen aus anderen Attributen der Objektmodelle als den geometrischen gefolgert oder zusammen mit Angaben über Beleuchtung usw. berechnet werden.

Bisher mussten die Modelle in einem langwierigen Prozess entweder manuell oder aber durch Training by Showing eingegeben werden. Hier ergeben sich neue Möglichkeiten, indem CAD-Modelle verfügbar gemacht werden, aus denen wenigstens im Prinzip die in den obigen Fällen benötigten Modelle automatisch abgeleitet werden können. Heute werden sowohl in der Robot Vision als auch im CAD eine Vielzahl verschiedener Modellier-Systeme angewendet. Es ist klar, dass die Beschreibung von Objekten, wie sie im CAD benutzt wird, wiederum modifiziert bzw. angepasst werden muss, um Information extrahieren zu können, die die effiziente und automatische Analyse von industriellen Szenen ermöglicht. Es ist eine Art Cross-Compiler zwischen den Repräsentationen von CAD und Computer Vision notwendig oder aber eine universale Darstellung der Objekte, aus der jeder der angesprochenen Bereiche direkt oder mit minimaler Umformung die für seinen Bereich optimale Darstellung erzeugen kann. Dieses junge Forschungsgebiet befindet sich in voller Expansion.

Methoden, die mit Aspekten arbeiten

Sowohl ein normales Grauwertbild als auch ein Tiefenbild zeigen nur die sichtbaren Oberflächen eines Objekts, so dass sein Aussehen (Aspect) sich mit dem Beobachterstandort ändert. Da dieser nicht vorausgesagt werden kann, muss das Objektmodell Information über das Aussehen des Objekts von allen Standorten aus enthalten. Es wurde der Vorschlag gemacht, als Objektmodell eine (strukturierte)

Kollektion typischer Ansichten des Objekts zu verwenden. Eine Ansicht wird dadurch charakterisiert, dass sich in ihr das qualitative Aussehen des Objekts nicht ändert. Eine Ansicht definiert also eine Äquivalenzklasse von Beobachterstandorten. Das Bild 3 zeigt die Aspekte einer dreidimensionalen Version des Buchstabens L. Die Linien, die die Aspekte voneinander trennen, werden *Ereignisse* genannt. Auf ihnen sieht man «entartete» Ansichten, Singularitäten. Noch entartetere Ansichten findet man auf den Kreuzungspunkten zweier Linien. Eine Darstellung als Graph mit den Aspekten als Knoten und den Ereignissen als Kanten drängt sich auf. Eine Objekterkennung wird dann als Matchen eines Subgraphen, der aus der Szene extrahiert wurde, mit einem Teil des Aspektgraphen formuliert. Dazu muss auch noch ein Ähnlichkeitsmass definiert werden.

Aspect Graph Matching ist gut geeignet für Polyeder, da die geraden Kanten zwischen den Flächen es leicht machen, Aspekte zu definieren. Objekte mit Freiformflächen (ohne Kanten) machen grosse Schwierigkeiten. Ausserdem gibt es bis jetzt nur Untersuchungen über einzelne Objekte, nicht über Szenen mit mehreren Objekten. Obwohl die auf Aspektgraphen beruhende Methode sehr interessant ist, ist sie (noch) nicht praktikabel, vor allem wegen der rechnerischen Komplexität: Vergleiche müssen mit vielen Objektmodellen, und für jedes Objektmodell für viele Aspekte durchgeführt werden.

Multisensor-Fusion

Weiter oben wurden die neuen Möglichkeiten, die die Tiefendaten bringen, hervorgehoben. Aber es muss gesagt werden, dass mit ihnen nicht alle Probleme der Robot Vision gelöst werden können. Sie werden mit Vorteil mit anderen Datenarten zusammen genutzt. Es ist zum Beispiel möglich, ein Grauwertbild aufzunehmen, das genau auf ein Abstandsbild passt. Man sagt dann, die beiden Bilder seien miteinander registriert. Die aus beiden Bildarten stammenden Informationen ergänzen sich gegenseitig. Beim Projekt COR wurde die Information, die aus drei Kamerabildern von verschiedenen Standorten stammte, erfolgreich integriert. Die Integration verschiedener Sensor-Modalitäten (z.B. visuelle, sei es schwarzweiss oder Farbe, taktile und Abstands-Daten) wird entscheidende

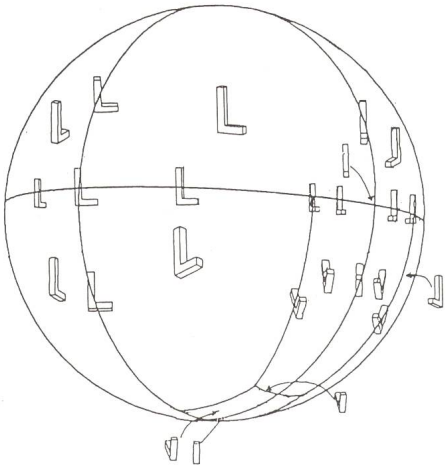


Bild 3 Aspekt-Kugel des Buchstabens L

Fortschritte bezüglich der Erkennungssicherheit von Merkmalen und Objekten ermöglichen. Die Integration kann auf verschiedenen Niveaus erfolgen, z.B. auf der Datenebene oder auf einem höheren symbolischen Niveau.

Farbe

Der überwiegende Teil der Computer Vision basierte bisher auf Grauwertbildern und Tiefenbildern. Vereinzelte Arbeiten mit Farbbildern und sogar industrielle Anwendungen mit solchen gibt es bereits seit langem. Massiv haben sie sich jedoch nicht durchgesetzt, was an den hohen Preisen für gute Farbkameras, an der grösseren Datenmenge und der höheren Komplexität der Algorithmen liegen mag. Unbestritten ist, dass die Einbeziehung von Farbinformation die Robustheit der Objekterkennung (im Sinne der Multisensor-Fusion) erhöhen kann. Insbesondere ist Farbe eine starke Heuristik für Objektzusammenhang.

Methoden ohne Objektmodelle

Man kann sagen, dass auf dem Gebiet der Objekterkennung das objektmodellbasierte Vorgehen dominiert. Es ist viele Jahre hindurch gründlich erprobt worden und wurde an viele verschiedene Aufgabenstellungen angepasst. Es besteht jedoch auch das Bedürfnis nach Vision-Systemen, die fähig sind, mit Objekten umzugehen, für die kein Modell abgespeichert wurde. Die einzige Information, über die das System in diesem Fall verfügen kann, ist die äussere Form der Objekte bzw. ihre Raumerfüllung. Hier braucht man also Methoden, die es gestatten, Formen zu analysieren, sie im Computer darzustellen und dann aufgrund dieser Darstellung Schlussfolgerungen zu ziehen, z.B. die Greifposition für eine Roboterhand zu bestimmen. Die Zerlegung einer aus mehreren Objekten bestehenden Szene ist hier noch schwieriger als im objektmodellbasierten Vorgehen. Die leichtere Verfügbarkeit von Tiefendaten, in denen die Form der Oberflächen besonders direkt zugänglich ist, hat zum gegenwärtigen Boom von Arbeiten über Formdarstellung beigetragen. Im folgenden werden einige wenige der angebotenen Methoden besprochen.

Formbeschreibung durch parametrisierte Flächen

Im Zweidimensionalen sind Verfahren zur Zerlegung beliebiger geschlossener Konturen in elliptische harmonische Fourier-Komponenten bekannt [7]. Die entsprechende Reihenentwicklung hat die Form

$$\mathbf{r}(s) = \begin{pmatrix} x(s) \\ y(s) \end{pmatrix} = \sum_{l=0}^{\infty} \mathbf{a}_l \cos\left(\frac{2\pi l s}{S}\right) + \sum_{l=0}^{\infty} \mathbf{b}_l \sin\left(\frac{2\pi l s}{S}\right) \qquad (1)$$

wobei s die Bogenlänge als Parameter, S der Umfang der geschlossenen Kontur ist. Die $\boldsymbol{a}_l$ und $\boldsymbol{b}_l$ sind Fourierkoeffizientenvektoren. Das Bild 4 zeigt eine Originalkontur sowie Rekonstruktionen mit abgebrochenen Reihen mit $l = 1$ und $l = 5$. Die Fourierkoeffizienten können als Merkmale von 2D-Formen betrachtet wer-

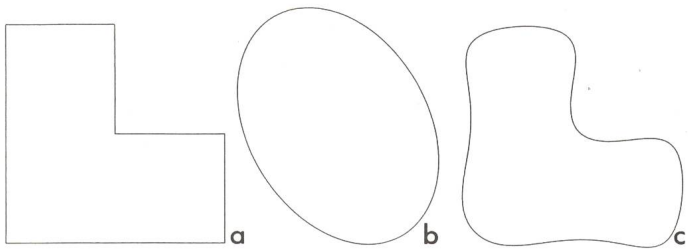


Bild 4 Approximation geschlossener Kurven durch elliptische harmonische Fourier-Komponenten: Original und Approximationen mit $l = 1$ und $l = 5$

den, wie sie zum Beispiel schon zum Erkennen von Buchstaben und Ziffern eingesetzt wurden.

Es ist verführerisch, eine Übertragung dieses Verfahrens auf drei Dimensionen zu versuchen. Dies ist kürzlich gelungen [8]. Damit ist es möglich geworden, jede beliebige geschlossene, einfach zusammenhängende Oberfläche als Reihenentwicklung auf der Basis von Kugelfunktionen darzustellen, das heisst die Orte $\boldsymbol{r}(\vartheta,\phi)$ auf der ursprünglichen Oberfläche lassen sich als Reihe

$$\mathbf{r}(\vartheta,\phi)=\begin{pmatrix} x(\vartheta,\phi) \\ y(\vartheta,\phi) \\ z(\vartheta,\phi) \end{pmatrix}=\sum_{l=0}^{\infty}\sum_{m=-l}^{l}\mathbf{c}_m^l\, Y_m^l(\vartheta,\phi) \quad (2)$$

darstellen. Auch hier braucht man nur so viele Glieder dieser Reihe mitzunehmen, wie es für den Verwendungszweck der Darstellung notwendig ist (Bild 5). Zum Beispiel wird man zur Planung des Greifens eines Objekts mit einer Roboterhand mit einer relativ groben Darstellung auskommen. Die Methode ist allerdings nur auf geschlossene, einfach zusammenhängende Oberflächen anwendbar. Auch wird das Problem der Zerlegung einer Szene in Einzelobjekte gar nicht angesprochen.

Zerlegung von Flächen in Dreiecksflächen

Daten, die von einem Tiefensensor stammen, können als eine Menge von *(x, y, z)*-Tripeln aufgefasst werden, die eine ausgedehnte, deformierte Punktwolke bilden, die aufgrund des überlagerten Rauschens eine gewisse Dicke aufweist. Diese Punktwolke kann in mehr oder weniger ausgedehnten Gebieten durch Flächenstücke (Patches) approximiert werden. Neben der damit erzielten Datenreduktion erhält man als Nebeneffekt noch eine Glättung der Daten. Wegen ihrer Einfachheit ist die Approximation durch aneinander anschliessende Dreiecke sehr verbreitet. Da jedem Dreieck auch seine Normalenrichtung mitgegeben werden kann, hat man damit auch die notwendige Information zum Berechnen von Greifpositionen. Von Bedeutung ist dieses Verfahren auch bei den Projekten zum Bau eines autonomen Landvehikels (ALV), an dem verschiedene Forschungsinstitute arbeiten, wo es die zur Kollisionsvermeidung notwendigen Daten liefert.

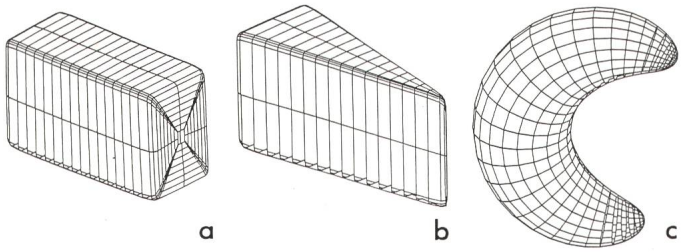


Bild 6
Deformierbare Superquadriken
a. Quader
b. Torte
c. Gipfel

Superquadriken

Wenn man Grund zu der Annahme hat, dass eine bestimmte Untermenge eines Tiefendatensatzes von *einem* Objekt kommt, kann man versuchen, diese durch eine implizite Funktion, das heisst, eine Gleichung der Form $f(x, y, z) = 0$ zu approximieren und dadurch zu beschreiben. In letzter Zeit sind Superquadriken, insbesondere deformierbare Superquadriken bekannt und häufig verwendet worden. Die Gleichung einer Superquadrik lautet:

$$\left(\left(\frac{x}{a}\right)^{\frac{2}{2\varepsilon_2}}+\left(\frac{y}{b}\right)^{\frac{\varepsilon_2}{\varepsilon_1}}+\left(\frac{z}{c}\right)^{\frac{2}{\varepsilon_1}}\right)-1=0 \quad (3)$$

Mit $\varepsilon_1 = 1$ und $\varepsilon_2 = 1$ reduziert sich die obige Gleichung zu der eines Ellipsoids, mit $\varepsilon_1 \neq 1$ und $\varepsilon_2 \neq 1$ wird das Ellipsoid eckiger.

Durch bestimmte Koordinatentransformationen kann man auch Verjüngungen und Biegungen erhalten. Die Bilder 6a–c geben eine Ahnung von der Vielfalt möglicher Formen. Die Qualität des Fits (Anpassung der Daten an Modell) hängt stark davon ab, ob die zu fittenden Tiefendaten von allen Seiten des Objekts kommen oder ob es nur eine einseitige Auswahl ist.

Schlussbetrachtung

Die Computer Vision scheint eine Midlife Crisis überstanden zu haben. Die Neubesinnung verspricht eine bewusste Hinwendung zur Arbeit an der Lösung gut definierter konkreter Einzelprobleme, wodurch die Robot Vision nur profitieren kann. Dieser Beitrag skizzierte eine Anzahl von aktuellen Problemen, die zeigen, dass Robotik und Robot Vision dynamische Forschungs- und Entwicklungsgebiete sind. Er konnte den Stand der Robot Vision nicht vollständig wiedergeben, sondern musste sich notwendigerweise mit der Darstellung einiger Highlights zufrieden geben.

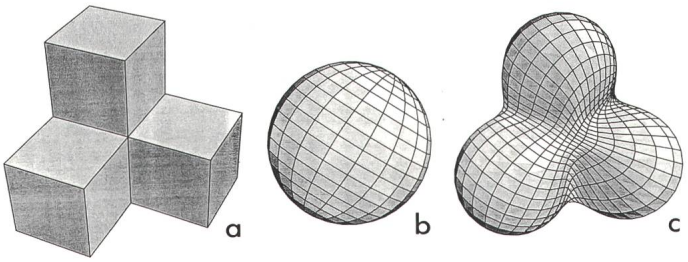


Bild 5
Approximation geschlossener Flächen durch Entwicklung nach Kugelfunktionen:
Original und Approximationen mit $l = 1$ und $l = 5$

Literatur

[1] *Vischer D.:* Cooperating Robot with Visual and Tactile skills. Proc. 1992 IEEE International Conference on Robotics and Automation, Nice, 10.–15.5.1992, pp. 2018–2025.

[2] *Ade F., Peter M., Rutishauser M., Trobina M., Ylä-Jääski A.:* Vision for a 3-D Object Manipulation System, Proc. 14. DAGM Symposium, Dresden, 14.–16.9.1992, pp. 117–124.

[3] *Rechsteiner M., Schneuwly B., Guggenbühl W.:* Roboter im Postdienst – Ein schnelles robustes Vision-System zur Paketvereinzelung. Bulletin SEV/VSE 83(1992)21.

[4] *Mulgaonkar P.G., Cowan C.K., De Curtins J.:* Understanding Object Configurations Using Range Images. IEEE Trans. Pattern Analysis and Machine Intelligence, 14(1992)2, pp. 303–307.

[5] *Kriegman D.J., Binford Th.O., Sumanaweera Th.:* Generic Models for Robot Navigation, Proc. DARPA Image Understanding Workshop, Cambridge, Massachusetts, 6.–8. April 1988, pp. 453–460.

[6] *Gigus Z., Malik J.:* Computing the Aspect Graph for Line Drawings of Polyhedral Objects. IEEE Trans. Pattern Analysis and Machine Intelligence, 12(1990)2, pp. 113–122.

[7] *Kuhl F.P., Giardina Ch.R.:* Elliptic Fourier Features of a Closed Contour. CGIP (1982)18, pp. 236–258.

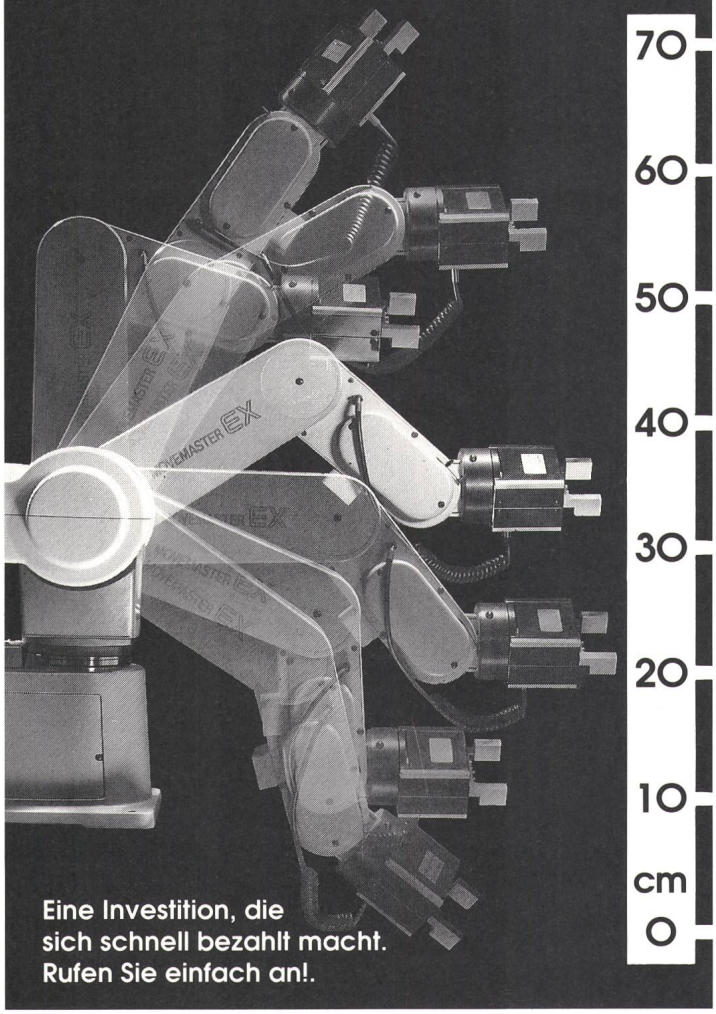
[8] *Brechbühler Chr., Gerig G., Kübler O.:* Surface Parametrization and Shape Description, Proc. Conf. on Visualization in Biomedical Computing, Chapel Hill, N.C., Oct 13–16, 1992.

5-Achsen-Roboter in 2 Leistungsklassen...

von Mitsubishi Electric

Typ	RV-M1	RV-M2
Arbeitsradius	602 mm	755 mm
Last	1,2 kg	2,0 kg
Genauigkeit	0,3 mm	0,1 mm

1.000 fach bewährt in den Bereichen Handling, Prüfen und Testen, Labor-Automation, Pick and Place, Montage und Oberflächenbearbeitung.

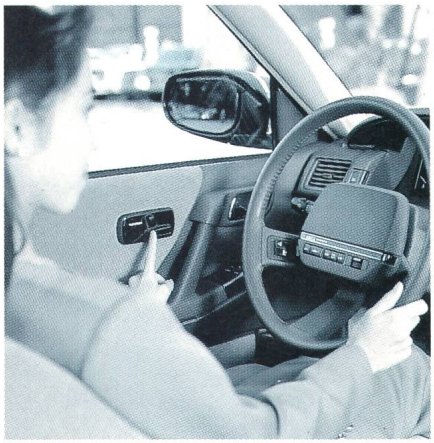


Eine Investition, die sich schnell bezahlt macht. Rufen Sie einfach an!.

ECONOTEC
INDUSTRIE AUTOMATION AG

CH-8309 NÜRENSDORF
POSTFACH 282
FAX 01/836 73 43
☎ 01/836 40 90

TAKE IT EASY



...dank innovativer, moderner Automatisierungstechniken! Wie stünden wir heute da, wenn sich Garagentore und automatische Türen nicht wie von selbst öffnen, Rolltreppen und -bänder uns nicht bequem transportieren oder Geld und Benzin nicht jederzeit zur Verfügung stehen würde? Nicht nur in diesen alltäglichen Bereichen sondern auch in der Industrie erleichtert innovative Technik von OMRON das tägliche Leben oder macht Automatisierungsprozesse erst möglich.

INNOVATIVE TECHNIK

OMRON

OMRON ELECTRONICS AG

Gewerbestrasse 7
CH-6330 Cham

Telefon (042) 41 64 41
Telefax (042) 41 64 45

OMRON

McCANN-ERICKSON

Für Unternehmer, die mit Teamwork gross werden. Ascoline.

Ob Ihre Mitarbeiter gut zusammenspielen, ist oft eine Frage der richtigen Einstellung. Deshalb fördert das Kommunikationssystem Ascoline den Teamgeist mit flexiblen Team- und Stellvertreterschaltungen, damit auch in hektischen Situationen keine Anrufe verlorengehen. Und dank einer gehörigen Portion High-Tech, bspw. die akustische Bedienerführung mittels Sprechtexten, stehen Sie mit dem System jederzeit auf du und du. Wie Sie also in Unternehmen mit 30 bis über 300 Mitarbeitern bedeutend mehr vom Telefon haben, erfahren Sie von der Ascom Business Systems AG. Rufen Sie noch heute eines unserer Regionalzentren in Ihrer Nähe an: Zürich: 01/823 14 14, Bern: 031/999 44 93, Lausanne: 021/641 42 11. Oder kontaktieren Sie Ihre zuständige Fernmeldedirektion, Telefon 113. **TELECOM** PTT

Das formschöne Brigit 100 mit Lautsprecher, Display und Message-Anzeige für wirtschaftlichen Telefonkomfort.

Teilnehmervermittlungsanlagen: **ascom** *denkt weiter.*