

„Computer lernen Sehen“

Statistische Lernverfahren bringen Bildverarbeitung voran – kommentiert von Prof. Dr. Cremers



Prof. Dr. Daniel Cremers ist Leiter der Computer Vision Gruppe an der Universität Bonn

Webcams, Kameras in Handys und Kameras in Autos: Nicht nur in unserer Alltagswelt, sondern auch in der Industrie, in der Robotik und in der Medizin wird es zunehmend wichtig, Maschinen in die Lage zu versetzen, dass sie Bilddaten in ähnlicher Weise interpretieren können wie der Mensch. Prof. Dr. Daniel Cremers von der Universität Bonn über die Entwicklungen und Perspektiven.

Trotz rapide wachsender Leistungsfähigkeit, Gigabytes an Speicherplatz und Gigahertz an Taktfrequenz können Computer bis heute nur kaum mit den visuellen Fähigkeiten biologischer Systeme mithalten. Das menschliche Sehsystem arbeitet

scheinbar mühelos. Aus einer Ansammlung farbiger Lichtquanten auf der Netzhaut stellt es alle Informationen bereit, die man zum Lesen dieser Zeilen, zur Orientierung in einem Raum oder zum Greifen nach einer Kaffeetasse benötigt.

Zu einem großen Teil beruhen derartige Fähigkeiten biologischer Sehsysteme darauf, dass vorher gelerntes Wissen über die Umwelt ausgenutzt wird. Schließlich verfügen Menschen über umfangreiches Weltwissen – Wissen darüber, wie Tassen typischerweise aussehen, dass Tassen oftmals im Küchenschrank zu erwarten sind, und dass man gefüllte Tassen tunlichst mit Öffnung nach oben transportiert. Derartiges Weltwissen haben wir uns durch einen mehrjährigen Lernprozess angeeignet. Und es ist Voraussetzung dafür, dass wir uns in unserer Umwelt zurechtfinden.

In mehr als drei Jahrzehnten hat sich die digitale Bildverarbeitung weit entwickelt. Viele der klassischen Herausforderungen für die maschinelle Analyse digitaler Bilder sind heute bereits teilweise gelöst: Z. B. gibt es Digitalkameras, die automatisch Gesichter in entsprechenden Bildregionen detektieren, um Beleuchtung und Autofokus anzupassen. Es gibt automatische Spurerkennungssysteme im Automobil, die anhand von Kamerabildern ein akustisches Warnsignal abgeben, sobald das Fahrzeug die Randmarkierungen zu überfahren droht. Im Bereich der medizinischen Bildverarbeitung gibt es Verfahren, die anhand einer Ultraschallbildsequenz online berechnen, wie viel Blut ein schlagendes Herz pro Minute pumpt. Die Chancen und Möglichkeiten, die in derartigen Technologien stecken, sind offensichtlich enorm.

Dennoch fallen die Fähigkeiten bisher eingesetzter Bildverarbeitungsmethoden oft weit hinter denen menschlicher Beobachter zurück, insbesondere dann, wenn die Komplexität der beobachteten Szene und der Raum möglicher Interpretationen größer werden. Z. B. sind für die Bildverarbeitung im Straßenverkehr Herausforderungen wie das zuverlässige Verfolgen anderer Verkehrsteilnehmer und das sichere Erkennen von Hindernissen im Fahrkorridor bis heute nicht zufriedenstellend gelöst. Hier bleibt die menschliche Fähigkeit zur Bildanalyse die weiterhin unerreichte Messlatte.

In den vergangenen Jahren ist es allerdings gelungen, die Leistungsfähigkeit von Computer Vision Verfahren deutlich zu

verbessern – nämlich durch die Integration von Weltwissen. In der Bonner Computer Vision Gruppe sind beispielsweise statistische Methoden entwickelt worden, mit denen ein Computer ähnlich dem Menschen die Fähigkeit erlangt, Wissen über Objekte in einem Lernprozess anzusammeln. Anschließend findet die Maschine die in Bezug auf dieses Weltwissen wahrscheinlichste Interpretation eines

„Computer-Vision-Verfahren erlangen zunehmend Weltwissen“

gegebenen Bildes. Unter anderem erlauben es diese Methoden, die zeitliche Veränderung vertrauter Formen – z. B. laufender Menschen – statistisch zu lernen. Mit derartigem Wissen kann die Maschine laufende Menschen zuverlässig in Videos verfolgen, und zwar auch dann noch, wenn der Rauschpegel so hoch ist, dass selbst menschliche Beobachter das Objekt nur mit Mühe erraten können.

Zunehmend finden derartige statistische Methoden Einzug in die industrielle Bildverarbeitung. Umgekehrt treiben die Anforderungen der Industrie die Entwicklung neuer Methoden in der Grundlagenforschung. Die Bonner Arbeitsgruppe kooperiert zu diesem Zweck mit verschiedenen Industriefirmen: In Kooperation mit der Firma DaimlerChrysler sind Optimierungsverfahren entwickelt worden, die pixelgenau Hindernisse im Fahrkorridor erfassen. In Zusammenarbeit mit Siemens Corporate Research sind statistische Methoden entwickelt worden, die drei-dimensionale Herzenmodelle aus Ultraschall-Bildsequenzen extrahieren. In Zusammenarbeit mit Microsoft Research werden Methoden entwickelt, um von Bildern und Bildsequenzen in möglichst unüberwachter Weise zu einer Szeneninterpretation zu gelangen. Mit statistischen Lernverfahren dringen Maschinen in die Bereiche der menschlichen Sehfähigkeiten vor. So werden werden sie mittelfristig auch für komplexere Herausforderungen unserer Alltagswelt zunehmend Einsatz finden.

UNIVERSITÄT BONN

WWW

www.vfmz.de/170002