Bordcomputer mit 7. Sinn

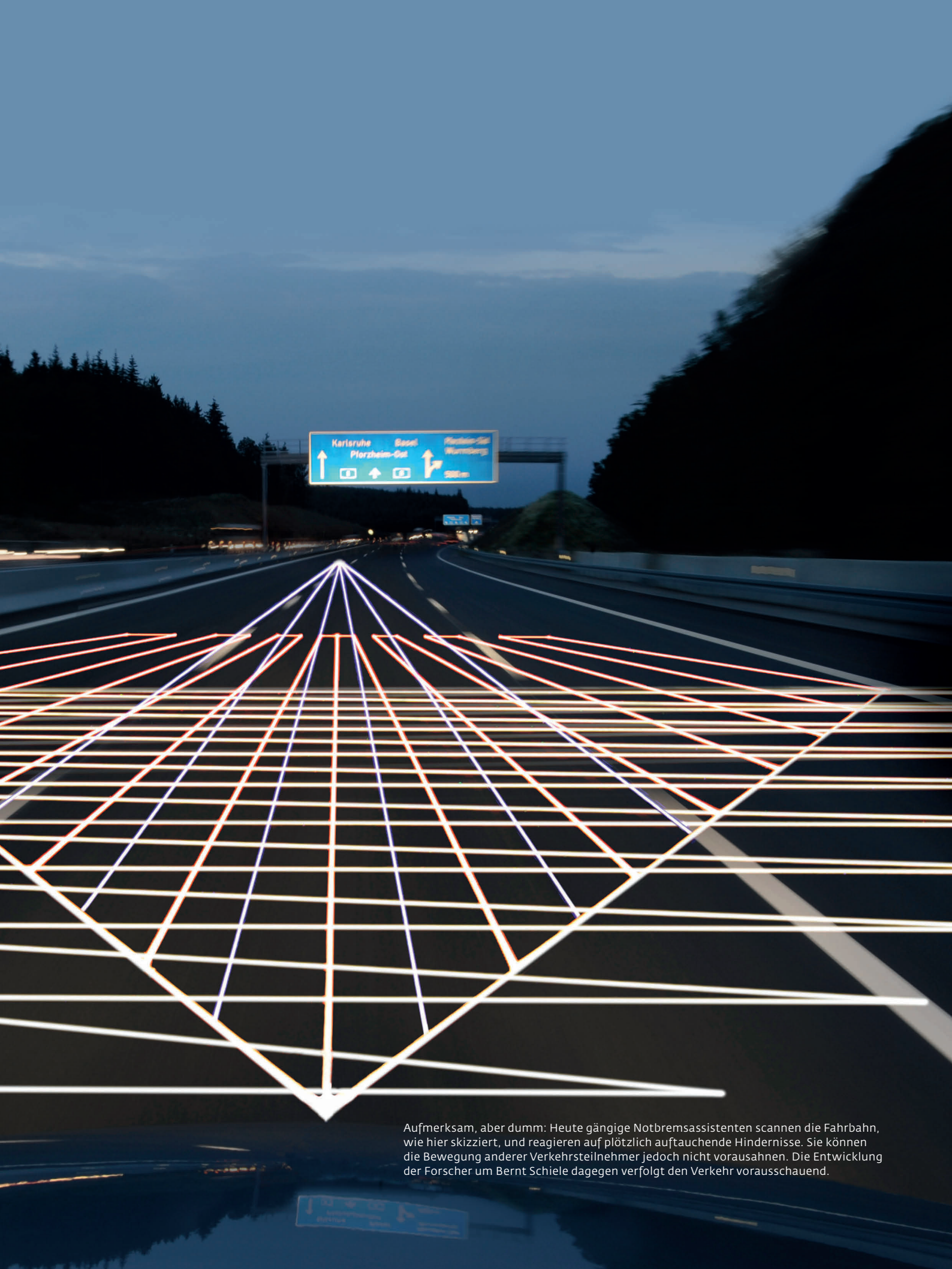
Notbremssysteme verhindern heute schon manchen Unfall im Verkehr, einen richtigen Überblick über das Geschehen auf der Straße haben elektronische Assistenten aber noch nicht. Das will **Bernt Schiele**, Direktor am **Max-Planck-Institut für Informatik** in Saarbrücken ändern. Er bringt Computern bei, die Wege von Fahrzeugen und Fußgängern vorauszuahnen.

TEXT **TIM SCHRÖDER**

Sie heißen Forward Alert, PreSafe oder PreSense und finden sich vor allem an Bord teurer Autos. Bei den Systemen mit den kernigen englischen Namen handelt es sich um sogenannte Notbremsassistenten. Läuft plötzlich ein Kind auf die Straße oder braust bei Rotlicht ein Raser über die Kreuzung, dann geht der Assistent von allein in die Bremsen – selbst wenn der Fahrer noch starr vor Schreck ist. Mit kleinen Infrarotkameras oder Radarsensoren tasten diese Systeme die Straßenszene vor dem Auto ab. Objekte auf Kollisionskurs erfasst der Notbremsassistent in Sekundenbruchteilen, zum Beispiel einen Fußgänger, der gedankenverloren auf die Straße tritt. In Millisekunden analysiert der Mikroprozessor die Lage: Wie schnell ist das Objekt? Kommt es zum Zusammenstoß? Achtung. Stopp!

Die Fahrzeughersteller betonen, dass die Notbremsassistenten allein in Europa jedes Jahr Tausende von Unfällen verhindern. Die Technik funktioniert und ist in Schrecksekunden schneller als der Fahrer. Doch ist sie auch heute noch vergleichsweise dumm. Notbremsassistenten und vergleichbare Sensorsysteme im Auto können nur auf die letzte Sekunde reagieren. Sie konzentrieren sich voll auf ein herannahendes Objekt – mehr aber nicht. Ein ausgeruhter Autofahrer hingegen hat die ganze Szene im Blick. Den Rettungswagen, der sich von weit rechts nähert zum Beispiel, oder ein flottes Auto, das an einer mehrspurigen Kreuzung kurz von einem Lkw verdeckt ist und dann über die Kreuzung saust. Kurz: Notbremsassistenten können im letzten Moment einschreiten, aber nicht vorausschauend fahren. >

Foto: Daimler AG



Aufmerksam, aber dumm: Heute gängige Notbremsassistenten scannen die Fahrbahn, wie hier skizziert, und reagieren auf plötzlich auftauchende Hindernisse. Sie können die Bewegung anderer Verkehrsteilnehmer jedoch nicht vorausahnen. Die Entwicklung der Forscher um Bernt Schiele dagegen verfolgt den Verkehr vorausschauend.

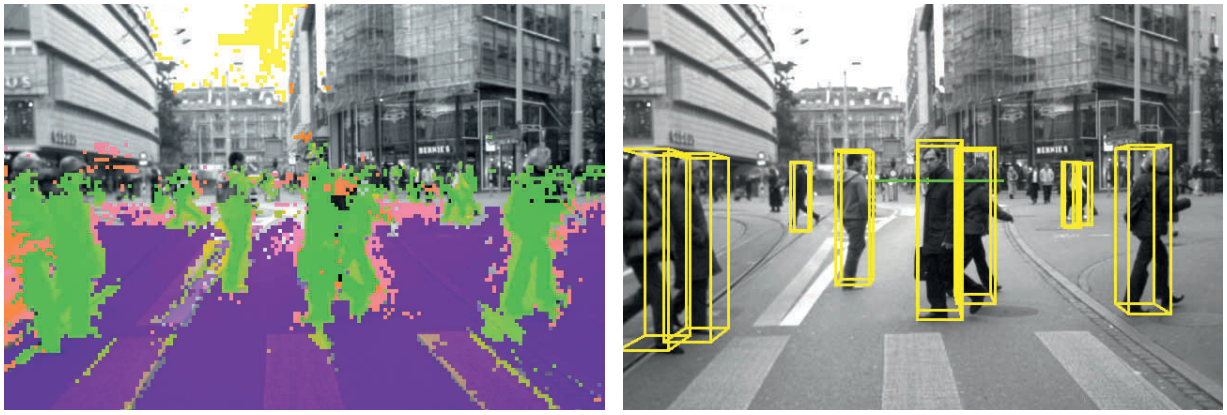
Das Gleiche gilt für andere Automaten, für Assistenzroboter zum Beispiel. Sie sind heute durchaus in der Lage, ein Tablett vom Küchentisch zu nehmen und an Hindernissen vorbei ins Wohnzimmer zu tragen. Eine komplexe Szene, in der der Hund durch die Küche springt und Kinder vor den Schränken toben, bringt die blechernen Helfer aber noch völlig aus dem Konzept. Das Ziel der Forscher ist es, ihren Maschinen beizubringen, ganze Szenen zu erfassen und vor allem Bezüge zwischen den Objekten herzustellen. Bremst ein Auto, dann bremst auch das nachfolgende. Eine so simple Logik beherrschen die modernen Automaten meist nicht. Das ist ein Grund dafür, warum es bis heute keine Assistenzkameras für Blinde gibt. Zu komplex sind die Zusammenhänge. Fußgänger weichen einander aus, queren plötzlich den Weg. Und halb verdeckte Fußgänger oder Autos werden von Automaten regelmäßig übersehen.

EIN AUTO SOLL AUTONOM DURCH DEN VERKEHR NAVIGIEREN

Assistenzsysteme klüger zu machen, das ist das Ziel von Bernt Schiele, Direktor der Abteilung Computer Vision and Multimodal Computing am Max-Planck-Institut für Informatik in Saarbrücken. Er will Computern beibringen, eine Szene wie ein Mensch zu verstehen – und entsprechend zu handeln. Schieles Arbeitsgruppe hat raffinierte Rechenvorschriften entwickelt, die eine Straßenszene vollständig analysieren, alle Objekte erfassen, Fußgänger, Autos und Lkw, ganz gleich, ob diese gut zu sehen oder halb verdeckt sind. Wenn Schiele und seine Mitarbeiter dürften, wie sie wollen, dann würden sie ein Auto mit ihrer Software autonom durch den Verkehr navigieren lassen. In den Bildern einer Videokamera, die hinter der Windschutzscheibe montiert ist und den Verkehr beobachtet, können sie bereits gan-

Fußgänger von rechts: Die Software der Saarbrücker Informatiker analysiert die Szene. Sie erkennt die Fahrzeuge und den Passanten, verfolgt die Bewegung aller Verkehrsteilnehmer und berechnet, ob ein Zusammenprall droht. Bei Gefahr reagiert sie autonom.

Foto: istockphoto



Wie ein Computer unübersichtliche Verkehrssituationen versteht: Ein Classifier ermittelt, mit welcher Wahrscheinlichkeit es sich bei einer bestimmten Pixelwolke um einen Fußgänger handelt. Detektoren tragen die Ergebnisse in einer Art Landkarte (links) ein, in der die grünen Bereiche mit hoher Wahrscheinlichkeit Fußgänger darstellen. Die identifizierten Verkehrsteilnehmer rahmt die Software farbig ein (rechts), um sie zu verfolgen und ihre Bewegungen zu antizipieren.

ze Straßenszenen erfassen. „Autonomes Fahren wird kommen, ist aber natürlich – noch – nicht erlaubt“, sagt Schiele.

Szenen verstehen. Das klingt so simpel. Dank unserer Erfahrung erfassen wir Menschen die Situation an einer Kreuzung sofort. Ampel rot: Die Leute bleiben stehen. Ampel grün: Ich darf gehen. Uns ist es egal, wie viele Menschen hin und her eilen. Was ein Computer aus dem Bild einer Kamera herausliest, ist etwas gänzlich anderes. Er sieht Tausende von Bildpunkten, helle, dunkle, rote, grüne, und muss erst lernen, was diese überhaupt zu bedeuten haben.

Für die Szenenanalyse braucht man deshalb ein ganzes Bündel an ausgeklügelten Algorithmen, mit denen der Computer Stück für Stück analysiert, was Sache ist. Zunächst Algorithmen, die bestimmte Strukturen erkennen. Fußgänger sind länglich und haben eine bestimmte Höhe. Sie besitzen zwei Arme und zwei Beine. Autofronten sind flach, Lkw-Fronten hoch. In einem nächsten Schritt muss die Software herausfinden, ob und wohin sich die Objekte bewegen. Und drittens muss der Computer logische Schlüsse ziehen: Hält Auto eins an der Ampel, so tut das aller Voraussicht nach auch Auto zwei dahinter.

Grundlage all dieser Analysen ist die Wahrscheinlichkeitsrechnung. Mit Milliarden von Rechenoperationen pro Sekunde fragt das Programm ab, mit welcher Wahrscheinlichkeit ein Haufen Bildpunkte tatsächlich ein Gegenstand ist. „Probabilistische 3D-Szenenanalyse“ nennt Schiele diese komplexe Form des Computerverstehens. Bemerkenswert ist, dass Schiele für die dreidimensionale, also räumliche Analyse eine einzige Kamera benutzt. Menschen haben zwei Augen, um räumlich sehen zu können. „Wir berechnen die dreidimensionale Information aus dem zweidimensionalen Computerbild“, sagt Schiele.

DIE SOFTWARE LERNT, WIE AUTOS ODER FUSSGÄNGER AUSSEHEN

Doch zunächst mussten die Saarbrücker Forscher ihre Algorithmen anlernen: Sie fütterten den Computer mit Trainingsdaten – mit Hunderten Bildern von Fußgängern, Autos und Lkw. Die Software lernte damit nach und nach, wie ein Auto oder Fußgänger aussieht. „Classifier“ heißen diese Erkennungsprogramme, die in den Pixelwolken des Kamerabildes nach bestimmten Objekten forschen. Kanten von Objekten zum Beispiel detektieren sie anhand abrupter Veränderungen der Farbe oder der Helligkeit benachbarter Pixel. Am Ende werfen sie „Scores“ aus – Werte, die angeben, wie wahrscheinlich es ist, dass es sich bei einer bestimmten Pixelstruktur tatsächlich um ein Objekt handelt. Für die komplexen Straßenszenen benötigt Schiele eine ganze Reihe verschiedener Classifier, zum Beispiel solche, die Umgebungsstrukturen wie eine Straße oder einen Baum erkennen, und solche, die im Pixeldurcheinander diskrete Objekte wie Autos, Lkw und Fußgänger erspähen. Dabei reichen Classifier, die nur ganze Objekte erkennen, nicht aus. Zum Einsatz kommen außerdem spezielle Classifier, die Schiele und seine Mitarbeiter mit Objektteilen trainiert haben: einem Arm, einem halben Rücken, einer Kühlerhaube. Denn nur so lassen sich später verdeckte Objekte sicher detektieren. Die Ergebnisse der Classifier, die Scores, werden von größeren Algorithmen, den Detektoren, ausgewertet. Die Detektoren erstellen für jedes einzelne Bild einer Videosequenz eine Art Landkarte, eine Score-Map, auf der für jedes Pixel verzeichnet ist, mit welcher Wahrscheinlichkeit es zu einem bestimmten Objekt gehört.

Um festzustellen, ob das Detektorergebnis plausibel ist, gleicht der Computer die Score-Map-Werte mit seinem Weltwissen ab. Anhand der Trainingsdaten hat er gelernt, wie eine Straße oder ein Auto aussieht. Hinzu kommt das dreidimensionale Wissen. Autos, die entgegenkommen, wirken umso kleiner, je weiter weg sie sind. Darüber hinaus befinden sich weiter entfernte Autos im Kamerabild weiter oben als

Fotos und Grafik: MPI für Informatik

» Ein wesentlicher Unterschied zum klassischen Auto-Assistenzsystem liegt darin, dass die Software die Bewegung der Objekte permanent verfolgt.

nahe Autos. Gemäß erlerntem Weltwissen könnte beispielsweise ein großes Auto nicht am oberen Bildrand erscheinen. Eine Straßenlaterne wiederum ist kein Fußgänger, weil sie deutlich größer ist. So prüft der Computer für jedes Videokamerabild die Plausibilität seiner Analyse: Wie wahrscheinlich ist es, dass die Objekte in dieser Szene tatsächlich der realen Szene entsprechen?

So trainiert und mit Weltwissen ausgestattet, musste sich die Software einem Test stellen: Die Forscher spielten ihr reale Videosequenzen vor, die sie im fahrenden Auto aufgenommen hatten und in denen Fußgänger zu sehen sind, die beispielsweise durch die Straßen von Zürich eilen. ETH-Loewenplatz, ETH-Linthescher oder ETH-PedCross2 heißen diese Bildfolgen, die Forscher von der ETH Zürich aufgenommen haben, an der Schiele einige Zeit verbracht hat.

DAS SYSTEM AHNT BEWEGUNGEN DER OBJEKTE VORAUS

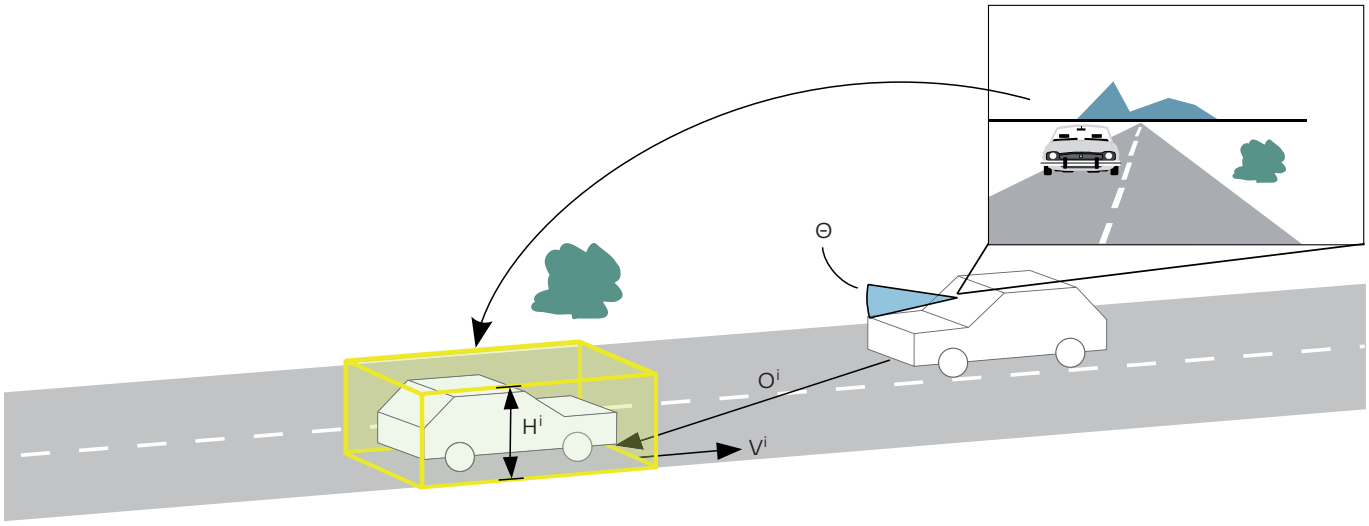
Dabei zeigte sich, dass die Classifier und Detektoren oftmals danebenlagen, wenn sie nacheinander jedes Bild einzeln auswerteten. Vor allem verdeckte Objekte wurden häufig übersehen. Das änderte sich, wenn die Algorithmen etwa fünf aufeinanderfolgende Bilder miteinander abglichen. Eine flimmerfreie Filmsequenz besteht aus mindestens 24 Bildern in der Sekunde. Bewegte Objekte verändern ihre Position wie in einem Daumenkino von Bild zu Bild minimal. Berücksichtigen die Algorithmen mehrere aufeinanderfolgende Bilder, erkennen sie vor allem die verdeckten Objekte besser. „Die Bilderkennung wurde sehr viel robuster“, sagt Schiele. „Tracklets“ nennt Schiele diese analytisch verschmolzenen kurzen Bildfolgen.

Ein wesentlicher Unterschied zum klassischen Auto-Assistenzsystem liegt darin, dass die Software die Bewegung der Objekte von Tracklet zu Tracklet permanent verfolgt. Notbremsassistenten nehmen Gefahren wahr, die blitzartig auftauchen. Bei Schiele hingegen „propagieren“ die Objekte: Hat die Software auf dem Bildschirm ein Objekt erfasst, wird es farbig umrahmt. Von Tracklet zu Tracklet bewegt sich der farbige Rahmen mit dem Objekt mit, bis dieses aus der Szene verschwindet. Wird es auf der Straße voll, bewegen sich Dutzende dieser Rahmen über das Videobild. Dank des Weltmodells kann das System sehr genau vorausahnen, wie sich ein Objekt bewegt. Ein Fußgänger wird bei Grün nicht plötzlich auf Autogeschwindigkeit beschleunigen. Und ein Fahrzeug, das kurz hinter einem Gegenstand im Vordergrund verschwindet, bewegt sich im Gedächtnis der Software weiter, wenn alle anderen Fahrzeuge mit derselben Fahrtrichtung ihren Weg ebenfalls fortsetzen.

Beachtlich ist das Tempo der Analyse. Es geht nicht allein um 24 Bilder pro Sekunde. Wer den fließenden Verkehr im Auge behalten und in Echtzeit ana-

Bringen dem Bordcomputer den 7. Sinn bei: Bernt Schiele (oben), Bojan Pepik, Jan Hosang und Mykhaylo Andriluka (unten, von links) entwickeln eine Software, die es Autos sogar ermöglicht, autonom zu fahren.

Fotos: MPI für Informatik - Manuela Meyer



Aus einem zweidimensionalen Bild, wie es rechts in der Grafik dargestellt ist, gewinnt der Bordcomputer des rechten Autos ein dreidimensionales Verständnis der Verkehrssituation. Eine Kamera filmt die Straßenszene mit dem Winkel Θ. Anhand ihres antrainierten Weltwissens berechnet die Software aus der Höhe H^i den Abstand O^i zu einem nahenden Fahrzeug und aus der Veränderung der Höhe dessen Geschwindigkeit V^i.

lysieren will, muss schneller sein. So läuft die ganze komplexe Wahrscheinlichkeitsrechnung in Millisekunden ab. Diese permanente schnelle Analyse hat den Vorteil, dass sie vor bösen Überraschungen schützt. Ein Auto, das verdeckt durch eine Schlange stehender Fahrzeuge in die Kreuzung rast, wird vom Notbremsassistenten möglicherweise übersehen. Mit der 3D-Szenenanalyse dürfte es frühzeitig entdeckt werden, wenn es in den Lücken zwischen den stehenden Fahrzeugen kurz auftaucht.

VIDEOS VON EINER KAMERA IM RÜCKSPIEGEL

Fahrzeugen ein vorausschauendes Verständnis für das beizubringen, was um sie herum geschieht, interessiert natürlich auch Autohersteller. Tatsächlich kooperiert Schiele seit Jahren mit solchen Unternehmen. Für die Aufnahme der Auto-Videosequenzen wurde ihm und seinen Mitarbeitern beispielsweise ein Fahrzeug mit einer kleinen Kamera am Rückspiegel zur Verfügung gestellt. „Doch es geht hier ja nicht nur um Autos“, sagt Schiele. Die probabilistische 3D-Szenenanalyse eigne sich vielmehr für die Analyse sehr verschiedener Filmsequenzen – zum Beispiel auch für die Bilder aus den Kameraaugen eines Roboters im Haushalt oder in einer Fabrik.

Einige der von Schiele und seinen Mitarbeitern entwickelten Softwarebausteine werden demnächst erstmals in den USA in einem autonom fahrenden Auto zum Einsatz kommen. Zunächst sollen die Fußgänger- und Objektdetektoren zeigen, was sie leisten können. Dabei will Schiele möglichst praxisnah testen, wie die Detektoren mit Radar- und Laserscannern zusammenarbeiten. Ein Ziel ist es, mit möglichst wenigen und vor allem handelsüblichen Radar- und Kamerasystemen auszukommen, wie sie heute bereits in Autos eingesetzt werden. Denn nur wenn sich das System mit vertretbarem technischem Aufwand realisieren lässt, kann es einem Fahrer künftig vielleicht einen 7. Sinn für den Verkehr geben oder irgendwann gleich ganz das Steuer übernehmen. ◂

AUF DEN PUNKT GEBRACHT

- **Notbremssysteme stoppen im letzten Moment, wenn ein Kind oder ein Fahrzeug vor einem Auto auftaucht. Das Verhalten anderer Verkehrsteilnehmer können sie weder analysieren noch vorausahnen.**
- **Die automatische 3D-Szenenanalyse erkennt anhand einer Wahrscheinlichkeitsanalyse andere Teilnehmer im Straßenverkehr, auch wenn diese zeitweise verdeckt sind, und kann ihre Bewegungen im Voraus berechnen.**
- **Vorausschauende Assistenten lassen sich mit geringem technischem Aufwand realisieren und könnten ein Auto autonom steuern; sie ermöglichen Robotern aber auch die Bewegung in einer komplexen Umwelt.**

GLOSSAR

Classifier: Eine Software, die Objekte in Verkehrssituationen erkennt. Der Classifier entscheidet anhand von Wahrscheinlichkeitswerten, sogenannten Scores, ob es sich bei einer bestimmten Pixelansammlung im Bild einer realen Szene um das Objekt handelt, auf das er mit Bildern – etwa von Autos oder Fußgängern – trainiert wurde. Für jede Objektklasse sind spezielle Classifier nötig.

Detektor: Ein Programm, das die Ergebnisse der Classifier in einer Score-Map zusammenführt. Es verzeichnet für jedes Pixel eines Bildes, wie hoch die Wahrscheinlichkeit ist, dass es zu einem bestimmten Objekt gehört.

Tracklet: Abfolge von etwa fünf Bildern einer Videosequenz, die für die Auswertung einer Szene zusammengefasst werden. Da sich von Tracklet zu Tracklet größere Sprünge in einer Bewegung ergeben, erkennt der Detektor vor allem teilweise verdeckte bewegte Objekte zuverlässiger.

Grafik: MPI für Informatik