Autos gehen **die Augen auf**

Einen Wagen mit Chauffeur könnte es irgendwann für jeden geben, wenn nämlich ein Roboter das Steuer übernimmt. Damit Autos auch ohne großen technischen Aufwand autonom fahren können, müssen Computer unübersichtliche Verkehrssituationen jedoch mindestens genauso gut beurteilen wie der Mensch. Dafür entwickeln **Andreas Geiger** und seine Mitarbeiter am **Max-Planck-Institut für Intelligente Systeme** in Tübingen die nötige Software.

Foto: Wolfram Scheible

TEXT **CHRISTIAN J. MEIER**

Die Technik hat ihre Augen heute fast überall. Webcams gibt es für ein paar Euro; Smartphones enthalten oft mehrere Kameras, und in vielen Oberklassewagen erfassen Stereokameras Szenen räumlich, ähnlich wie Menschen. Immer billigere Bildsensoren werden so immer allgegenwärtiger im Alltag, und immer mehr Situationen des Lebens werden auf Bild oder Video gebannt. Sekündlich landet neues Videomaterial von insgesamt 48 Stunden Dauer bei Youtube. Instagram, ein Onlinedienst zum Teilen von Fotos, zählt täglich 20 Millionen neue Bilder.

Vielen Menschen öffnen die allgegenwärtigen Kameras neue Fenster in die Welt. Für Andreas Geiger vom Max-Planck-Institut für Intelligente Systeme in Tübingen bedeuten sie aber noch mehr: Er betrachtet Kameras als die Augen von Computern. Als einen ihrer wichtigsten Sinne, um die Welt zu erkennen und zu verstehen.

„Wahrnehmung ist ein essenzieller Teil von Intelligenz“, sagt der Informatiker und verdeutlicht dies an einem Beispiel: „Wir Menschen geben Dingen oft auffallende Farben und Formen, zum Beispiel Verkehrsschildern, um uns in unserer Welt zurechtzufinden.“ Weil Computer sich in der Welt der Menschen zukünftig immer besser orientieren und, etwa als Haushaltsroboter oder selbst fahrende Autos, autonom bewegen sollen, müssen sie wie der Mensch zunächst lernen, ihre Umgebung wahrzunehmen.

Doch es gibt ein Problem. Computer verstehen Bilder nicht, für sie handelt es sich dabei um ein chaotisches Mosaik von Millionen verschiedenfarbiger Pixel

Objekte erkannt: Eine Art Weltwissen hilft einer Software zum einen, Personen und Autos zu identifizieren, auch wenn diese teilweise verborgen sind. Zum anderen ermöglicht es, das Verhalten von Verkehrsteilnehmern vorherzusagen.

Mit Stereobildern zum Modell: Um Entfernungen zu schätzen, sucht eine Software die beiden korrespondierenden Punkte auf zwei Bildern, die aus unterschiedlichen Blickwinkeln aufgenommen wurden, und rekonstruiert eine Szene auf diese Weise mit Tiefeninformation. Für die weißen Stellen gibt es keine Bildinformation, weil sie für die Kamera verdeckt sind.

und nicht um eine Szene mit Häusern, Bäumen, Autos oder Bordsteinen. Im Gegensatz dazu erkennen Menschen Objekte, können komplexe Situationen erfassen, Bewegungen vorausahnen und Entfernungen abschätzen. „Davon sind Computer noch weit entfernt", sagt Geiger. „Viele Schätze, die in der Bilderflut schlummern, bleiben Computern bislang verborgen", meint der Informatiker.

Um etwa ein Auto ohne Hilfe des Fahrers durch den Stadtverkehr zu lotsen, müssten Computer beurteilen können, ob der Vordermann im nächsten Moment abbiegt oder nicht oder ob ein Kind am Straßenrand auf die Fahrbahn rennt oder nicht. „Daher entwickeln wir Systeme, die wie ein Mensch wahrnehmen und entsprechend reagieren können", sagt Geiger.

Dinge zu erkennen und Szenen zu interpretieren müssen Computer erst mühsam lernen. „Sie müssen das eingefangene Licht in Bedeutung umwandeln", wie Andreas Geiger es ausdrückt. Zu diesem Zweck muss eine Software zunächst die dreidimensionale Welt rekonstruieren, die auf Bildern in nur zwei Dimensionen eingefangen wurde. Für Aufgaben wie diese entwickeln Andreas Geiger und seine vierköpfige Forschergruppe die nötige Software.

Nun lassen sich Objekte wie Autos, Tische, aber auch der menschliche Körper mitsamt seinen komplexen Bewegungen heute schon in der Sprache der Computer darstellen. So existieren im virtuellen Raum dreidimensionale Modelle von Menschen, Monstern oder Formel-1-Rennwagen. In Computerspielen treffen solche Modelle aufeinander, bekämpfen sich, rennen gegeneinander, sprich: Der Computer simuliert hochkomplexe Szenen in einer räumlichen virtuellen Realität.

MEHRDEUTIGKEITEN IN ZWEIDIMENSIONALEN BILDERN

Der Spieler nimmt das aber nicht wahr. Er sieht nur zweidimensionale Bilder. In jedem Moment projiziert die Grafikkarte die komplizierte dreidimensionale Modellwelt des Spiels auf den flachen Bildschirm. „Das räumliche Modell einer Welt in ein zweidimensionales Bild umzurechnen funktioniert bereits erstaunlich gut", stellt Geiger fest. Die Aufgabe bestehe nun darin, den umgekehrten Prozess zu ermöglichen: aus zweidimensionalen Kamerabildern ein Modell der dreidimensionalen Realität zu berechnen.

„Dabei haben wir das Problem, dass sich Mehrdeutigkeiten ergeben", sagt Geiger. Ein Bild, auf dem ein dicker Baumstamm zu sehen ist, kann ein Computer auf verschiedene Weisen erklären. Bei dem dicken Stamm könnte es sich in Wirklichkeit um einen dünnen Stamm handeln, welcher näher am Betrachter steht. Zwei verschiedene 3-D-Modelle – eines mit einem entfernten dicken Stamm und eines mit einem nahen dünneren Stamm – würden ein ähnliches Bild in der Kamera erzeugen.

Weil einem zweidimensionalen Bild die Tiefe fehlt, lässt sich zwischen den beiden Alternativen nicht sicher unterscheiden. Daher verwenden Computer wie wir Menschen Stereobilder, um Entfernungen abzuschätzen und die räumliche Struktur einer Szene zu erkennen. Doch auch dabei können Mehrdeutigkeiten auftreten. Das verdeutlicht Geiger anhand zweier Bilder einer von Altbauten gesäumten Wohnstraße, an deren beiden Seiten Autos parken. Die Aufnahmen zeigen dieselbe Szene aus leicht unterschiedlichen Blickwinkeln, ähnlich wie die beiden Augen eines Menschen sie sehen. Dessen Gehirn erzeugt aus zwei Blickwinkeln einen räumlichen Eindruck.

Grafik: Andreas Geiger/MPI für intelligente Systeme

Eine Software kann auf ähnliche Weise Entfernungen schätzen, indem sie misst, wie weit ein Merkmal, etwa ein Fensterrahmen, auf der einen Aufnahme verschoben scheint, verglichen mit der anderen. Ist die Verschiebung im Bild groß, liegt das Objekt nah an der Kamera. Ist das Merkmal nur wenig verrückt, entspricht dies einem großen Abstand zum Objekt. Ähnliches kann man selbst beobachten, wenn man sich einen nahen Gegenstand ansieht und dabei abwechselnd das linke und das rechte Auge zukneift. Der Gegenstand wird vor dem Hintergrund hin- und herrücken. Diese Verschiebungsinformation rechnet der Computer um in den tatsächlichen Entfernungswert, angegeben in Metern.

Dazu vergleicht der Computer die einzelnen Pixel auf den beiden Bildern. Er sucht für jedes Pixel des ersten Bildes das Pendant im zweiten –also jenes Pixel, das dem gleichen Punkt in der realen Szene entspricht. Zu diesem Zweck analysiert er die Farbwerte der Pixel.

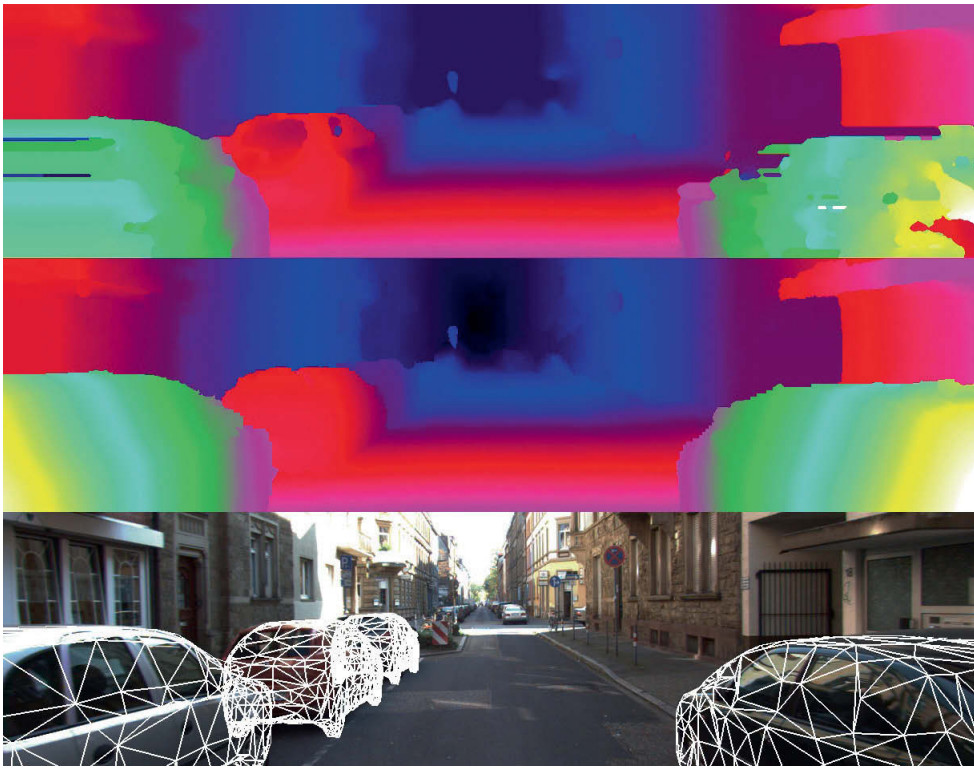
„Kanten wie ein Fensterrahmen lassen sich auf diese Weise leicht orten", sagt Geiger. Denn sie zeigen einen abrupten Übergang von einer Farbe zur anderen, der sich auf dem zweiten Bild leicht wiedererkennen lässt. Der Lack an der Autotür hingegen ist meist einfarbig, alle Pixel besitzen einen ähnlichen Farbwert. Dann gibt es für jedes Pixel in dem einen Bild sehr viele Kandidaten im zweiten Bild, die als Partner infrage kommen. Vor diesen Mehrdeutigkeiten kapitulieren existierende Verfahren zur Berechnung des Tiefenbildes. Im schlimmsten Fall kommt es zu Fehlschätzungen der Tiefe, was in einem System, das für die Sicherheit relevant ist, fatale Folgen haben kann.

Geiger veranschaulicht das Problem mit dem Bild einer Szene, in der die Tiefe durch Falschfarben dargestellt wird. Vorne dominiert Grün, weiter hinten Violett und Rot, während alles, was weit weg ist, blau erscheint. Die Umrisse der Autos sind auf dieser Tiefenkarte zwar zu erkennen – aber viele bunte Flecken zeigen sich dort, wo sich die Autotüren befinden. „Hier konnte der Rechner die Entfernung nicht schätzen, oder es kam zu Fehlmessungen", sagt Geiger.

OBJEKTWISSEN HILFT, ENTFERNUNGEN ZU SCHÄTZEN

Damit der Computer es dennoch schafft, Entfernungen zuverlässig zu schätzen, geben die Tübinger Forscher ihrer Bilderklärungssoftware Wissen über das dargestellte Bild mit, sogenanntes Objektwissen. Sie machen also aus einer Ansammlung von Bildpixeln eine Szene mit Objekten, wie auch der Mensch sie wahrnimmt. Es gibt lernfähige Software, die anhand von vielen Beispielbildern Autos als solche erkennt und zuverlässig in neuen Bildern die Stellen markiert, an denen sich Autos befinden. Der Computer erfährt somit, wo im Bild Autos sind und wo nicht.

Geiger nennt das Objektwissen Mid-Level-Wissen, also etwa „Wissen mittlerer Abstraktionsstufe". Denn es hilft, die Szene, aufbauend auf pixelbasierten Low-Level-Merkmalen wie etwa dem erwähnten Fensterrahmen, in verschiedene Dinge aufzuteilen, ähnlich wie ein Mensch in einer Wohnung Tische, Stühle und Schränke erkennt. >



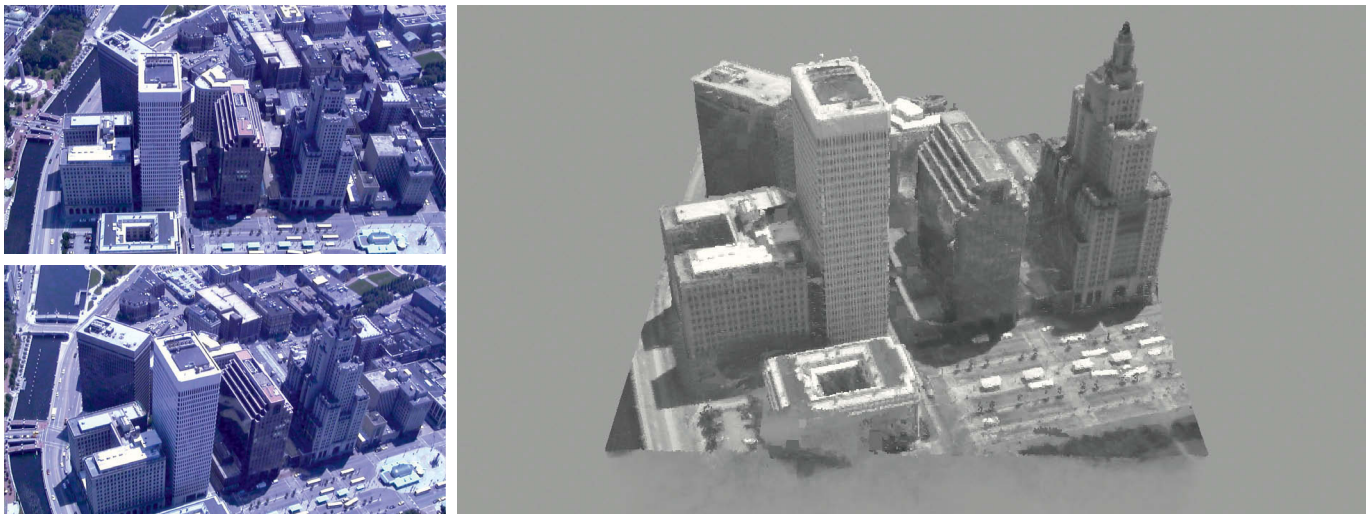
Oben: In einer Tiefenkarte sind Entfernungen durch verschiedene Farben codiert (Gelb – nah; Blau – entfernt).

Unten: Bei der Einschätzung von Distanzen hilft der Software das Wissen über die Geometrie von Objekten wie etwa Autos, von denen in der Software Modelle hinterlegt sind.

Grafik: Andreas Geiger/MPI für intelligente Systeme

Rechte Seite | Auf Wahrscheinlichkeiten gebaut: Osman Ulusoy, Joël Janai und Andreas Geiger (von links) diskutieren den Algorithmus, mit dem sie aus Stereobildern 3-D-Modelle rekonstruieren. Das Bild im Hintergrund zeigt ihnen, wie sicher der Algorithmus Tiefeninformationen für das Capitol in Providence einschätzt. Bei weißen Bildpunkten ist die Schätzung ziemlich sicher, bei schwarzen Punkten nicht. Im zweiten Fall nutzt der Algorithmus mehr Vorwissen, etwa über die generelle Form von Gebäuden.

Unten | Downtown Providence steht Modell: Aus in unterschiedlichem Winkel aufgenommenen Luftbildern (links) berechnet Osman Ulusoy eine 3-D-Rekonstruktion seiner Heimatstadt im US-amerikanischen Rhode Island. Diese ermöglicht es dann unter anderem auch, Ansichten der Innenstadt aus anderen Perspektiven als auf den Ausgangsbildern zu erzeugen (rechts).



Die Software des Teams nutzt nun 3-D-Geometriemodelle von Autos, um die Szene virtuell nachzustellen. Es entsteht eine 3-D-Simulation mit hintereinanderstehenden virtuellen Autos. Mithilfe moderner Grafikkarten lassen sich solche Szenen in perfekte Tiefenkarten umrechnen. Diese enthalten dann keine Lücken an den Autotüren, da sie auf kompletten 3-D-Modellen basieren.

Ganz eindeutig ist die Sache allerdings noch nicht. Die Fotos lassen nicht klar erkennen, wie viele Autos an den Straßenrändern stehen und wie die Fahrzeuge orientiert sind: ob sie parallel zur Bordsteinkante stehen oder nicht. Es gibt somit Tausende von Simulationen mit unterschiedlich vielen Autos und Ausrichtungen der Wagen, die das Foto der Straßenszene mehr oder weniger gut reproduzieren.

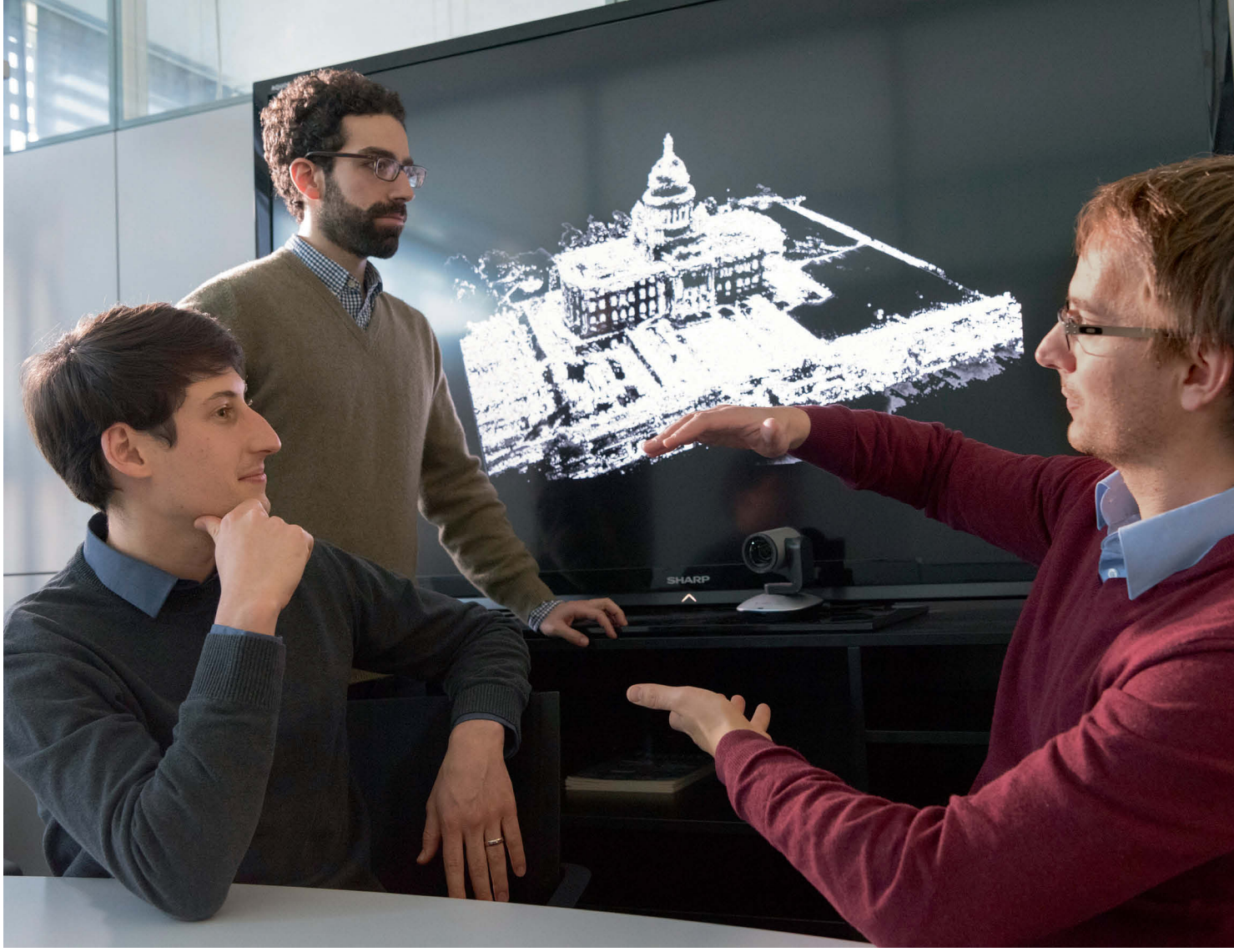
All diese Varianten testet das Programm der Tübinger auf ihre Übereinstimmung mit den aufgenommenen Bilddaten. So vergleicht es beispielsweise die Tiefenkarte, die sich aus der Simulation ergibt, mit der ausschließlich anhand des Pixelvergleichs, also ohne Weltwissen, erstellten Tiefenkarte. Zudem misst die Software, wie gut das künstliche Bild die Bereiche reproduziert, in denen sich im realen Bild Fahrzeuge befinden. „Auf diese Weise wird die wahrscheinlichste Hypothese herausgefiltert“, sagt Geiger. Die Methode liefert somit zwar keine letzte Gewissheit, aber eine konsistentere und sinnvollere Interpretation des Bildes.

AUS LUFTBILDERN ENTSTEHT DAS 3-D-MODELL EINER STADT

Ähnliches demonstriert Geigers Mitarbeiter Osman Ulusoy anhand von Luftbildern seiner Heimatstadt Providence im US-Bundesstaat Rhode Island. „Aus Luftbildern mit unterschiedlichen Blickwinkeln lässt sich ein 3-D-Modell der Innenstadt erstellen“, sagt Ulusoy. Doch spiegelnde Fassaden etwa kann ein Computer schwer rekonstruieren, da Reflexionen ihn bei der Schätzung der Entfernung in die Irre führen.

„Wir geben dem Computer sogenanntes A-priori-Wissen, um die Lücken zu schließen“, sagt Osman Ulusoy. Gemeint ist damit eine Art Weltwissen darüber, wie die Dinge im Allgemeinen beschaffen sind. Spiegelfassaden sind demnach in der Regel glatt. Die Software kann so das Stadtmodell trotz mehrdeutiger Beobachtungen vervollständigen. „Interessant könnte das für Stadtplaner sein“, sagt der Informatiker. „Man könnte die Entwicklung der Stadt in drei Dimensionen dokumentieren.“

Auch Innenraumszenen lassen sich virtuell nachbauen, wie Andreas Geiger anhand des Bildes eines Zimmers mit Bett, Stuhl und Schrank zeigt. „Das Modell kennt Formen und Größen typischer Einrichtungsgegenstände“, er-



klärt Geiger. Es erkenne einen Stuhl auch dann, wenn auf dem Bild nur die Stuhllehne von der Seite zu sehen sei. Auch hier stecken die Forscher A-priori-Wissen in die virtuelle Nachstellung der Szene. „Schränke, Betten oder Sofas stehen in der Regel an der Wand“, erklärt Geiger. Zudem durchdringen die Gegenstände sich nicht gegenseitig. Dieses Wissen beschränkt, ähnlich wie bei der Szene mit den parkenden Autos, die Zahl der möglichen Hypothesen auf ein Maß, das der Computer in kürzerer Zeit durchtesten kann.

Von Nutzen können virtuelle Rekonstruktionen von Innenräumen für Roboter sein, die in einem Haushalt sicher manövrieren sollen. Sie könnten aber auch Architekten und Designern helfen, meint Geiger, um etwa realitätsnähere Entwürfe zu erstellen oder ergonomische Designs zu entwickeln.

Indem der Computer Wissen über Objekte nutzt, lernt er also, das Gesehene zu erkennen. „Dabei ist es wichtig, dass man das Problem als Ganzes betrachtet und nicht nur seine einzelnen Bestandteile“, sagt Geiger.

HIGH-LEVEL-WISSEN FÜR DIE INTERPRETATION DER BILDER

Die Objekte auf einem Bild miteinander in Beziehung zu bringen gelingt den Tübingern, indem sie dem Rechner sogenanntes High-Level-Wissen geben, also Wissen hoher Abstraktionsstufe. Dazu gehört die erwähnte Annahme, dass Möbel einander nicht durchdringen oder an der Wand stehen.

Erst das High-Level-Wissen ermöglicht es dem Computer, nicht nur statische, sondern auch bewegte Bilder sinnvoll zu interpretieren. Geiger spricht hier von „3-D-Szenenfluss“, was für die Schätzung der dreidimensionalen Bewegung aller Objekte in der Szene steht. Sein Team versucht zum Beispiel, das Beste aus der etwas ungünstigen Perspektive herauszuholen, die ins Auto eingebaute Kameras auf Verkehrsszenen haben, etwa an einer innerstädtischen Kreuzung zweier viel befahrener Straßen.

Um eine solche Situation zu verstehen, wäre eine starre Vogelperspektive ideal. Denn darauf würden sich nur die Fahrzeuge bewegen, und es wäre zugleich ersichtlich, auf welchen Spuren sie das tun, welche Ampeln es an der Kreuzung gibt und wie sich die Ampelphasen abwechseln. „Aus 1,60 Meter Höhe, in der die Stereokameras typischerweise am Auto angebracht sind, ist die Ableitung solchen Wissens deutlich schwieriger und mit größeren Unsicherheiten behaftet“, sagt Geiger. Oft sehe die starr eingebaute Kamera nicht einmal, ob eine Ampel für das eigene Fahrzeug gerade Rot oder Grün zeigt.

Die Tübinger Forscher wollen Autos trotz solch unvollständiger und unsicherer Informationen autonom ma-

Foto: Wolfram Scheible

Zwei Teile, ein Mensch: Andreas Geiger führt selbst vor, welche Szenen Computer nicht auf Anhieb verstehen. Sie wissen nämlich nicht, dass auf dem Bild nur ein Forscher zu sehen ist und nicht zwei. Diesen Schluss zu ziehen, bringt Geigers Team einer Software bei.

chen – durch mehr Intelligenz des Bordcomputers: indem dieser lernt, den Szenenfluss richtig zu erkennen und zu interpretieren.

WENIGER MODELLE DANK DER STARRHEIT VON OBJEKTEN

Erstes Problem: die anderen Verkehrsteilnehmer auszumachen. Für den Computer handelt es sich bei der Straßenszene zunächst einmal um einen Schwarm sich bewegender Pixel. Wir Menschen hingegen wissen, dass viele Szenen, die wir beobachten, insbesondere auch im Verkehr, aus einigen wenigen starren Objekten bestehen. Autos nehmen nicht plötzlich eine andere Form an, sondern bewegen sich als ein kompaktes Ganzes.

Außerdem gibt es selbst auf einer viel befahrenen Kreuzung nicht Hunderte von Fahrzeugen, sondern in jedem Moment nur einige wenige. „Wir sagen dem Computer: Zerlege die Szene in möglichst wenige starre Einzelteile“, erklärt Geiger. Starre Gegenstände haben weniger Freiheit, sich zu bewegen, als etwa ein menschlicher Körper: Sie können sich entlang dreier Richtungen fortbewegen: vor und zurück, nach links und rechts sowie nach oben und unten. Außerdem können sie sich um drei Achsen drehen, während die komplexe Bewegung eines Körpers mit Hunderten Variablen beschrieben wird, zum Beispiel mit den Drehwinkeln aller Gelenke.

„Die Annahme der Starrheit schränkt das Modell der Szene daher stark ein“, erklärt Geiger. Der Computer muss weniger Varianten auf ihre Plausibilität testen und kann Mehrdeutigkeiten besser auflösen. Zudem schließt das Gebot, möglichst wenige Objekte zu identifizieren, viele weitere Hypothesen aus, etwa dass ein Auto, das durch einen Laternenmast zweigeteilt erscheint, als zwei Objekte fehlinterpretiert wird. Die Starrheit ist somit ein einfaches Kriterium mit großer Wirkung.

Nachdem Geigers Software die einzelnen Fahrzeuge auf einer Kreuzung ausgemacht hat, verfolgt sie diese für eine gewisse Zeit. Fahren sie geradeaus? Biegen sie ab? Dabei hilft eine Technik namens maschinelles Lernen. Anhand von vielen Beispielbildern lernen Computer, bestimmte Bildelemente zu erkennen. Wird ein Rechner etwa mit Tausenden Abbildungen von Gesichtern trainiert, kann er schließlich selbstständig Gesichter auf neuen Fotos erkennen.

KAMERAS UND INTELLIGENZ ERSETZEN TEURE TECHNIK

Das Tübinger Programm lernt auf ähnliche Weise, unter anderem aus der Gesamtheit des Verkehrsflusses und anhand der Fahrbahnmarkierungen, auf die Geradeaus- und Abbiegespuren zu schließen und wie die Ampeln angeordnet sein müssen. „Es gibt verschiedene Typen von Ampelkonfigurationen, die mit einer bestimmten Abfolge der Ampelphasen verbunden sind“, erklärt Geiger. „Bei uns lernen Computer diese Abfolgen, basierend auf großen Mengen von Messdaten, und nutzen sie, um

Foto: Wolfram Scheible

Verkehrsteilnehmer besser miteinander in Bezug setzen zu können."

Auch die Umgebung der Kreuzung wird untersucht: Wo stehen Gebäude, wie sind die Straßen orientiert? Mit all dieser Information rekonstruiert der Computer eine digitale Karte der Kreuzung und lässt einen virtuellen 3-D-Film ablaufen, der die von den Kameras eingefangene Szenerie auf das Wesentliche reduziert. Darauf aufbauend, kann das autonome System die richtigen Entscheidungen ableiten. Und das macht es ad hoc für jede neue Kreuzung, auf die ein Fahrzeug zusteuert.

„Wenn autonome Fahrzeuge Kameras und Intelligenz kombinieren würden, kämen sie ohne die teure Technik aus, die heutige Prototypen mit sich führen, etwa Laserscanner oder Radar", meint Geiger. Auch hochpräzise Satellitennavigation und aufwendig erstellte digitale Karten, auf denen aktuelle Systeme basieren, seien nicht nötig. Für eine Übergangszeit, in der es nur wenige selbstständig fahrende Autos auf den Straßen gebe, sei auch nicht mit intelligenter Infrastruktur zu rechnen, die autonome Pkws unterstützt.

Mit der Software, die komplexe Szenen analysiert, gibt es derzeit allerdings noch ein Problem: Sie macht noch relativ viele Fehler. Ein Sofa hält sie fälschlicherweise für ein Bett, oder einen Flügel erkennt sie als Tisch. Bei Szenen von Kreuzungen patzt die Software unter anderem, weil sich das maschinelle Lernen hier schwieriger gestaltet als etwa bei der Gesichtserkennung. Für das Training braucht sie sehr viele Daten, doch es gibt deutlich weniger Bildsequenzen mit Autos als Fotos von Gesichtern. Darüber hinaus müssen die Trainingsdaten von Menschen mit Information versehen werden, sie zeigen dem Rechner zum Beispiel, wo auf den Bildern Gesichter sind. „Solche Annotationen sind bei Kreuzungsszenen sehr aufwendig", sagt Andreas Geiger.

Die Tücken der Digitalfotografie bedeuten für die Tübinger Forscher eine weitere Hürde. Die Sonne etwa kann die Bildsensoren blenden, Bäume können die Szene verstellen, oder große Unterschiede zwischen Hell und Dunkel machen es unmöglich, das Geschehen fotografisch zu erfassen. In solchen Fällen leidet die Genauigkeit der virtuellen Rekonstruktion, oder sie wird ganz unmöglich.

DIE AKZEPTANZ FÜR DIE TECHNIK WIRD KOMMEN

Auch dieser technischen Schwierigkeit wollen die Forscher mit A-priori-Wissen begegnen. „Bei Häusern in einer Siedlung kann man davon ausgehen, dass sie einander ähneln", erklärt Geiger. Die Annahme der Ähnlichkeit hilft dabei, eine ganze Wohnstraße virtuell zu rekonstruieren, auch wenn entlang der Straße viele Bäume stehen oder die Kamera häufig in die Sonne blickt.

Man kann sich das in etwa so vorstellen: Von einem Haus zeichnet das System die Vorderfront auf, vom anderen die linke Außenwand und von einem dritten die rechte. Weil die Häuser als ähnlich angenommen werden, lässt sich aus den drei Puzzleteilen ein typisches Haus dieser Straße zusammenfügen. „Das Modell ist so flexibel, dass es Geometrien extra- und interpolieren kann", sagt Geiger. Das heißt, es kann Häuser generieren, die nie beobachtet wurden, aber von ihrem Erscheinungsbild perfekt in die Siedlung passen.

Doch auch wenn die Software immer besser wird, Milliarden von Pixeln in Bedeutung zu verwandeln, wird es sich bei dem, was die Computer in Bildern erkennen, immer um Schätzungen handeln. Und selbst die wahrscheinlichste Hypothese ist nur eine Hypothese und keine Gewissheit. Aber ist im Verkehr nicht genau das nötig: Gewissheit?

„Auch ein guter Autofahrer kann nur einschätzen, wie sich der Vordermann verhalten wird", entgegnet Geiger. Allerdings sei der Computer darin noch nicht so gut wie ein Autofahrer, räumt er ein. „Die Akzeptanz für eine solche Technik wird kommen, sobald die Systeme deutlich weniger Fehler machen als ein Mensch." ◂

AUF DEN PUNKT GEBRACHT

- **Für Computer bestehen Bilder zunächst einmal nur aus bedeutungslosen Pixeln. Andreas Geiger und sein Team am Max-Planck-Institut für Intelligente Systeme bringen ihnen daher bei, Bilder vor allem von komplexen Verkehrssituationen zu verstehen und das Verhalten der Verkehrsteilnehmer zu antizipieren.**
- **Wenn eine Software aus zweidimensionalen Bildern ein dreidimensionales Modell einer Straßenszene berechnet, ergeben sich Mehrdeutigkeiten etwa bei der Abschätzung von Entfernungen. Deshalb stellen die Forscher den Programmen Wissen mittlerer Abstraktionsstufe zur Verfügung. Dieses hilft Computern etwa, einzelne Objekte wie Autos zu erkennen.**
- **Um die einzelnen Objekte in einem Bild miteinander in Beziehung zu setzen, nutzt die Software Wissen hoher Abstraktionsstufe. Demnach können sich Gegenstände zum Beispiel nicht gegenseitig durchdringen.**
- **Wenn Computer mithilfe des maschinellen Lernens viele Verkehrssituationen analysiert haben, können sie den Verkehrsfluss etwa an Kreuzungen vorhersagen.**