

Links

<http://cvpr.in.tum.de>
www.youtube.com/user/cvprtum

Computervisionen

Am Lehrstuhl für Bildverarbeitung und Mustererkennung der Fakultät für Informatik lernen Maschinen, das menschliche Sehsystem nachzuahmen. Im komplexen Zusammenspiel aus Wahrnehmung und Interpretation gelingt es ihnen immer besser, Bilder zu „verstehen“. Es tun sich verblüffende Möglichkeiten auf. Computer können uns schon heute helfen, in neue Dimensionen vorzustoßen

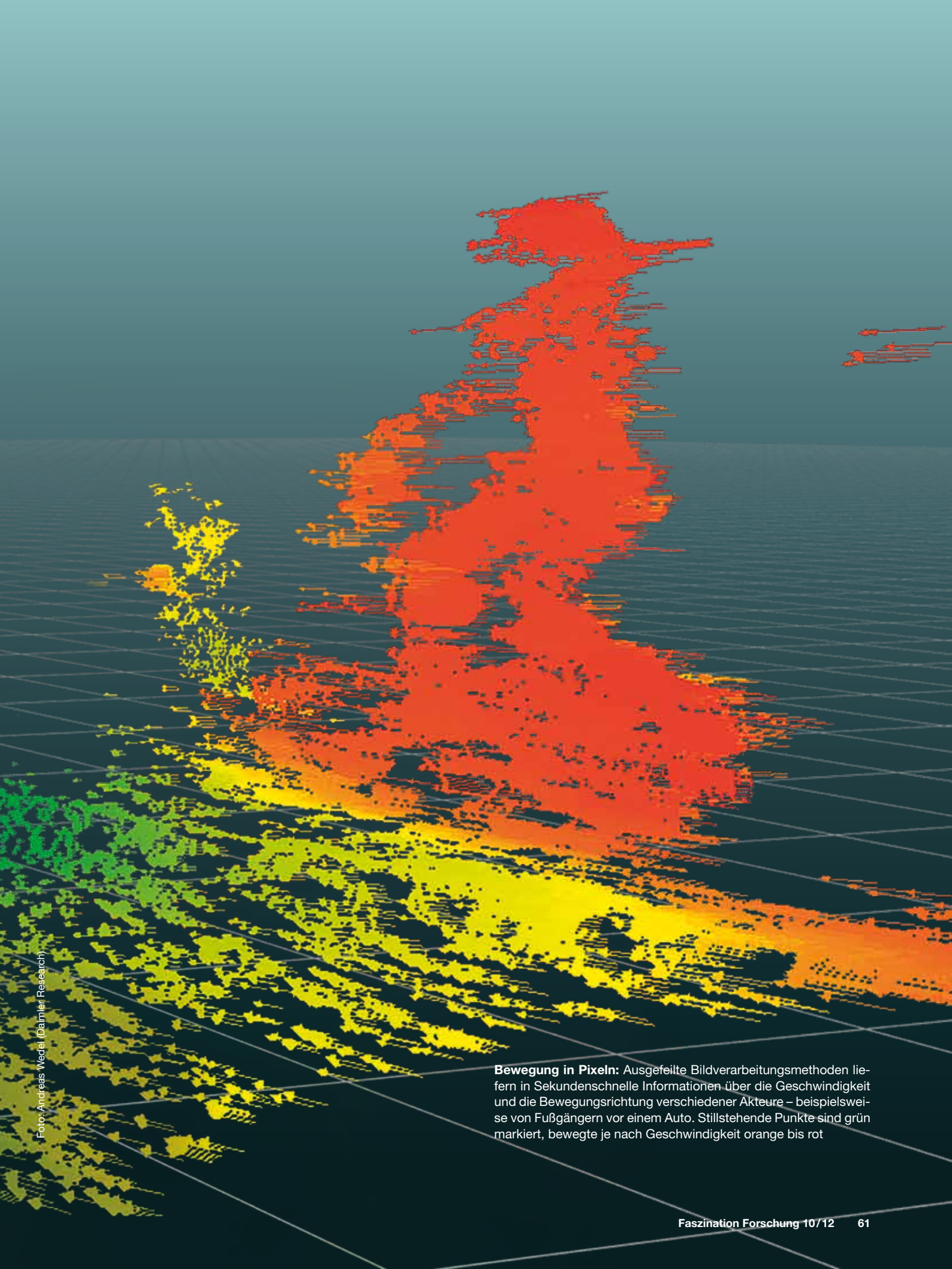


Foto: Andreas Wedel (Damier Research)

Bewegung in Pixeln: Ausgefeilte Bildverarbeitungsmethoden liefern in Sekundenschnelle Informationen über die Geschwindigkeit und die Bewegungsrichtung verschiedener Akteure – beispielsweise von Fußgängern vor einem Auto. Stillstehende Punkte sind grün markiert, bewegte je nach Geschwindigkeit orange bis rot

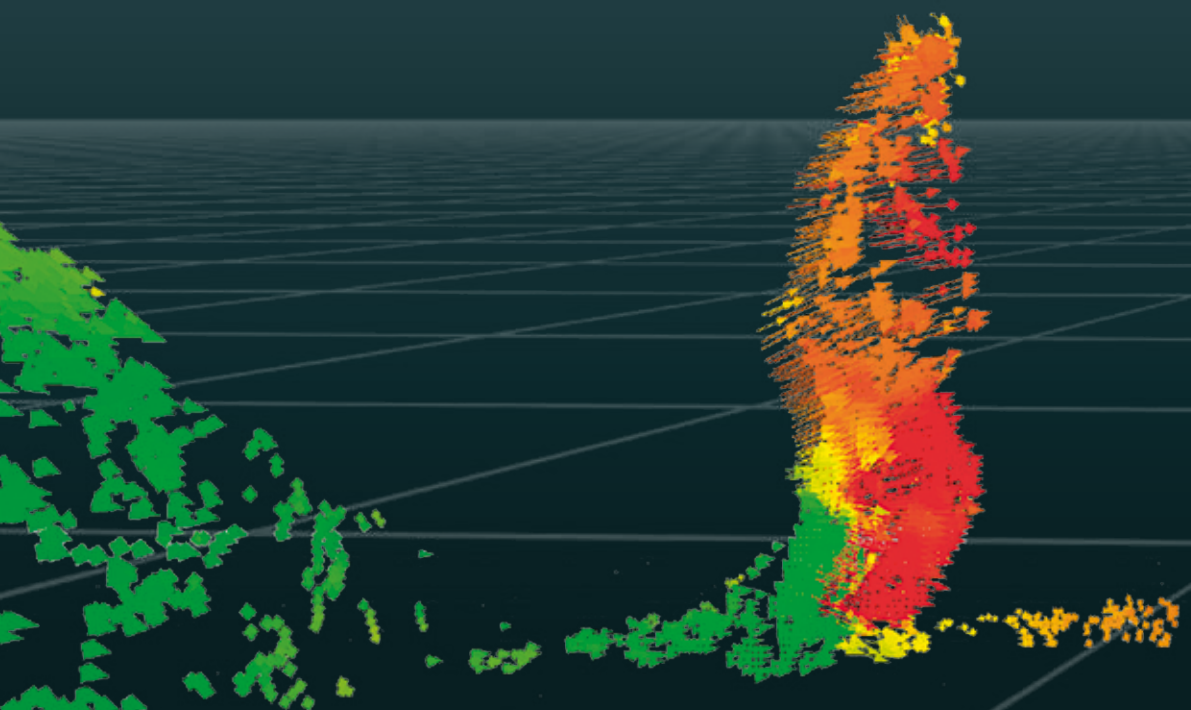
Maschinen, die Sinnesleistungen imitieren oder verbessern können, sind seit geraumer Zeit im Fokus der Wissenschaft. Unsere Computer werden immer schneller und leistungsfähiger. Doch trotz Terabytes an Speicherplatz und Gigahertz an Taktfrequenz können sie sich bis heute kaum mit biologischen Systemen messen. Scheinbar mühelos arbeitet zum Beispiel der menschliche Sehapparat. Aus einer Ansammlung farbiger Lichtquanten auf der Netzhaut stellt er alle Informationen bereit, die man zum Lesen dieses Artikels oder zur Orientierung in einem Raum braucht. Prof. Daniel Cremers, Inhaber des Lehrstuhls für Bildverarbeitung und Mustererkennung an der TUM, weiß, welche Leistung dahinter steckt. Er sagt: „Der Mensch verwendet 50 Prozent

seiner Gehirnkapazität auf die Bildverarbeitung. Daran sieht man, dass es uns nur scheinbar leichtfällt, unsere Umgebung zu verstehen und uns darin zu orientieren. So etwas auch nur in Teilen maschinell nachzumachen, ist aufwendig und kompliziert.“ Mit seiner Forschergruppe hat Cremers diese Herausforderung angenommen.

Gefahren schneller erkennen

Ein wichtiges Anwendungsgebiet des Maschinensehens ist die Entwicklung von Hilfen für den Straßenverkehr. Viele Autofahrer haben es schon erlebt: Wie aus dem Nichts erscheint beim Abbiegen plötzlich ein Radfahrer. Oder ein Kind rennt unvermittelt auf die Fahrbahn. Mit Glück gehen solche Situationen meist ohne Unfall >

Zwei Kameras im oder am Auto erstellen ein dreidimensionales Bild der Umgebung. Jeder Punkt im Umfeld des Autos wird mit einer Bewegungsinformation versehen, bewegte Gegenstände werden rot, stationäre grün markiert. So lässt sich zum Beispiel darstellen, wenn plötzlich ein Kind hinter dem Auto auftaucht





Der Assistent kennt keine Schrecksekunde: Kinder, die vor fahrende Autos laufen, sind eine klassische Gefahrensituation. Der Rechner des Assistenzsystems kann in Echtzeit aus den Bildern der Kameras ermitteln, wie schnell und wohin sich das Kind bewegt. In Zukunft könnte ein solches System eigenständig eine Notbremsung auslösen

aus. Reagiert der Fahrer allerdings nicht schnell genug, kommt es zum Zusammenstoß mit teilweise katastrophalen Folgen. Die Ergebnisse eines Projekts, das Forscher des Lehrstuhls Bildverarbeitung kürzlich zusammen mit Daimler bearbeiteten, werden helfen, die Straßen sicherer zu machen. Das Projekt beschäftigte sich mit der Schätzung von Geometrie und Bewegung der beobachteten Strukturen mit dem Ziel, ein effizientes Frühwarnsystem zu ermöglichen. Die Wissenschaftler der TUM gehören in der Berechnung von Bewegungsprognosen seit Jahren zu den führenden Kräften weltweit. Das Gemeinschaftsprojekt mit dem industriellen Partner war für den Deutschen Zukunftspreis 2011 nominiert.

Um vom Auto aus zu erkennen, dass ein Kind in der nächsten Sekunde auf Kollisionskurs geht, muss ein Rechner an Bord des Pkw in Echtzeit sowohl die Position des Kindes als auch dessen Geschwindigkeit und Bewegungsrichtung ermitteln. „6D-Vision“, wie Daimler die Technik nennt, verwendet dazu das Prinzip des Stereosehens. Dabei berechnet das System zunächst anhand der Daten zweier Kameras die Position einzelner Bildpunkte im Raum. Diese werden anschließend von Bild zu Bild verfolgt. Da im nächsten Zeitschritt wieder die 3-D-Position ermittelt wird, lässt sich ein Geschwindigkeitsvektor für jedes Pixel ableiten (engl. scene-flow). 3-D-Position und 3-D-Bewegung ergeben zusammen die sechs Dimensionen des Verfahrens. Die vom Computer berechneten Bewegungsfelder werden farblich codiert. Dabei entspricht jede Geschwindigkeit einem bestimmten Farbton. Ändert sich die Bewegung, erkennt der Algorithmus von Daniel Cremers und seinen Mitarbeitern dies sofort und stellt sie in einer anderen Farbe dar. Er kann so abschätzen, wohin sich ein Objekt bewegen wird. Dieses System ermöglicht Prognosen für potenzielle Verkehrshindernisse in Entfernungen von bis zu 50 Metern. Zum Vergleich: Die menschliche Stereowahrnehmung endet bei zehn bis zwölf Metern.

Die neue sechsdimensionale Sicht benötigt nur ca. 0,2 Sekunden, um drohende Kollisionen im gesamten Sichtfeld der Kameras zu erkennen. So kann sie nicht nur Notbremsysteme unterstützen, sondern auch automatische Ausweichmanöver. Cremers betont: „Die Herausforderung an die Bildverarbeitung ist hier, dass alles sehr

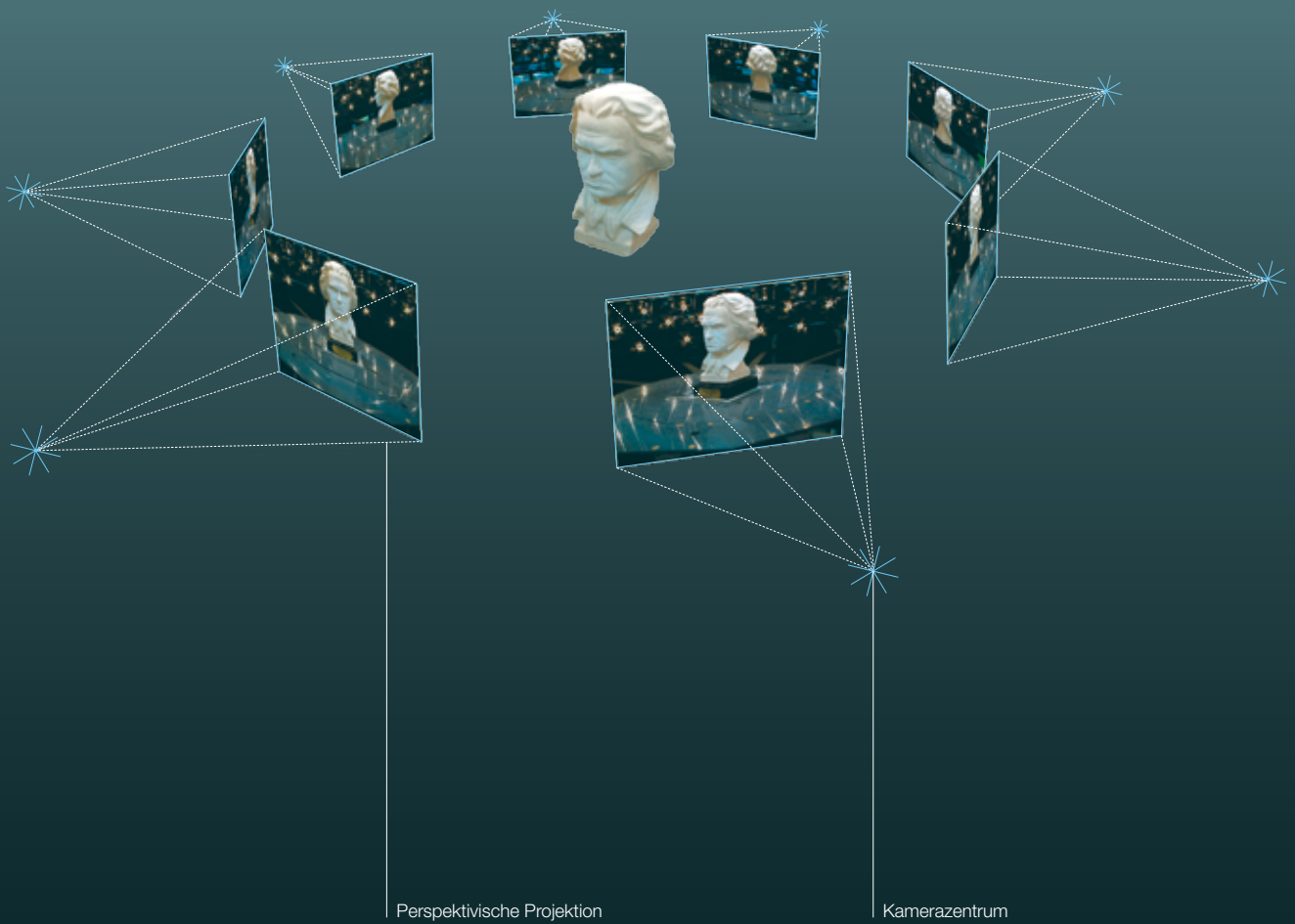


Die 3-D-Rekonstruktion von altrömischen Niobidenstatuen anhand von Fotos soll klären, wie die Figuren im Ensemble ursprünglich zueinander ausgerichtet waren

schnell gehen und auch bei schlechten Sichtverhältnissen und hoher Fahrzeuggeschwindigkeit funktionieren muss. Die Fahrerassistenz erfordert Warnungen in Echtzeit. Es macht keinen Sinn, wenn sie erst eine halbe Stunde später kommen.“

3-D-Rekonstruktionen

Ein weiteres Spezialgebiet des Lehrstuhls Bildverarbeitung ist die Rekonstruktion der dreidimensionalen Welt auf der Grundlage zweidimensionaler Bilder. „Umgekehrt zur grafischen Animation, die die Welt ins Bild setzt, kommen wir von einem Bild auf ein dreidimensionales Modell“, erklärt Cremers. In Kooperation mit dem Akademischen Kunstmuseum der Universität Bonn hat seine Gruppe altrömische Niobidenstatuen anhand von Fotos virtualisiert, um Kunsthistorikern bei der Klärung der Frage zu helfen, wie die Figuren ursprünglich im Ensemble zueinander ausgerichtet waren. Auch hier gehören die Rekonstruktionsverfahren von Cremers und seinen Mitarbeitern nachweislich zu den präzisesten weltweit. Gegenüber Laserscans haben derartige bildbasierte ▶



Visualisierung des 3-D-Rekonstruktionsprozesses für eine Statue: Algorithmen berechnen aus einer Menge von Fotos, wo und in welche Richtung jede Kamera aufgenommen hat und wie die aufgenommene Geometrie aussah



Eines von zwei bei einer 3-D-Rekonstruktion verwendeten Luftbildern der Stadt Graz

Verfahren Vorteile, zum Beispiel dass sie kostengünstig zu realisieren sind und dass sie neben der Geometrie auch die Oberflächenfarbe liefern. Darüber hinaus sind sie auch auf Formen anwendbar, die sich im Zeitverlauf ändern, was die Rekonstruktion ganzer Handlungen in 3-D ermöglicht.

Mit der Rekonstruktion von Geometrie hat auch die Frage zu tun, wie man aus zwei leicht versetzten Luftaufnahmen eine präzise Höhenkarte der beobachteten Welt berechnen kann. Auch darauf haben die jungen Wissenschaftler eine Antwort. Bilder, die mit Algorithmen des Lehrstuhls Bildverarbeitung berechnet wurden, verdeutlichen, dass man auf diesem Wege die Welt bis hin zu den Details der Dachformen und einzelner Laternenpfähle rekonstru- ▶

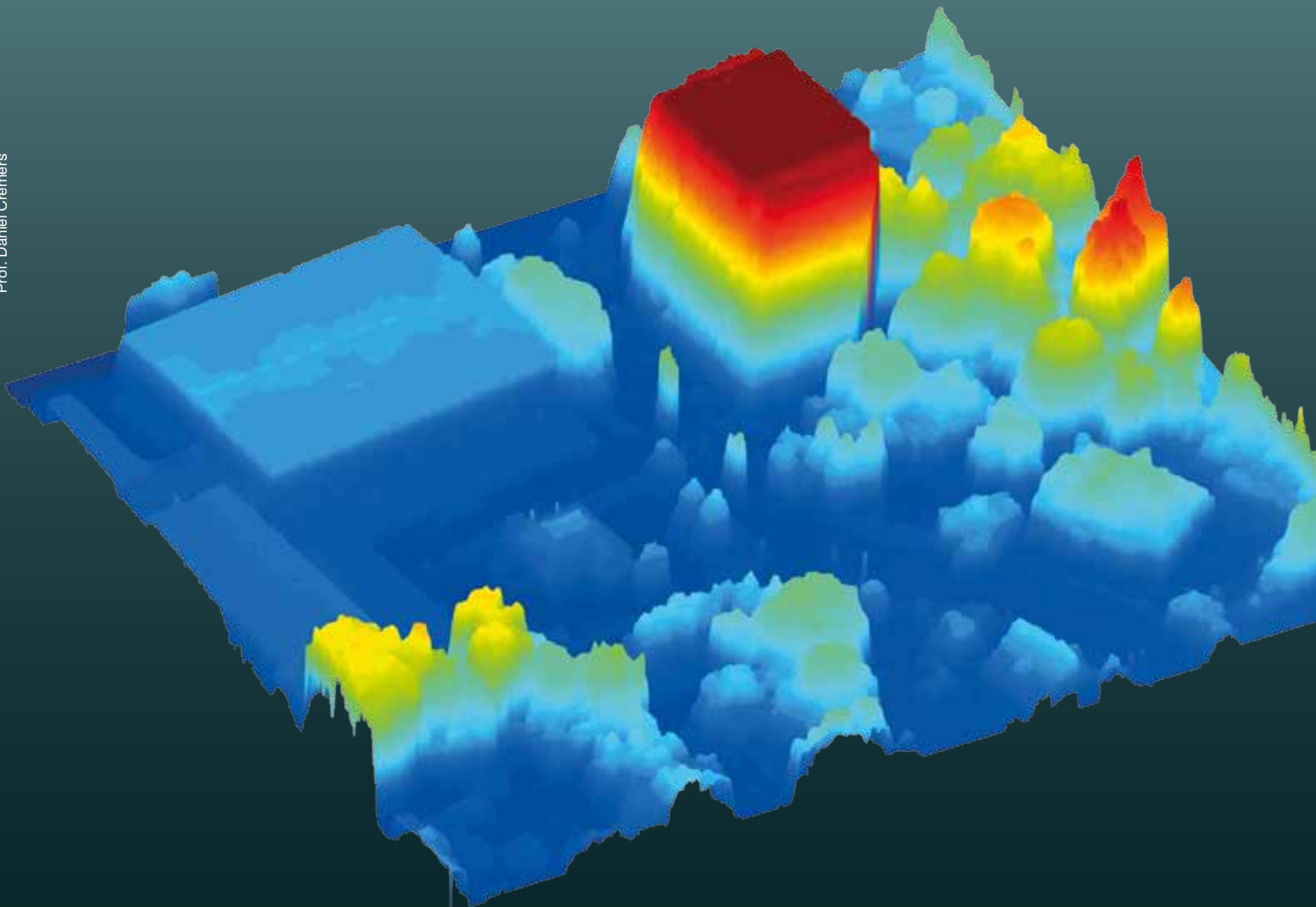
Bild: Dr. Pock, TU Graz

Forscher vieler Fachrichtungen

Daniel Cremers ist seit 2009 Ordinarius für Bildverarbeitung und Mustererkennung an der TUM. Eine Rangliste des Wirtschaftsmagazins „Capital“ zählte ihn im Jahr 2010 zu Deutschlands „Top 40 Forschern unter 40“. In seiner Forschungsgruppe arbeiten derzeit 15 Doktoranden und drei Postdocs. Cremers ist Physiker, seine Mitarbeiter kommen unter anderem aus der Robotik, Mathematik und Informatik. Gemeinsam gehören sie seit Jahren zu den Top-Beiträgern der International Conference on Computer Vision (ICCV), dem renommiertesten Forum ihres Fachgebiets. Den Erfolg der Gruppe erklärt Cremers so: „Unser Rezept besteht darin, dass wir anspruchsvolle mathematische Formalismen entwickeln und auf praxisrelevante Anwendungen zuschneiden.“



Bild: Dr. Thomas Pock (TU Graz) und Prof. Daniel Cremers



Farbcodiertes Höhenprofil rekonstruiert aus Luftbildern (siehe Seite 66). Rote Bereiche sind hoch, blaue Bereiche sind niedrig gelegen

ieren kann. Cremers argumentiert: „Mit derartigen vollautomatischen Verfahren können Projekte wie Google Earth digitale Höhenmodelle der Erde erstellen.“

Die maschinelle Analyse von Bildern macht immer größere Fortschritte und an vielen dieser Entwicklungen ist die TUM beteiligt. Neue Algorithmen erlauben es etwa, die Auflösung unscharfer, beschädigter oder stark gealterter Foto- und Filmaufnahmen zu erhöhen. Wenn sie weiter optimiert werden, könnten die sogenannten Super-Resolution-Verfahren bald in der Filmrestaurierung oder in der Kriminologie, etwa bei der effizienteren Auswertung von Bildmaterial aus Überwachungskameras, breite Anwendung finden. Eine verlockende Aussicht für Filmfans schildert Cremers: „Ich glaube, es dauert nicht mehr allzu lan-

ge, bis wir eine neue Art des 3-D-Fernsehens bekommen. Dann kann der Zuschauer vielleicht in Filmwelten eintauchen und zum Beispiel einmal in „Vom Winde verweht“ wie einst Scarlett O’Hara die Treppe hinunterschreiten.“

Die Zukunft der Bildbearbeitung hat begonnen

Mittlerweile gelingt es bereits, aus einem einzelnen Bild eine dreidimensionale Welt zu formen. Die Wissenschaftler sprechen von Single View Reconstruction. Der Software-Hersteller Microsoft will auf dieser Basis neue Wege der Bildbearbeitung beschreiten und kooperiert bei der Optimierung der Verfahren eng mit den Münchner Pionieren der Bildverarbeitung. Bald soll es möglich sein, am PC aus Bildern einzelne Objekte herauszugreifen, ihre

Das aus einem einzigen Bild einer Teekanne berechnete 3-D-Modell lässt sich vielfach manipulieren. Hier wurden die Reflexionseigenschaften der Oberfläche so geändert, dass das Objekt aussieht, als würde es aus Porzellan oder aus Metall bestehen



Größe zu verändern und sie leicht gedreht in ein anderes Bild einzusetzen. Das am Lehrstuhl Bildverarbeitung entwickelte Verfahren ist denkbar einfach: Der Computer berechnet, wie ein Ballon aussehen würde, der in die Silhouette eines Objekts eingepasst und mit Luft aufgeblasen wird. Der Anwender kann mehr oder weniger Luft in den Ballon pumpen und damit dickere oder dünnere Rekonstruktionen des Objekts erzeugen. „Auch wenn die entstehende Geometrie nicht mit der unbekanntem übereinstimmt“, erklärt Cremers, „so lassen sich dennoch Ansichten des Objekts von neuen Blickwinkeln berechnen, die erstaunlich realistisch aussehen.“

Das von den Forschern der TUM entwickelte Optimierungsverfahren berechnet innerhalb von etwa einer Se-

kunde aus einem einzigen Bild ein plausibles dreidimensionales Modell. Das Objekt, beispielsweise ein Flugzeug, ein Vogel oder eine Teekanne, lässt sich anschließend interaktiv aus völlig neuen Blickwinkeln darstellen. Zudem kann man auf Knopfdruck die Reflexionseigenschaften der Oberfläche derart manipulieren, dass der Eindruck zustande kommt, das Objekt bestehe aus unterschiedlichen Materialien – im Fall der Teekanne etwa Porzellan, Metall oder Aluminium. Mit Sicherheit ein schönes Werkzeug für Designer. Dieses Projekt der Wissenschaftler hat auf der Asian Conference on Computer Vision 2010 einen Preis gewonnen. Es wurde in den USA auf dem diesjährigen Microsoft TechFest am Firmensitz in Redmond, Washington, präsentiert.

Karsten Werth

Beispiel einer Single View Reconstruction: In nur einer Sekunde erstellt ein Rechner aus dem Foto eines auf dem Ast sitzenden Vogels ein manipulierbares 3-D-Modell eines Tieres und generiert daraus weitere räumliche Ansichten

