FOTO: MIKKELWILLIAM/GETTY IMAGES

Mit dem Avatar durch das Metaverse

Der eigene digitale Zwilling

Vielleicht werden wir künftig als 3D-Avatare durch digitale Räume wandern. Forscher entwickeln gerade Algorithmen, die solche Ebenbilder automatisch generieren können – am besten aus Videos von der Smartphone-Kamera

VON DR. THOMAS BRANDSTETTER

Geht es nach den großen Techkonzernen, soll der virtuelle Raum in Zukunft eine Art zweite Heimat für uns werden. In ihrer Vision pendeln wir in Form einer digitalen Person munter zwischen Besprechungsraum, Multiplayer-Game und Kleideranprobe hin und her. Damit das klappt, sind wir in der realen Welt mit VR-Brillen ausgestattet und werden durch 3D-Kameras getrackt, die Gesichtsausdrücke und Körperbewegungen auf einen digitalen Zwilling übertragen. Dieser Avatar kann, muss aber nicht aussehen wie wir selbst. Wer will, kann sein Selbstbild im künftigen Metaverse von Mark Zuckerberg gerne auch aufpolieren – sei es zum Muskelprotz oder zur Diva inklusive der passenden Stimme.

Soweit die Vision, aber wie sieht die Realität aus? Grundsätzlich haben wir alle technischen Voraussetzungen für das Erstellen eines persönlichen Avatars, doch es ist noch ein langer Weg, bis die Technologie einfach und zuverlässig genug sein wird, um beim User anzukommen.

Damit sich ein Gefühl echter, sozialer Interaktion einstellt, sollen die digitalen Figuren so realistisch wie möglich aussehen. In Computerspielen vermitteln gut gemachte Avatare einen solchen Eindruck, doch wurden diese nicht auf Knopfdruck von einem Algorithmus erzeugt, sondern ein menschlicher Designer hat sie mit viel Zeit- und Arbeitsaufwand entworfen.

„Man kann natürlich jederzeit jemanden beauftragen, der für Tausende Euro händisch und basierend auf Fotos den perfekten Avatar von einem selbst generiert“, sagt Professor Andreas Geiger, der an der Universität Tübingen die Gruppe für Autonomes Maschinelles Sehen leitet. „Wir wollen diesen Prozess aber automatisieren“. Dafür soll als Input auch kein einstündiges Video erforderlich sein. Das wäre für die User viel zu umständlich.

Heute nur mit Hightech-Scanner, künftig reicht die Handy-Kamera

Für ihre Studien greifen Geiger und sein Team auf die Bilder eines Hightech-Scanners zurück, der mit über 20 Tiefenkameras ausgestattet ist. In ihm kann sich eine Versuchsperson sehr frei bewegen und unterschiedliche Posen einnehmen. Ähn-

lich wie der 3D-Scanner Kinect der Xbox Spielekonsole projizieren die Kameras ein unsichtbares Lichtmuster auf die Person, mit dessen Hilfe sie die räumliche Geometrie erfassen. Das Ergebnis ist ein exaktes, dreidimensionales Gitternetz des gesamten Körpers, das als Input für den Algorithmus dient. Und am Ende kann dieser daraus auch Posen generieren, die er vorher noch nie gesehen hat.

„Wenn man das später beim Endnutzer machen möchte, muss das natürlich einfacher gehen“, sagt Geiger. Idealerweise sollte eine kurzen Videosequenz einer herkömmlichen Webcam oder besser einer Tiefenkamera ausreichen, wie sie bereits in vielen Smartphones enthalten ist, um ein Gittermodell zu erstellen. Die wenigen Daten, die dabei erzeugt werden, würden allerdings nur sehr ungenaue Ergebnisse liefern. Deshalb greift der Algorithmus auf seine „Erfahrungen“ zurück. In seinem Training mit den exakten Daten einer Vielzahl unterschiedlicher Personen aus dem Hightech-Scanner hat er bereits gelernt, wie Menschen typischerweise aussehen und wie sie sich bewegen.

Den Tübinger Forschern ist es in Zusammenarbeit mit dem Advanced Interactive Technologies Lab der ETH Zürich gelungen, den nötigen Input auf etwa eine Minute lange Aufnahmen einer Tiefenkamera zu reduzieren. Darin muss die Person verschiedene Posen einnehmen und sich mindestens einmal um die eigene Achse drehen, damit der Algorithmus ihre Rückenansicht erfasst. Das Ergebnis ist nicht ein simples, dreidimensionales Körpermodell, sondern umfasst auch die Deformation der Kleidung. Schließlich wirft ein weiter Pullover bei unterschiedlichen Körperhaltungen immer andere Falten.

Dem so generierten Avatar liegt eine Art digitales Skelett in Form eines Strichmännchens zugrunde, das sich nach Belieben animieren lässt, damit der Avatar neue Posen einnehmen kann. Der Input für die Steuerung darf sogar aus einer einfachen Kamera stammen, welche die Bewegung des Users trackt und auf den Avatar überträgt. Auf dieselbe Art lässt sich auch jeder beliebige andere Avatar steuern. Der Fülle unterschiedlicher Identitäten wären im Metaverse also keine Grenzen gesetzt.

„Die Steuerung ist kein Problem mehr“ sagt Andreas Geiger. „Der schwierige Teil ist, den Avatar auch echt aussehen zu lassen“. Einfache comichafte Darstellungen lassen sich schon jetzt mehr oder weniger problemlos generieren. Die automatische Erzeugung realistischer Avatare wie der von Geigers Team sind dagegen ein Feld der aktuellen Forschung.

Vom Foto bis zum 3D Modell
Forscher des ETH Zürich und des Max Planck Instituts für Intelligente Systeme in Tübingen haben eine Software entwickelt, die aus einfachen Fotos dreidimensionale Figuren extrahiert

FOTO: ANNETTE CARDINALE

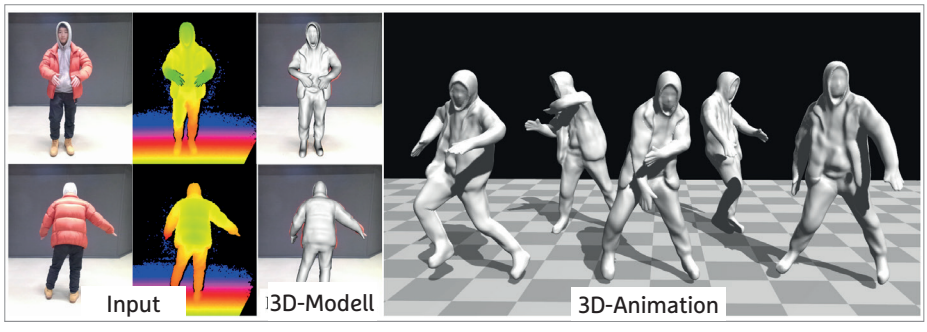
„Wir wollen die Erzeugung von echt aussehenden Avataren automatisieren.“

Prof. Dr. Andreas Geiger
Universität Tübingen

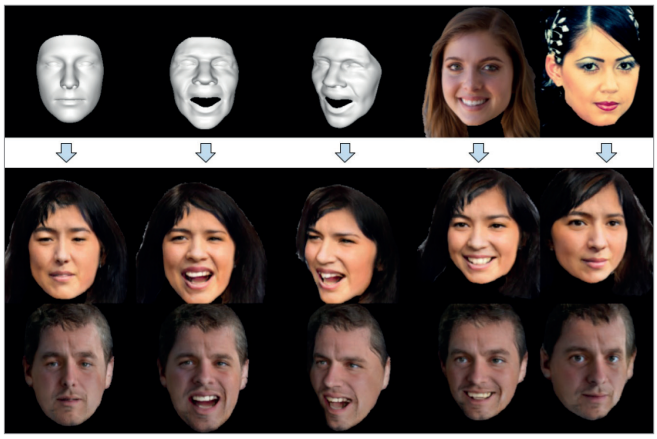
Virtuelle Masken für die täglichen Videokonferenzen erstellen

Besonders wichtig für die soziale Interaktion in der virtuellen Welt ist das Gesicht eines Avatars. Hier gibt es vor allem für den einfachen Einsatz bei Videokonferenzen bereits seit längerem Software, die Mimik und Lippenbewegungen einer Person in Echtzeit auf das Gesicht einer anderen Person überträgt.

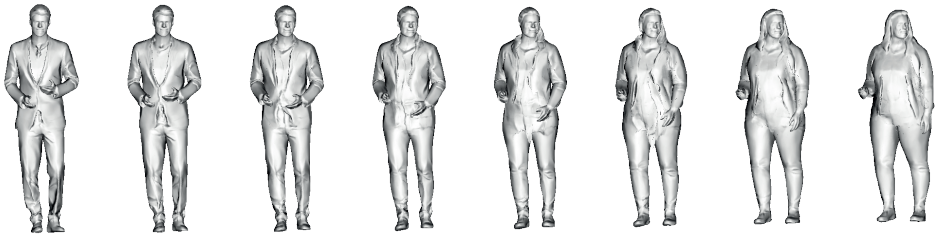
Vom Ablauf her ist das vergleichbar mit einem Puppenspieler, der eine Marionette steuert. Dazu reicht es, den „Puppenspieler“ während des Sprechens mit einer handelsüblichen Webcam zu filmen. Bevor es losgehen kann, muss die Software die „Marionette“ anhand eines kurzen Videos analysierten. Nur so kann sie auch ein 3D-Modell des Gesichts erstellen und die



Dreidimensionale Avatare aus Videoclips
Die Software PINA (Personalized Implicit Neural Avatars) des Advanced Interactive Technologies Lab der ETH Zürich generiert aus kurzen Videosequenzen Avatare samt Kleidung



Zum Avatar das passende Gesicht
Die von Google und der ETH Zürich entwickelte Software VariTex kann einem Avatar ein beliebiges Gesicht aufsetzen und verschiedene Gesichtsausdrücke dazu generieren



Wenn Avatare verschmelzen
Mit gDNA (Towards Generative Detailed Neural Avatars) der Universität Tübingen, ist es möglich, aus der Verschmelzung zweier Avatare einen völlig neuen entstehen zu lassen

FOTO: JANA DÜNNHAUPT

„Es ist schwieriger, Stimmen auszutauschen, als Gesichter zu manipulieren“

Dr. Ingo Siegert
Universität Magdeburg

verschiedenen Mundbewegungen lernen, die sie dann später während des „Puppenspiels“ zeigen soll.

Wird dagegen in einer Filmsequenz das gesamte Gesicht einer Person durch ein anderes ersetzt, spricht man von den sogenannten „Deepfakes“. Dafür lernt die Künstliche Intelligenz anhand von möglichst vielen Bildern, ein Gesicht auf seine wesentlichen Merkmale zu reduzieren. Sie extrahiert unter anderem Informationen darüber, ob die Augen offen, die Mundwinkel nach oben gerichtet oder der Blick zur Seite gewandt ist. Was genau das neuronale Netz als „wesentlich“ betrachtet, entscheidet es dabei selbst und ist für seine menschlichen Trainer oft überhaupt nicht nachvollziehbar. Wichtig dabei ist nur, dass es alle Gesichter auf das gleiche Set von Merkmalen reduziert. Denn damit kann es anhand der gewonnenen Daten ein anderes Gesicht mit dem exakt gleichen Gesichtsausdruck rekonstruieren.

Künstliche Stimme: So klingen wie ein Hollywood-Star

„Das Manipulieren von Gesichtern funktioniert schon sehr gut. Stimmen auszutauschen ist allerdings deutlich schwieriger“, sagt Ingo Siegert, Juniorprofessor am Institut für Informations- und Kommunikationstechnik der Universität Magdeburg. Im Gesicht müsse man nur Dinge wie Blickrichtung, Öffnen und Schließen der Augen, Brauen, Mundwinkel und Nase tracken. „Sprache ist aber viel reichhaltiger. Schließlich kann man ein und denselben Inhalt auf sehr unterschiedliche Arten ausdrücken“, sagt Siegert, der sich aktuell mit der KI-gestützten Anonymisierung von Stimmen beschäftigt, bei der Emotionen und der Persönlichkeitsausdruck des Sprechers erhalten bleiben sollen.

In den Achtzigerjahren haben Forscher noch versucht, die Sprachproduktion mit rechnerischen Modellen nachzubilden, um Computerstimmen zu erzeugen, wie man sie etwa von Stephen Hawking kennt. Später ging man dazu über, beispielweise für Durchsagen auf Bahnhöfen, Sprachsamples in ihre Grundlaute, die Phoneme, zu zerschnippeln und zu neuen Worten und Sätzen zusammenzufügen.

Heute setzen Forscher ähnliche Machine-Learning-Ansätze wie in der Bildverarbeitung auch zur künstlichen Erzeugung von Sprache ein. Nach den Prinzipien von Imitation und Mustererkennung trainieren die intelligenten Algorithmen anhand von Sprachproben, Laute zu bilden und können dadurch Stimmen imitieren oder völlig neue Stimmen erzeugen.

Mit dieser Technik erhält der eigene Avatar die Stimme einer anderen Person. Wer im künftigen Metaverse gerne wie sein Lieblingsschauspieler klingen will, etwa die schön tiefe Reibeisenstimme von Robert de Niros Synchronsprecher – kein Problem, sofern genug Stimmproben des Zielsprechers vorhanden sind. Diese muss der User nachsprechen, um ein neuronales Netz zu trainieren, das die akustischen Charakteristika der beiden Sprecher identifiziert und in der Folge umwandelt.

„Aktuell werden für solch ein Training insgesamt etwa 10 bis 15 Minuten nachgesprochener Samples benötigt, um ein halbwegs vernünftiges Ergebnis zu bekommen“, sagt Siegert. Das gilt allerdings nur für gelesene Sprache mit neutraler Betonung. Spontane Sprache, wie wir sie im Alltag häufig gebrauchen, ist wesentlich schwieriger zu erzeugen.

Während das Training eines solchen Stimmumwandlungs-Systems einiges an Zeit und Rechenleistung erfordert, sollte die eigentliche Anwendung später auf einem PC oder Handy in Echtzeit funktionieren. Bis ein normaler User ohne Vorwissen ein solches System auf die eigene Stimme trainieren und einsetzen kann, wird es Siegert zufolge noch einige Jahre dauern. Aber schließlich wird auch das Metaverse noch etwas Zeit brauchen, um so richtig in Schwung zu kommen.

redaktion@chip.de