Voller Körpereinsatz

Dank ausgeklügelter Sensoren und Algorithmen lassen sich Computerspiele mit Bewegungen steuern – neuerdings sogar völlig ohne Gamecontroller.

Die Zeiten, in denen man vom Daddeln nur müde Augen und Daumen bekam, sind vorbei. Wer heute ein Computerspiel an einer Konsole spielt, die mit den entsprechenden Zusatzsensoren ausgestattet ist, kommt schnell ins Schwitzen, weil er hüpfen, rennen oder werfen muss. Jede der drei stationären Spielekonsolen, die derzeit auf dem Markt sind, arbeitet bei der Erkennung von Gesten nach einem anderen Prinzip. Bei Nintendos Wii, dem Pionier des Konzepts, hält der Spieler eine Art Fernbedienung in der Hand, den Gamecontroller, über den er das Spiel mit seinen Handbewegungen steuert. Im Gamecontroller stecken ein lithografisch gefertigter Beschleunigungssensor, wie er auch in Fahrzeugen oder Smartphones[&)] verwendet wird, sowie ein Infrarotsensor, der die Signale von Leuchtdioden in einem Zusatzgerät registriert, das beim Bildschirm stehen muss. So kann die Spielekonsole Orientierung und Beschleunigung des Gamecontrollers bestimmen. Alternativ gibt es für die Wii eine Fußmatte mit Sensorik, um Bewegungen der Beine in der Spielszene am Bildschirm umzusetzen. Auch bei Sonys Konsole Move benötigen die Spieler einen Gamecontroller, der auf der einen Seite in einer bunten Kugel mündet – für jeden Spieler in einer anderen Farbe. Eine Kamera in der Konsole verfolgt die Position der Kugeln und liefert das Signal für die Steuerung des Spiels. Für ein möglichst realistisches Spielerlebnis geben solche Gamecontroller ein haptisches Feedback: Der Spieler spürt Vibrationen oder einen leichten Widerstand, wenn er in einem Spiel zum Beispiel den Tennisball mit dem Schläger trifft. Im Gegensatz dazu erfasst die dritte Konsole, Microsofts Xbox mit dem Sensor Kinect, die Körperbewegungen der Spieler direkt auf optischem Wege – ganz ohne Gamecontroller.

Microsoft

Spielekonsolen beziehen die Spieler körperlich stärker in das Geschehen mit ein. Microsofts Zusatzsensor Kinect erfasst sogar vollständig die Körperbewegungen und -haltungen aller Mitspieler, nicht nur die Bewegungen der Arme und Hände. Möglich wird dies durch strukturiertes Licht.

Die Hardware der Kinect besteht aus einer optischen und einer Infrarotkamera, einer Infrarotlaserdiode, vier Mikrofonen sowie der Steuerungselektronik für diese Komponenten (**Abb. 1**). Ein weiterer wichtiger Baustein ist ein Mikrochip, der aus dem Infrarotbild in Echtzeit eine Tiefenkarte berechnet, welche die Bewegungen der Spieler enthält. Dieses Prinzip ist als strukturiertes Licht bekannt: Mit der Laserdiode projiziert die Kinect ein zeitlich konstantes, pseudozufälliges Muster aus Punkten in den Raum. Sobald es auf einen Spieler fällt, verzerrt es sich auf charakteristische Weise. Die in die Kinect integrierte Infrarotkamera nimmt dieses verzerrte Muster auf. Da sie durch den seitlichen Versatz zur Laserdiode einen anderen Blickwinkel hat, lassen sich auf diesem Wege räumliche Informationen über die Spieler gewinnen.

Um die Tiefenkarte zu erstellen, berechnet die Kinect eine Kreuzkorrelation für jedes Pixel zwischen dem reflektierten und dem ursprünglich projizierten Muster. Die resultierende Tiefe eines bestimmten Bildpunkts liefert zusammen mit der Aufnahme der optischen Kamera ein Bild, das die Tiefeninformationen enthält.

Microsoft hat die Technologie der Kinect von der Firma Primesense zugekauft. Die zugrundeliegenden Prinzipien des Sensors sind schon länger bekannt, Primesense ist es aber gelungen, die Techno-

Microsoft

Abb. 1 Die Hardwarekomponenten der Kinect sind (von links nach rechts) eine Laserdiode, die das strukturierte Licht erzeugt, eine RGB-Kamera sowie eine Infrarotkamera. Die vier, nicht sichtbaren Mikrofone befinden sich an den beiden Enden.

&) vgl. Physik Journal, April 2011, S. 42

logie erstmals mit sehr billigen Hardwarekomponenten und mit sehr guten Spezifikationen zu implementieren: Das Verfahren funktioniert für Tiefen zwischen 0,8 und drei Meter mit einer durchschnittlichen Verzögerungszeit bei voller VGA-Auflösung (640 × 480 Pixel) von 40 Millisekunden. Bei einem Abstand von zwei Metern zwischen Sensor und Objekt erreicht die laterale räumliche Auflösung drei Millimeter. Das erfasste Gesichtsfeld liegt horizontal bei 58 Grad, vertikal bei 45 Grad und diagonal bei 70 Grad. Microsoft macht keine Angaben zu den Kenndaten der Kinect, immerhin nennen Mitarbeiter in einem Konferenz-Paper eine Bildwiederholrate von 30 Frames pro Sekunde mit einer räumlichen Auflösung von einigen Zentimetern.[5)]

Wald aus Entscheidungsbäumen

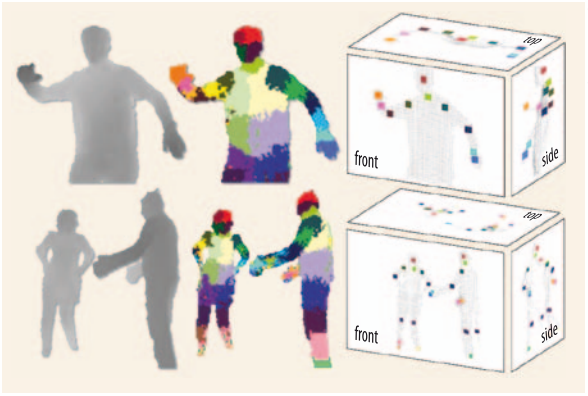
Doch die Tiefenkarte ist erst die halbe Miete für das Spielerlebnis. Entscheidend ist auch der Algorithmus, mit dem der Sensor die Bewegungen mehrerer Personen erfasst und sie jeweils in ein Skelettmodell umsetzt, durch das sich nicht nur eine Armbewegung des Spielers, sondern sämtliche Körperbewegungen in Echtzeit erkennen lassen (**Abb. 2**). Die Körperteile der Spieler werden dazu aus dem Tiefenbild segmentiert, um sie den unterschiedlichen Gelenken und Körperabschnitten in einem Skelettmodell zuzuordnen. Dieses ähnelt einer Gliederpuppe, die der Algorithmus mit Verfahren der Wahrscheinlichkeitsrechnung schließlich wiederum im realen Raum verortet. Die Kinect führt diese Berechnungen pixelweise durch, weil das den Rechenaufwand im Vergleich zu kombinatorischen Ansätzen reduziert und trotzdem ausreichend genau ist.

Damit der Algorithmus funktioniert, wurde er zuvor trainiert. Soll heißen, die Entwickler klassifizierten mit dem Algorithmus bekannte Tiefenbilder von Menschen, die sich in Körperbau, Größe und Körperhaltung stark voneinander unterschieden. In diesen Bildern war jedes Pixel bereits dem richtigen Körperteil zugeordnet. Da die Klassifikation der zu einem Menschen gehörenden Körperteile für einen einzigen Entscheidungsbaum zu komplex ist, erfolgte sie mit einem „Random Forest", einer Sammlung mehrerer unkorrelierter Entscheidungsbäume. Dabei wählt der Algorithmus bei jeder Entscheidung eine zufällige Gruppe von Fragen aus allen Entscheidungsbäumen aus, die ihm zur Verfügung stehen, um aus der resultierenden Wahrscheinlichkeitsverteilung eine möglichst korrekte Klassifikation eines Pixels abzuleiten.

Drei Entscheidungsbäume mit einer Tiefe von jeweils 20 Entscheidungen anhand von einer Million Bilder zu trainieren, dauerte auf einem Rechner-Cluster mit tausend Prozessorkernen etwa einen Tag. Die optimierte Implementierung des Algorithmus auf der Xbox klassifiziert ein Bild in weniger als fünf Millisekunden. Das ist eine Größenordnung schneller als bei vergleichbaren existierenden Ansätzen auf dem PC.

In die Spielszenen auf dem Bildschirm werden die Bewegungen der Spieler praktisch ohne Verzögerung auf die Figur übertragen. Duckt sich ein Spieler, macht das die Figur im Spiel ebenfalls in Echtzeit. Springt ein Spieler über einen imaginären Abgrund, überwindet auch die Figur das Hindernis. Allerdings erfordert das Konzept der Kinect mehr Aktivität von den Teilnehmern: Während es besonders bequeme Zeitgenossen schaffen, mit dem Gamecontroller der Wii zum Beispiel im Sitzen Kegeln zu spielen, geht es bei der Kinect nicht mehr ohne vollen Körpereinsatz.

Michael Vogel



J. Shotton et al.

Abb. 2 Ein Algorithmus ermittelt aus jedem Tiefenbild der Kinect für jedes Pixel, zu welchem Körperteil eines Spielers es gehört. Hieraus leitet der Algorithmus für jeden Spieler ab, wie er sich durch ein Skelettmodell im Raum darstellen lässt. Alle Berechnungen erfolgen in Echtzeit.

5) *J. Shotton* et al., IEEE Computer Vision and Pattern Recognition, Juni 2011