

Virtual Assistant for Endoscopic Surgery in Simulation Environment

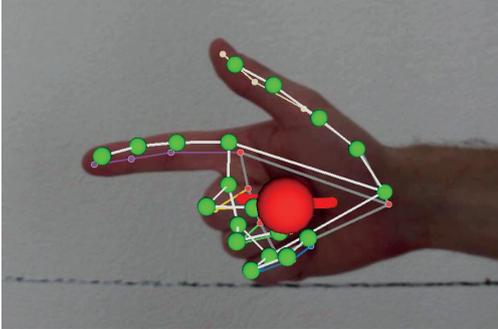


Abb. 1: 3D-Handpose mit virtuellem Ball

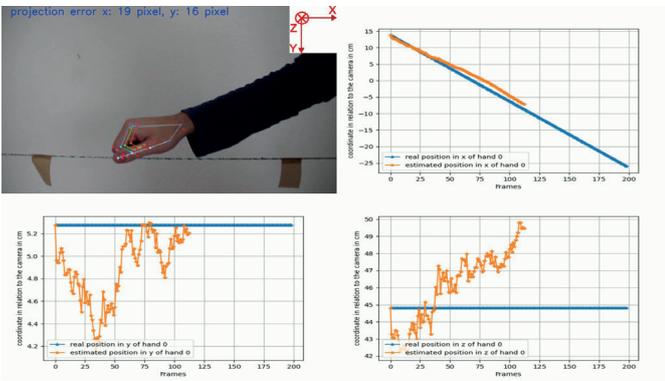


Abb. 2: Bewegungsbestimmung der Hand im dreidimensionalen Raum

Confusion Matrix

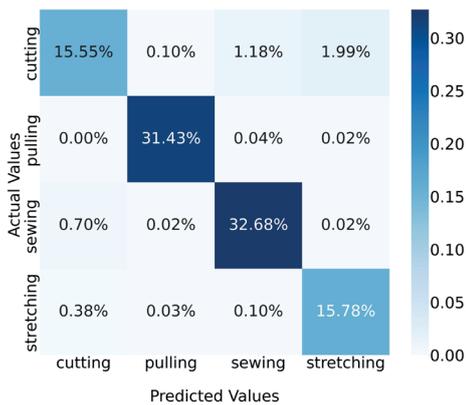


Abb. 3: Testergebnisse des Deep-Learning-Netzwerks

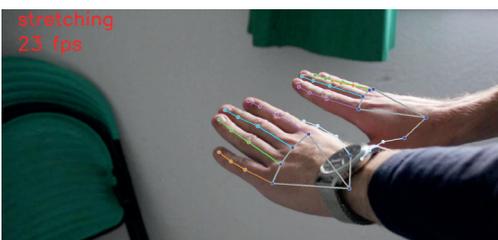


Abb. 4: Gestenklassifikation in einem Video

Problemstellung

Die Firma VirtaMed AG produziert Simulatoren für die chirurgische Ausbildung. Mittels virtueller Realität wird eine hochrealistische Simulation ermöglicht. Bei manchen operativen Eingriffen wird ein Endoskop mit integrierter Kamera für die Sicht auf die Operationsstelle im Körper benötigt. Das Endoskop wird in der Realität wie auch momentan beim Simulator von einer Assistenzperson gesteuert. Damit für eine Trainingseinheit am Simulator der Chirurg auch ohne Assistenz trainieren kann, soll die Kameraführung mithilfe der Handbewegungen automatisiert werden. Da es sich um eine virtuelle Kamera in der Simulationsumgebung handelt, werden keine mechanischen Komponenten benötigt. In dieser Arbeit wird untersucht, wie mit Machine Learning die ausgeführten Gesten und die Bewegungsrichtung der Hände im dreidimensionalen Raum ermittelt werden können.

Lösungskonzept

Mit dem Machine-Learning-Algorithmus MediaPipe wird die dreidimensionale Handpose in einem Video detektiert. Ein trainiertes Deep-Learning-Netzwerk verwendet die Informationen von MediaPipe, um die ausgeführte Geste zu bestimmen. Dabei kommt Zeitreihenklassifikation zum Einsatz.

Realisierung

Mehrere Python-Skripts ermöglichen ein effizientes Erstellen von Trainingsdaten und neuronalen Netzwerken für neue Handgesten. In einem zweiten Schritt wird MediaPipe eingesetzt, um mithilfe von Computer Vision die Bewegungsrichtung und zurückgelegte Strecke der Hand im Raum abzuschätzen. Verschiedene Verfahren werden angewandt, um die Genauigkeit der Schätzung zu erhöhen.

Ergebnisse

Beim Messen der im Raum zurückgelegten Distanz werden auf Bildebene Toleranzen von 13% erreicht. Die Bildtiefe ist wenig konstant und stark von der geschätzten Handpose durch MediaPipe abhängig. Dabei entstehen Abweichungen von 10 cm und teilweise mehr. Die Gestenklassifikation mit Deep Learning kombiniert mit MediaPipe erzielt gute Resultate. Handgesten können in einem Video mit einer hohen Genauigkeit erkannt werden. Ein erstelltes neuronales Netzwerk mit 25'000 Parametern besitzt dabei eine Testgenauigkeit von 95%. In wenigen Stunden kann mit den geschriebenen Skripten ein Netzwerk mit ähnlicher Zuverlässigkeit auf neue Gesten trainiert werden. Der limitierende Faktor der Ausführungszeit ist dabei MediaPipe, welcher ohne Grafikkarte die Verarbeitung von 20 Bildern pro Sekunde erlaubt.



Diplomand
Waber Ismael

Dozent
Prof. A. Rumsch

Themengebiet
Mechatronik/Automation/Robotik

Projektpartner
VirtaMed AG

