

Arbeit, Heim, Komfort

3 Kanal Video (HD), Farbe, Untertitel, 4 Kanal Sound, 122 min.
Michael Mönnich, 2020

„Computer vision has a great potential to help our daily lives by searching for lost keys, watering flowers or reminding us to take a pill. To succeed with such tasks, computer vision methods need to be trained from real and diverse examples of our daily dynamic scenes. [...] We propose a novel Hollywood in Homes approach to collect such data. Instead of shooting videos in the lab, we ensure diversity by distributing and crowdsourcing the whole process of video creation from script writing to video recording and annotation. Following this procedure we collect a new dataset, Charades, with hundreds of people recording videos in their own homes, acting out casual everyday activities.“¹

Die Arbeit an Datensätzen ist prädestiniert für Outsourcing und wird auf Internetseiten wie *Amazon's Mechanical Turk* ausgelagert.² Auf digitalen Arbeitsmarktplattformen wie dieser schreiben Unternehmen Aufgaben wie zum Beispiel die Kategorisierung oder Generierung von Daten an eine global konkurrierende Arbeiter*innenschaft aus.

Maschinelles Sehen findet beispielsweise Anwendung in der Gesichtserkennung von Smartphones oder dem autonomen Fahren. All dieses Sichten von Bildern soll automatisch stattfinden. Dafür muss eine Kamera jedoch erst angelernt werden, bevor sie dann eine berechnete Aussage treffen kann. Es erfordert große Mengen an klassifiziertem Bildmaterial, um die Bedeutung bestimmter Konzepte automatisiert erfassen zu können. Was also ist eine Tür, was ist ein Lächeln, wie sieht Kaffee trinken aus.

Arbeit, Heim, Komfort beschäftigt sich mit einem 68,8 Stunden langen Video-Archiv, welches für Forschungszwecke im Internet frei verfügbar ist. Dieser Datensatz ist einzigartig, weil er in seiner Produktion vollständig ausgelagert wurde - nicht nur die

Beschriftung von Bildern werden von global verteilten Individuen in ihrem Zuhause bearbeitet, sondern sie erstellen auch die Bilddaten selbst.

Die Arbeitskräfte bekommen als Aufgabe gestellt, ein Skript zu verfassen, das bestimmte Objekte und Handlungen enthält. Dieses Skript sollen sie dann vor der Kamera ausführen, zum einen aus der "subjektiven" Ego-Perspektive, zum anderen aus der Perspektive einer dritten Person. Die beschrifteten 30-sekündigen Videoclips senden sie zurück und erhalten nach Abnahme einen Dollar von den Auftraggebenden.

In der Videoarbeit sind die Skripte und Aufnahmen zweier Arbeitskräfte, den Subjekten P2J3 und CC19 zu sehen, die vom selben Ort aus arbeiten. Beide Subjekte führen jedes Skript doppelt aus den zwei unterschiedlichen Perspektiven auf. Die Erstellung der Untertitel wurde von Michael Mönnich selbst ausgelagert. Der Arbeiter Rakesh Raman übersetzte die im Video hörbare Sprache ins Englische und legte die Untertitel an.

Der private Raum des Zuhauses ist in diesem Setting von zweifacher Bedeutung. Einerseits wird er auf *Mechanical Turk's* Website als ein Argument angeführt, sich für Crowdwork zu entscheiden (»im Komfort des Zuhauses arbeiten«). Andererseits ist er zentral für die Informatiker*innen, deren Fokus der Forschung auf der Entwicklung von intelligenten Kameras liegt, die uns daran erinnern sollen unsere Blumen zu gießen und Schlüssel zu finden.

Anna Weidendahl, 14/02/2020

¹ Gunnar A. Sigurdsson Et al., Computer Vision – ECCV 2016, vol. 9905, 2016.

² <https://www.mturk.com/>