Detektion von Handlungen in anonymisierten Videos

Vor rund einem Monat haben wir einen Beitrag über die Berücksichtigung des Datenschutzes bei intelligenter Videosensorik veröffentlicht.

In der dort vorgestellten Lösung wird statt dem Original-Kamerabild ein statischer Hintergrund verwendet, in dem für jede Person des Original-Videos ein „Strichmännchen" (Skelett) eingezeichnet wird.

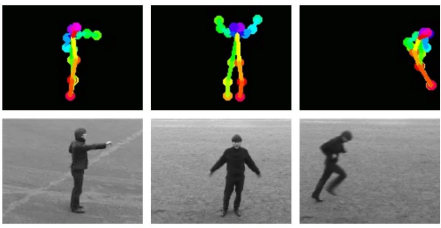

Abb. 1: Anonymisierung mithilfe von Strichmännchen

Wie in Abb. 1 ersichtlich, besitzen die Strichmännchen im anonymisierten Video die gleiche Körperhaltung wie die ursprünglichen Personen.

Im Rahmen einer Bachelorarbeit wurde im EnBW TechLab untersucht, ob diese Informationen zur Detektion von Handlungen verwendet werden können.

Eine Handlungserkennung kann unter anderem zur Erhöhung der persönlichen Sicherheit genutzt werden, beispielsweise indem bei Stürzen oder Streit automatisch um Hilfe gerufen wird. Auch eine statistische Auswertung von Handlungen kann interessant sein. Kommunen könnten beispielsweise die primär auf einem Platz ausgeführten Aktionen identifizieren, um die weitere Stadtentwicklung auf das Verhalten der Bürger anzupassen.

Im Rahmen der Arbeit wurden zunächst bisherige Ansätze und Implementierungen untersucht. Das Ergebnis dieser Untersuchung: Betrachtet man einen zeitlichen Verlauf von Körperhaltungen, so lassen sich Rückschlüsse auf die ausgeführten Handlungen schließen.

Um diese theoretische Erkenntnis zu untermauern, wurde ein Prototyp entwickelt, der Bilder einer Webcam anonymisiert und Handlungen in den anonymisierten Videos erkennt (siehe Abb. 2).

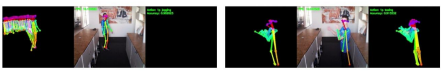

Abb. 2: Prototyp zur Handlungserkennung

Dieser erste Prototyp unterscheidet einfache Handlungen wie „springen", „winken" oder „gehen", wenn genau eine Person frontal vor der ausgewerteten Kamera steht.

Aus technischer Sicht kombiniert der Prototyp eine Human Pose Keypoint Estimation mit einem Machine Learning Bildklassifikator. Um nicht nur statische Bilder, sondern Videosequenzen in Echtzeit auswerten zu können, werden die einzelnen Skelett-Bilder überlagert und die Überlagerungen als Bilder klassifiziert.

Durch eine Weiterentwicklung der Algorithmen sollen in Zukunft auch mehrere parallel stattfindende Handlungen mehrerer Personen ausgewertet können.

Für die Erkennung von Körperhaltungen wurde der Pose Estimator von Ildoo Kim verwendet. Dieser basiert auf Tensorflow und Python und ist modular aufgebaut, weshalb Funktionen mit einer hohen Geschwindigkeit direkt aus dem in diesem Projekt entwickelten Python-Code angesprochen werden können.

Der Code zur Überlagerung der Körperhaltungen wurde im Rahmen dieser Arbeit selbst geschrieben.

Zum Klassifizieren der Skelett-Bilder wurde ein Bildklassifikator der Tensorflow-Entwickler um Funktionen zur Genauigkeitsmessung erweitert und mit eigenen Daten trainiert. Der Bildklassifikator basiert auf dem Inception v3 Netz, das zur Unterscheidung von Objekten des ImageNet Datensatzes vortrainiert ist.

Für unsere Zwecke wurde dieses neuronale Netz um eine weitere Schicht ergänzt, die auf den KTH-Datensatz trainiert wurde. Dieser beinhaltet Videos von sechs verschiedenen Handlungen wie Winken, Springen, Boxen oder Laufen. Die Kombination des Handlungs-Datensatzes mit einem allgemeinem Datensatz wie ImageNet verringert die Trainingszeit und erhöht die Genauigkeit des Klassifikators, da grundlegende Zusammenhänge nicht von null auf gelernt werden müssen.

Bei diesem Datensatz erreicht der Handlungs-Klassifikator eine Genauigkeit von 85,7%.

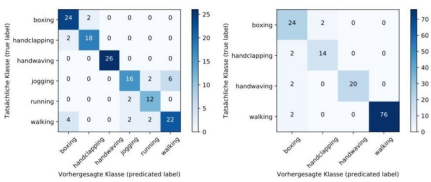

Abb. 3: Wahrheitsmatrizen des zweiten Prototyps beim KTH Datensatz. Es ist erkennbar, dass deutlich weniger Verwechslungen stattfinden, da nur wenige der klassifizierten Bilder außerhalb der Hauptdiagonale liegen.

In Abb. 3 sind zwei Wahrheitsmatrizen dargestellt. Diese zeigen, wie viele Videos einer Klasse korrekt klassifiziert wurden und zeigen darüber hinaus, welche Klassen oft verwechselt werden. Je weniger Werte außerhalb der Hauptdiagonale liegen, desto genauer ist der Klassifikator.

Bei Vereinigung der ähnlichen Klassen *Running*, *Jogging* und *Walking*, die auch Menschen nur schwer unterscheiden können, erhöht sich die Genauigkeit des Klassifikators auf 96,3%.

Genauere Details der technischen Implementierung sowie eine Untersuchung bisheriger Ansätze und Datensätze können in der Bachelorarbeit nachgelesen werden, die auf der Projektseite des Autors heruntergeladen werden kann.

Der vorgestellte Code zur Handlungserkennung kann auf GitHub angesehen werden. Dort befindet sich auch eine Installationsanleitung für alle, die selbst mit einer Handlungserkennung experimentieren möchten.