

Robust Real-World Crowd Pose Estimation

Das Fraunhofer-Institut für Optronik, Systemtechnik und Bildauswertung IOSB ist eines der größten Institute für angewandte Forschung auf dem Gebiet der Bildgewinnung und Bildauswertung in Europa. Die Abteilung Videoauswertesysteme (VID) beschäftigt sich mit der automatischen Auswertung von Signalen bewegter bildgebender Sensorik in komplexen, ggf. nichtkooperativen Szenarien. Diese Sensorik wird beispielsweise im Aufklärungs- und Überwachungsbereich als integrierte Komponente in fliegenden, weltraumgestützten oder mobilen landgestützten Plattformen verwendet. VID entwickelt und integriert hierfür Bildauswertelgorithmen für autonome oder interaktive Systeme.

Beschreibung

Die Analyse des Verhaltens Einzelner in Menschenmengen (engl. Crowds) oder Menschengruppen an öffentliche Plätze hat in diesem Jahr, beispielsweise durch Abstandsgebote, enorme an Bedeutung gewonnen. Dabei werden oft Körperposen als Abstraktion benutzt um die Grundrechte des Einzelnen nicht einzuschränken.

Die derzeitigen Methoden zur Schätzung der menschlichen Posen lassen sich in Top-down-Methoden und Bottom-up-Methoden unterteilen. Top-down-Methoden verwenden typischerweise zuerst einen Personendetektor, um jeweils Personen mit einer Bounding Box zu detektieren und reduzieren dann das Problem auf eine Einzel-Person Posenschätzung. Im Gegensatz dazu Bottom-up-Methoden starten mit der Lokalisierung identitätsfreier Landmarken für alle Personen in einem Eingabebild gefolgt von deren Gruppierung in Personeninstanzen. Diese Strategie ermöglicht Bottom-up-Methoden schnellere und echtzeitfähigere Crowd-Posenschätzung.



Abbildung 1: Posenschätzung für eine Crowd [1].

Aufgabenstellung

In einem ersten Schritt wird mithilfe eines Top-Down Posenschätzer einen Video-Datensatz für Crowd-Posenschätzung in der realen Welt erstellt. Dieser Datensatz bildet anschließend die Grundlage weiterer Untersuchungen und Experimente. In einem zweiten Schritt soll ein Bottom-Up Verfahren [2] für Crowd-Posenschätzung erweitert und optimiert werden. Abschließend wird das Verfahren gegenüber State-of-the-Art-Verfahren hinsichtlich Leistungsfähigkeit und Geschwindigkeit evaluiert.

Voraussetzungen

- Studienfach: Informatik, Mathematik, Angewandte Physik oder vergleichbar
- Sehr gutes Verständnis für die (theoretischen) Grundlage von Deep Learning
- Sehr gute Programmierkenntnisse (idealerweise Python)
- Erfahrung mit dem Deep Learning Framework Pytorch vorteilhaft.
- Fähigkeit zum selbstständigen Arbeiten
- Bereitschaft, sich in neue Themengebiete einzuarbeiten und Freude am Einbringen eigener Ideen

Bei Interesse senden Sie uns bitte Ihre Bewerbungsunterlagen (kurzes Anschreiben, tabellarischer Lebenslauf, aktueller Notenauszug) in elektronischer Form sowie den gewünschter Starttermin.

Literatur

- [1] Golda, Thomas, et al. "Human pose estimation for real-world crowded scenarios." 2019 16th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS). IEEE, 2019.
- [2] Cheng, Bowen, et al. "HigherHRNet: Scale-Aware Representation Learning for Bottom-Up Human Pose Estimation." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020.

Kontakt

Mickael Cormier, M. Sc.
Fraunhofer-Institut für Optronik, Systemtechnik und Bildauswertung
Fraunhoferstraße 1, 76131 Karlsruhe
Tel.: 0721 / 6091-634
mickael.cormier@iosb.fraunhofer.de