

Effiziente Erkennung von Handlungen und Gesten durch videobasierte Analyse von Handbewegungen [Bachelor/Master]

Insbesondere im Bereich feinmotorischer Tätigkeiten (z.B. manuelle Elektronikfertigung) geben die Bewegungen der Hände wesentliche Hinweise über die Handlung der Person. Auch für eine gestenbasierte Mensch-Computer-Interaktion [1] ist die computergestützte Analyse von Handbewegungen unabdingbar. Videobasierte Verfahren eignen sich besonders aufgrund der berührungslosen und damit nicht hinderlichen Einsatzmöglichkeit. Um eine latenzarme Reaktion des Systems generieren zu können, ist die zeitliche Effizienz ein wichtiges Kriterium.

Verschiedene Veröffentlichungen behandeln eine videobasierte Handlungserkennung, wobei häufig Objekt-Detektoren wie YOLOv3 [2] [3] [4] auf Einzelbilder angewendet werden [5]. Zur Extraktion von bewegungsabhängigen Merkmalen kann zusätzlich eine Berechnung des optischen Flusses erfolgen [5] [6]. In vielen Veröffentlichungen bezieht sich das Anwendungsfeld jedoch auf die Detektion und Analyse gesamter Personen sowie einer groben Unterteilung von Aktivitäten (z.B. Unterscheidung verschiedener Sportarten, vgl. *Abbildung 1*). Für die Analyse von Handbewegungen ist hingegen eine feine Unterteilung von Handlungen/Gesten sowie eine zeitliche Abgrenzung erforderlich.

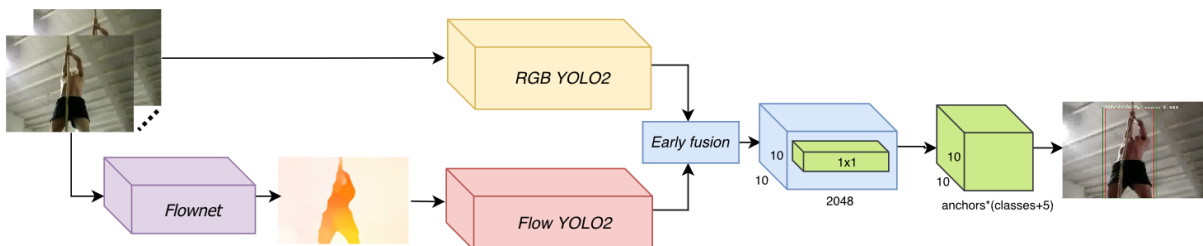


Abbildung 1: Verarbeitungspipeline zur videobasierten Handlungserkennung von Personen nach El-Nouby et al. 2018 [6].

Ein möglicher aber nicht zwingend notwendiger Zwischenschritt besteht in der Erkennung der Handposen (vgl. *Abbildung 2*), sodass aus dem zeitlichen Verlauf der Gelenkpositionen Merkmale für die Handlungs- bzw. Gestenerkennung abgeleitet werden können. Ein Beispiel für die Erkennung von 3D-Handposen aus RGB-Bildern führt [7] auf.

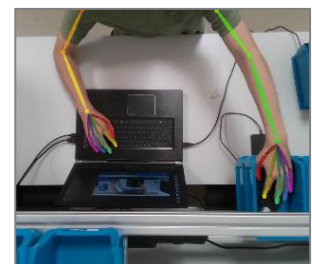


Abbildung 2: Detektion von 2D-Hand-Posen am ZBS-Versuchsarbeitsplatz mittels OpenPose [8].

Unter Berücksichtigung verschiedener Gesichtspunkte (siehe Aufgabenschwerpunkte) soll ein Verfahren zur effizienten Erkennung von Handlungen und Gesten durch videobasierte Analyse von Handbewegungen umgesetzt werden.

Aufgabenschwerpunkte

(individuelle Anpassung kann je nach Vorkenntnissen und Art der Arbeit erfolgen)

- Ergänzende Recherche bzgl. Verfahren zur Handlungs- und Gestenerkennung aus Videos. Ausgangsbasis kann die unten aufgeführte Literatur darstellen.
- Auswahl eines geeigneten Verfahrens oder Entwicklung eines eigenen Lösungsansatzes unter Berücksichtigung von Gesichtspunkten wie zu erwartende Geschwindigkeit, zu erwartende

Genauigkeit, Realisierbarkeit, Robustheit gegenüber unterschiedlichen Hintergründen und Kamera-Anordnungen, Generalisierbarkeit (Eignung für Handlungs- und Gestenerkennung).

- Umsetzung des ausgewählten Verfahrens.
- Auswertung des umgesetzten Verfahrens.
- Dokumentation der Ergebnisse.

Ansprechpartner am ZBS e. V.

David Reese (david.reese@zbs-ilmenau.de)

Literatur

- [1] S. S. Rautaray und A. Agrawal, Vision based hand gesture recognition for human computer interaction: a survey., Artificial Intelligence Review, 2015.
- [2] J. Redmon, S. Sivvala, R. Girshick und A. Farhadi, You only look once: Unified, real-time object detection, IEEE, 2016.
- [3] J. Redmon und A. Farhadi, YOLO9000: better, faster, stronger, 2016: arXiv preprint.
- [4] J. Redmon und A. Farhadi, „YOLOv3: An Incremental Improvement,“ arXiv preprint, 2018.
- [5] G. Singh, S. Saha, M. Sapienza, P. Torr und F. Cuzzolin, Online real-time multiple spatiotemporal action localisation and prediction., ICCV, 2016.
- [6] A. El-Nouby und G. W. Taylor, Real-Time End-to-End Action Detection with Two-Stream Networks, arXiv preprint, 2018.
- [7] P. Panteleris, I. Oikonomidis und A. Argyros, Using a single RGB frame for real time 3D hand pose estimation in the wild, arXiv preprint, 2017.
- [8] T. Simon, H. Joo, I. Matthews und Y. Sheikh, Hand Keypoint Detection in Single Images using Multiview Bootstrapping, CVPR, 2017.