

Aufgabenstellung:

Deine Aufgabe besteht darin, ein Computer-Vision-System zur Echtzeiterkennung und Übersetzung von Gebärdensprache in Text zu entwickeln und den gesamten Entwicklungsprozess in einem Projektbericht zu dokumentieren und zu reflektieren. Zunächst formulierst Du klare Projektziele und -fragen und erstellst einen detaillierten Projektplan. Du wählst und beschaffst geeignete Datensätze wie RWTH-PHOENIX-Weather 2014 T und bereitest diese durch Markierung relevanter Frames und Annotation der Gebärden vor. Im nächsten Schritt entwickelst Du ein Subsystem zur Erkennung und Verfolgung der Handbewegungen und Gesichtsausdrücke in Videostreams, basierend auf Algorithmen wie Mediapipe, und extrahierst aussagekräftige Merkmalsvektoren. Du implementierst ein Modell zur Prädiktion mit Zeitreihendaten, z. B. ein tiefes neuronales Netzwerk basierend auf LSTM-Architekturen, das für die Verarbeitung von Sequenzen geeignet ist, und trainierst sowie validierst dieses Modell. Anschließend integrierst Du das trainierte Modell in eine echte Anwendungspipeline zur Echtzeiterkennung und Übersetzung, optimierst diese hinsichtlich Latenz und Genauigkeit, und dokumentierst den technischen Entwicklungsprozess. Die abschließende Evaluierung erfolgt durch Tests mit verschiedenen Videos zur Analyse der Genauigkeit und Reaktionszeit; mögliche Fehlerquellen werden optimiert und die Ergebnisse werden in Relation zu theoretischen Erwartungen bewertet. Schließlich dokumentierst Du alle Schritte umfassend in einem Projektbericht, erstellst eine Präsentation sowie ein Demo-Video zur Funktionsweise des Systems, das Du auf einer Cloudplattform wie OneDrive oder einem anderen sicheren Cloud-Anbieter hochlädst und den Link in dem Bericht integrierst.

Die Erwartung ist nicht die Entwicklung eines fertigen Produkts, sondern die prototypische Umsetzung zentraler Bausteine (z. B. Handtracking + Training eines einfachen LSTM Sequenzmodells) und die Dokumentation des Vorgehens, wo es Projektplanung mit klaren Zielen und Forschungsfragen gibt. Also der Schwerpunkt liegt auf Methodenverständnis und systematischem Vorgehen, nicht auf einem fertigen, fehlerfreien Endprodukt. Wichtig aber dabei auch ist, das Modell mit aussagekräftigen Metriken zu evaluieren, wie Genauigkeit, False Positive, Latenz unw.

Einführende Literatur:

Koller, O., Forster, J. & Ney, H. (2015). Continuous sign language recognition: Towards large vocabulary statistical recognition systems handling multiple signers. *Computer Vision and Image Understanding*, 141, 108–125.

<https://search.ebscohost.com/login.aspx?direct=true&AuthType=sso&db=edselp&AN=S1077314215002088&site=eds-live&scope=site&custid=s6068579>

Hinweise:

Im Folgenden findet sich geeignete Software.

- RWTH-PHOENIX-Weather (2014). *Dataset*. <https://www-i6.informatik.rwth-aachen.de/~koller/RWTH-PHOENIX/>
- OpenCV (2023). *Open Computer Vision Library*. <https://opencv.org/>
- MediaPipe (2020). <https://chuoling.github.io/mediapipe/>
- Keras (2024). <https://keras.io/>
- Tensorflow (2024). <https://www.tensorflow.org/>