

Entwicklung eines 2D- und 3D- Lippenbewegungssensors

Maschine lernt Lippenlesen: Das Lippenlesen ist ein Informationskanal, der für elektronische Geräte interessant sein könnte. Doch der Wunsch nach maschinellem Lippenlesen stellt bis heute grosse technische Hürden dar. Menschen können durch die Beobachtung der Lippenbewegungen visuell viele Silben und Worte wiedererkennen. Die Form und Haltung der Lippen und des Kiefers haben bei vielen Lauten einen wesentlichen Einfluss auf ihre akustische Art und den Klang.

Aufnahmen in 3D: Ziel dieser Arbeit war, Gesichter mit Lippenpositionen beim Sprechen von Vokalen in Deutsch in 2D und 3D aufzunehmen. In diesen Aufnahmen wurden Gesichtsmerkmale ausgemacht und ausgewertet, indem sie verglichen und diskriminiert werden. Die Idee war, mittels 3D-Aufnahmen an die Tiefeninformationen des sprechenden Gesichtes zu gelangen und damit die daraus resultierenden Vokale besser zu erkennen. Darum wurden verschiedene Kriterien zur Wahl einer passenden 3D-Kamera untersucht.

Gesichtserkennung: Diese Arbeit zeigt die Vorteile der 3D-Erfassung gegenüber üblichen 2D-Messverfahren sowie die entstandenen Herausforderungen. In der Arbeit werden Techniken des maschinellen Lippenlesens diskutiert und an einem eigenen Ansatz getestet. Mithilfe eines eigens aufgebauten Messplatzes für die Aufnahme von Audio- sowie Video-Signalen in 2D und 3D wurden experimentelle Daten für die Vokale in Deutsch aufgenommen. Ein Algorithmus markiert in den Gesichtern typische Merkmale, wie Mund, Auge oder Nase. Aus diesen Messdaten wurde die Lippenposition und -haltung in den Bildern identifiziert und die für Vokale typischen Eigenschaften hergeleitet. Dies erlaubte eine Zuordnung neuer Aufnahmen zu bekannten Vokalen.

3D hat das Sagen: Die Auswertung mit 3D-Bildern hat eine um 13 % bessere Erkennung geliefert als banale 2D-Aufnahmen ohne Tiefeninformationen. Besonders wenn die Aufnahme des Gesichtes leicht schräg, anstatt frontal ausfiel, zeigte das 3D-Verfahren seine Stärken. In einer Testreihe mit ausschliesslich frontalen Aufnahmen machte der Erkennungsgewinn der 3D- gegenüber der 2D-Technik 10 % aus, bei ausschliesslich schrägen Aufnahmen hingegen sogar 19 %.

Die 3D-Variante verbessert also die optische Spracherkennung.



Diplomierende
Lukas Meier
Simon Marco Müller
Daniel Zinsstag

Dozent
Julien Rey



Dank kompakter 3D-Kamera und intelligenter Software sollen Gehörbeeinträchtigte in Zukunft auch ohne Gebärden auskommen können.