

Gebaarherkenning van de SMOG-gebarentaal voor VR-belevingen

Yara Mijnendonckx

master IIW Informatica

Introductie

Het onderzoek richt zich op het oefenen van Spreken Met Ondersteuning van Gebaren (**SMOG**). Bij SMOG wordt de **gesproken taal ondersteund door gebaren**, terwijl gebarentaal op zichzelf staat zonder een gesproken boodschap. De combinatie van gebaren en gesproken taal kan mensen helpen bij het begrijpen van gesproken taal alsook om zichzelf beter uit te drukken [1]. Figuur 1 toont bij wijze van voorbeeld de lijntekening van het SMOG-gebaar voor toilet.

Het aanleren en oefenen van SMOG is echter een **arbeidsintensief** proces, aangezien één begeleider telkens slechts één persoon kan bijstaan. *Virtual reality* (VR) heeft potentieel om SMOG-gebaren doeltreffender en speelser in te oefenen en tegelijkertijd de werklast van begeleiders te verminderen. Daarom onderzoekt deze masterproef in welke mate het mogelijk is om een **accuraat en kostenefficiënt gebaarherkennings-systeem** voor VR te ontwikkelen met behulp van *machine learning*.



Figuur 1: SMOG-gebaar voor toilet [2]

Methode

In eerste instantie werden verschillende **sensoren** om gebaren waar te nemen geselecteerd op basis van criteria zoals gebruiksvriendelijkheid en kostenefficiëntie. De gekozen sensoren omvatten de dieptecamera RealSense D435f van Intel (zie Figuur 2), de webcam van een HP laptop 15s-eq2047nb en de smartphonecamera van een Samsung Galaxy A50. Aangezien er geen

bestaande dataset beschikbaar was voor SMOG-gebaren, werd een **eigen dataset samengesteld**. Zeven specifieke gebaren werden vastgelegd, door 13 verschillende uitvoerders. Figuur 3 toont een uitvoerder die het SMOG-gebaar voor "toilet" uitbeeldt. Deze video's werden vervolgens gebruikt om een *machine learning*-model te trainen, dat met slechts een paar voorbeelden kan leren



Figuur 2: Intel RealSense D435f [3]



Figuur 3: SMOG-gebaar voor toilet uitgebeeld door een uitvoerder

Resultaten

Tabel 1: Nauwkeurigheid van de *transfer learning*-modellen

	1 epoch	2 epochs	4 epochs
laptopwebcam	47,9%	82,9%	95,0%
smartphone-camera	53,6%	84,3%	99,3%

Uit de literatuur blijkt dat **transfer learning** een geschikte methode is voor het trainen van een model met een beperkte hoeveelheid data. Bij gebruik van een **smartphonecamera** - de sensor die het best scoort op het vlak van gebruiksvriendelijkheid en kostenefficiëntie - vertoonde *transfer learning* met **MoViNets** bij 4 epochs de hoogste nauwkeurigheid voor het correct classificeren van gebaren, namelijk 99,3%, vergeleken met 95,0% bij de laptopwebcam (zie Tabel 1).

Conclusie

De resultaten tonen aan dat het mogelijk is om gebruikers in een VR-omgeving te filmen, terwijl zij gebaren uitvoeren en deze gebaren met een **hoge nauwkeurigheid correct te classificeren**. Deze masterproef geeft hiermee de aanzet tot het

uitbouwen van een **virtuele leeromgeving** in combinatie met **immersive storytelling**. Met deze aanpak kunnen alle SMOG-gebaren geïntegreerd worden, zodat jongeren de gebaren spelenderwijs kunnen leren en er actief mee aan de slag kunnen gaan.