PROJECT BASED LEARNING

Visual Speech Recognition with State-of-the-Art Optical Flow Camera

Short Description

This project aims to develop a visual speech recognition system with a state-of-the-art optical flow camera.

Introduction

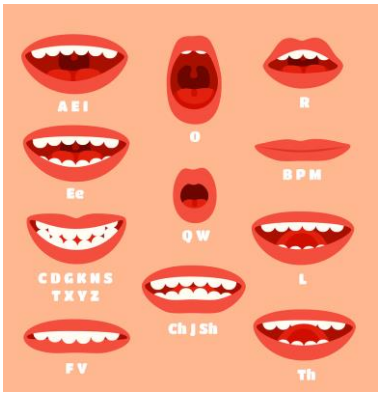

Figure 2 Mouth shape for alphabets

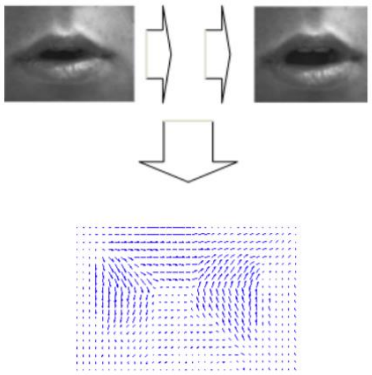

Figure 1 Optical flow of mouth [Makkook & Basir, SMC 2007]

Humans understand speech not only based on auditory information but also based on *visual information*. When we speak, we should move our mouth properly according to the pronunciation of words. Therefore, the shape of our mouth provides visual information for listeners to correctly understand the word. Prior work has proposed several machine learning-based approaches for visual speech recognition (also known as lipreading). One of the approaches is to extract the optical flow of mouth from images and use the optical flow information for visual speech recognition[1][2]. This project is to develop a visual speech recognition system with a state-of-the-art optical flow camera from ST. With the state-of-the-art optical flow camera, it is unnecessary to extract optical flow from mouth images, enabling fast visual speech recognition.

Application Scenario

The visual speech recognition system can help people to recognize speech even without audio. Therefore, the system can be helpful when the audio is not clear. For example, when people cannot hear a speech in a video due to loud noise, the system can help people to recognize the speech better. Furthermore, as future work, it is possible to enhance speech recognition by fusing auditory and visual information. In other words, we can develop a robust and efficient audio-visual speech recognition system[3] with the state-of-the-art optical flow camera.

Goal & Tasks

- Analyze the optical flow of mouth in speech
- Implement a visual speech recognition system with the optical flow camera
- Evaluate the visual speech recognition system

Prerequisites

- Programming skills in Python
- Experience with deep learning frameworks
- Background on recurrent neural networks (RNNs)

Supervisors

Dr. Michele Magno, D-ITET PBL , michele.magno@pbl.ee.ethz.ch
Dr. Seonyeong Heo, D-ITET PBL, seoheo@ethz.ch

Character

- Student team project
- Bachelor thesis (BA)
- Master semester project (SA)

[1] A. A. Shaikh, D. K. Kumar, W. C. Yau, M. Z. C. Azemin and J. Gubbi, "Lip reading using optical flow and support vector machines," 2010 3rd International Congress on Image and Signal Processing (CISP 2010), 2010.
[2] M. Makkook and O. Basir, "Visual speech understanding using independent component analysis," 2007 IEEE International Conference on Systems, Man and Cybernetics (SMC 2007), 2007.
[3] T. Afouras, J. S. Chung, A. Senior, O. Vinyals and A. Zisserman, "Deep Audio-visual Speech Recognition," in IEEE Transactions on Pattern Analysis and Machine Intelligence (IEEE TPAMI), 2018.