ICC'2020 Project Proposal

Team G'express

1. Project Synopsis

A lot of Human Machine Interactions are used In our daily life such as keyboard, mouse, tactical screen, voice control etc. There is one type of interaction which can provide a more natural way: Gesture. In recent days, there have been research and application of gesture control in smart homes, smart vehicles as well as VR and AR.

It is noticed that students, coworkers and businessmen start to use more and more online tools to connect with each other, especially since the pandemic of COVID-19. A lot of online platforms provide VIP courses for kids, with language, maths or other subjects. Unlike the traditional teaching, online courses lack the classroom environment and human interaction. Children are more likely to lose their attention to the teacher, at the same time, the teacher lacks the surveillance of the students. The project aims to improve the online user experience by more attractive interaction with the gestures.
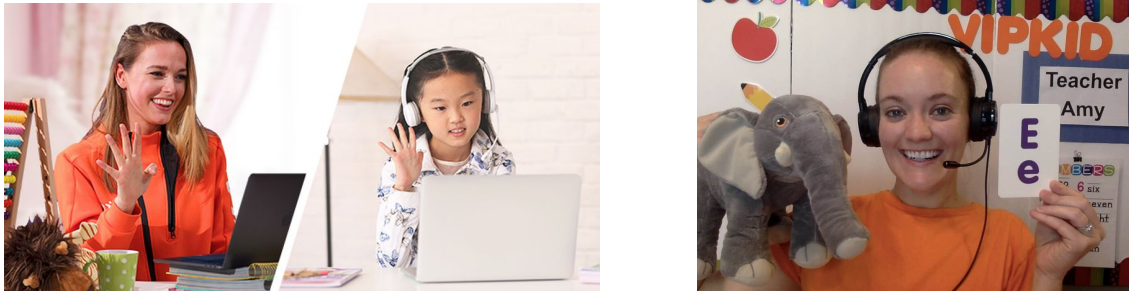

Image sources: Teaching After Hours: The Rise Of International Online Teaching
Work From Home Teaching English! A VIPKID Review!

Through our project G'express, by meaning of Gesture Express, we want to make a prototype of a gesture interaction system that can detect the gesture used by teachers or children and enhance the gesture expression by emoji, sound and animation added on the screen to avoid the disconnection of natural interaction through the online video communication. The users can make the pre-setting gestures such as Raising Hand, Great, etc. or other gestures that are registered by the users, then some visual effects will be triggered on the streaming window.

2. The Team

We have three members with an Engineering background studying or working in France and very passionate about the AI field and their application. All the members have several professional experience in different industries such as IT, automotive, electronic, transport etc. Here are our profiles and our contribution to the project:

Di AI

Di pursues an AI specialized Post-Master's degree at MINES ParisTech, and is currently working at INRIA (French Institute for Research in Computer Science and Automation) as a research engineer intern in Deep Learning and Computer Vision.

He's passionate about machine learning & deep learning algorithms, mastering Python programming skills, and experienced in IT project management. His role in the project group is deep learning network design, Python development as well as project management.

Long JIAO

Long is an Engineering student in Data Science and AI at MINES ParisTech, Currently working as a research intern at EDF R&D Center on AI application of Human Movement Recognition. The role in the project group is use case study, algorithms research, architecture building and python programming.

Zaiwei ZHANG

Currently working at Brake Business Unit of Hitachi Automotive Systems, Zaiwei is a bench and tools development engineer, who is in charge of developing continuous integration and DevOps tools for electric parking brake software as well as developing testing methods and testing bench. Also intrigued by AI technologies, her role in this group is the user interface creation.

3. Project Description

Gestures

After analysing the user experience and the situation encountered through the online courses, we proposed several gestures and the effects that can be associated to enhance the expressions.

- **Raising Hand**: When a child student raises the hand in front of the camera, a visual or sound alert will be sent to the teacher.
- **Great**: When the teacher gives a "Great" gesture, a corresponding Emoji will be added to the screen.
- **Fish Animation**: When the teacher are telling a story, he/she may make the fish swimming gesture, a fish animation will be added to the screen at the same position of the hand.
- **Snow Animation**: When the teacher is doing the snow fall gesture, a full screen animation of snow will be added to the screen to give more visual effect.

Gesture	Raising Hand	Great	Fish swimming	Snowing
Original Screen Output				
Enhanced Screen Output				

In addition, we would also like to let the user record their own gesture and name the new class, the algorithm can learn the new class through one or few shot learning. Which can give the users more freedom to add personal preference to the recognition system.

Proposed method

To achieve the goal, we will propose a gesture recognition method using a web camera, without any additional hardware and sensors.

Firstly, we will use the open-source and light-weight hand tracking deep learning frame MediaPipe to extract the hand keypoints (skeletal data) of each gesture sample video. Some examples of MediaPipe output are as-below:
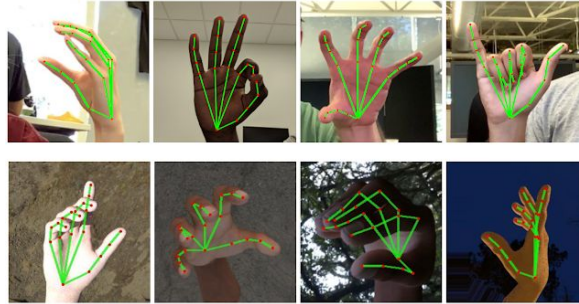

Image Source: [On-Device, Real-Time Hand Tracking with MediaPipe](#)

Then we will train an auto-encoder to generate embedding of hand keypoints. The output of auto-encoder will be compared with ground-truth hand keypoints (output of MediaPipe) to compute the loss of auto-encoder. In order to train the network with few-shot learning, we will use a Siamese network. Triplet loss will be computed between anchor, positive and negative gesture sample videos.

During inference, the model input will be real-time video. After extracting hand keypoints using MediaPipe and generating embedding using encoder, the model will compute the similarity of this the embedding with all pre-defined gestures' embeddings. The gesture class of the most similar embedding will be the model output. The framework of proposed method is illustrated in below diagram:
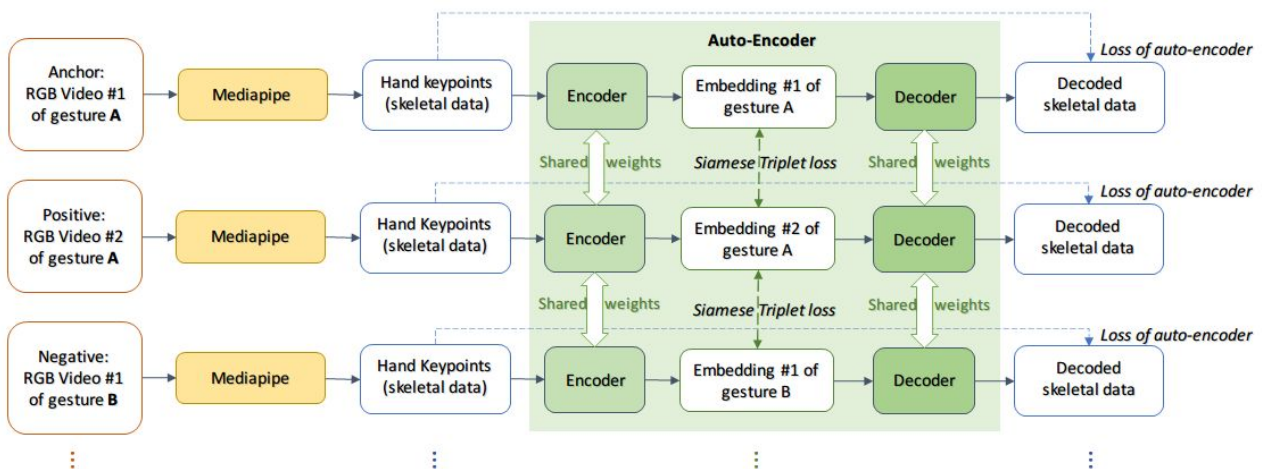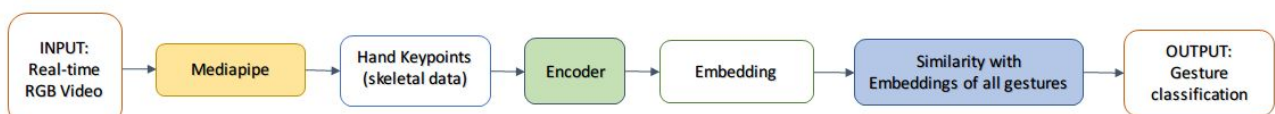
Framework of proposed method

Objective

What we want to achieve during the 9 days Hackathon program is to set up a prototype of this web camera real-time gesture recognition system through the proposed deep learning framework as well as a web-based user interface, in order to verify the feasibility of the ideas and try to improve the accuracy and to find the limitation of the system. This is a good start for us to build our own application of gesture recognition systems and to be able to apply to various domains in the future.

Planning

The planning of our 9 days can be organized as following:

- Define the user case and create the dataset of gestures (Day 1 and 2)
- Set up the pipeline of skeleton output through the Deep Learning algorithm (Day 3)
- Data collection and feature extraction/selection (Day 4)
- Set up the classification algorithms (Day 5 and 6)
- Visualisation of the recognition results (Day 7)
- Improve the classification algorithms by accuracy, latency and one shot learning (Day 8)
- Presentation of the prototype (Day 9)

Of course, some of the tasks can be done in parallel by different members of the team to gain more time.

4. Beyond ICC'2020

If the prototype is satisfying, we plan to create an API based on the prototype and make it usable for various platforms and applications for the developers. It would be better to introduce our work through the conferences or workshops and make them available to more people in the AI community and keep improving the prototype. We think the gesture interaction will be a big trend for the next generation of daily life technology such as AR glasses. The idea of creating more human based interaction will be a necessary element in the application of the near future.

5. Additional input

To be able to achieve the prototype, we need a good hardware environment that can be provided such as a camera and graphic card.