DOLL·E
Gesture-Controlled Automation For Remote Filmmaking

IRI Research Project 2021-2022

Presented by Olivia Loh

Introduction

Olivia Loh

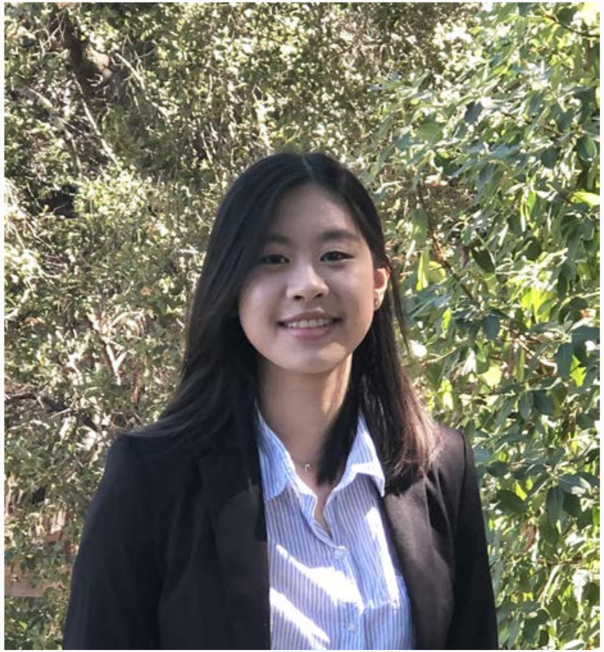
Year: Senior

Major: Computer Engineering

Minor: Film, Television, Digital Media (FTVDM)

Project: Gesture-Controlled Automation for Remote Filmmaking

Mentor: Professor Jeff Burke



DOLL·E

Conception

Motivation

- Effects of COVID on film industry:
  - Social, collaborative, and practical field
  - Film production slowdown due to social distancing
- Remote-filmmaking and virtual production mitigates this slowdown

- Enhance remote-filmmaking by integrating gesture-control to create a more intimate filming experience and convenient user interface

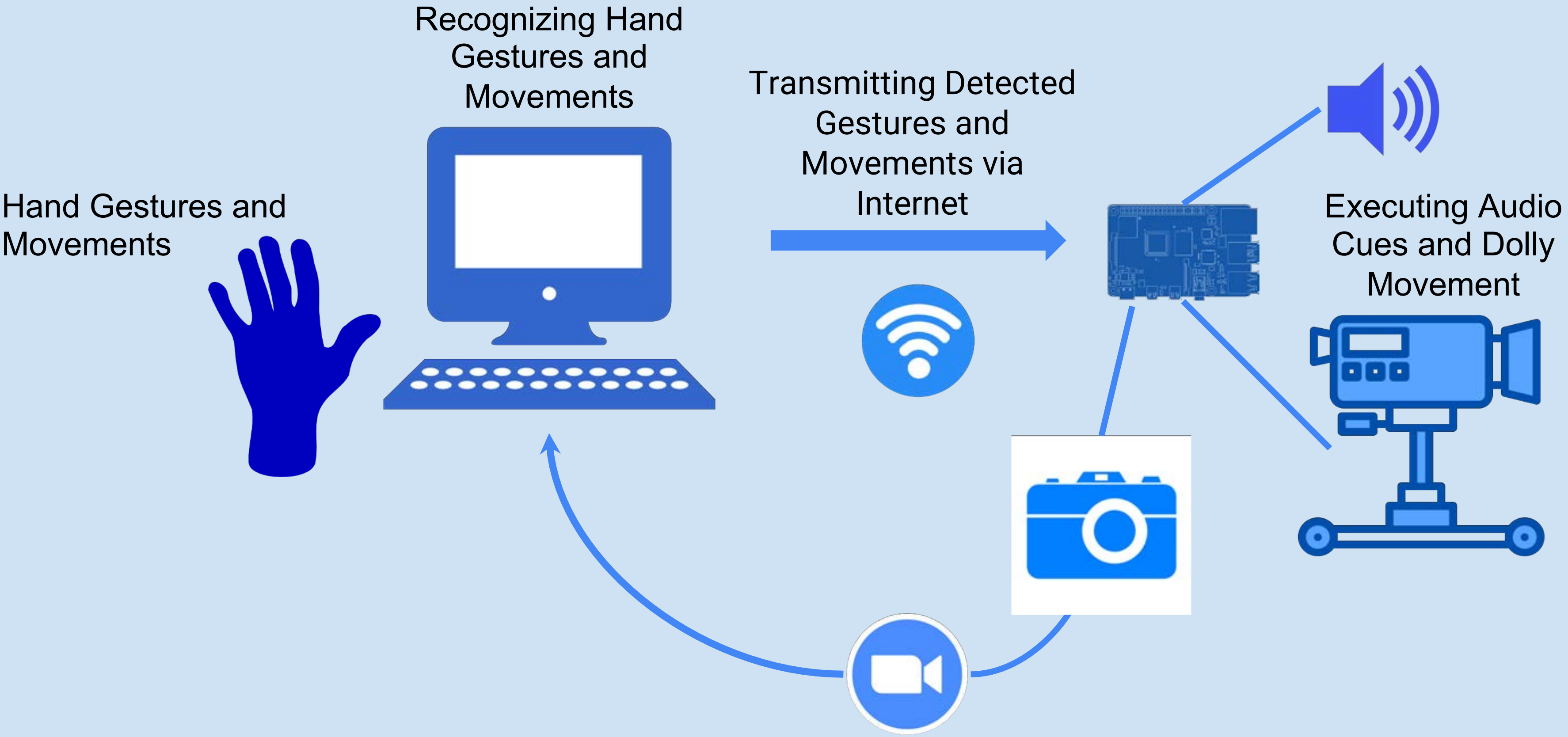
Background

- Current remote filmmaking software facilitates high-quality end-to-end live streaming service
- Gesture control implemented in automobiles and smartphone
- Ongoing research on gesture recognition using deep learning
- I propose to investigate the applications of gesture control as a new form of teleoperation for physical virtual work

Field Study and Use Cases

- When is gestural control useful?
  - The “subtlety and dynamic range of fingers” as opposed to buttons or voice control
- When is remote control useful?
  - During the pandemic. When crew members need to quarantine, they can still be “present” on set
  - Shooting at overseas locations. Flying less crew to locations cut travel costs.
  - Operating large equipment: Lighting equipment, Jibs, Dolly, Cranes, Mechanical Effects
  - Filming in difficult situations, such as aerial shots, underwater shots, cold or hot climate
- Use Cases for Different Film Set Roles:
  - Cinematographers
    - Smoothly control fluid head tripod to produce subtle moves in shot.
  - Actors
    - Can more naturally drive their own action instead of automated mechanical effects,
  - Directors
    - Communicate with their “splinter” (second) crew

DOLL·E

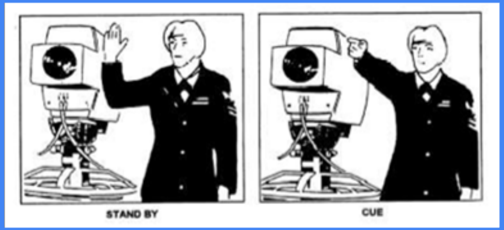

Hand Gestures and Movements
Recognizing Hand Gestures and Movements
Transmitting Detected Gestures and Movements via Internet
Executing Audio Cues and Dolly Movement

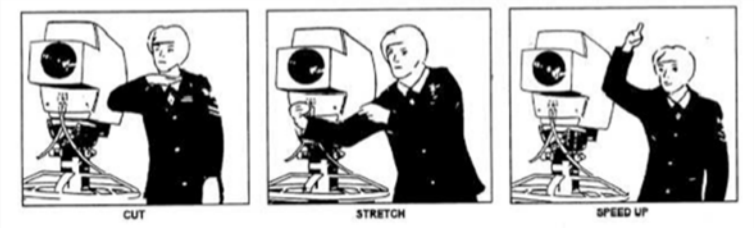
DOLL·E
Progress

Hand Gestures and Movements

Choosing Gestures

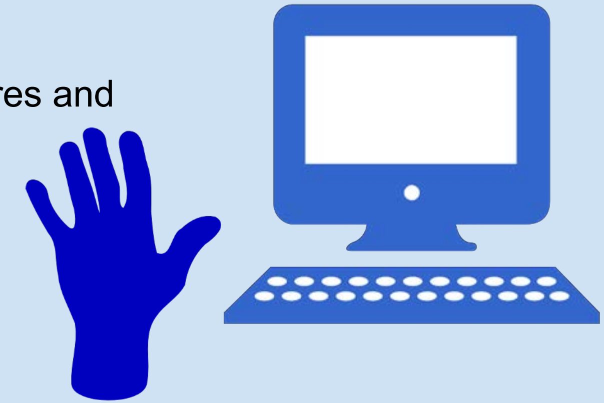
- Objectives:
  - Intuitive and Natural
- Many filmmakers with different roles on set (e.g. cameraman, lighting, etc.)
  - For the purposes of my project, I decided to focus on cinematographer
- Chose 4 hand signals:
  - Cue → Camera Rolling
  - Palm Up → Move dolly: Forwards, Backwards, Left, Right
  - Fist → Stop dolly
  - Cut → Stop





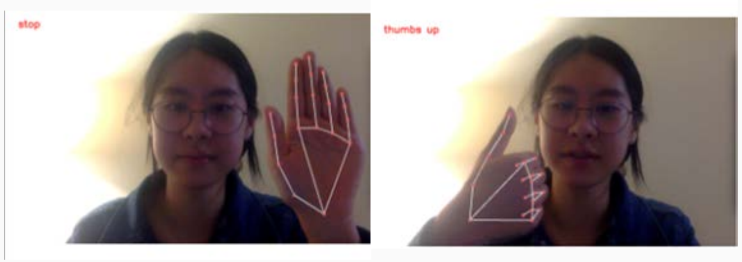
Recognizing Hand Gestures and Movements

Hand Gestures and Movements



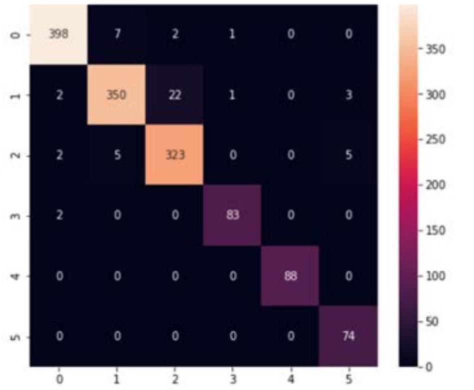
Hand Gesture Recognition: Skeletal Approach

- Computer Vision:
  - Pros:
    - Completely contactless HCI
    - A simple webcam would suffice
  - Cons:
    - Change in lighting conditions
    - Occlusion
    - Background colors (depend on vision technique)
- Skeletal method:
  - Perform hand segmentation by calculating 3D connections and Euclidean distance over hand skeleton pixels
  - Good for dynamic hand gesture recognition



Hand Gesture Recognition: Skeletal Approach

- MediaPipe API
  - Uses regression (direct coordinate prediction) to robustly locate 21 3-D points of hand. Dataset of ~30K labelled images serves as ground truth
- Tensorflow Library
  - Multi-layer perceptron network. Takes in vector input and uses two ReLU hidden layers and one softmax final layer to output class probability score
- MediaPipe and Tensorflow Open-Source Example: https://github.com/kinivi/hand-gesture-recognition-mediapipe/
  - Came with pre-trained model and dataset of poses
  - Added additional pose data and re-trained model:
    - Training data: 4 different poses with 1000 sets of 21 hand points each

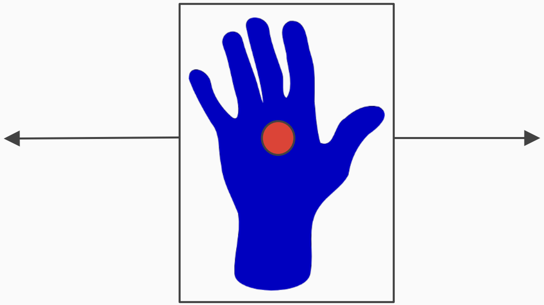


Classification Report

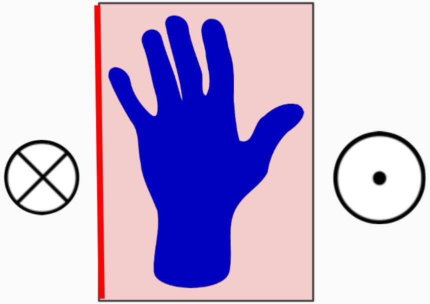
	precision	recall	f1-score	support
0	0.99	0.98	0.98	408
1	0.97	0.93	0.95	378
2	0.93	0.96	0.95	335
3	0.98	0.98	0.98	85
4	1.00	1.00	1.00	88
5	0.90	1.00	0.95	74
accuracy			0.96	1368
macro avg	0.96	0.97	0.97	1368
weighted avg	0.96	0.96	0.96	1368

Metrics for Movement Detection: Two-Axis

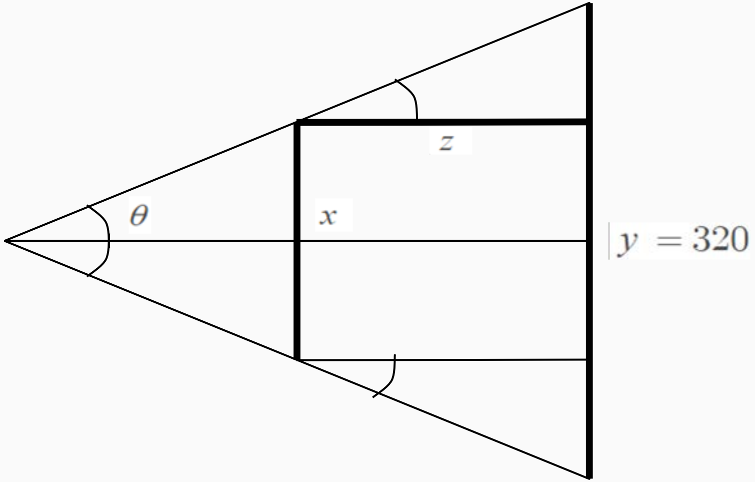
- X-axis
  - Midpoint of bounding box



- Z-axis
  - Length of bounding box
  - Area of bounding box

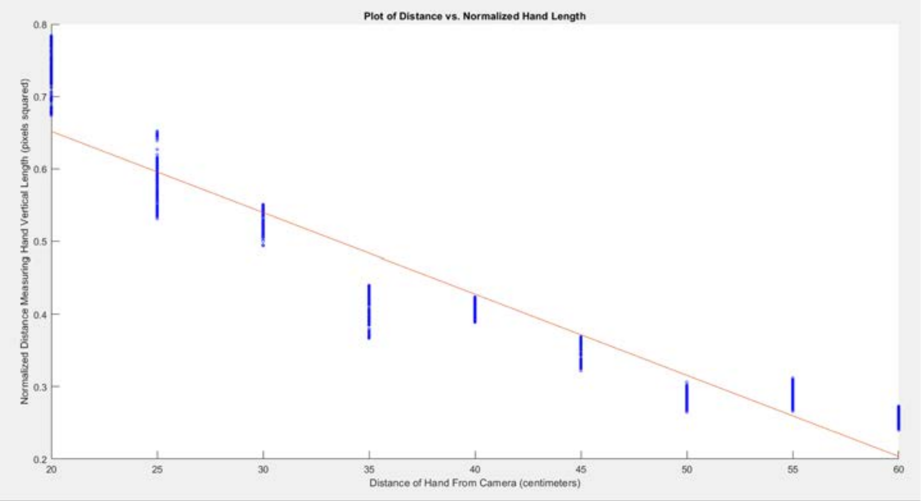


Least Squares Regression: Curve Fitting for Depth as a Function of Length



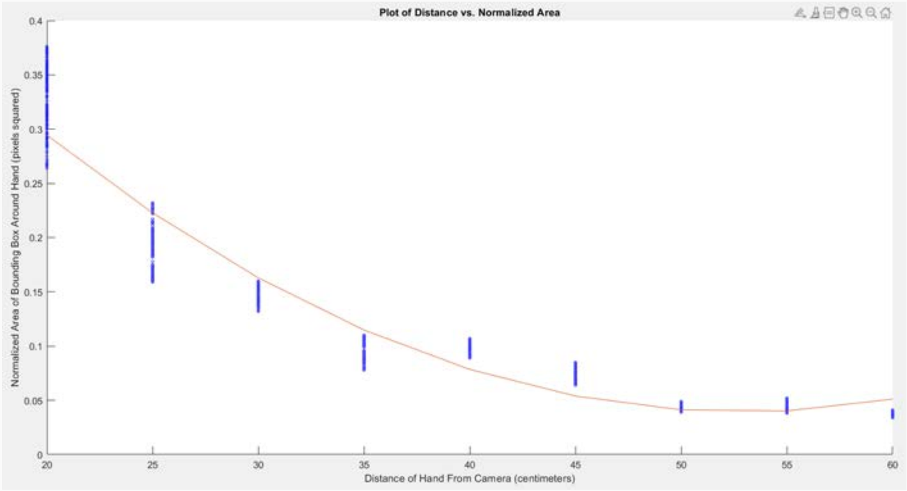
$$(x-y)^2 = 4z^2a^2 \text{ , where } a^2 = \left(\frac{1}{cos^2\theta} - 1\right)$$

$$x = y - 2az,\ y + 2az$$



Least Squares Regression:
Curve Fitting for Depth as a Function of Area

- Quadratic model
- Yields lower bias and variance

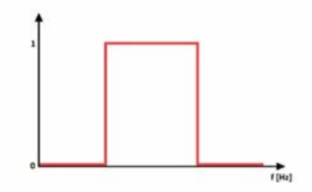


Noise Corrections

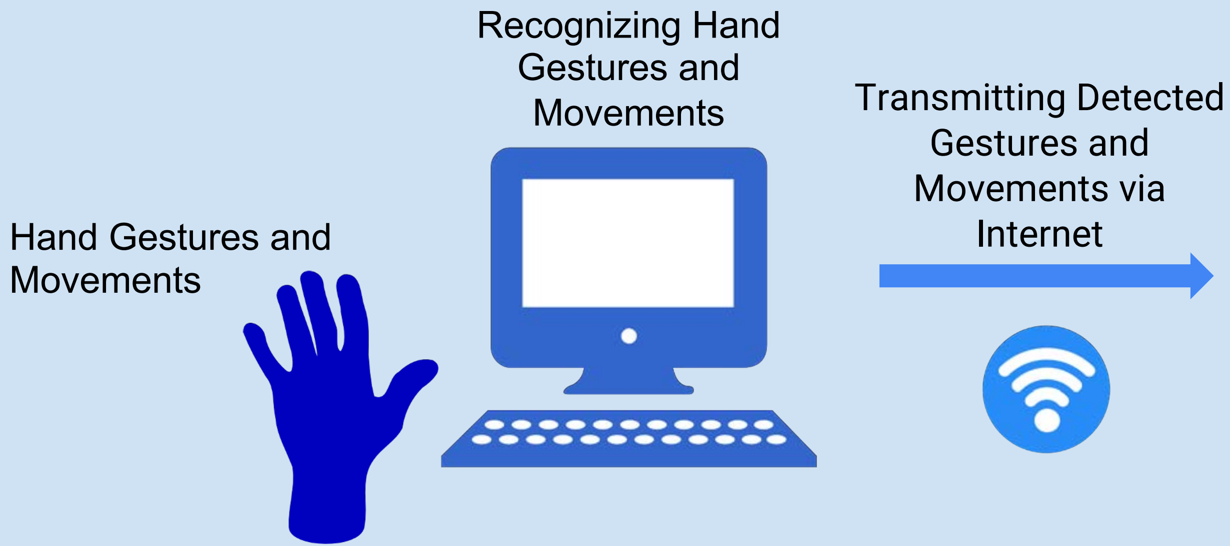
1. Utilize the area of bounding box and previously fitted function to determine instantaneous velocity.

$$\frac{dA}{dt} \cdot \frac{dz}{dA} = \frac{dz}{dt}$$

1. Attenuating resulting values that surpass the lower and upper threshold.

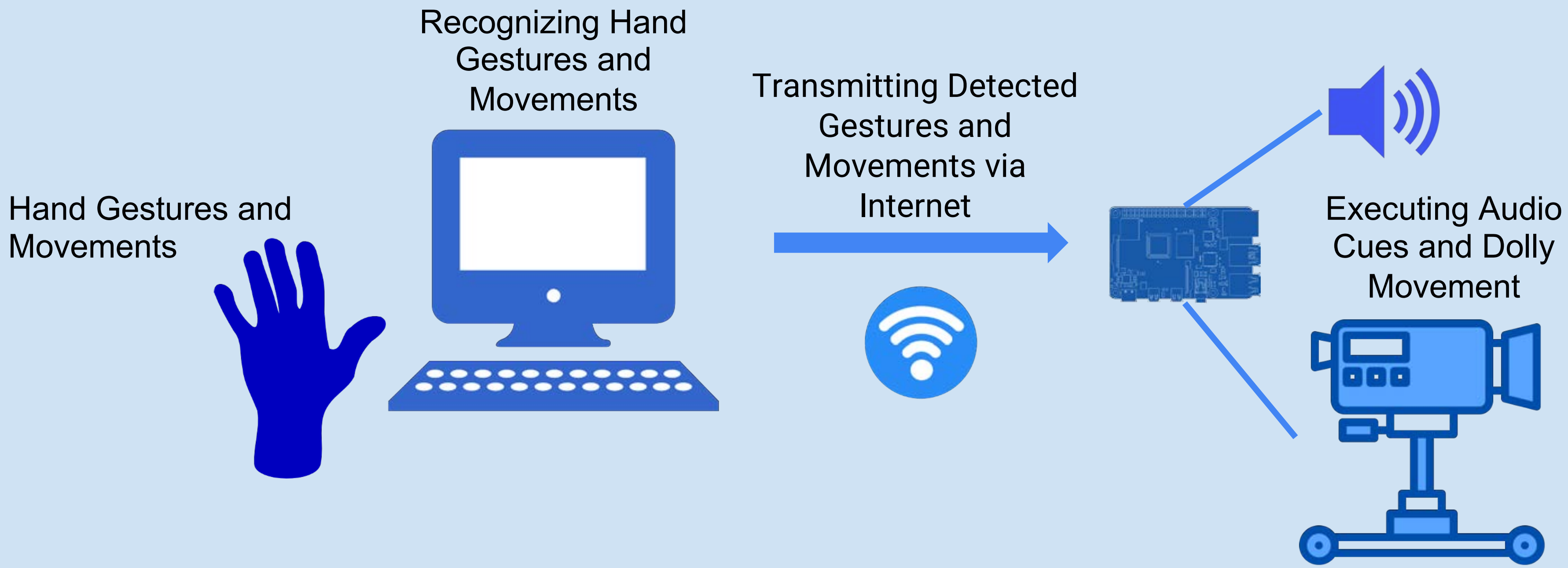


1. Map into a suitable value for varying speed: [-0.4, 0.4] → [-30, 30]
2. Baseline speed + this value: [70 - 30, 70 + 30]


Hand Gestures and Movements
Recognizing Hand Gestures and Movements
Transmitting Detected Gestures and Movements via Internet

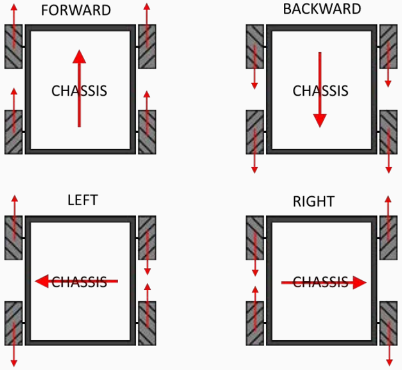
Communications

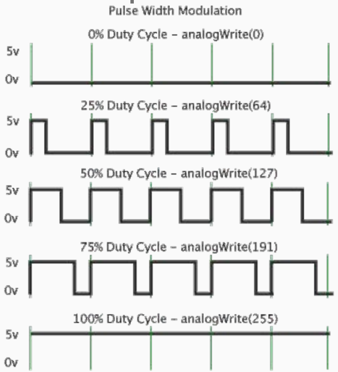
- Uni-directional Communication
    - Traditional Server Client Model
    - MQTT Protocol:
        - Publish/Subscribe to organized topics
        - Suitable for controlling IoT devices
        - Lightweight, easy to implement for prototyping
        - Low power consumption
        - (also capable of bi-directional communication)


Hand Gestures and Movements
Recognizing Hand Gestures and Movements
Transmitting Detected Gestures and Movements via Internet
Executing Audio Cues and Dolly Movement

Executing Dolly Movement

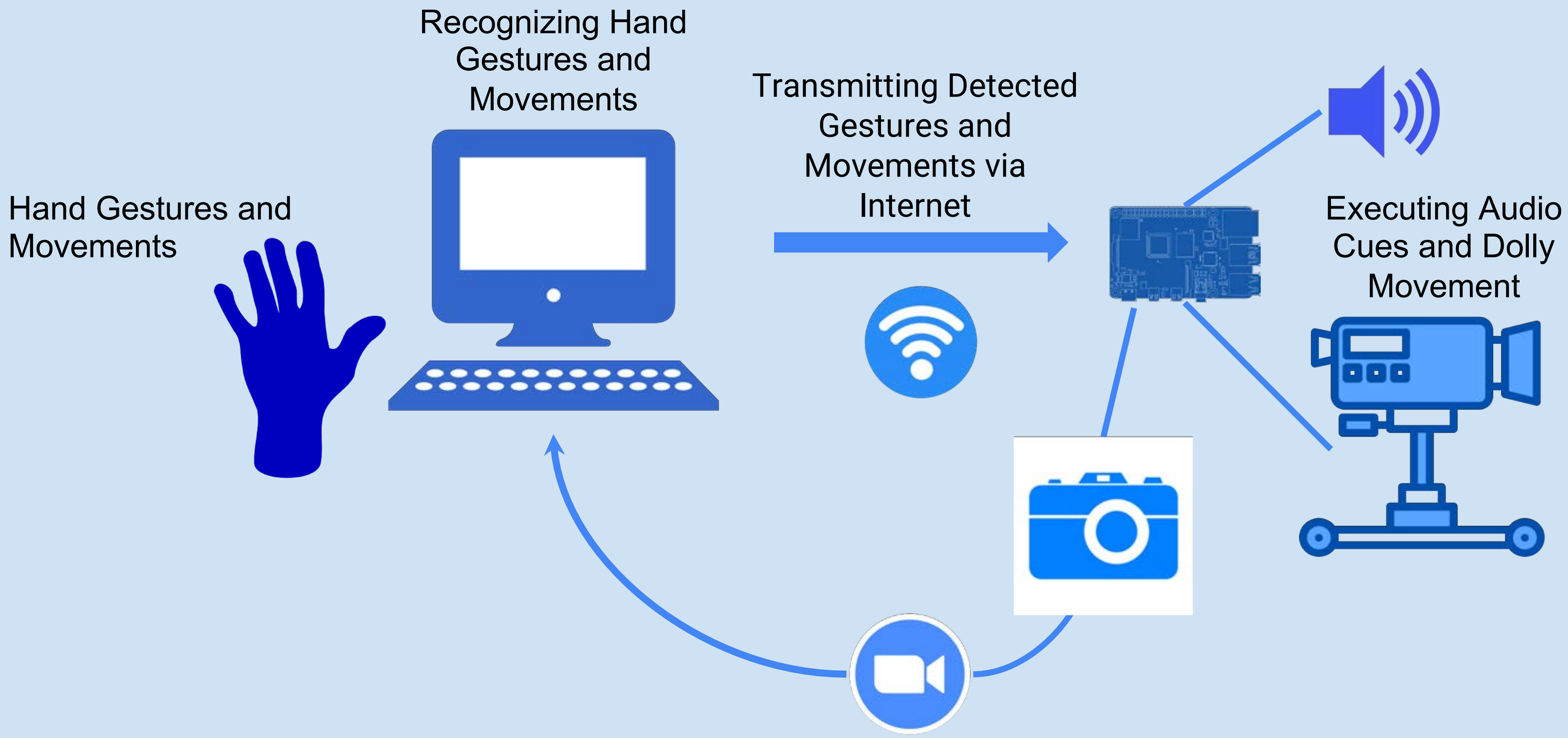
- Robot Design:
  - Simulating a real-life film dolly
  - Mecanum wheels for smooth forwards, backwards, right, left movement. No turning required
  - Motors driven by PWM pins to control speed





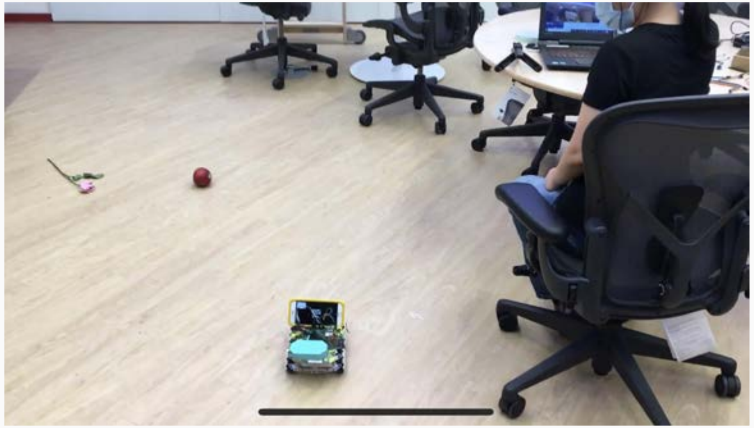
DOLL·E
Results

DOLL·E


Recognizing Hand Gestures and Movements
Transmitting Detected Gestures and Movements via Internet
Hand Gestures and Movements
Executing Audio Cues and Dolly Movement

User Study

- Users were asked to move the dolly around to frame a shot of an apple and a rose lying next to each other



Demo Video

Survey Results

- Tasks Timing:
  - Each user took around 40s to 80s to complete the tasks
- Gestural Control & UI
  - Easy to Learn Controls: 4.83
  - Intuitive Interface: 4.33
  - “Natural-ness”: 4.5
  - Lag: 4.167
- Remote Control & Mechanical Automation
  - Speed of robot reflects hand speed: 2.5
  - Direction of robot reflects hand direction: 3.833
  - Lag: 4
- Filming with the Robotic Dolly
  - Fluidity of robot motion, “cinematic-ness”: 3.417
  - Time spent setting up a shot: 3.667

DOLL·E
Conclusion

Conclusion

- Using the Mediapipe open source tool for skeletal-based computer vision yielded noise that affected the program's performance.
- Area as a more reliable metric than distance.
    - Although this could have been be affected by noise
- Users agree that hand gestures for controlling the robotic dolly felt natural, but the speed of dolly movement did not fully reflect the speed of their hand motion.
- Speed determined by pixel distance or area was harder to model than I expected
    - Longer sample window → higher accuracy
        - Instantaneous velocity is not accurate due to jitter from inaccurate samples
    - Trade-offs:
        - Shorter sampling window → faster processing and message transmission

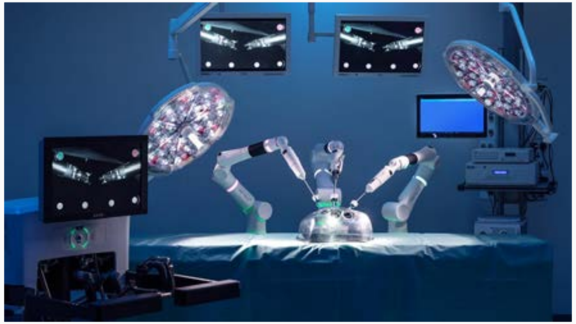
Next Steps

- Techniques for better sampling and noise elimination
- Issue of “resetting” hand motion upon hand reaching edge of screen
  - Most students found this unnatural and cumbersome
- Better hardware
  - Most students attributed dissatisfaction of mechanical automation to hardware limitations
  - Motors with more torque
  - PCB and soldered components instead of breadboard and loose wires
- Optimize algorithms for less lag and higher efficiency
- Implement a third axis (y-axis) to enable a tilt up and tilt down option on dolly’s camera

Other Applications

- Disabled
- Elderly
- Physical labor
- High risk construction work
- Surgical robotics
- Medical treatments



Special Acknowledgments

Professor Jeff Burke

Professor Leonard Kleinrock

Zhaoyu Wu

Professor Susan Littenberg

Professor William McDonald

Thank you! Any questions?