Head Movement and Facial Expression Detection: An Alternate Input System for Gamers with Muscular Disabilities

Felicity Huang Yujung "Jenny" Kim Alyssa Koh

Stanford University, Department of Computer Science

Introduction

Background:

- **Video games** provide social and emotional experiences and are widely used in education, rehabilitation, and health interventions.
- **Individuals with muscular disabilities** often face barriers in accessing video games, impacting their ability to participate in recreational and social activities.

Alternative Computer Input Systems for Accessibility:

- **Hands-free Human-Computer Interaction (HCI)** use devices like the Emotiv EPOC+, which allows control through motion sensors and brain signals. [Expensive, inaccessible, hard to initialize]
- **Eye movement-based wearables** detect gaze and blinks [Error- and noise-prone]
- **Mouth-based controllers** involve a wireless intra-oral module uses tongue-controlled buttons. [Expensive, discomfort over long time]

Head Position and Facial Behavior Analysis Tools:

- **Convolutional neural networks (CNNs):** Facial features -> head pose estimation [Significant computational power, intensive training]
- **Multi-view learning algorithms:** Information from various camera angles -> reduce errors from occlusions or extreme head poses [Complicated setup, expensive]
- **OpenFace (open source):** Geometric features -> real-time facial landmark detection, head pose estimation, facial action unit recognition.

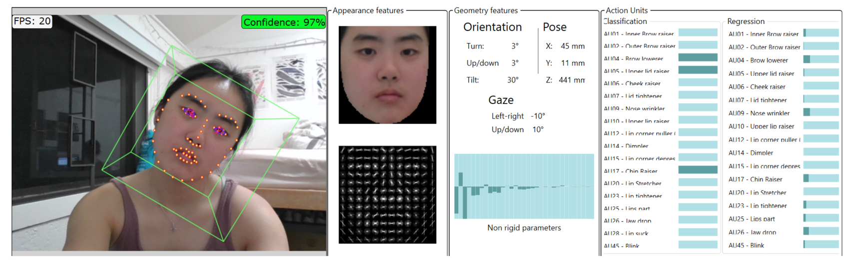
Problem Statement

- **Enhance accessibility** in gaming to improve quality of life for those with disabilities.
- **Develop a deep neural network-based, hands-free gaming approach** using facial movements captured using a simple webcam and processed with the OpenFace model.
- **Ensure low latency, high precision, and enjoyable gameplay** through testing

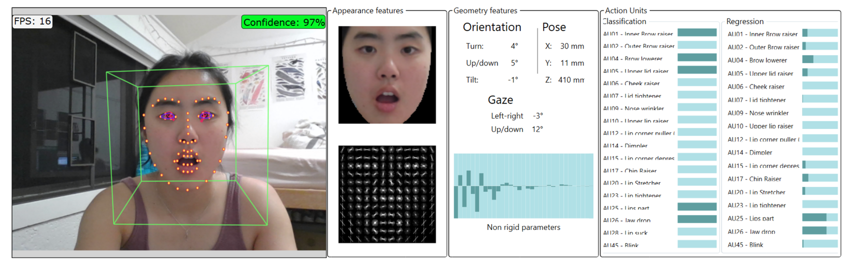
Test Game: Mario Kart Wii [Well known game, clear and simple gameplay, requires precise control]

Dataset

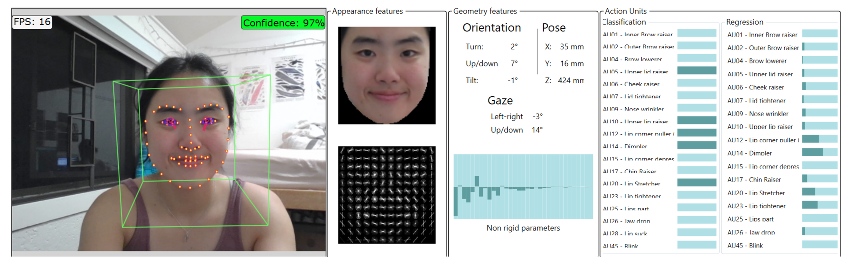
Webcam captures of head tilts and facial expression units to be translated using OpenFace to mouse/keyboard inputs for Mario Kart Wii control.



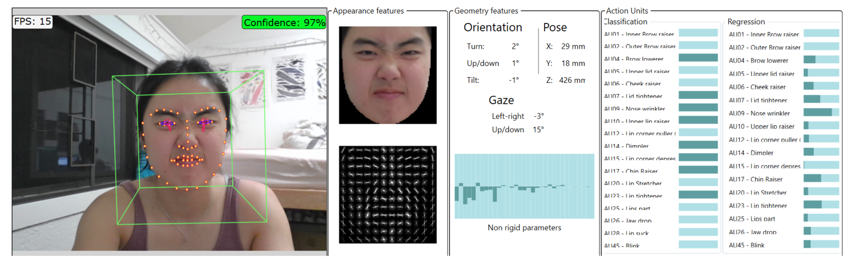
Tilt: Left-and-Right Continuous Control



Open-Mouth: Acceleration through Button "A" Press



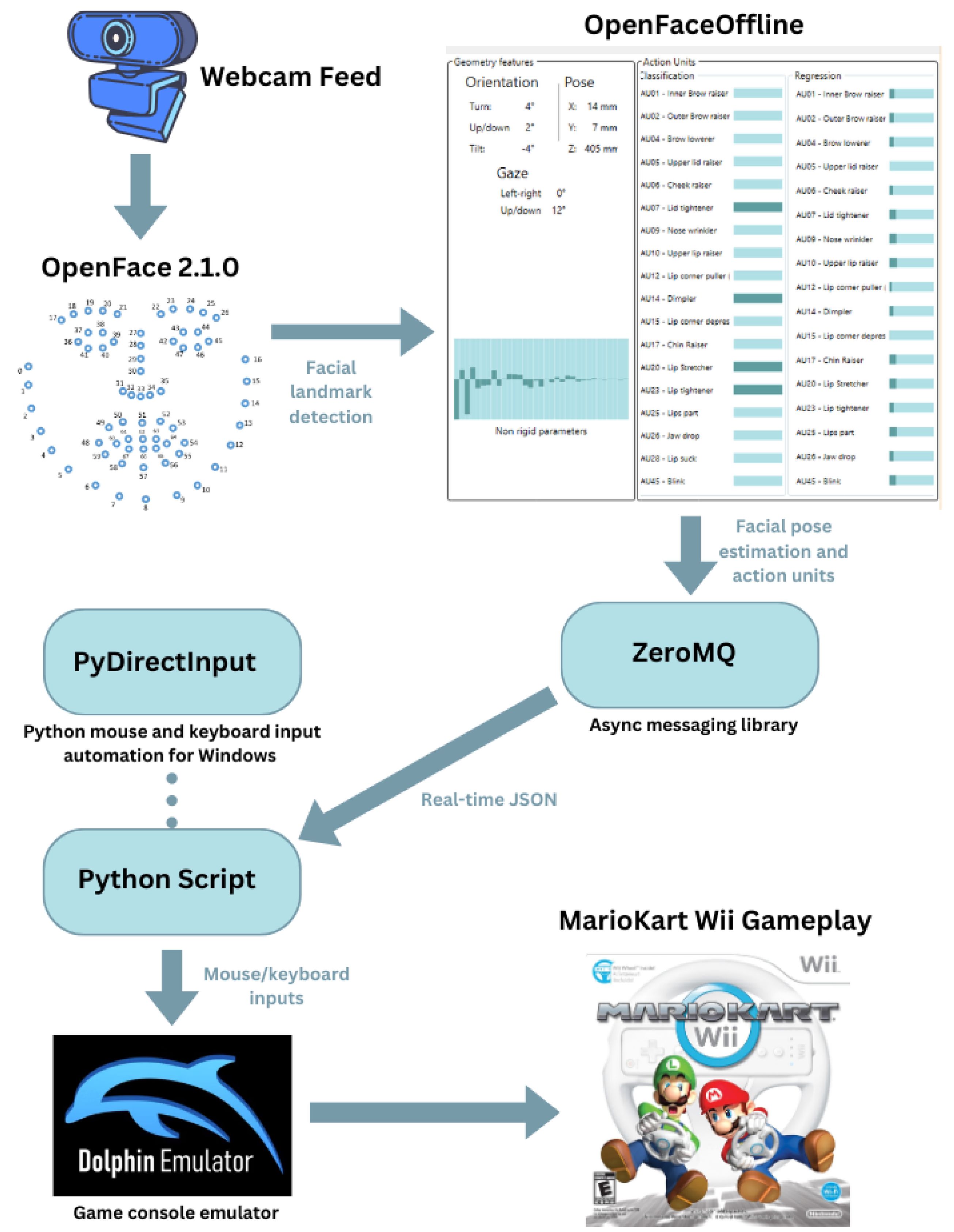
Smile: Brake/Drift through Button "B" Press



Nose Wrinkle: Item Use through Button "Z" Press

Process

End-to-end Pipeline Diagram



OpenFace:

- **Facial Landmark detection**: Conditional Local Neural Fields - projects shape variations of 3D facial landmarks onto face with orthographic camera projection.
- **Head Pose Estimation**: Uses landmarks
- **Facial Action Unit (AU) recognition**: Linear kernel SVM (AU presence) - person-specific recognition by subtracting median value of features observed so far from current frame estimates.

Facial Input to Game Control Mappings

Player Actions	JSON Action Units	Mouse/Key Inputs	Mario Kart
Head Tilt (L/R)	Pose_Rz (-/+)	Mouse drift (L/R)	**Turn left/right**
Open Mouth	Lip part	Mouse click	**Accelerate**
Smile	Upper lip raise, lip corner puller, dimpler	Mouse right click hold	**Reverse/Drift**
Nose Wrinkle	Lid tightener, nose wrinkler, upper lip raiser	"z" key press	**Use item**
Eyebrow Raise	Inner brow raiser, outer brow raiser, upper lid raiser	"c" key press	**Look back**

Experimentation

Subjects: 5 Participants (Burmese, Chinese, Hispanic/Latino, Indian, and White-Asian; 3 female, 2 male)
Testing tools: Laptop webcam, OpenFace output, Python mouse/key log

1. **Facial Tracking Accuracy** - With different lighting (dim, bright) and people
   - Players record synchronized timed sequence of head movement and facial expressions
   - Examine OpenFace Accuracy of videos
2. **Gameplay Reliability** - Consistent execution of inputs
   - Complete 1 lap of Mario Kart track "Yoshi Falls"
   - Players review gameplay footage and mark inaccuracies:
     - False Negatives: Inputs not registering
     - False Positives: Unintended actions occurring
     - Delays
     - Other difficulties

Results

Average Accuracy of OpenFace Tracking in Dim and Bright Lighting

Avg Accuracy	Player A	Player B	Player C	Player D	Player E	Lighting Avgs
Dim	0.981	0.917	0.996	0.912	1.0	0.961
Bright	0.984	0.918	1.0	0.920	1.0	0.964
Player Avgs	0.982	0.918	0.998	0.916	1.0	**0.963**

Classification of Each Recorded Action Types, along with Avg Lap Times for Each Player

Player	Action False Pos	Action True Pos	Action False Neg	Avg Lap Time Face	Avg Lap Time Mouse/Keyboard
Player A	6	27	2	0:59:421	0:38:033
Player B	9	29	3	1:13:082	0:36:033
Player C	10	33	5	1:02:155	0:49:833
Player D	7	24	2	1:46:372	1:18:147
Player E	7	27	4	1:10:543	0:41:364
Total Action Counts	39	140	16	N/A	N/A
Overall Avg	7.8	28	3.2	1:17.115	0:38.482

Average Similarities of Game Inputs to Participant Inputs Across Players, out of Total Actions Taken

	Participant Performed Action	Participant Did Not Perform
Game Performed Action	0.7179	0.2000
Game Did Not Perform	0.0821	N/A

Conclusion

Our Program can serve as an fun, functional substitute for traditional analog and button inputs. However, significant issues include:

- Responsiveness of facial inputs (jitter/noise -> higher rate of false positives)
- Results vary with individual differences (tracking, intuitiveness of controls)
- Latency of Program

Future Directions

- Adaptation to other video games or for general computer use
- Comprehensive user testing involving actual persons with muscular disabilities and model fine-tuning