



Multimodal Features for Group Dynamic-Aware Agents

Iliana Castillon*, Videep Venkatesha*, Hannah VanderHoeven, Mariah Bradford,
Nikhil Krishnaswamy, Nathaniel Blanchard

* These authors contributed equally.



Introduction

- For an AI agent to facilitate Collaborative Problem Solving (CPS) tasks, the agents needs to both follow the team's progress through a task and interpret the teams interactions.
- The agent needs access to multimodal communication features
 - Speech
 - Pose
 - Gesture
 - Expression.



Implementation

- **Toolbox:** We examined the available tools which aid in interpreting features present in group work and filtered based on accessibility and ease of implementation
- **Collaboration:** Our ultimate aim is to implement this agent in the real world K-12 classrooms and hence need the feedback from the education community.
- **Privacy:** We also acknowledge the inevitable privacy concerns that accompany data collection and the tradeoff between improving the agents performance and the privacy of the data



A Toolbox for Providing Information to AI Agents

- For an AI agent to decide how to guide an interaction, it needs to consider multiple relevant information channels, such as video and audio
- We outline several categories of features as well as current technologies that can be used to automatically provide said features



Language and Auditory Features

- Language and Auditory features can provide insight into the nature of the groups participants and task progress .
 - Utterances
 - Automatic speech recognition
 - Speaking length and time
 - Prosodic speech features (Energy and Pitch)
- Wavesurf and OPENsmile
- Spacy



Visual Features

- Grants access to the visual world and interactions between groups, CPS task objects and potentially CPS task progress
- 6D Object Pose Estimation can track key CPS task objects.

Visual Features

- Grants access to the visual world and interactions between groups, CPS task objects and potentially CPS task progress
- 6D Object Pose Estimation can track key CPS task objects.





Visual Features

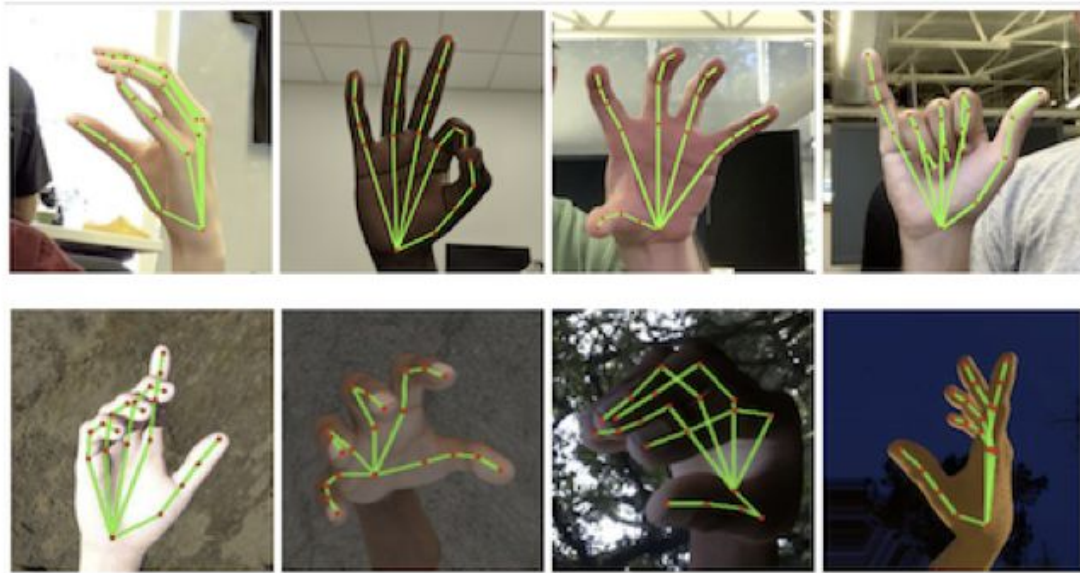
- Grants access to the visual world and interactions between groups, CPS task objects and potentially CPS task progress
- 6D Object Pose Estimation can track key CPS task objects.
- Gestures of participants such as pointing can be passed on to the agent.
 - OpenPose
 - MediaPipe
- Face Detection can help with understanding the participant's attention

OpenPose





MediaPipe





Discussion

- Educational stakeholder input is essential for identifying key classroom features we have overlooked
- The toolbook we covered works out of box
- However, these tools will only allow us to create an initial agent
 - Improving this agent has important privacy implications



Agent Failure

- Why do agents fail?
 - Feature incorrect
 - Processing incorrect
 - Out of training domain
- How should we debug live failures that disrupt the classroom?
 - Saved videos
 - Who has access
 - Agent trainers
 - Teachers
 - Parents
 - Legal ramifications



Thank you!





Questions?