**Introduction:**

Hello, my name is Jeffrey Hoang and I am going to be demonstrating my EasyMocap Web Application. The purpose of my project is to create an automated solution to calibrate cameras using the checkerboard method, and then use those intrinsic and extrinsic parameters to capture and reconstruct challenging human motion from a small number of calibrated cameras. I am currently using an existing Github project by EasyMocap and Python’s Django software to build my web application and automate the process.

**Demonstration:**

So here is the web application. Below is the documentation that gives users instructions on how to set up their dataset and run it through the pipeline. Once the user’s dataset is ready, they must first enter the number of cameras used. The minimum number of cameras is set to 4 because a sufficient number of angles must be used to obtain accurate results. To obtain the intrinsic camera parameters, users must upload their checkerboard videos and specify the number of rows and columns of each checkerboard, as well the square size. The same process goes for extrinsic camera calibration. Users can check the accuracy of their camera parameters by clicking on ‘check calibration’. Images of a cube will display onto the checkerboard. Ensure that the cube aligns with the checkerboard and that dimensions are correct. Click any button on your keyboard to loop through each image and once finished, you will be prompted to move onto the next step. With accurate camera parameters, the last step is to obtain the 3D mesh that captures and reconstructs your motion. I am going to skip through the loading process because it takes a long time but here is an example of the input and output. A summary of how this works is that we use YOLO and HRNet softwares to estimate 2D keypoints of the human body. We then perform multi-view triangulation to obtain accurate 3D key points of the human body and fit the pre-trained SMPL model onto the 3D key points to obtain a mesh that reconstructs the human motion.

**Future Implications:**

This project is not yet complete. Currently, it relies on clap detection for video synchronization, a visible checkerboard to determine external camera parameters, and multiple cameras to capture and reconstruct human motion. While this approach ensures accuracy within a controlled studio setting, it becomes less practical for use in uncontrolled outdoor environments due to the inconvenient setup process. My goal is to extend the applicability of this project to more diverse, real-world scenarios where users can input any images or videos and obtain a 3D mesh that accurately reconstructs their motion.

To achieve this, three key areas of work are required. First, I need to develop a method for synchronizing videos that does not depend on clap detection. Second, I must implement self-calibration or structure-from-motion (SfM) techniques, allowing external camera parameters to be estimated based on multiple views of a scene rather than relying on known patterns like a checkerboard. Finally, I need to create a deep learning algorithm capable of accurately estimating and reconstructing human motion from a single camera view. Successfully addressing these issues will make the web application more versatile, enabling accurate motion capture and 3D reconstruction in a wide range of environments, especially those where traditional methods are not practical.