TEAM PLAN REPORT

Driving Photo-Realistic Avatars using Drone

Team G

Annanya
Andrew Zhang
Ming Feng Li
Naman Menezes
Saksham Kukreja

Industrial Sponsor

Meta

13th February 2024

Carnegie Mellon University
THE ROBOTICS INSTITUTE
Master of Science
Robotic Systems Development

Table of Contents

1. Introduction

This document outlines the series of tests to be conducted by Team G - Sputnik during the Spring 2023 semester. The primary objective of these tests is to validate that the system requirements established during the Fall 2022 semester are successfully fulfilled at various stages of the sprint cycles. Team G - Sputnik will communicate the outcomes of these assessments through progress reviews and the Spring Validation Demonstration. By the time of the Spring Validation Demonstration, we aim to have a fully integrated autonomous human-following drone for avatar driving, seamlessly integrated with Meta Quest 3, meeting all the specified requirements detailed in Appendix A.

2. Logistics

The equipment and test sites required for the project are enumerated below.

- **DJI Mavic 3 Enterprise:** The hardware platform which will be used for human tracking and image capture.”.
- **Drone Cage at AI MakerSpace:** This drone cage serves as the main testing site for our subsystems and the Spring Validation Demonstration. We can utilize this facility for scenarios involving hovering drones.
- **MRSD Lab:** Located on the B-level of Newell Simon Hall, this facility is designated for unit and subsystem bench tests.
- **Onboard SDK:** This software development kit facilitates the streaming of data from the drone to our computational system.
- **Lambda Machine:** This system is utilized for intensive computations required for generating avatar models.
- **Nvidia Jetson Nano:** Serving as the onboard central compute platform for the drone.
- **Meta Quest 3:** Used for visualizing the test results.
- **Meta Quest Pro:** This device provides improved visualization of the subject’s face in cases where camera images may not yield satisfactory results.
- **Wifi Module:** Enhances communication speed between the drone and the system.

3. Schedule

Identifier	Capabilities	Test(s)	Requirement(s)
Progress Review 1 (16 February)	- Human Detection and 3-D pose estimation code works to get results - Avatar code is tested	- T1 - T2 - T3	
Progress Review 2 (1 March)	- Fully functional communication pipeline from drone to the lambda machine - Streaming data from drone to the processing unit - Integrated Human detection and 3-D pose to produce results at 45 FPS and < 50 MAPS	- T4 - T5 - T6	- M.P. 3 - M.P. 4
Progress Review 3 (22 March)	- Get depth estimation with error less than 0.1 m of human with respect to to drone - Integrated system to get 3-D pose estimation and depth estimation - Producing results from the integrated system from real-time data from drone at 45 FPS	- T7 - T8 - T9 - T10	- M.P. 5 - M.P. 6 - M.P. 7
Progress Review 4 (5 April)	- Rendering Avatar at the rate of 36 FPS -Visualization of the driving avatar at 36 FPS with PSNR > 30dB - Hovering Drone which captures images and streams on the server for further operation	- T11 - T12 - T13	- M.P. 1 - M.P. 2
Progress Review 5 (18 April)	- Reconstruct Avatars with Unseen Motion PSNR and Novel View PSNR ≥ 28 dB - Total MSE of avatar over 10 frames will be < 100 cm^2 and FPS should be at least 36 - Percentage of frames capturing the Human should be ≥ 95%	- T14	- M.P. 5 - M.P. 4 - M.P. 6 - M.P. 7 - M.N. 1 - M.N. 2 - M.N. 3

4. Tests

Table 1: Test No. 1: Human Detection Test

Test No. 1: Human Detection Test	
Objective	Validate Human Detection Code Module
Elements	Perception Subsystem
Location	AI Maker Space
Equipment	1. Laptop 2. Webcam
Personnel	Andrew
Procedure	1. Connect the laptop with the Webcam and start live streaming from the camera. 2. Visualize the human detection bounding box.
Verification Criteria	1. The human can be constantly detected even when faced backward. 2. The speed of bounding box generation needs to be higher than 45 frames per second (fps).

Table 2: Test No. 2: 3-D Pose Estimation

Test No. 2: 3-D Pose Estimation	
Objective	Validate 3-D pose estimation model
Elements	Perception Subsystem
Location	AI Maker Space
Equipment	1. Webcam 2. Lambda Machine 3. Personal Laptop
Personnel	Annanya
Procedure	1. Prepare the 3-D pose estimation model for validation. 2. Utilize different datasets for testing the pose estimation. 3. Switch to the lambda machine for performance evaluation. 4. Measure the Frames Per Second (FPS) of producing 3-D pose estimation on the lambda machine.
Verification Criteria	1. 3-D pose estimation with an error of less than 50 Maps on datasets. 2. FPS of producing 3-D pose estimation using the lambda machine at 35 for a video.

Table 3: Test No. 3: Avatar 3-D Reconstruction Test

Test No. 3: Avatar 3-D Reconstruction Test	
Objective	To test the 3-D avatar model provided by Meta
Elements	Perception and Localization Subsystem
Location	AI Maker Space
Equipment	1. Lambda Machine 2. Code provided by Meta 3. Personal Laptop
Personnel	Justin
Procedure	1. Deciphering Meta's code for analysis. 2. Executing model inference on the given dataset. 3. Display the results by rendering the 3-D avatar.
Verification Criteria	1. Obtaining a rendered 3-D avatar of the individual.

Table 4: Test No. 4: Testing the ROS node connections between the drone and lambda machine

Test No. 4: Testing the ROS node connections between the drone and lambda machine	
Objective	Validate the communication module
Elements	Communication Subsystem
Location	AI Maker Space
Equipment	1. Lambda Machine 2. Drone SDK 3. Personal Laptop
Personnel	Saksham and Annanya
Procedure	1. Start the ROS nodes on the Lambda Machine. 2. Establish communication between the drone and the Lambda Machine via ROS topics. 3. Send data between the drone and Lambda Machine to verify the connection. 4. Monitor the ROS node connections for stability and reliability.
Verification Criteria	1. Reliable transmission of camera data and commands between the drone and Lambda Machine. 2. Stable performance of the ROS nodes during testing without interruptions or disconnects.

Table 5: Test No. 5: Test on integrated human detection and 3-D pose estimation on drone camera's data

Test No. 5: Test on integrated human detection and 3-D pose estimation on drone camera's data	
Objective	Validate the pipeline for the perception stack
Elements	Perception Subsystem
Location	AI Maker Space
Equipment	1. Drone 2. Lambda Machine 3. Laptop
Personnel	Andrew and Annanya
Procedure	1. Integrate human detection and 3D pose estimation code. 2. Run the code with the pre-recorded drone camera's video through the ROS node and visualize the real-time result. 3. Review the result frame by frame.
Verification Criteria	1. Human detection detects the human correctly using the drone camera. 2. Outputs of human detection get into the 3D pose estimation module successfully. 3. 3D pose generates correct pose estimation using the output of the human detection module. 4. The FPS of the integrated system should be higher than 36.

Table 6: Test No. 6: Streaming Data FPS real-time

Test No. 6: Streaming Data FPS real-time	
Objective	Validate the communication and streaming speed
Elements	Perception and Localization Subsystem
Location	AI Maker Space
Equipment	1. DJI Mavic 3 2. Lambda machine 3. Personal Laptop
Personnel	Naman and Saksham
Procedure	1. Interface E-Port with the drone 2. Establish a stable connection between the drone and the lambda machine 3. Stream the camera data on the server
Verification Criteria	1. Receive images from the drone on the server 2. The FPS of the received data matches our functional requirement.

Table 7: Test No. 7: Depth Estimation using a Monocular Camera

Test No. 7: Depth Estimation using a Monocular Camera	
Objective	Validate the depth estimation for drone with respect to human
Elements	Perception and Localization Subsystem
Location	AI Maker Space
Equipment	1. DJI Mavic 3 2. OnBoard Camera 3. Personal Laptop 4. Lambda Machine 5. Optitrack System
Personnel	Saksham
Procedure	1. Measure the distance of the target from the camera through Optitrack System 2. Measure the distance from the depth estimation Network 3. Compare the two distances.
Verification Criteria	1. The error in the two distances is less than the set threshold

Table 8: Test No. 8: Control drone through SDK

Test No. 8: Control drone through SDK	
Objective	Validate SDK functionallity
Elements	Drone Control Subsystem
Location	AI Maker Sapce
Equipment	1. DJI Mavic 3 2. Lambda machine 3. Personal Laptop
Personnel	Saksham
Procedure	1. Design a trajectory for the drone to follow 2. Establish connection with drone SDK through ROS 3. Run the script and follow the trajectory autonomously
Verification Criteria	1. Drone follows the trajectory 2. The Root Mean Square Error of the trajectory is equal to our functional requirement

Table 9: Test No. 9: Integration of Perception Module and Drone Streaming Pipeline

Test No. 9: Integration of Perception Module and Drone Streaming Pipeline	
Objective	Validate Subsystem Integration
Elements	Perception and Localization Subsystem
Location	AI Maker Space
Equipment	1. DJI Mavic 3 2. Lambda Machine 3. Personal Laptop
Personnel	Annanya and Andrew
Procedure	1. Interface the drone with the lambda machine 2. Live-streaming the data for human detection model 3. The results are inputted into 3-D Pose Estimation
Verification Criteria	1. The system could produce accurate 3-D pose estimations of humans from a real-time drone camera. 2. The FPS of the whole system could achieve the system requirement.

Table 10: Test No. 10: Human following algorithm using CrazyFly

Test No. 10: Human following algorithm using CrazyFly	
Objective	Validate the algorithm for Human following algorithm
Elements	Perception and Localization Subsystem
Location	AI Maker Space
Equipment	1. Personal laptop x 2 2. CrazyFly drone 3. Optitrack camera system
Personnel	Saksham
Procedure	1. Get the current pose of human using Optitrack 2. Establish ROS connection and run target tracking script. 3. Control the drone to follow tracking waypoints
Verification Criteria	1. Drone maintains a fixed distance from the human in every pose 2. Error between waypoints and drone pose is less than a threshold

Table 11: Test No. 11: Integration between Avatar and Perception Module

Test No. 11: Integration between Avatar and Perception Module	
Objective	Validate Integration of sub-systems
Elements	Perception and Avatar Subsystem
Location	AI Maker Space
Equipment	1. Desktop
Personnel	Andrew
Procedure	1. Taking images with a human and passing them into the perception module. 2. Pass the result from the perception module into the avatar driving module. 3. Visualize the avatar.
Verification Criteria	1. The pipeline works 2. The generated avatar behaves the same as the image inputs

Table 12: Test No. 12: Visualization of driving avatar in real-time

Test No. 12: Visualization of driving avatar in real-time	
Objective	Validate Integration of subsystems
Elements	Integration
Location	AI Maker Space
Equipment	1. DJI Mavic 3 2. Payload SDK 3. Lambda Machine
Personnel	Justin and Naman
Procedure	1. Connect the Drone with the Desktop through payload SDK for real-time camera streaming. 2. Start real-time 3D pose estimation using the streaming data. 3. Reconstruct and visualize the 3D avatar using the poses.
Verification Criteria	1. The pipeline works with sufficient FPS. 2. The quality of the avatar is great, and it behaves correctly.

Table 13: Test No. 13: Fully Integrated System Test

Test No. 13: Fully Integrated System Test	
Objective	Validate fully integrated system
Elements	Integration
Location	AI Maker Space
Equipment	1. DJI Mavic 3 2. Payload SDK 3. Lambda Machine
Personnel	All Team
Procedure	1. Setup hardware and communication modules 2. Run the fully integrated script 3. Visualize the driving avatar on a monitor
Verification Criteria	1. The drone hovers at a fixed height 2. The code runs without any errors 3. Visualization produces desirable results 4. All the promised functional requirements are met

Table 14: Test No. 14: SVD

Test No. 14: SVD	
Objective	Demonstrate fully functional system
Elements	Integrated System
Location	AI Maker Space
Equipment	1. Lambda System 2. DJI Mavic 3 3. Personal laptop 4. Wifi for communication
Personnel	All team
Procedure	1. Ready the Software and Hardware 2. Start the Integrated script to launch ROS nodes 3. Acquire the results from the system and visualize the avatar on a monitor 4. Compare the results with the promised functional requirements
Verification Criteria	1. The drone hovers at a fixed height 2. The code runs without any errors 3. Visualization produces desirable results 4. All the promised functional requirements are met for the scenario of hovering

A. Appendix

Table 15: Mandatory Functional Requirements and Performance Metrics

ID	Functional Requirement	Performance Metric
M.P.1	Reconstruct Avatars	Unseen Motion PSNR and Novel View PSNR will be $\geq$ 28 dB and 30 dB respectively.
M.P.2	Control Avatar	Total MSE over 10 frames should be below $100\,\text{cm}^2$. FPS should be at least 36
M.P.3	Detect and Track the Human	Percentage of frames capturing the Human should be $\geq 95\%$.
M.P.4	Estimate Human 3D Poses	Mean Average Precision (mAP) $\geq 80\%$.
M.P.5	Estimate Human Position	$\pm 0.4\,\text{m}$ for vertical, ± 0.4 m for horizontal and for depth ± 0.3 m
M.P.6	Plan Trajectories for Drone	The average distance between the drone with ego human and obstacle should be between $3\,\text{m}$ to $5\,\text{m}$ and $0.6\,\text{m}$ to $0.8\,\text{m}$ respectively.
M.P.7	Control Drone	MSE between drone's actual positions and the predefined waypoints $\leq 25\,\text{cm}^2$.
M.P.8	Detect Obstacles	False Negative Rate (FNR) $\leq 8\%$.