UCONN CSE Senior Design Group 26

Shared Environmental Layout Data Structure of Human Position and Posture Representation for Human Robotic Interaction Control

Nick Durkin, Allan Feng, Max Kessler, Bude Su

Sponsored by MOVIA Robotics
November 30, 2021

Summary

The purpose of the project is to create a system that can take video data from two input devices and combine the data into one global environment. This will allow MOVIA to supplement their Kebbi Air Robot Platform with data from additional devices, improving the ability of the system to interact with subjects. The project will be implemented using Unity and Microsoft's world locking tools. The system will have a module that is responsible for tracking subjects' positions and locations using fixed real world objects (glyphs). This module will be installed on multiple devices, and will feed all the data to a second module. The second module will combine that information using coordinate transformations and create a unified global coordinate space. The global coordinate space will check for duplicate data or errors, and will create a single representation of each subject in the environment. These virtual representations will be displayed as the output on a laptop. This document outlines the exact details and design decisions for the system described above, and will also present the action plan for completing the project.

Table of Contents

Introduction:

Context

The intent of this project is to develop an integrated system for tracking body and face movements of human subjects which can take input from multiple sources and combine them into a composite representation of reality. This system will be built to augment the capabilities of MOVIA's Kebbi Air Robot Platform by improving its reaction to communicative cues. The Kebbi robot is an android robot that uses artificial intelligence to identify and react to human subjects. By using multiple input sources, the Kebbi will be able to better react to the subject and keep them more engaged. This can be used to assist teachers and parents to improve child learning.

One example of the application of this technology is MOVIA's Robot Assisted Instruction. An article from Autism Spectrum News [1] details how this technology improves the learning opportunity for autistic children. It states that autistic children are able to naturally connect with the robot in a way that they may not be able to with conventional learning methods. Interacting with the robot can help to develop socialization skills and make communication routine. This project will improve the data that the robot platform can take in so that it can be more effective with assisting educators. Figure 1 shows a Kebbi responding to a child subject in a learning environment.

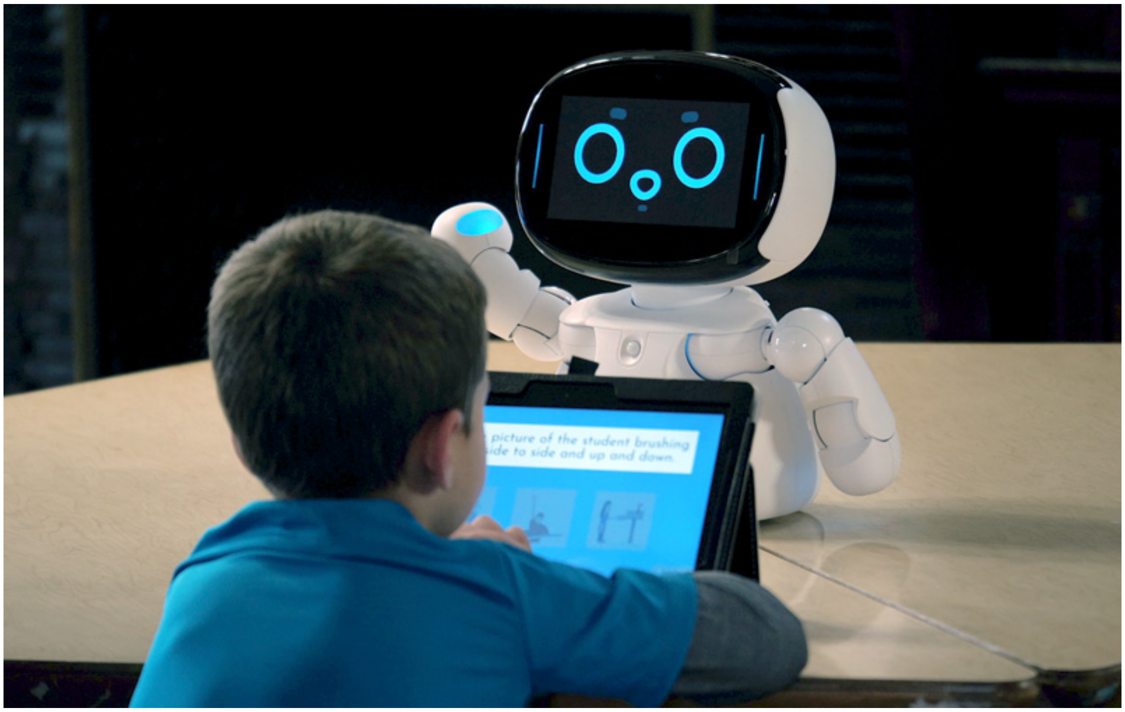

Figure 1: The Kebbi Air Robot interacting with a child [1]

This system will have a variable number of input sources (cameras) which can track body position and facial expressions, as well as their location in the world. The input sources will track the relative positions of real world objects (glyphs) to calibrate all the objects from camera sensors into one coordinate system. These calculations will be done on the individual devices using a basic script that will perform simple linear algebra in order to align all coordinates. These devices will send this information to a controller device (laptop) which will be responsible for building the composite and rendering the virtual environment.

The primary goal of this project is to get this system working for two devices, the tablet and the Kebbi. Initially this will be done in 2D space in order to simplify the procedure. That is to say the devices, objects, and glyphs will all be on the same plane. Once an effective 2D prototype is created the project will be expanded to 3D space. From here the project's focus can be turned to expanding the software to work for multiple devices as opposed to just two.

Because the final product will be used by children and non-technical savvy adults it is important that the system reboots very quickly and easily. This software will be useless to the users if it requires technical know-how to calibrate the devices. For this reason the aim of this project is to create software that is above all else easy to use. It is acceptable if the program crashes as long as it reboots quickly and resumes running. This detail is very important as it will guide the development of the software throughout next semester.

Existing Architecture

The largest part of this project which already exists is the face and body tracking software. This code takes a video feed and will identify human subjects in it. The software continuously tracks the face and body positions of the subjects, and can identify additional information, such as if the subject has a hand raised. Figure 2 demonstrates an example of the tracking software running on a single subject. The software shows a coordinate map on the face which represents many distinct points it is tracking and a structure over the body which represents the joints it is tracking. It also demonstrates the ability to track complex information about the subject, such as if a hand is raised.

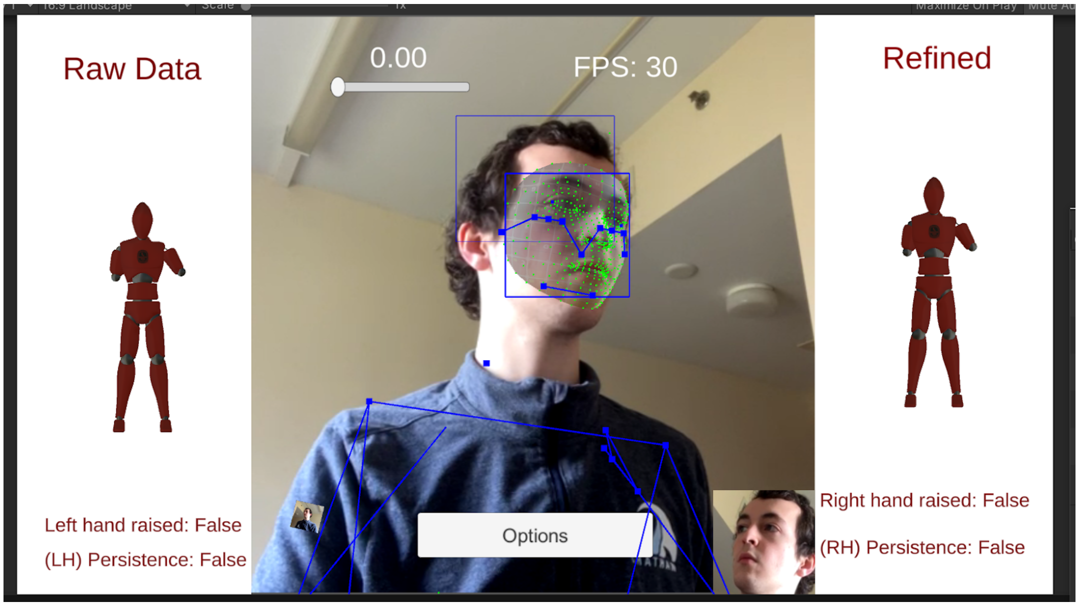

Figure 2: The body and face tracking software with a single subject

The major goal of this project will be to take the existing face and body tracking software, and expand it so that it can be used with multiple devices to create a unified composite representation of the environment.

Goals and Deliverables

The goals of this project are as follows:

1. Develop software to track the position of a camera relative to its environment
2. Using multiple cameras as input sources, transform the positions of the subjects they are tracking to a single global coordinate system
3. Render and output a composite representation of the environment and all the subjects located in it

To demonstrate these goals, the team will deliver a system that works with at least 2 input sources - the Kebbi robot and an android tablet - and outputs the environment on a laptop.

Procedure

Hardware

The main guiding principle that drove the choice of hardware is ease of use. The eventual purpose of this system is to be an educational tool. As such, we would like the system to not require any specialty hardware and to run on widely available consumer products.

The central devices of the project are the laptop, which will render the composite environment, and the Kebbi/tablet, which will be the primary input sources. This laptop will be any arbitrary Windows device. The Kebbi is essentially an android tablet and as such will run the same software as the tablet.

There will also be multiple glyphs. Glyphs or markers are real world objects that are used in augmented reality systems to place digital objects. These markers can also be used to find the distance between the camera and the marker itself. In this project we will be using these glyphs exclusively for the distance finding ability. The glyphs will likely be printouts of the MOVIA Robotics logo or a simple icon.

Software

The vast majority of this project will be software development. We have chosen to use Unity as the primary development environment which will be used for the face and body tracking as well as the glyph recognition. C# will be used for writing the scripts.

As the devices used in the project are all Windows or android devices, Unity provides an easy way to build code that can be packaged and run on any of the devices. The software will be divided into different modules that will run on each respective device. Unity will use the .NET platform to package one Sensor Module and allow it to run on both the tablet and Kebbi robot, as well as be able to expand to any additional devices that are added in the future.

Unity is also a good choice for this project as it has built in support for many of the features we will implement. The existing face and body tracking software has been written in Unity, and tools exist for building virtual environments and tracking the real world. More specifically, the Sensor Modules will use Microsoft's world locking and spatial anchor tools [2] to track the real world glyphs and identify the location of the device's camera in the environment.

Modules

The project can be split into two major modules. These are the Robot and Tablet Sensor Module, and the Information Space Module.

Robot and Tablet Sensor Modules

The Tablet Sensor Module will include the face and body tracking software provided by the sponsor. This detects movements and facial expressions of the subject. See figure 2 for an example of the face tracking software in action.

Next this existing software will be modified to allow it to output the coordinates of the subject in question. These coordinates will be at first relative to the sensor itself, that is the sensor will be the origin of the system. A body will be broken down into certain points, facial features like eyes and ridges, as well joints like elbows and shoulders. These points are already found via the given software and thus, they simply must be retrieved.

This module will also include the world locking software that will allow the user to find the sensor's place in relation to the glyph. After this is found the coordinates of the child or object will be converted to the global coordinates. Figure 3 shows the initial data points from the view of the sensor.

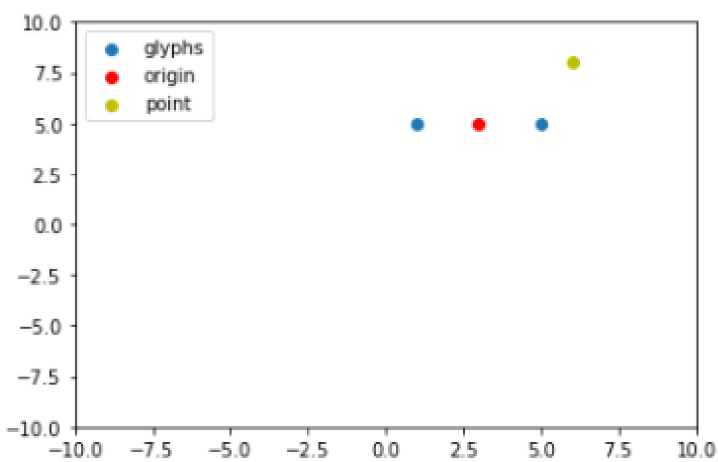

Figure 3: Example initial data

The conversion of the point from the sensor space to the global space can be done simply by subtracting the origin from the point. Figure 4 shows how this data looks once the points are transformed such that the "origin" is at (0,0). In this example, the origin is defined as the point directly between the two glyphs; however, in the implementation the origin could be some arbitrary point or one of the glyphs themselves.

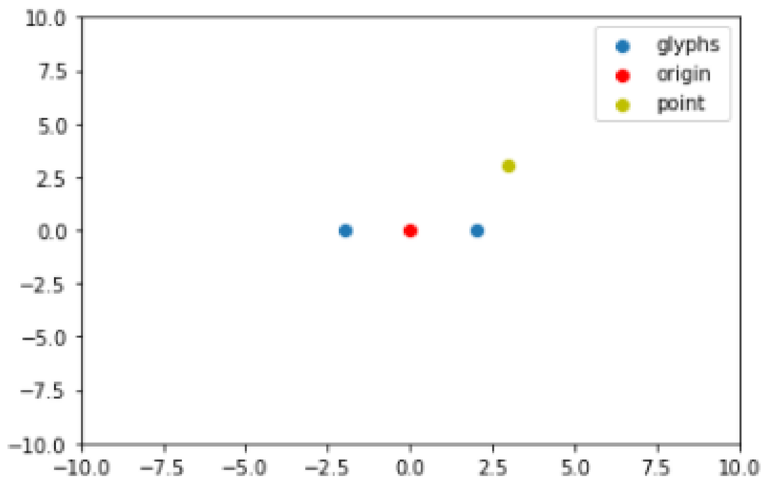

Figure 4: Example transformed data

Information Space Module

The Information Space Module (ISM) will be responsible for setting up the virtual environment, handling the connections and data from the Sensor Modules, and rendering the composite representation of the real world. This module will be running on the Windows laptop.

The first responsibility of the ISM will be handling the connections from the Sensor Modules. This project assumes that there will be two input devices, but the ISM will be designed with the intention of being able to add additional sensors in the future. The ISM must also provide a frictionless setup process and be able to handle devices disconnecting and reconnecting. This capability is important because the end users of the system may not be tech savvy, and will most likely not be familiar with the inner workings of the Kebbi Air Platform. Thus the system should be easy to set up, work robustly, and require minimal troubleshooting.

During operation, the first step of the ISM is to gather all the input data from the Sensor Modules. From each Sensor Module, The ISM will receive a set of coordinates that represents bodies and faces visible from the sensor's camera. Additionally, the ISM will receive the sensor's location relative to the glyphs.

Next, the ISM will do a coordinate transformation for each sensor so that the data from each Sensor will exist in a unified coordinate system. This alignment is possible due to the fact that the glyphs have a fixed real world location. As each sensor will be able to see the same glyphs, they can use these to calibrate. Figure 5 below shows a representation of an environment and how the transformations will take place from the sensor spaces to the global coordinate system.

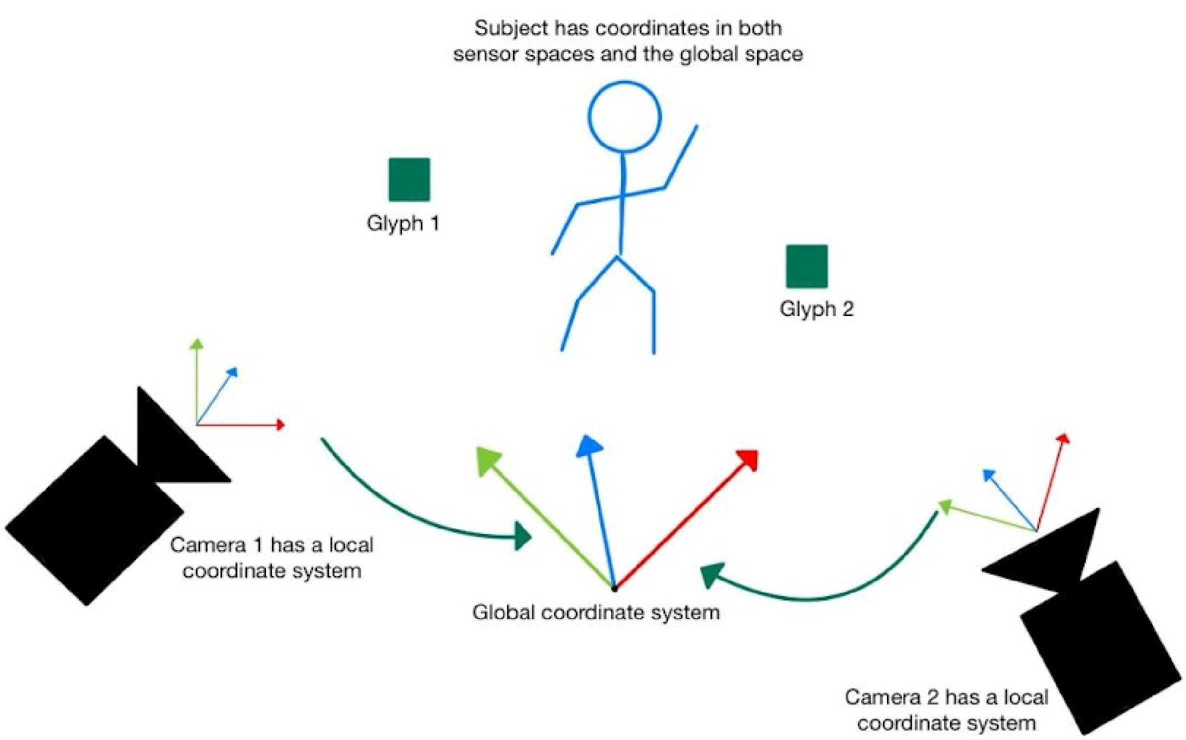

Figure 5: Visualization of the coordinate transformation from the sensors to the global space

Once the coordinates have been transformed, it is likely that different Sensor Modules will have tracked the same subject. In the composite environment, there will be two overlapping coordinate maps of the same object. The ISM will identify where two overlapping coordinates are the same subject and combine them into a singular object. To identify which coordinates are indeed the same subject, the ISM will use an algorithm that will determine how far apart the coordinates are physically located, as well as how similar in shape the objects appear. The exact parameters will be determined through testing. To actually combine the coordinates into one object, the ISM will either take the Sensor data which is most confident in its readings, or a weighted average of all the points corresponding to one object. As part of this process the ISM may have to identify if any of the data is invalid and should not be used. We will evaluate if this is necessary when testing the integrated system.

Finally, once a list of unique entities has been created, the ISM will render a representation of all the objects on the screen. This visualization will show the positions of the sensors in the environment, as well as entities being tracked. The visualization will also provide any information that has been tracked about the subjects such as whether their hands are raised or what their facial expressions are.

Data Flowchart

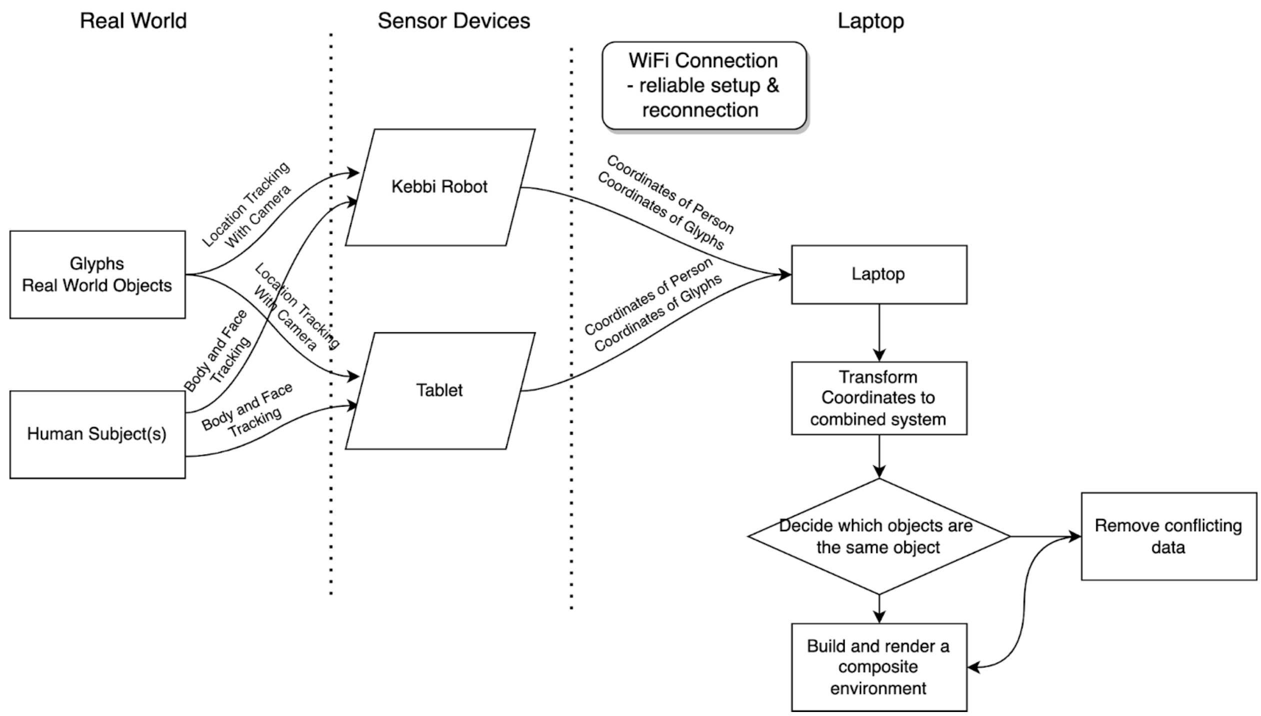

Figure 6: Summarized flowchart of system components and data

Figure 6 shows how the different components of the system interact and what data is being transferred. The first step is the real world, where data is gathered and sent to the Sensor layer. The Sensor layer does the processing described in the Robot and Table Sensor Module section, and then sends the data to the laptop over a WiFi connection. Finally the laptop does all the processing of the Information Space Module and displays the composite environment.

Collaboration

To work together and keep the project on schedule, we have divided the modules of the system into different components. Each component of the project has a team member that is responsible for leading development. We will work together on all of the components, but the leads will be responsible for keeping the task on schedule and determining if assistance is needed to complete development.

The responsibilities are divided up as follows:

Robot and Tablet Sensor Modules

Glyph detection and tracking	Allan
Determining location of the sensor relative to glyphs	Allan
Packaging face and body tracking data	Nick
Transmitting data to the Information Module	Bude

Information Space Module

Receiving information from the Sensor Modules	Bude
Transformation of all coordinates from local to global	Max
Identifying duplicate overlapping data	Allan, Max
Combining duplicate data to one entity	Allan, Nick
Displaying the final representation of all subjects	Max, Bude

To track progress on the project both as a whole and for individual components, we will use a Trello board. Trello will allow us to see which stage of development each component is at and if certain areas need additional resources to stay on track. An early stage of the Trello board is seen in figure 7 below. Each of the tasks will be broken down into more once development begins.

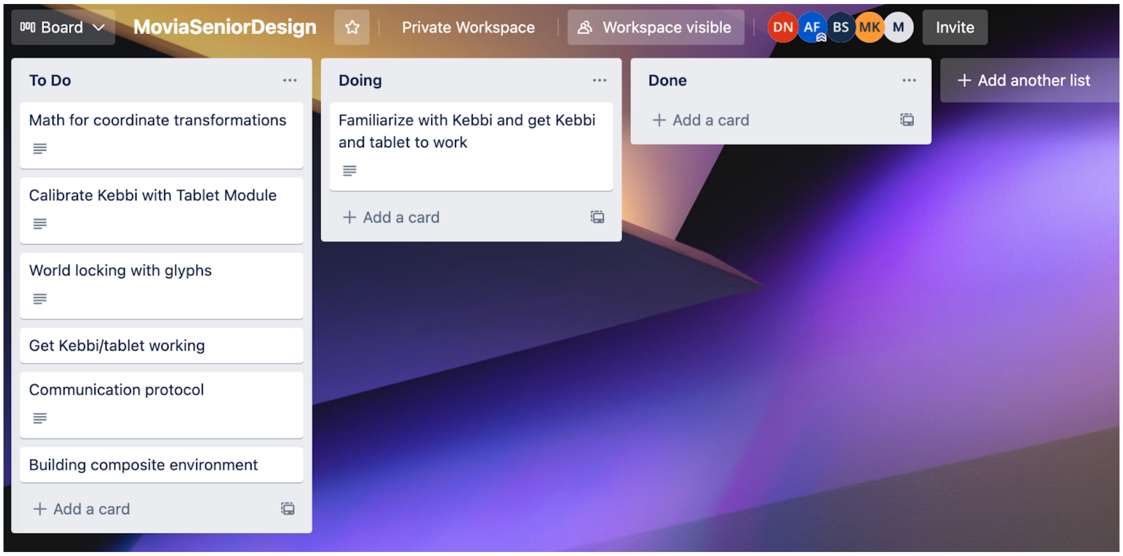

Figure 7: The Trello board at an early stage of development

In order for all of the team members to safely and efficiently contribute code, the team will be utilizing UConn's GitHub Enterprise to commit code. Best practices such as multiple branches and pull requests will be followed to ensure no work is being overwritten or lost.

<u>Timeline</u>

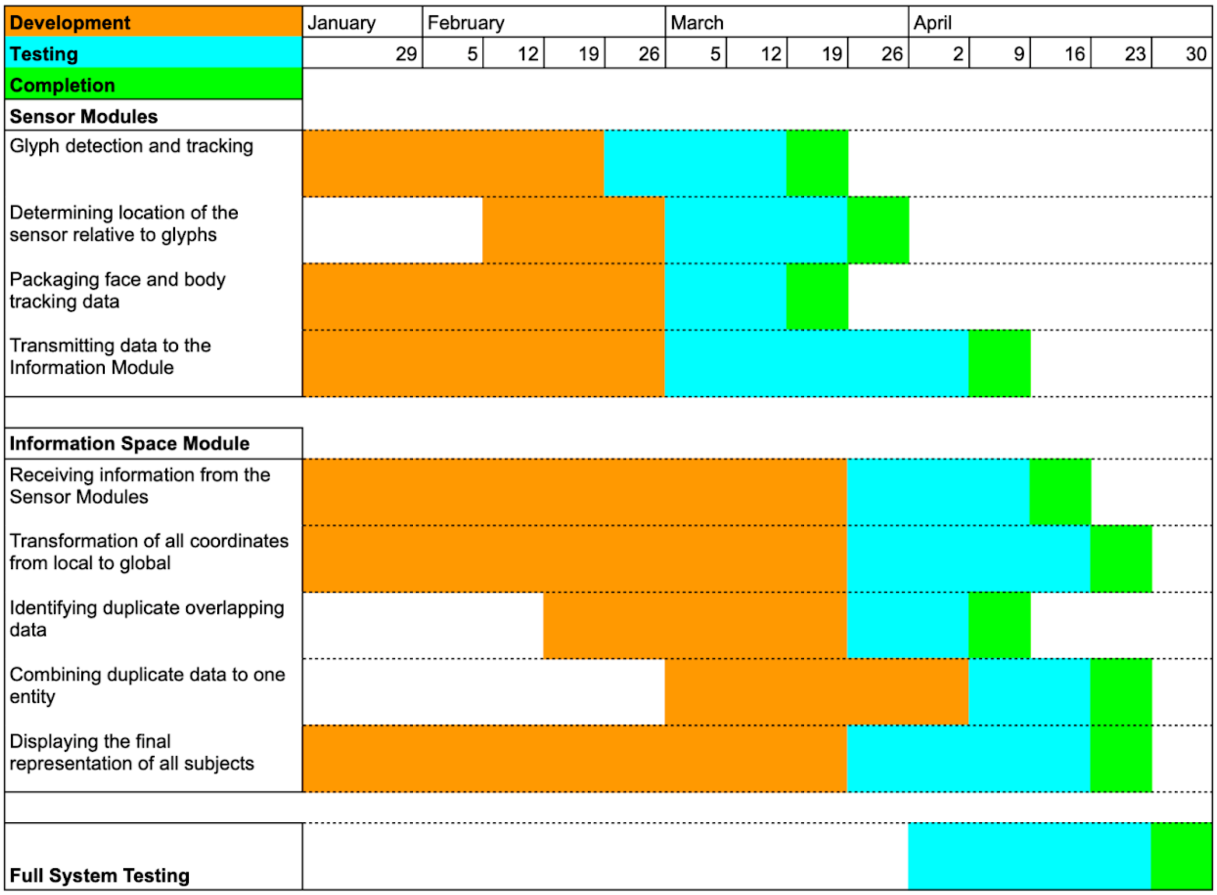

Figure 8: The project development schedule

Figure 8 presents the planned development timeline for this project. Many components can be developed in parallel, but there are a few specific areas that will require other parts of the project to be completed before they can advance. Most notably, identifying and combining duplicate data will need the Sensor Modules to be able to send preliminary data, and so those areas will begin development when they have the data necessary to begin.

The goal is to have the majority of the project functional by the end of March. This will provide lots of time for full system integration, testing, and troubleshooting issues that arrive.

This also provides padding in the event that there are delays during development of the modules, and will allow the project to still be completed on time. The team will meet weekly to discuss project progress and if the schedule needs adjustment.

Results

This semester we have had weekly meetings with founder and chief scientist Tim Gifford. In these meetings we have discussed the purpose and scope of the project. Namely the purpose of the project is to build software that can combine multiple coordinate spaces into one global space. Mr. Gifford also stressed to us the importance that the software be very easy for anyone to use. We also took a lot of time to discuss details about how the system will operate and to fully understand every aspect that is required for it to work.

We have also obtained the tablet device and Kebbi. We have tested these devices and have familiarised ourselves with how to use and program them. Programming the Kebbi was explained to us by the MOVIA team and we have also been given the contacts of several MOVIA employers who are experts with Kebbi.

Unity will be the major platform we code with and because of this the team has spent a lot of time training ourselves with the platform. We were given a demo of the face and body tracking software and have all become quite acquainted with the code. Additionally, the whole team has completed the Unity tutorials as well as read the documentation of the Microsoft world locking system.

One of the more important aspects of this design phase is the schedule and organizational planning. We created a Trello board on which tasks will be added and assigned. We also have a

GitHub repository for project files to be uploaded and modified. Loose roles for all team members have been assigned. These can be found in the Collaboration section. We have also developed a sophisticated timetable that can be seen in the Timeline subsection.

Conclusion

The objectives of this project are to develop a full system to track face and body movements from multiple sensors and combine the data into a composite representation of the environment. This project will take existing software to track subjects and package it to work on multiple devices and communicate together to build the composite.

This system will be broken up into two modules. The Sensor Module will run on a Kebbi robot and Android tablet, and will gather local coordinates for the subjects it is tracking as well as its location relative to glyphs. The Information Space Module will run on a laptop and will gather the data from the Sensor Modules. It will then apply coordinate transforms and build a virtual representation of the world which is displayed on the screen.

During this semester we have taken the time to understand the project, research the technologies that will be used, and put in place a framework so that development next semester can proceed smoothly.

This project will be used to improve the Kebbi Air Robot Platform by increasing its ability to react to human interactions. This will help to improve its capability as an educational tool for teachers and parents.

References

1. Lichtenstein, Lisa. "A Clinician's Perspective on Robot-Assisted Instruction for Autistic Children." *Autism Spectrum News*, Autism Spectrum News, 25 June 2021, https://autismspectrumnews.org/.
2. Ferrone, Harrison, and Mark Finch. "World Locking and Spatial Anchors in Unity - Mixed Reality." *Microsoft Docs*, Microsoft, 30 Aug. 2021, https://docs.microsoft.com/.

Acknowledgements

We would like to acknowledge our project sponsor MOVIA Robotics and specifically Tim Gifford for guiding us through the design process for this project. We are very grateful that Mr. Giffords took the time to meet with us weekly and explain the goals and requirements of the project.

We would also like to thank our faculty advisor Sheida Nabavi for helping us with project management.