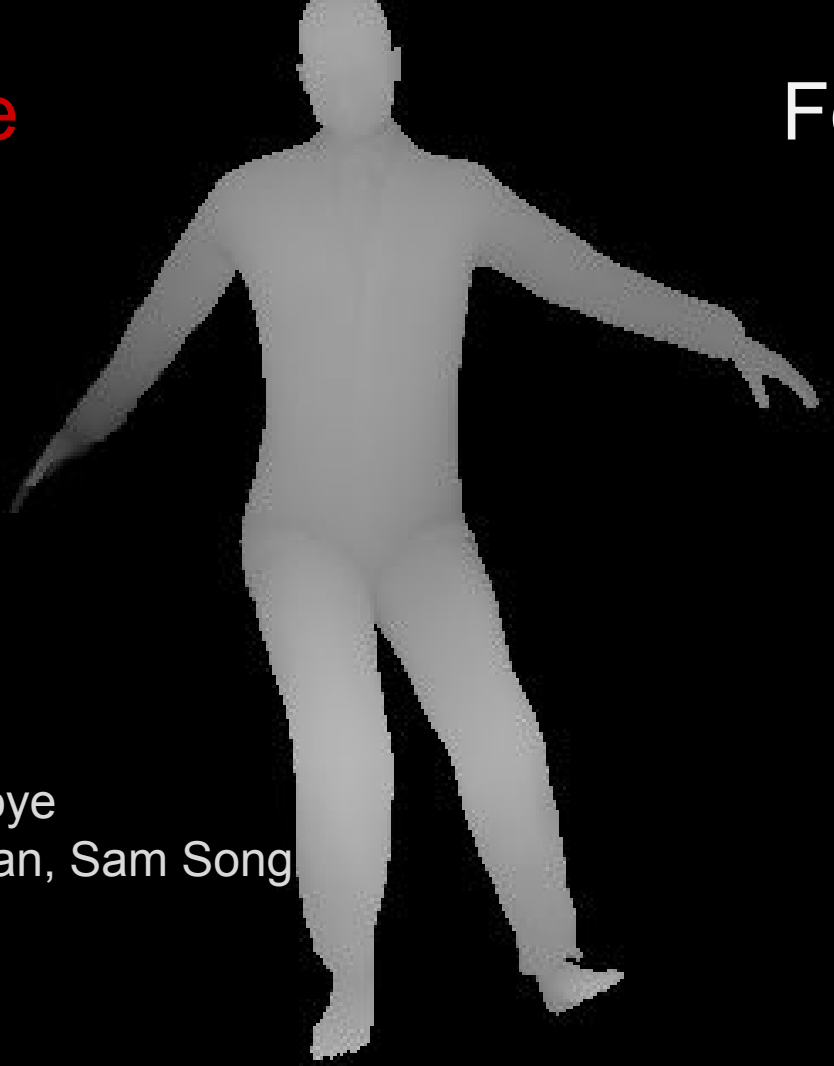
Human Pose Estimation

For Google Tango



Samuel Gbolahan Adesoye
Victor Vahram Shahbazian, Sam Song
cmps161 w-17
03-19-17

Introduction

GOAL

-To be able to track the human body using Google Tango enabled Devices.

WHY?

-Microsoft is a leader in this type technology, revolutionizing human tracking using the Kinect. We are using their approach to accomplish this project.

-We want to bring this to the Android platform. Before this wasn't possible, but now with depth sensing cameras on phones, the limitations have been lifted.

-This kind of technology opens up a door to many different possibilities in many different fields.

Implementation

-There are two ingredients that make Human Pose Estimation possible.

Depth Sensing Machine Learning

-For each pixel, pixels radially located to the original pixel are selected as feature vectors and then inputted into the Random Forest Tree to decide if the originally selected pixel is of the 25 labeled body parts.

-This is done by estimating the pose of different body parts. We are training the algorithm with 25 different labels. The labels are then used to localize the joints of the skeleton.

Random Decision Forest

-It’s simple and efficient for real time solutions that can run on a mobile devices.

-The Random Decision Forest is a forest of decision trees that decide whether or not something matches pre-trained data.

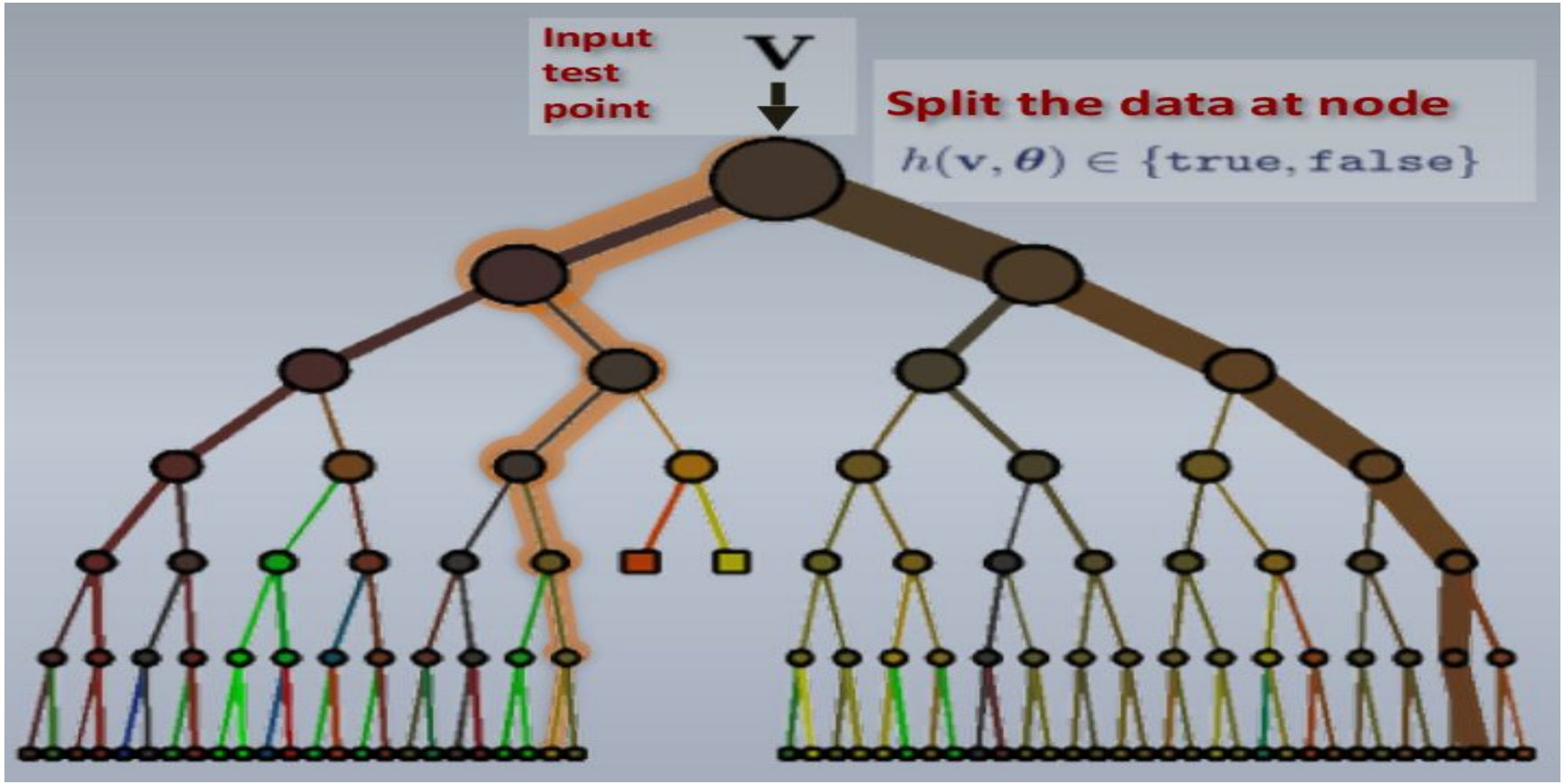
-Forest will take in a label vector for pixel and give the pixel a body part classification as output.

-Each tree produces a weighted probability for each label. Final decision is made by taking the label with the highest average probability.

Example of Tree output

Variables	T1	T2		
A	0.1	0.2	0.3	0.3/2=0.15
B	0.3	0.2	0.5	0.5/2=0.25
C	0.3	0.1	0.4	0.4/2=0.2
D	0.3	0.5	0.8	0.8/2=0.4

Decision Tree Example



Training Data

-A lot of our time spent on this project has involved us creating training data.

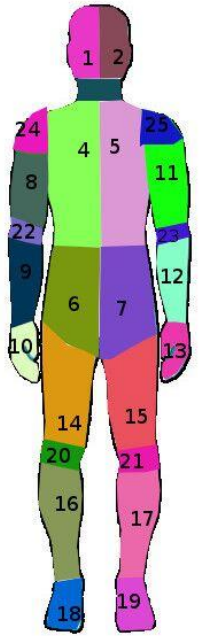
-Used Make Human to easily create human models.

-Importing into blender to be able to change the surface material of the object.

-All 25 parts of the body are labeled a specific color.

-When finished labeling, the object is rendered in several different poses using CMU MoCap scripts.

-This eliminates the need to take hundreds of depth images in person.



Moving Forward

-It took Microsoft a day to train a set of 1 million images on a 1000 cpu cluster system.

-GPU’s have many cores built within them making it possible to train a set of 2 million images on a single card within a week.

-For this presentation we have trained the application using 2 classes, hand and not a hand. Validating that the implementation works!

-The rest of the classes are currently being trained. This will take several days.