Final Report Walabot

Anton Lohner
Indrani Bhunia

Introduction

The Walabot is a small device about the size of a smartphone that looks like a portable hard disk. It utilizes RF radar wavelengths to deliver a visual portrayal of what lies behind a wide variety of relatively thin objects. Walabot has advanced imaging capabilities which allow for basic target detection behind thin walls, as well as target detection capabilities, 2D projections of 3D data, raw antenna pair and 3 dimensional coordinates, both spherical and cartesian, with an optional filter to remove static signals.The size of the arena detected by this sensor is set by the user through setting the bounds of the spherical coordinates, the quality of the snapshot and the step size between each measurement, with smaller volumes or lower quality snapshots requiring less time between capture. Something to keep in mind with this tool is that it measures signals in terms of intensity, with larger surfaces reflecting more wavelengths back. This means that at low enough resolutions multiple small surfaces can look nearly identical to a single, larger surface. In addition, there is an undocumented upper limit to the area that Walabot can capture per snapshot. The general workflow for Walabot is to perform initial calibration to set the area and filters, and then proceed to loop through the taking of several snapshots of data.

In previous works Walabot has been successfully used to track objects in motion, locate construction materials beneath walls, such as studs, electric wires and plumbing pipes, and detect basic human motions through thin walls. In this work data was collected for various hand movements using Walabot, with the data being graphed for visual comparison, as well as flattened and analyzed for statistical differences using a basic Pearson Correlation.

Related prior work

RF sensors have been used in previous works to identify specific human poses and objects, as well as specific movements. Additionally, a few previous projects have utilized the sensor to try and classify human expressions or gestures. One such example is Chen et al., which classified human expression by combining facial motion with previous image-based classification models. Dang et al., meanwhile, used heart rate to improve the model’s accuracy. Wang et al., utilized RF sensors to identify objects with particular shapes through walls, while Zhao et al. instead, relied heavily on custom hardware to detect human poses. Girard et al., meanwhile, investigated whether augmented reality visualizations could be made by combining 3D computer vision with

RF radar technology. However, previous works that utilized Walabot sensors frequently encountered the resolution issue. Shoaib suggested increasing resolution by utilizing the doppler dimension, which would reduce the number of antenna elements and higher frequency needed to achieve a higher level of resolution. To properly track human poses, custom hardware would have been required even then. The most promising approach, however, was Sluyter's et Al., who managed to train a model to recognize a wide variety of hand gestures with high accuracy utilizing Walabot itself by utilizing a fourier transform to reduce the dimensionality of the data, extracting several key pieces from the raw signals given by the antenna pairs. This approach would probably be best if attempts at facial recognition are the goal.

Approach/Experiments

The initial delve into investigating previous works on Walabot, as well as public github projects involving the sensor, seemed to indicate that the sensor did not have the necessary resolution for detecting subtle differences between expressions or gestures, so the initial approach was centered around trying to find a way to work around Walabot's limitations, by either hacking the inbuilt functions or utilizing a different sensing setting than the previous literature. Unfortunately initial delves into the Walabot code, including a decompiling of its dlls, indicated that the Walabot functions were out of our reach. Following that initial investigation, the next experiments with Walabot were performed utilizing 3D target tracking, as it seemed to handle basic gestures fairly well. The initial experiments were simply moving a finger up and down to see how each of these settings performed, and then examining the different variables collected, discarding variables which didn't seem to track the positioning of the finger and verifying we understood what each variable did. The data that walabot collected was saved into a csv file, which was then graphed into a simple animation for the 3D data. This allowed the identification of which settings would be ideal for data collection

As the hand motions became more complex and a visualization system was added for this setting, however, it quickly became apparent that this setting had several undisclosed limitations, with the biggest being its inability to detect more than three targets at once. Two of its other settings were 2 dimensional and made detection harder, so the two remaining settings to explore were the raw antenna pair data and the 3 dimensional point cloud data.

While one of the previous works by Shoaib had indicated a basic method for utilizing antenna pair data to create antenna arrays, this work had concluded that the resolution given was simply too low for any meaningful resolution. Without a proper background in radar mechanics, and after a couple of days trying to break down possible approaches, the 4D point cloud setting was instead pursued.

With the setting now set in stone initial experiments showed that having the sensor closer to the target caused a couple problems. The first of these was the physical limitation of the sensor: having a large object so close to the sensor meant that the θ and φ angles had to be incredibly large to compensate for the small R, which was impossible due to Walabot having an undocumented restriction on the total amount of volume the sensor could cover per capture. In order to compensate for this restriction, viable sets of θ, φ and r values, as well as their respective resolutions, were first calculated, with viability being defined by them properly covering a large majority of the population's hand dimensions. These sets would then be cycled into the Walabot code to see if they were accepted as valid or not, leading to a final set of angles, and radii and resolutions that were deemed as most likely to cover our requirements. Radii in particular varied highly between takes in order to see if closer gestures ended up being easier or harder to differentiate.

The second issue encountered was simply the proper method to visualize the data given by Walabot. This data had 5 dimensions in total, which coupled with inexperience handling point cloud data meant visualization took some effort to properly get working. This resulted in splitting the data into four 1 dimensional vectors per time period, which meant that it was relatively simple to do further analysis afterwards.

Once visualization and dimensionality were figured out four different radii and angle sets were chosen and data was collected for two simple, repeating gestures at these distances, with roughly 20 to 25 repetitions done for each gesture. All the data was collected in a csv file, and simple animations and statistical analysis was done for the data.

Results

The initial experiments were created primarily to help with efforts in cleaning and visualizing data, while the final experiment was aimed at helping with the analysis of data, though this ended up being quite limited. The bulk of our initial results were gathered primarily through walabot's sensor target detection mode, which tracked potential targets based on slower, high quality scans. The scan was taken every .06 seconds, though the duration between scans increased linearly if we wanted to capture other data in the same snapshot, like image energy or a 2D projection matrix of the data.

The first experiment was based on a finger moving left and right in front of the walabot, and was utilized both as our baseline for graphing the target detection data, and as a baseline for the amount of accuracy we could expect from the sensor itself. Something that was realized was that this mode would often create "ghost" measurements of targets in the center when walabot did not have a target to track initially, as well as whenever the target went out of walabot's range. This initial experiment was fairly promising, as though there was definitely some noise it was easy to see which motion had created this data. In order to try and remove some of this null data, the threshold for target detection was increased from 25 to 65. However, this ended up backfiring, as the threshold ensured that only large objects were classified as targets, leading to little to no meaningful data being collected.

After verifying that the null data couldn't be removed trivially, a second experiment was performed, which once again involved a simple motion with a single finger, this time moving in a circular motion. While messier than the first result, this second experiment verified that the walabot could track a single movement in short distances even in a different axis of dimension, though once again the amount of noise was a bit concerning.

This concern would be validated in the third experiment, which exposed some of the limitations this target detection had. The walabot was supposed to track two fingers doing a cutting motion periodically. This was a step up in complexity of the gesture walabot was supposed to track, and due to its target tracking focusing primarily on the amount of energy reflected back into the sensor and not movement, the amount of noise in the measurement jumped up significantly, leading to incredibly messy results. This experiment also revealed that there was a limit to the maximum number of targets the walabot could identify which was not in the documentation, which further complicated matters, as this meant any measurement could have a lot of significant information removed if any other part of the body was caught in the frame. Most of the measurements taken by target sensing had terrible resolution in the Z axis, with measurements only being split into 3 to 4 layers at most. All these factors meant we began to move away from target sensing and into 4D point clouds instead, though we also experimented with forcing a drop in the resolution of other axes in order to see if that improved or made it harder to detect the motion being performed.

The shift into 4D point clouds meant a rewrite was needed to handle the new data being fed into the CSV, as well as the visualization of the data. Though this visualization approach ended up being relatively slow and probably could be further optimized, it allowed for visual verification of what was being sensed by Walabot. This was followed up by the collection of four different data runs, which varied from being 20 centimeters away from Walabot to 50. To simplify the experiment only two gestures were used: the opening and closing of a fist and waving a finger left and right. In order to mark a change between doing one gesture and another, 15 seconds of empty space were used between gestures, which ended up being a mistake. This was due to the amount of empty data Walabot naturally collected. This might've been an inherent flaw with the sensor, the sensor having a smaller arena of measurement than what it was set to have, or a consequence of a single person having done all the measurements. Something similar happened with the measurements at higher distances, as though the position for the gestures was calculated using measuring tape the sensor didn't actually detect the

motions at all. Though the amount of data being recorded per frame would make most visual feedback impossible to properly visualize in the console, a summation of intensities might be fast enough while still allowing for the estimation of how much of the target the sensor is measuring.
Even with these flaws when collecting data, however, a basic pearson correlation on the datasets showed a similarity score of ~73% between a set of measurements of the two gestures.

Conclusion

In conclusion, initial forays into Walabot weren't very promising, as Walabot's API is closed off to tinkering and most of the example applications implemented so far are relatively simple in scope. The mistake of not revisiting known literature as we explored Walabot's capabilities meant that some of the more critical approaches other people had done were unknown, which limited the solutions proposed in this work. A lot of time in this project was also spent tinkering with and learning multidimensional graphing and animating. Target detection seemed relatively promising at first but ended up becoming too restrictive as the complexity of the task increased, while raw antenna pair data was mostly discarded due to the complexity in properly understanding the underlying mechanics of radar and doppler signal processing. Even with only the relatively limited data source of 4D point clouds, however, some promising results were achieved. Though there was a lot of statistical similarity between the two types of measurements collected, a large portion of this similarity is due to the sheer amount of empty space in the measurements.