Point Cloud based Gesture Recognition with Kinect 2

Anton Klarén, Jonathan Karlsson

Kinect v2

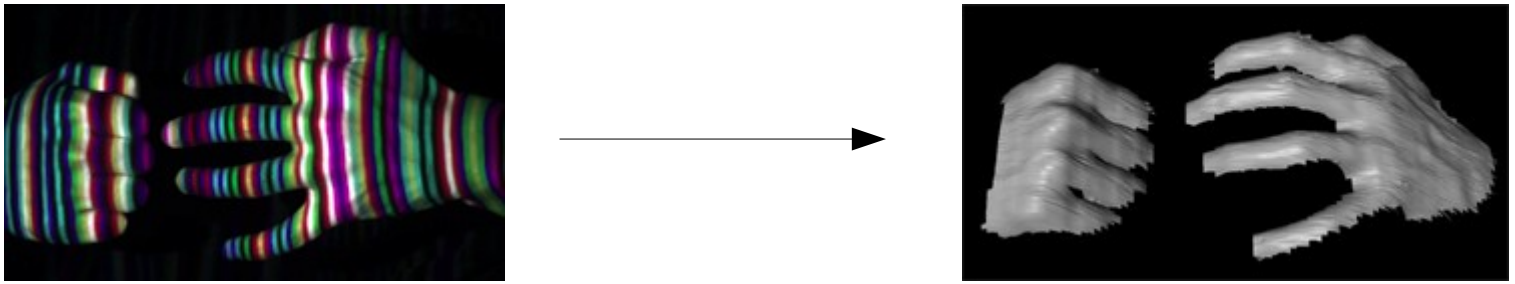
- 2.5D Sensor (Depth)
- Time-of-flight sensor
- Full HD color camera
- 4-microphone array
- 30+ frame rate

(Demo of Point Cloud)

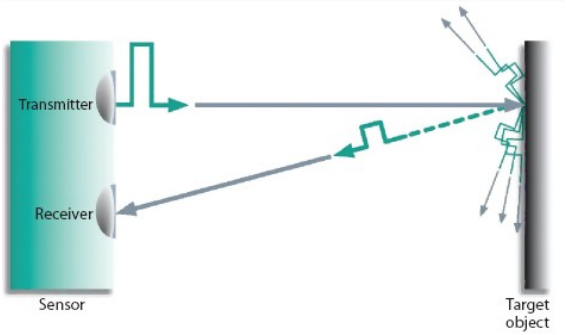
LUND UNIVERSITY

Kinect 1 vs Kinect 2

Structured Light



Time-of-flight



Same principles as a radar, only on smaller distances and in 2 dimensions

LUND UNIVERSITY

ROS – Robot Operating System

- Not actually an operating system
- Framework for building software for robots, on top of ubuntu LTS.
- Contains many tools and utilities that are common in robotic-software.
- Its core consists of a publisher-subscriber network for interoperability.

LUND UNIVERSITY

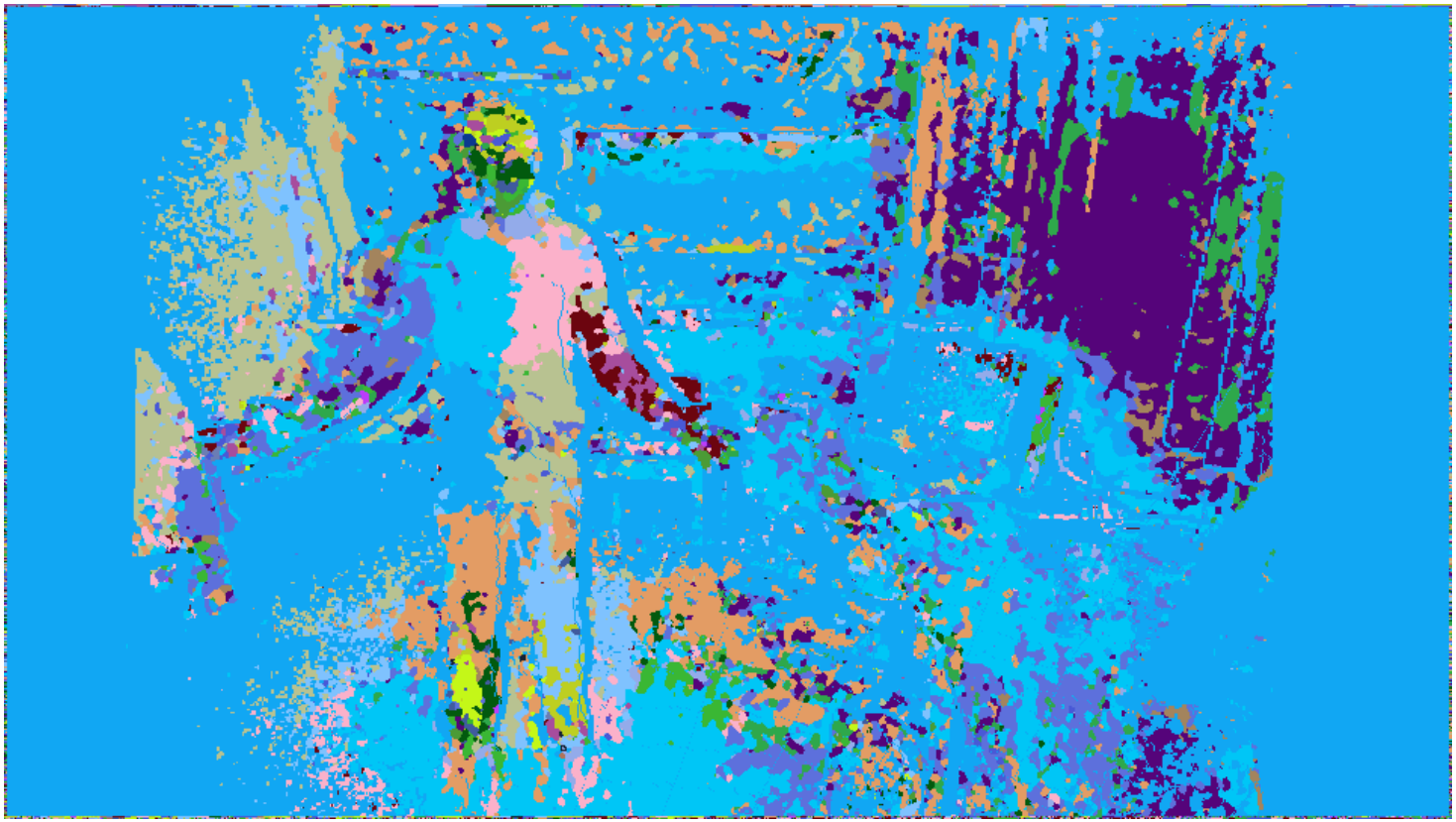
Detecting People

- Classify subparts of the point cloud (Random Decision Forest)
- Smooth classifications and clustering (Mean-shift)
- Try to fit a skeleton to the data, score skeleton based on ideal skeleton, select highest scoring
- Grow region from skeleton to extract the person (Approximated floodfill)

Random Decision Forest

- Uses the concepts of Machine Learning and Regression
- Forest – Consists of multiple trees that all gets a vote; a vote consists of a probability distribution of the confidence score
- Merge votes from all trees and return the top candidate
- Random – Too many possible questions, take a random subset
- Fast and effective

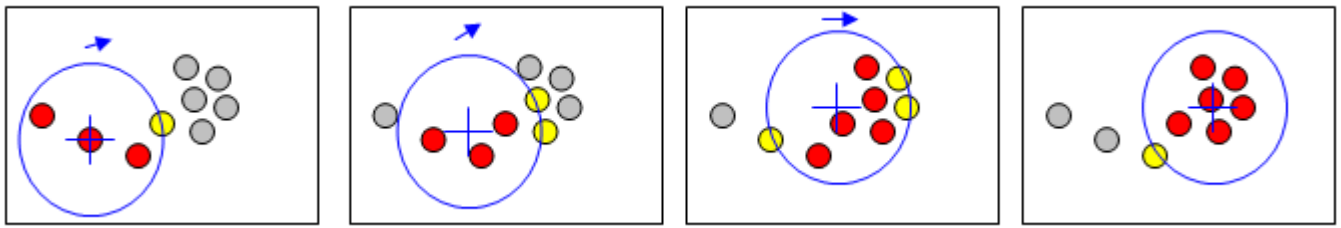
Random Decision Forest



LUND
UNIVERSITY

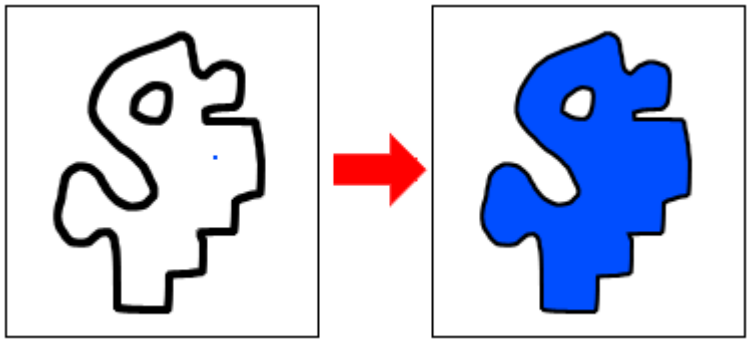
Mean Shift

- The objective is to find the densest region of a particular segment
- This is done with a sliding window that moves towards the mass-center (mean)
- Used to smooth the categorizations into segments



Approximated floodfill

- Perform edge detection on the depth data
- Groups segments by depth by ”filling”
- Used to extract interesting regions by masking with the filled regions



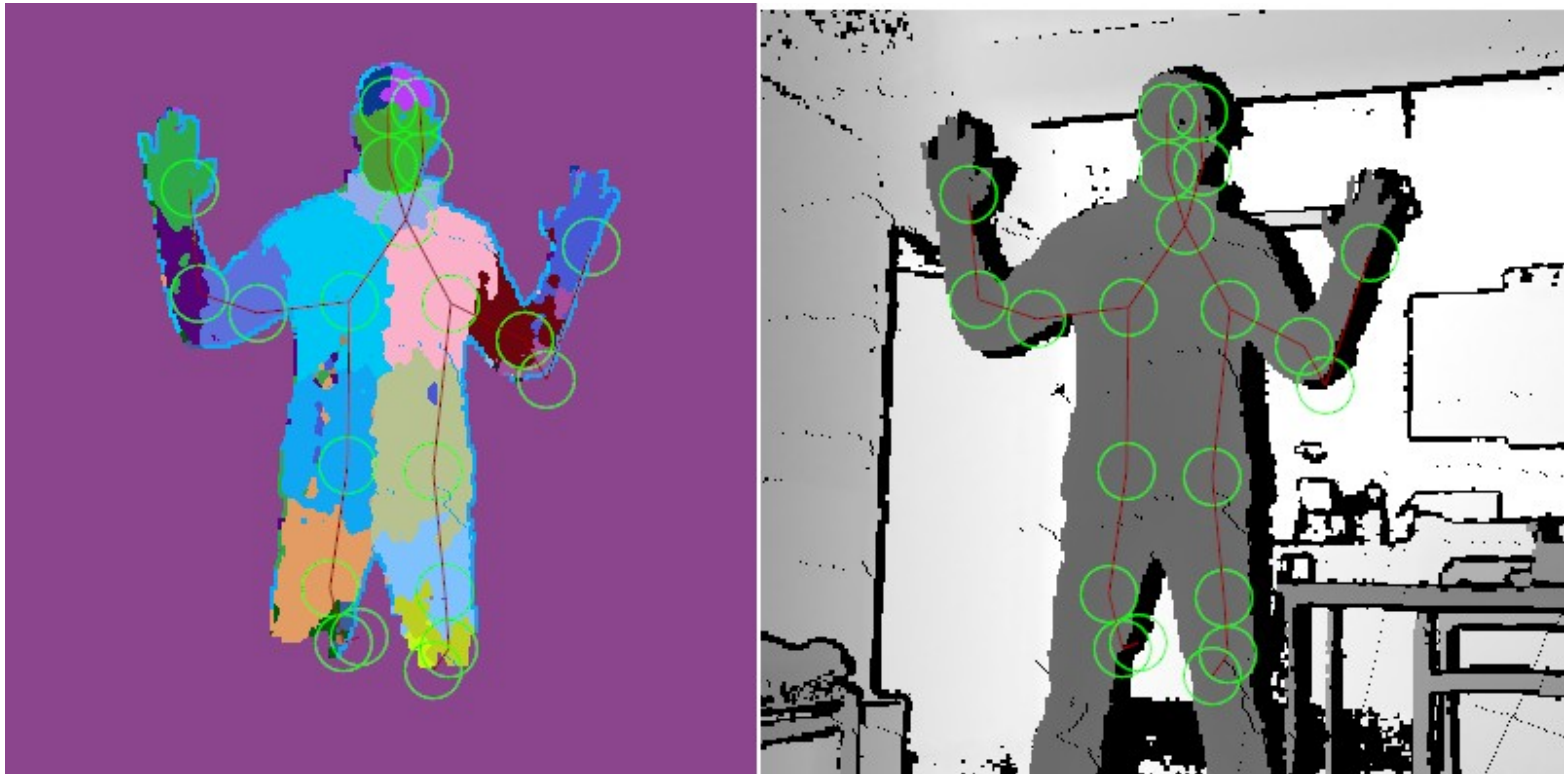
Mean Shift + Floodfill

LUND UNIVERSITY

Skeleton Fitting

- Start from a root node
- Find closest segment of child nodes
- Continue recursively until leaf nodes are found
- Discard improbable skeletons based on a score

Skeleton Fitting



(Demo Skeleton)

LUND UNIVERSITY

Recognizing Gestures

- AdaBoost (Adaptive Boosting – Classifier)
- Based on linear regression
- Less susceptible to overfitting than other similar methods
- Works with many dimensions (feature space)
- Can be parallelized over dimensions
- Example input: Joint positions, limb angle, etc..

Recognizing Gestures

- Random Forest Regression (RFR)
- Digests classifications from AdaBoost
- Emits classified gestures

> *A random forest is a meta estimator that fits a number of classifying decision trees on various sub-samples of the dataset and use averaging to improve the predictive accuracy and control over-fitting.*
>
> - Scikit-Learn developers

Future Improvements

- Implement AdaBoost and RFR
- Perform additional processing on hand-”blobs” to extract finger position
- Move parallelizable calculations to GPU to increase performance (> 3 fps)
- Tweak parameters and try with different hardware

LUND
UNIVERSITY