Object Position Prediction from human hand motion

Athanasios Christoforos Tsitos

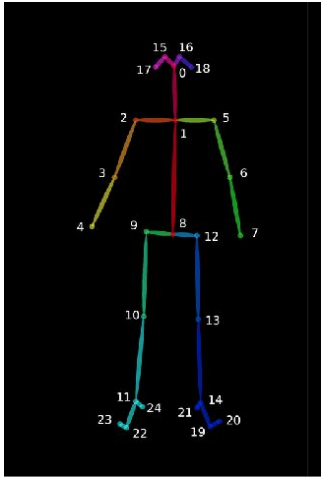
Contents

Introduction

- Prediction of the object location based on the motion of the human right wrist
- Five objects in known locations
- No object localization or object detection algorithms

Data Collection

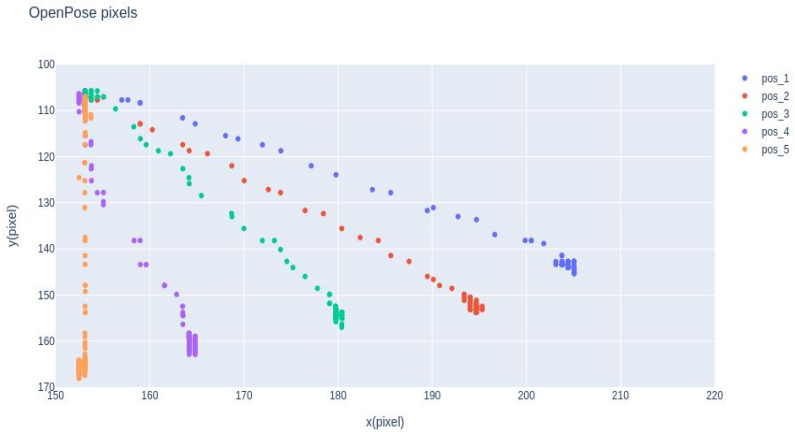
- RGB-D camera for human monitoring
- OpenPose for human wrist 2D pose estimation
- Predetermined and known object locations
- Predetermined and known starting position of the human wrist

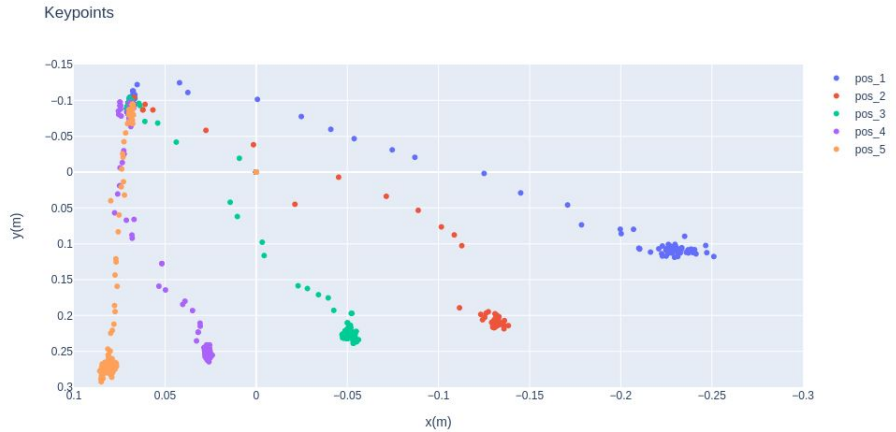


Data Collection

- RGB frames collected @ 60Hz (RGB-D sensor frequency)
- OpenPose data
  - ~22Hz (OpenPose frequency based on utilized GPU)
  - 60Hz
- Human motion representation
  - 2D OpenPose Pixels
  - 3D Cartesian Coordinates (OpenPose 2D pixels + PointCloud depth information)

Data Collection

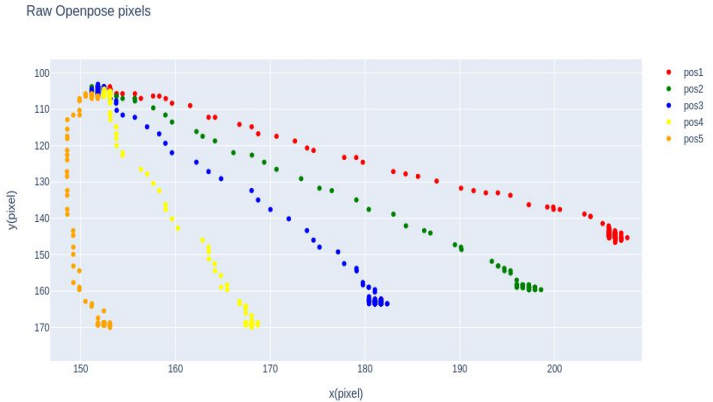


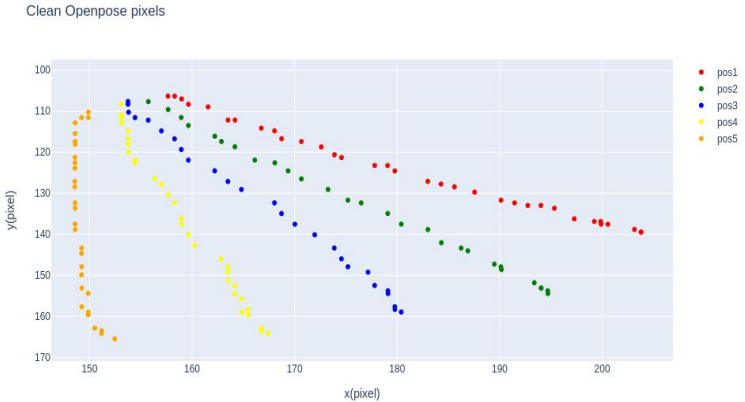


- OpenPose 2D pixels due to less noise

Data Preprocessing

- NaN values removal
- Outlier removal
- Static pixels removal





Feature Extraction

- Features: x, y OpenPose pixels
- Real-time behaviour
  - Pixels are obtained over time
  - Window of increasing length encapsulating available pixels at time t
    - $t=t_0 \Rightarrow [x_0, y_0]$, $t=t_1 \Rightarrow [x_0, y_0, x_1, y_1]$, ...
  - Class labels: $y_i=\{1, 2, 3, 4, 5\}$

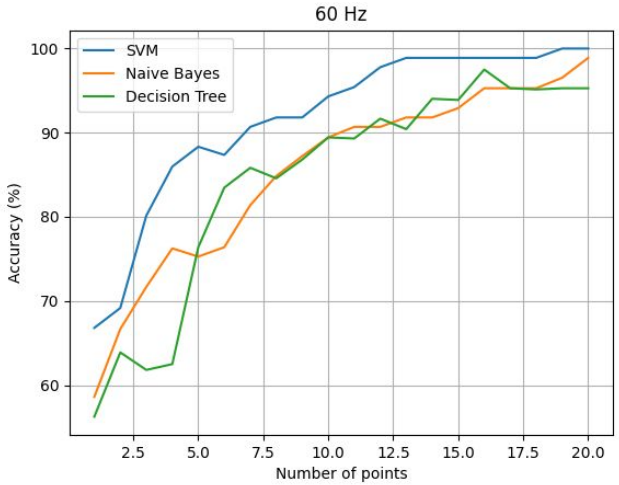
Training/Validation

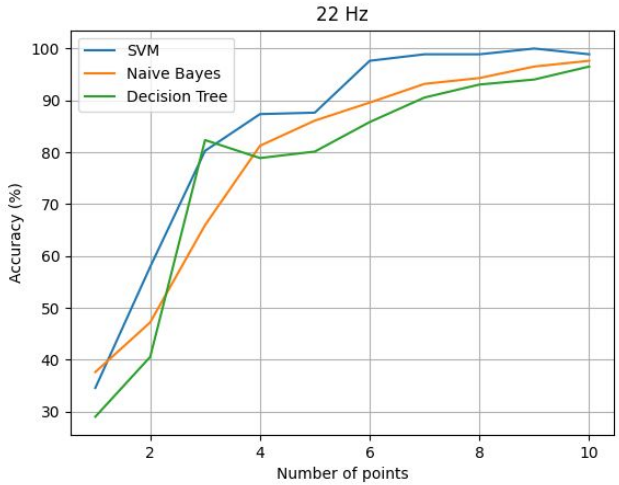
- Naive Bayes, Decision Tree, SVM
- Training/Testing
  - Dataset @ 22Hz (OpenPose frequency)
  - Dataset @ 60Hz
- KFold split (n=10)
- SVM
  - kernel={rbf, linear}
  - C={0.1, 1, 10}

Training/Validation

- Metrics
  - Accuracy: How good and at which time do the algorithms predict the human intended goal?
  - Confusion Matrix: In misclassification cases, what did the algorithm predict?

Training/Validation





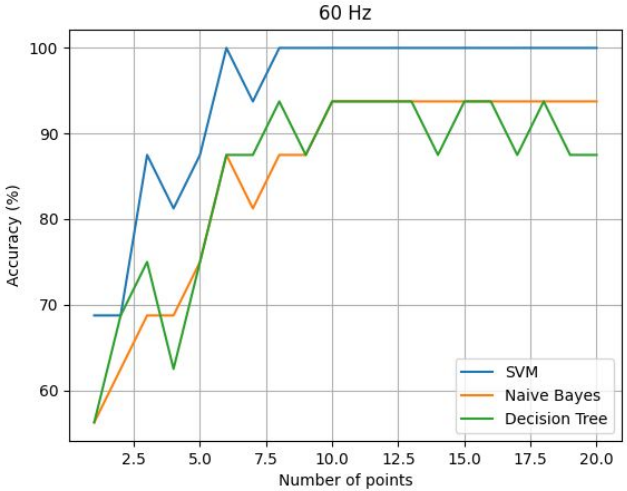
Training/Validation

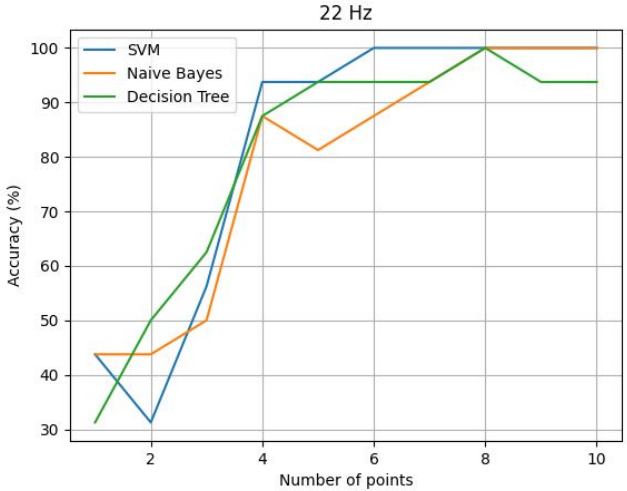
- SVM
  - Kernel: rbf
  - C=1
- SVM performs better than Naive Bayes and Decision Tree classifiers

Testing

- Training using the entire training-validation datasets
- Store the models for each classifier
- Test in testing dataset
- Metrics: same as in the training/validation phase

Testing

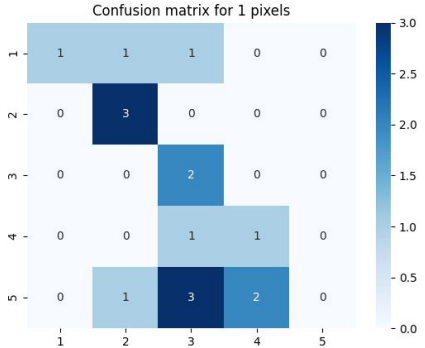

60 Hz
Accuracy (%)
Number of points
SVM
Naive Bayes
Decision Tree

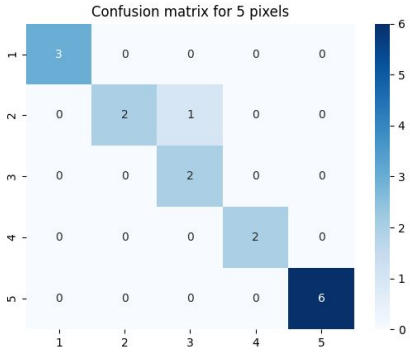

22 Hz
Accuracy (%)
Number of points
SVM
Naive Bayes
Decision Tree

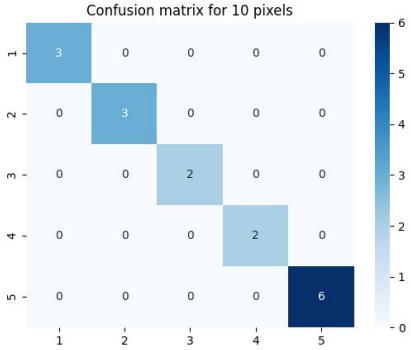
Testing

- SVM with linear kernel and C=1 produces the most accurate predictions
  - >90% accuracy in the first 188ms (22Hz) and 100ms(60Hz)
- Naive Bayes and Decision Trees
  - 85% accuracy in the first 188ms (22Hz) and 100ms (60Hz)
- Having available the pixels at 60Hz instead of 22Hz results in approximately 0.8-1.2 times faster prediction.

Testing


Confusion matrix for 1 pixels
1 1 1 0 0
0 3 0 0 0
0 0 2 0 0
0 0 1 1 0
0 1 3 2 0
1 2 3 4 5
3.0 2.5 2.0 1.5 1.0 0.5 0.0


Confusion matrix for 5 pixels
3 0 0 0 0
0 2 1 0 0
0 0 2 0 0
0 0 0 2 0
0 0 0 0 6
1 2 3 4 5
6 5 4 3 2 1 0


Confusion matrix for 10 pixels
3 0 0 0 0
0 3 0 0 0
0 0 2 0 0
0 0 0 2 0
0 0 0 0 6
1 2 3 4 5
6 5 4 3 2 1 0

Limitations

- Objects clearly separable
- Human initial position relatively far from the objects
- Complex human motions e.g. obstacles in the workspace