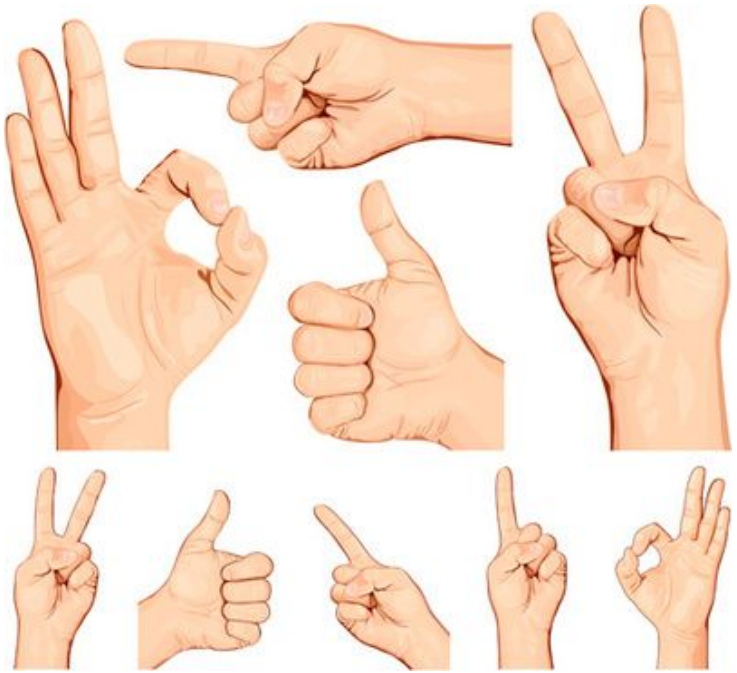
Interpreting Hand Gestures from a Monocular Camera

Bruno Georgevich Ferreira

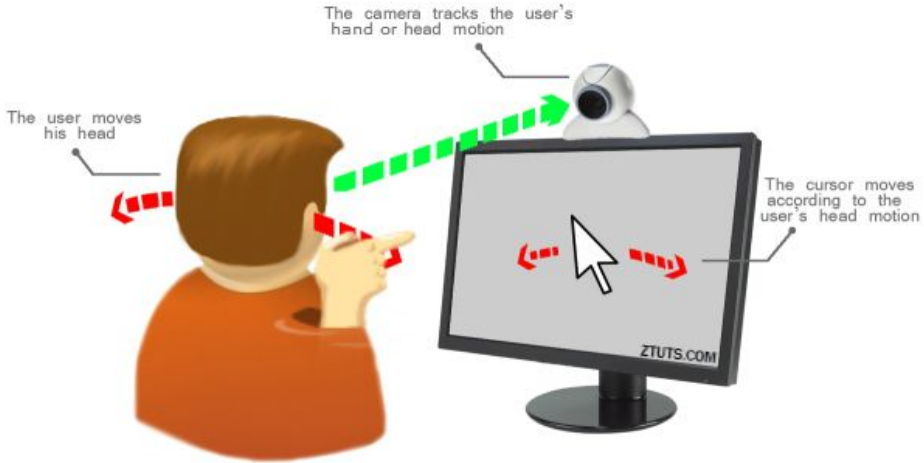
Motivation

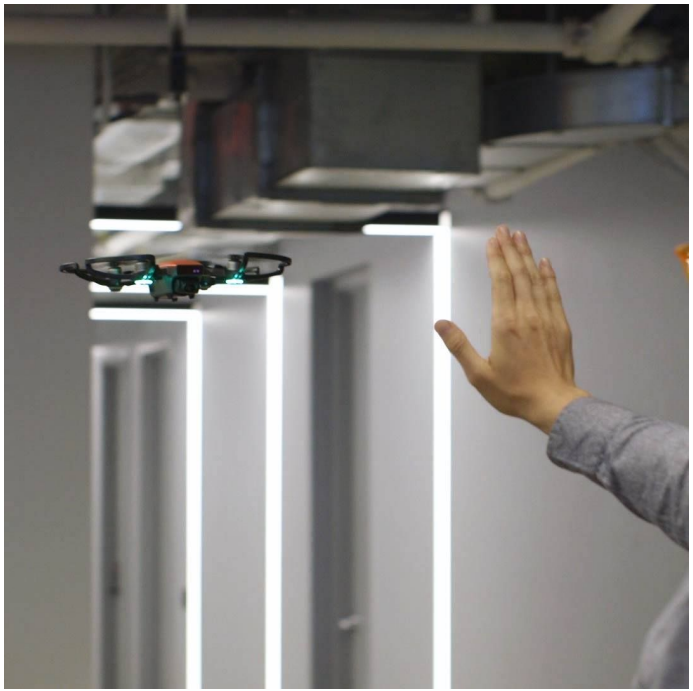
Control systems by performing hand gestures



Applications

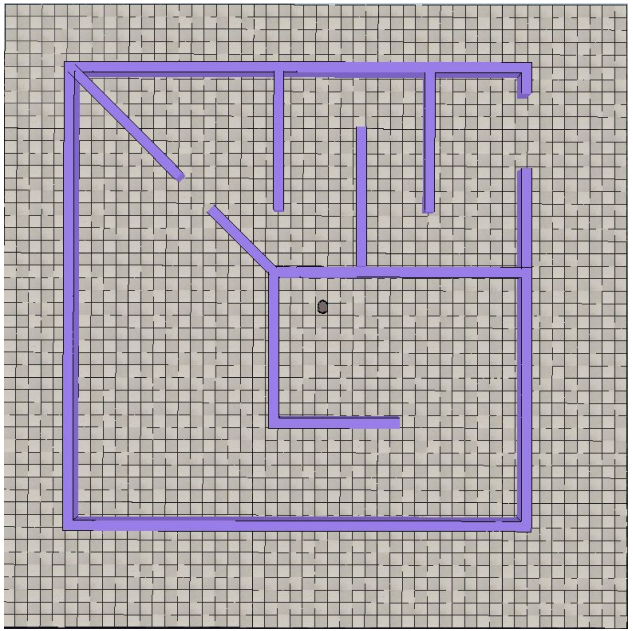
- Control Robots and Drones
- Remote Command Activation
    - Smart Home

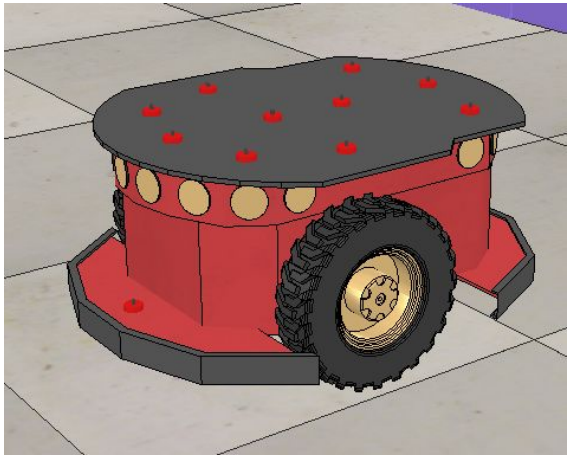




Proposal

- Develop a system capable of controlling a robot using only the hand to control it



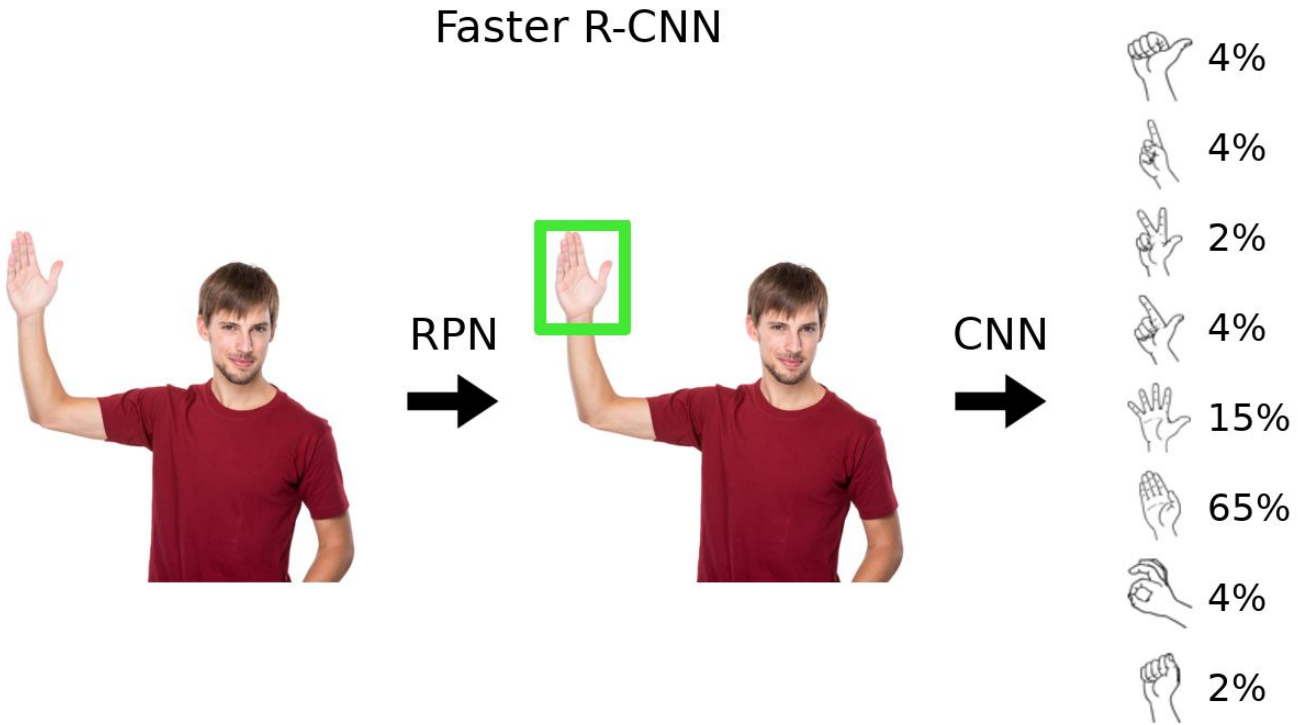


Processo


Gesture
Detection
Interprets the Gesture and
Translate in a Command
Feedback to the User

Scenarios

- Use of Faster R-CNN or Yolo for Hand Detection and Recognition

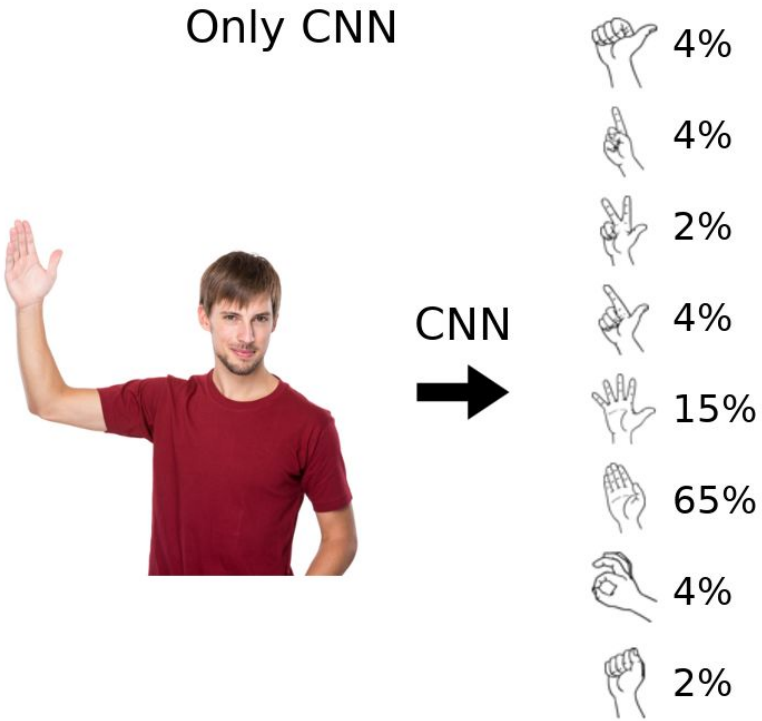


Important Notes

- Advantages
  - Good Detection
  - More than one hand simultaneously
  - Image Segmentation
- Disadvantages
  - Complex annotation process
  - Few datasets on the network

Scenarios

- Use of CNN Gesture Recognition

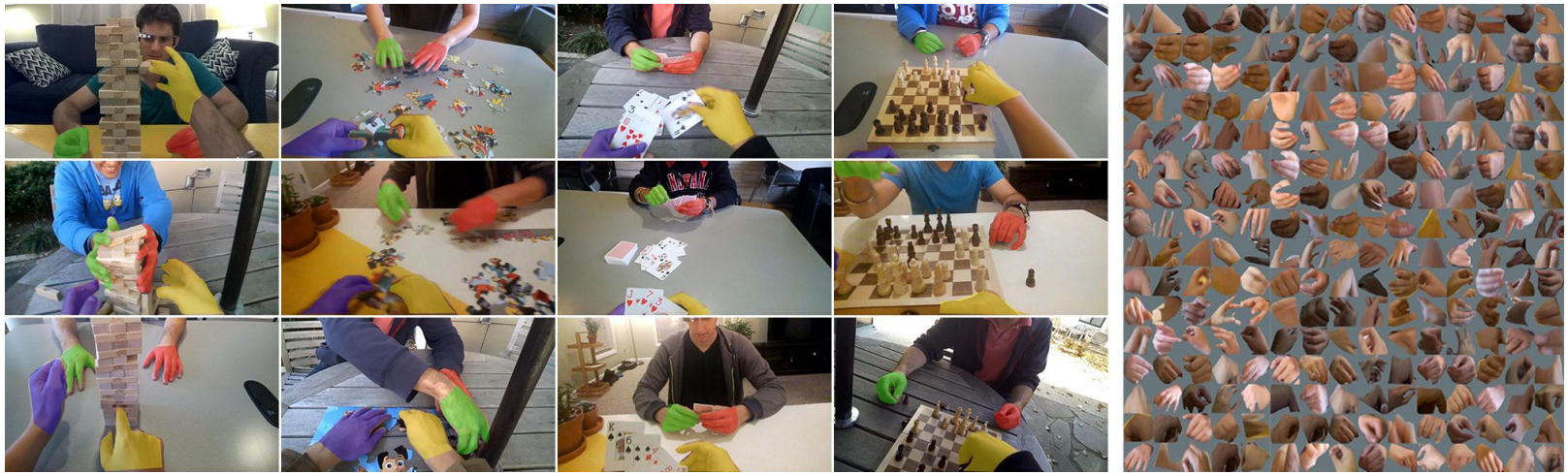


Important Notes

- Advantages
  - Good Classification
  - Simpler annotation process
  - Many datasets available in the network
- Disadvantages
  - Do not offer segmentation
  - Difficulty classify more than one hand

Datasets

- http://vision.soic.indiana.edu/projects/egohands/
  - Good for hand detection
  - Mask Faster R-CNN
  - Do not has predefined gestures
  - 4,800 images



Datasets

- https://www-i6.informatik.rwth-aachen.de/~koller/1miohands/
  - Has a good set of gestures (61)
  - Do not offer segmentation
  - 1 million of annotated images 18 GB)

Difficulties

- It would not be possible classify the detected hands
  - Has not dataset available for it
  - It would be necessary create a dataset for it
- It was not possible annotate the *1Million Hand Dataset* for segmentation
  - It would be hard to annotate 1,000,000 images

Chosen Model

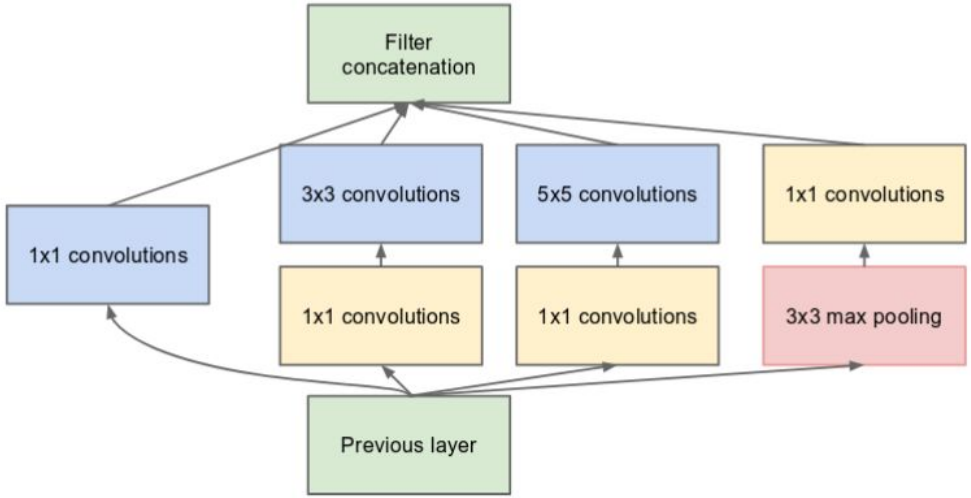
- Method: CNN
- Dataset: 1Million
- Motivation
  - *Well diversified dataset*
    - It achieves the projects needs
  - Due the complexity of annotate the dataset to use in a Faster R-CNN
  - Difficulty of create my own dataset

Process Steps

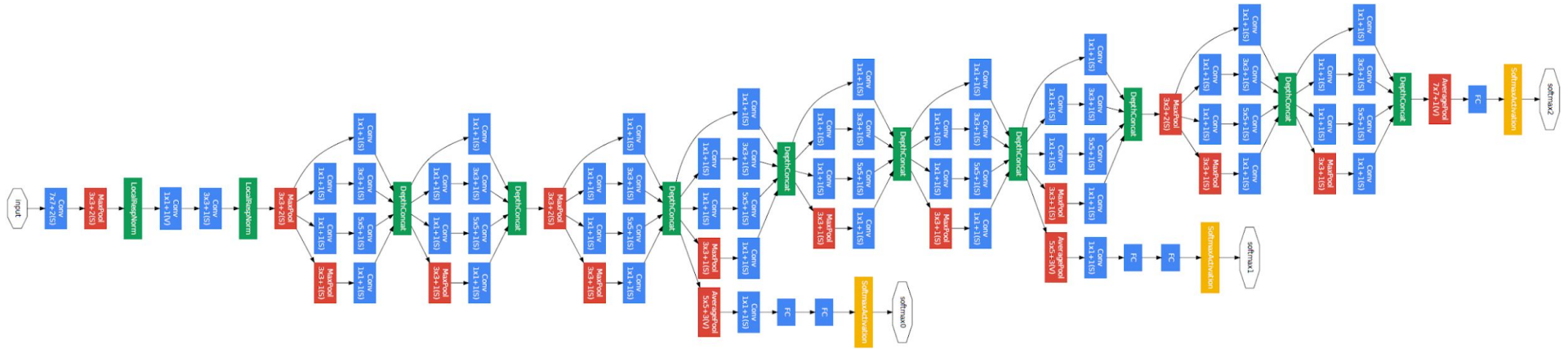
- Define the CNN Topology
- Train and Test

Network Topology

- Inception V1
  - Well defined
  - Scale Invariant



Network Topology



Training

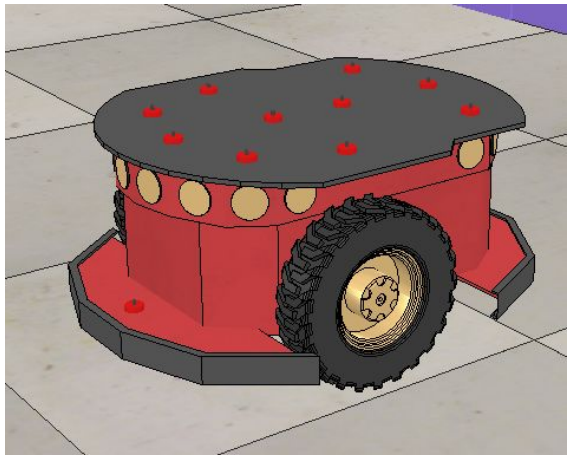
- 8 hours of training
  - 800,000 images for train
  - 100,000 to validate
  - 100,000 to test
- 61 classes
- Mean Filter
  - Extract Features

Execution

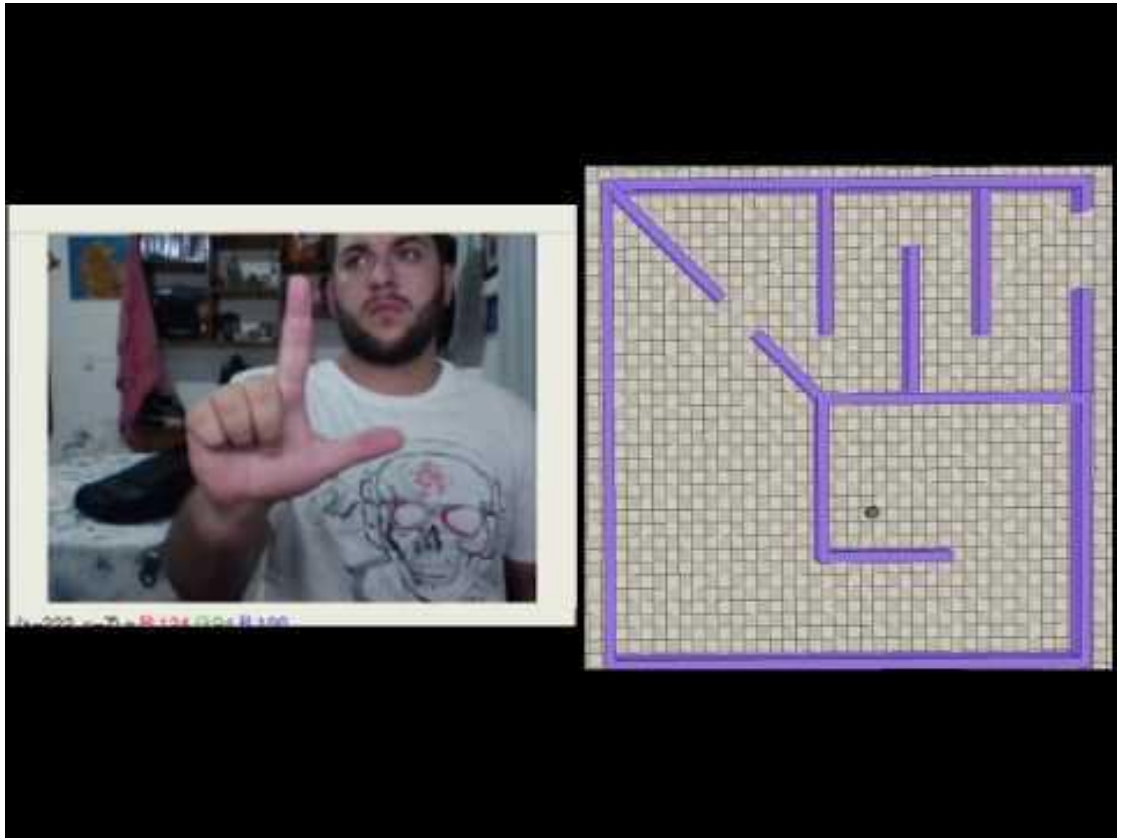
- Create a V-REP simulation and communicate the commands with the Robot via Python
  - Well documented

Execution

- Code: [Github]
- Methodology
  - 8 actions mapped to 8 gestures
    - Only gestures recognized above 85% accuracy
  - Robot performs the trajectory



Execution



Future Works

- Integrate more gestures
- Train a Faster R-CNN
  - Annotate the 1Million Dataset
- Use other robotic manipulators

Thanks!