

# Embedded Solution for Person Identification and Tracking

Master in Telecommunication Engineering

Ignacio Condés Menchén

Escuela Politécnica Superior  
Universidad Carlos III de Madrid

July 23, 2020

# Table of Contents

## 1 Introduction

Motivation  
Objectives

## 2 Proposed solution

Resources  
Design  
Perception Module  
Actuation Module

Results

Conclusions

## 3 Results

## 4 Conclusions

# Table of Contents

## 1 Introduction

- Motivation
- Objectives

## 2 Proposed solution

- Resources
- Design
  - Perception Module
  - Actuation Module

## 3 Results

## 4 Conclusions

Strong evolution in computer vision research and development.

Notable spreading of *deep learning* techniques on computer vision: CNNs (*Convolutional Neural Networks*).

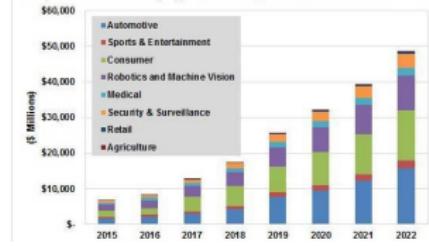
Multiple applications:

- Autonomous driving
- Medical diagnosis
- Surveillance

# Motivation I



Computer Vision Revenue by Application Market, World Markets: 2015-2022



## Motivation II



Powerful applications in robotics as well:

- Remote inspections on hazardousness
- Precision surgery
- Social robotics



# Objectives

Main goal: Autonomous robot that robustly follows a person.

## Introduction

Motivation

### Objectives

## Proposed solution

Resources

Design

Perception Module

Actuation Module

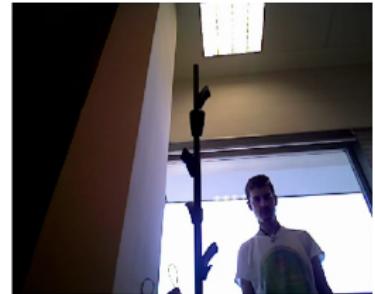
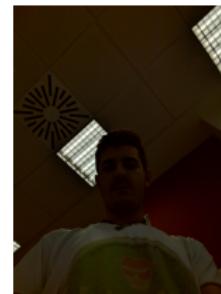
## Results

## Conclusions

# Objectives

Main goal: **Autonomous robot that robustly follows a person.**  
Subgoals:

- Implement a real-time person following behavior on affordable hardware.
- Use only robust CNNs for the inference tasks.
- Enhance the robustness combining the CNNs with optical tracking.



# Table of Contents

## 1 Introduction

Motivation  
Objectives

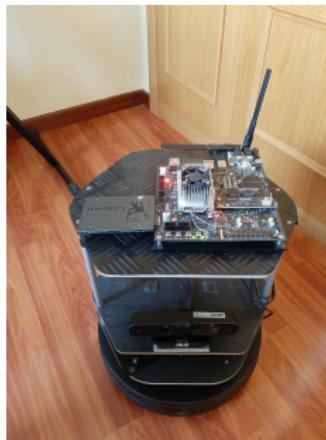
## 2 Proposed solution

Resources  
Design  
Perception Module  
Actuation Module

## 3 Results

## 4 Conclusions

## Hardware



- NVIDIA Jetson TX2
- Turtlebot2
- ASUS Xtion Pro Live

## Software



- Python 3.7
- ROS Melodic Morenia
- TensorFlow
- TensorRT

# Embedded Solution for Person Identification and Tracking

Ignacio Condés Menchén

## Introduction

Motivation

Objectives

## Proposed solution

Resources

Design

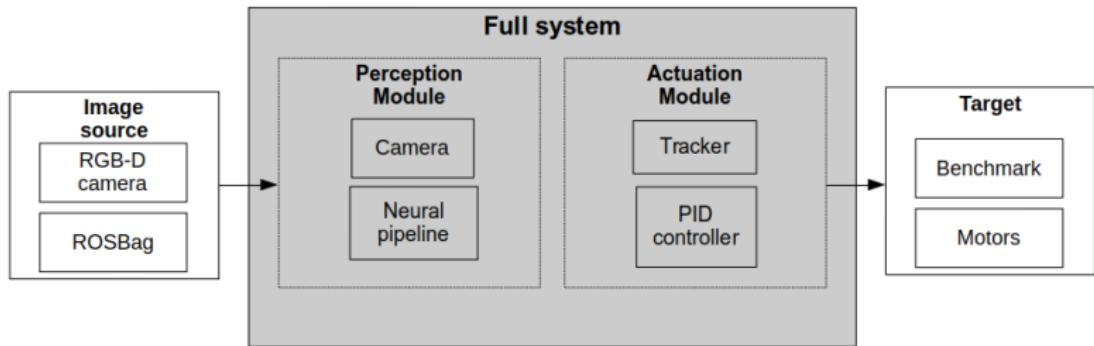
Perception Module

Actuation Module

## Results

## Conclusions

# Design



# Perception Module I

## Introduction

Motivation

Objectives

## Proposed solution

Resources

Design

Perception Module

Actuation Module

## Results

## Conclusions

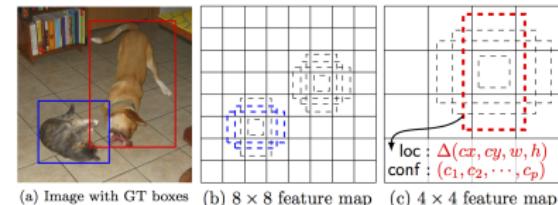
### Tasks to address:

- Person detection.
- Face detection.
- Face recognition.

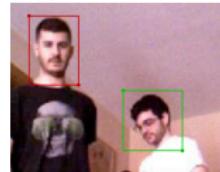
## Perception Module II

One devoted neural network for solving each task.

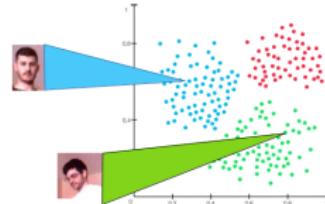
- **Person detection:** object detection CNNs: YOLOv3, SSD.



- **Face detection:** specific face detection CNN: faced.



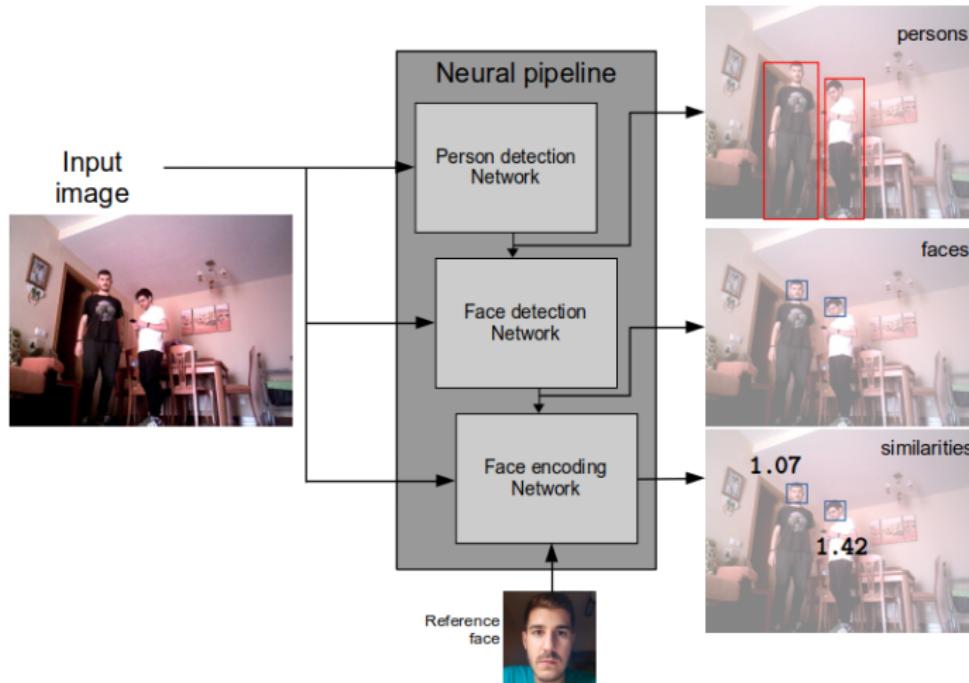
Latent space



- **Face recognition:** face encoder to latent  $\mathbb{R}^{128}$  space: FaceNet.

# Perception Module III

TensorRT used to optimize the networks across the neural pipeline.



# Actuation Module I

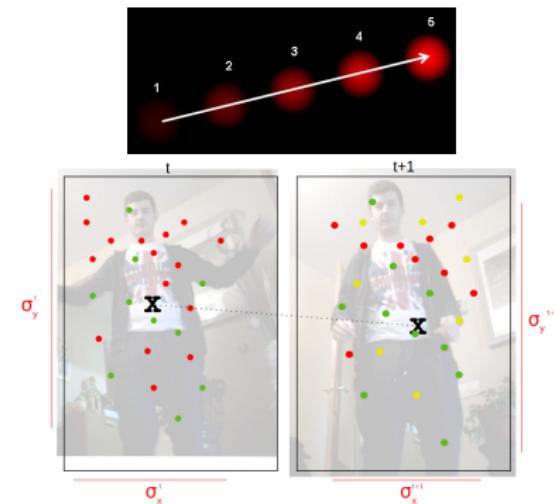
## Tasks to address:

- Track the persons.
- Move the robot.

# Actuation Module II

## Person tracker:

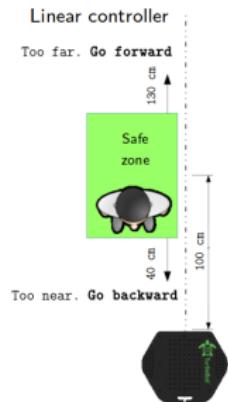
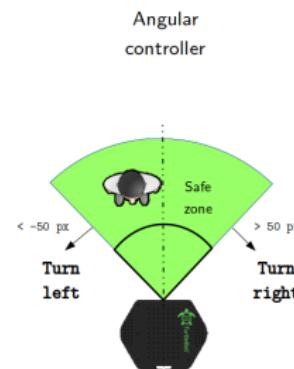
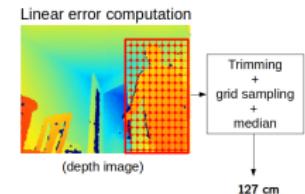
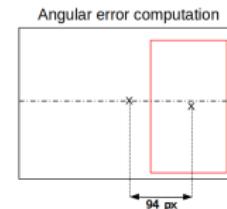
- Based on optical flow estimation (Lucas-Kanade).
- Shift and transform the bounding box.
- Patience for alleviating occlusions.



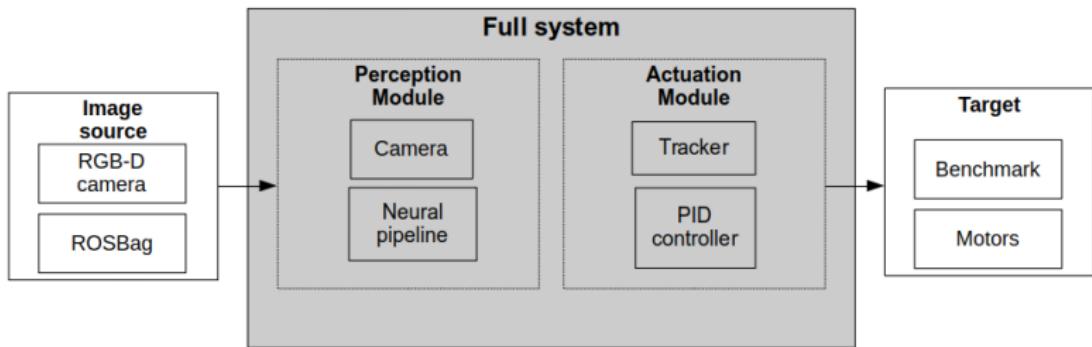
# Actuation Module III

## PID Controllers:

- Closed-loop control system.
- One per degree of freedom.
- Reactive responses.
- Safe zones for smooth response.



# Design summary



# Table of Contents

## 1 Introduction

Motivation  
Objectives

## 2 Proposed solution

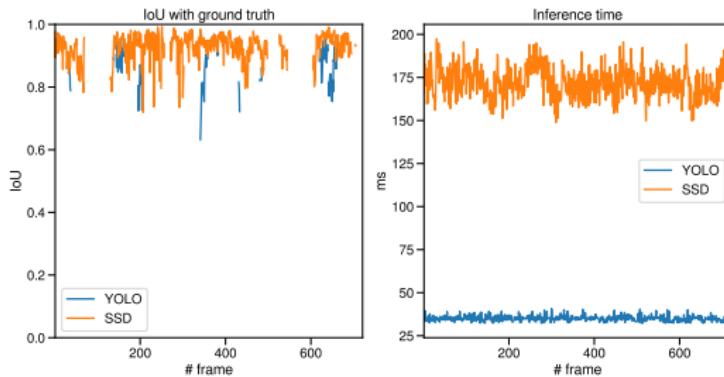
Resources  
Design  
Perception Module  
Actuation Module

## 3 Results

## 4 Conclusions

# Person detection

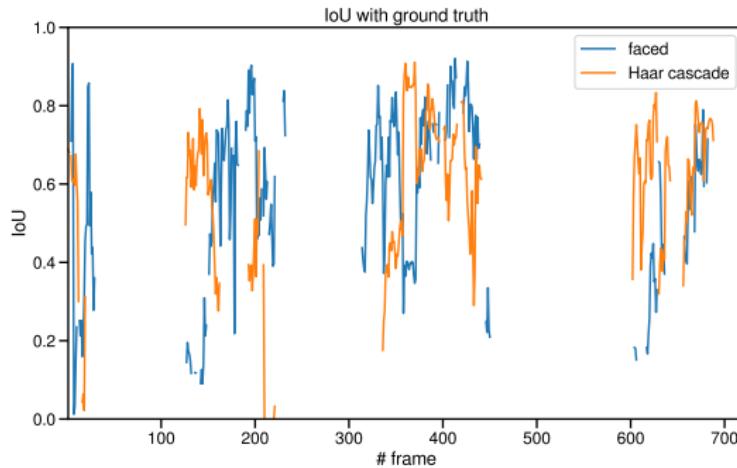
## Precision and inference time: SSD vs. YOLOv3 tiny



	YOLO	SSD
IoU	$0.858 \pm 0.068$	$0.926 \pm 0.044$
Inf. time (ms)	$35.003 \pm 1.503$	$172.237 \pm 8.791$
Frames with detection	123 (17.06%)	533 (73.93%)

# Face detection

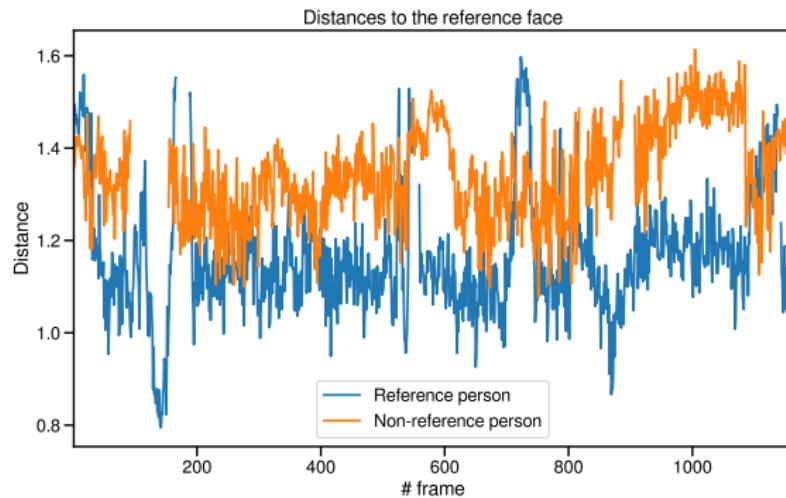
Precision: neural detector (`faced`) vs Haar-features classifier.



	haar	faced
IoU	$0.579 \pm 0.202$	$0.559 \pm 0.221$
Frames with detection	248 (34.40%)	266 (36.89%)

# Face recognition

Distance between two faces and the reference face.



	Ref. person	Non-ref. person
IoU	$1.160 \pm 0.128$	$1.344 \pm 0.102$

## Introduction

Motivation

Objectives

## Proposed solution

Resources

Design

Perception Module

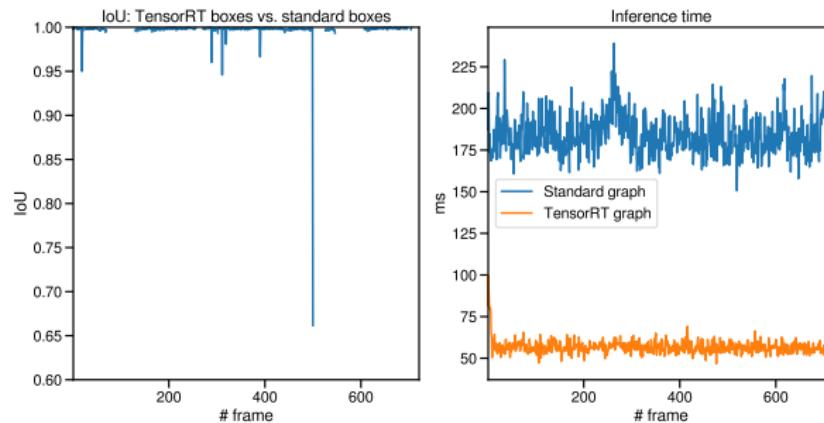
Actuation Module

## Results

## Conclusions

# TensorRT Optimizations

## Loss in precision and inference time.

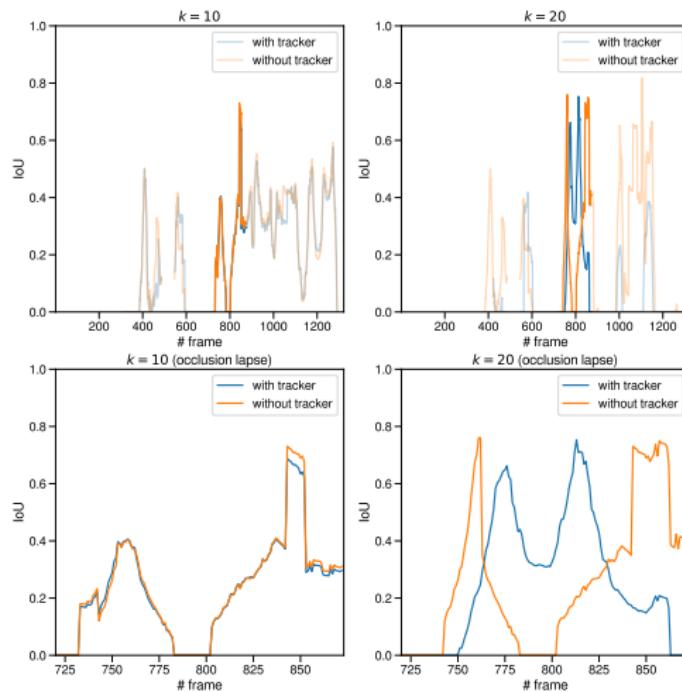


	Original graph	TensorRT graph
Inference time (ms)	$184.477 \pm 11.827$	$56.769 \pm 4.148$
IoU with original graph	-	$0.997 \pm 0.015$

# Motion tracker

Gain in robustness to occlusions using the tracker.

IoU with ground truth



# Embedded Solution for Person Identification and Tracking

Ignacio  
Condés  
Menchén

## Introduction

Motivation

Objectives

## Proposed solution

Resources

Design

Perception Module

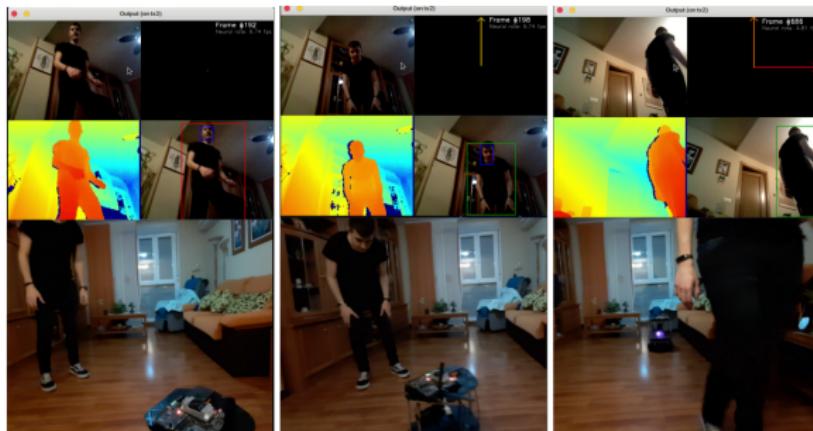
Actuation Module

## Results

## Conclusions

# Full system

Following the person.



# Table of Contents

## 1 Introduction

Motivation  
Objectives

## 2 Proposed solution

Resources  
Design  
Perception Module  
Actuation Module

## 3 Results

## 4 Conclusions

# Conclusions I

Accomplishment of the outlined objectives:

- ✓ Real-time following behavioral.
- ✓ Affordable/educational hardware.

# Conclusions I

Accomplishment of the outlined objectives:

- ✓ Real-time following behavioral.
- ✓ Affordable/educational hardware.
- ✓ Robust detection relying on deep learning.

# Conclusions I

Accomplishment of the outlined objectives:

- ✓ Real-time following behavioral.
- ✓ Affordable/educational hardware.
- ✓ Robust detection relying on deep learning.
- ✓ Enhanced softness using an optical tracker.

## Conclusions II

Main drawbacks found:

- ✗ Lighting affected by the camera position.

# Conclusions II

Main drawbacks found:

- ✖ Lighting affected by the camera position.
- ✖ Persons crossing each other may be confounded.

## Conclusions II

Main drawbacks found:

- ✖ Lighting affected by the camera position.
- ✖ Persons crossing each other may be confounded.
- ✖ Collisions with obstacles may occur.

# Conclusions III

Future lines to address:

- Sensor fusion to enhance the perception.

# Conclusions III

Future lines to address:

- Sensor fusion to enhance the perception.
- Probabilistic tracker to model trajectories.

# Conclusions III

Future lines to address:

- Sensor fusion to enhance the perception.
- Probabilistic tracker to model trajectories.
- Navigation algorithms for obstacle avoidance.

# Embedded Solution for Person Identification and Tracking

Ignacio  
Condés  
Menchén

## Introduction

Motivation

Objectives

## Proposed solution

Resources

Design

Perception Module

Actuation Module

## Results

## Conclusions

# Questions

# Embedded Solution for Person Identification and Tracking

Ignacio  
Condés  
Menchén

## Introduction

Motivation

Objectives

## Proposed solution

Resources

Design

Perception Module

Actuation Module

## Results

## Conclusions

# Thank you