Max Planck Institute for Intelligent Systems
Perceiving Systems Department

MAX-PLANCK-GESELLSCHAFT

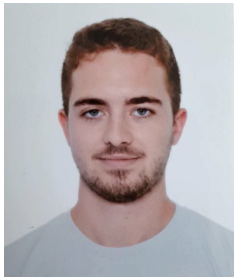
Deep Neural Network-Based Cooperative Visual Tracking through Multiple Flying Robots

Aamir Ahmad
Research Scientist

Eric Price
PhD Student

Guilherme Lawless

Roman Ludwig



Igor Martinović

Interns and Student Assistants

Michael Black
Director

Robot Perception Group

A Westpac Little Ripper Lifesaver drone rescuing swimmers in Australia. Courtesy: The Guardian via YouTube

Observation 1

- Emergencies, Accidents, or an Outdoor Event
  - Aerial support system.
  - Human responders.
  - Autonomous robots.

Observation 2

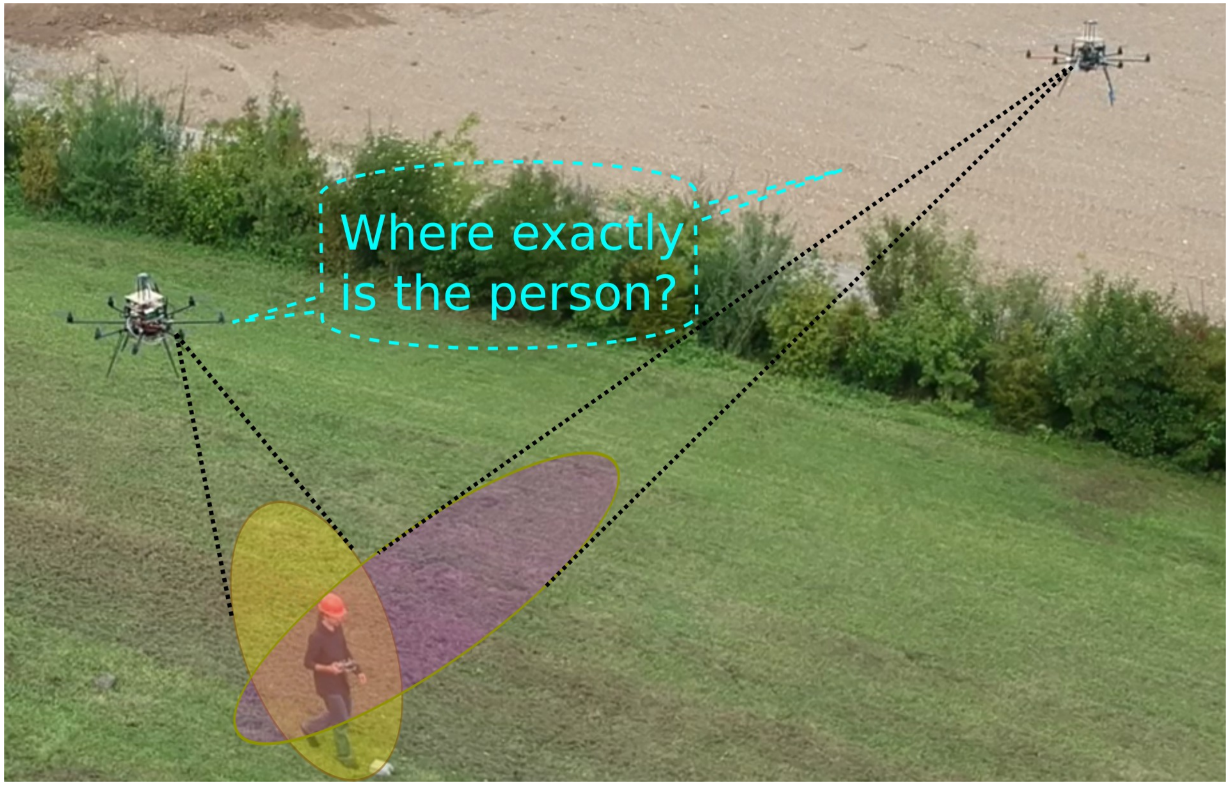
- Emergencies, Accidents, or an Outdoor Event
  - Aerial support system.
  - Human responders.
  - Autonomous robots.
- Response Time
  - Autonomously search, detect and rescue.

Observation 3

- Emergencies, Accidents, or an Outdoor Event
  - Aerial support system.
  - Human responders.
  - Autonomous robots.
- Response Time
  - Autonomously search, detect and rescue.
- Reliable Navigation
  - Experts not always present.
  - Unfriendly environment.
  - Unstable systems.

- Emergencies, Accidents, or an Outdoor Event
  - Aerial support system.
  - Human responders.
  - Autonomous robots.
- Response Time
  - Autonomously search, detect and rescue.
- Reliable Navigation
  - Experts not always present.
  - Unfriendly environment.
  - Unstable systems.

AirCap: Aerial Outdoor Motion Capture



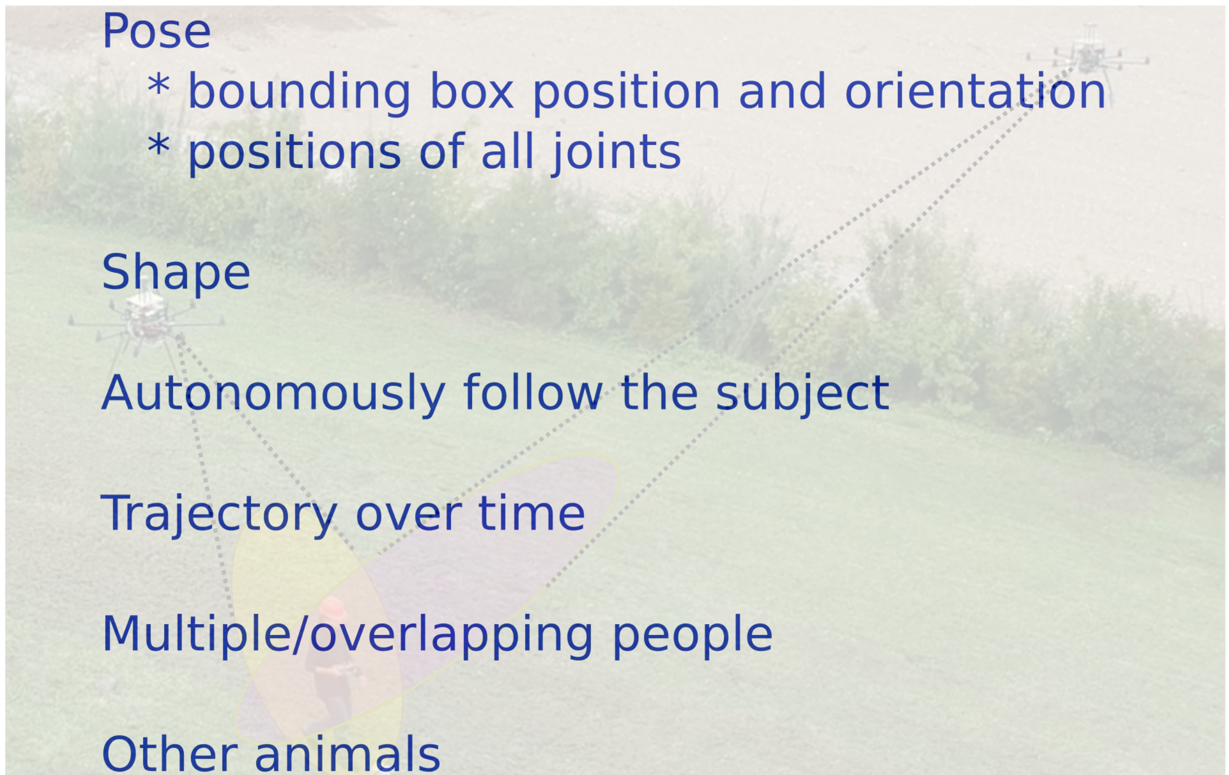
Online and On-board pose estimates for

Realtime Applications

Offline pose, shape and scene estimates for

Post-Analysis Applications

AirCap: Aerial Outdoor Motion Capture



Pose
* bounding box position and orientation
* positions of all joints

Shape

Autonomously follow the subject

Trajectory over time

Multiple/overlapping people

Other animals

Online and On-board pose estimates for

Realtime Applications

Offline pose, shape and scene estimates for

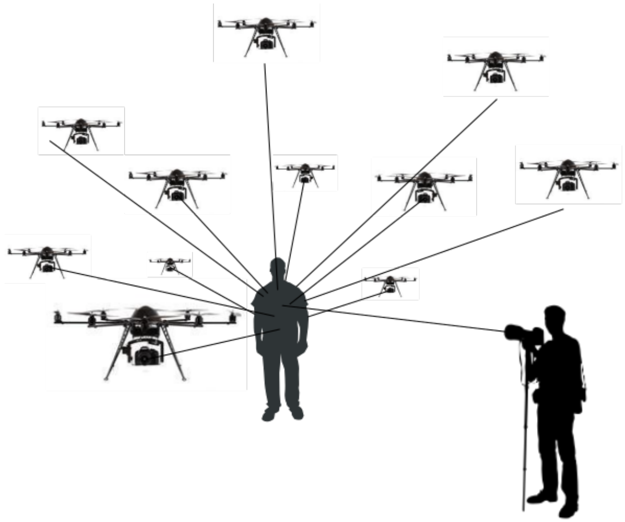
Post-Analysis Applications

Indoor MO-CAP: 4D Scanner at MPI

- 22 stereo camera pairs
- 22 full color cameras
- pattern projectors

- 60 meshes per second
- 150k points per mesh
- ~$(1.5m)^2$ scan platform

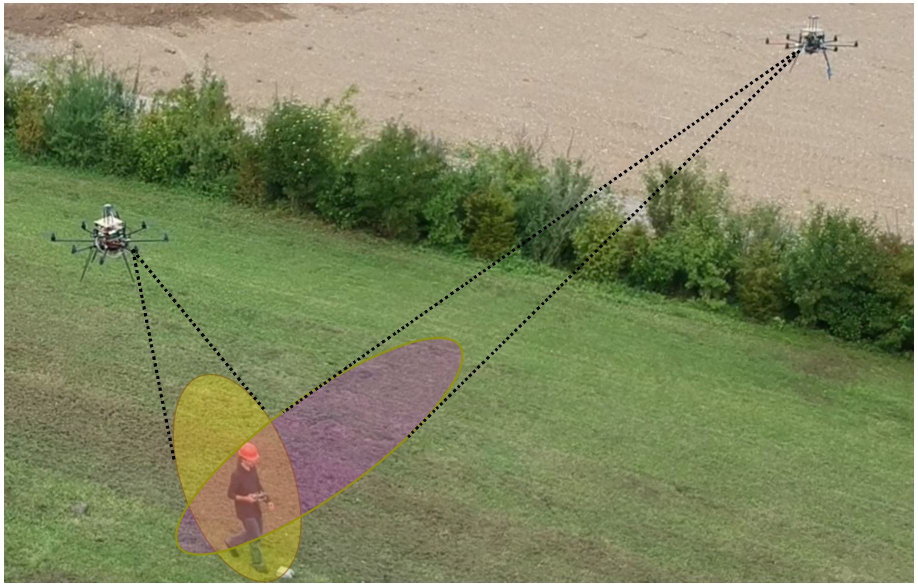
Can we do MO-CAP in outdoor scenarios?



- A team of flying robots.
- Cameras mounted on each robot.
- Robots follow the subject.
- Robot formation covers all important viewpoints.

Outdoor MO-CAP using UAVs –
What are the Research Challenges?

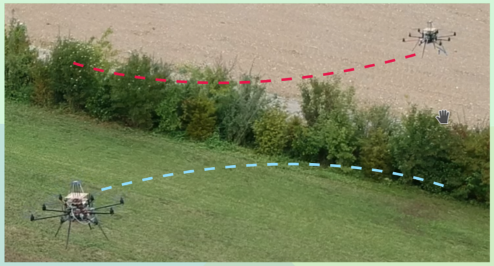
- Real-time performance.
- Detect and track people.
- Self-localization.
- Multi-robot Information fusion.
- Inter-robot collision avoidance.
- Obstacle detection and avoidance.
- Formation control.



Max Planck Institute for Intelligent Systems
Perceiving Systems Department

Our Unique Approach

Bayesian Sensor Fusion & Optimal Control



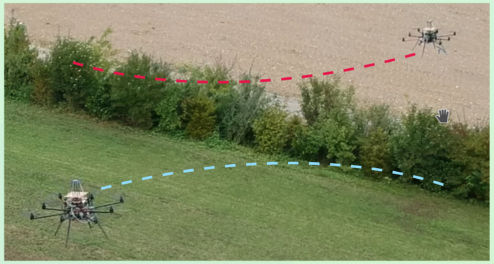
Deep Learning for Vision

Novel Robotic Hardware

Bayesian Sensor Fusion & Optimal Control

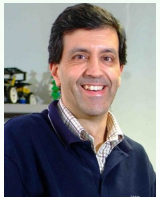
Our Unique Approach

Bayesian Sensor Fusion & Optimal Control



Deep Learning for Vision

Novel Robotic Hardware



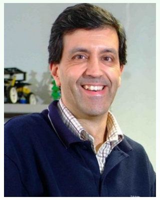
Integrated Bayesian Approach to Cooperative Perception (CP)

What is CP?

- Self pose
- Teammate poses
- Poses of dynamic targets/objects
- Map of the environment

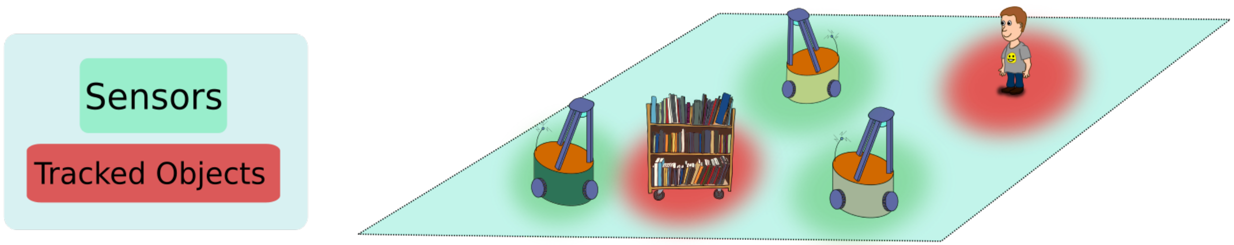
Why do we need it?

- Increase the accuracy and precision of the local estimates.
- Improve robustness to individual sensor failures.
- Increased coverage area.
- Improves coordination when performing teamwork!



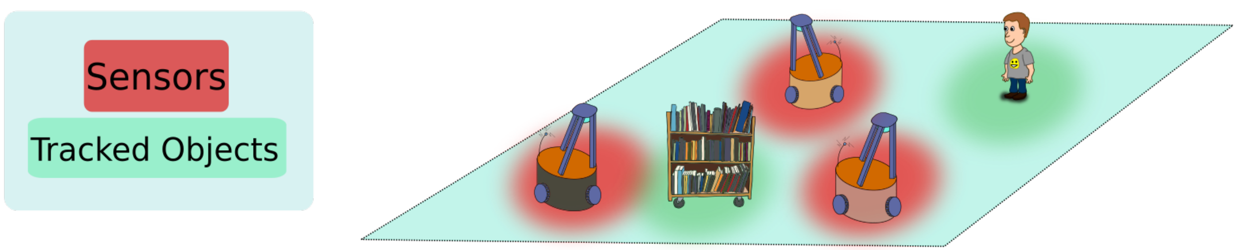
Bayesian Sensor Fusion & Optimal Control

Assumptions and problem classes in CP

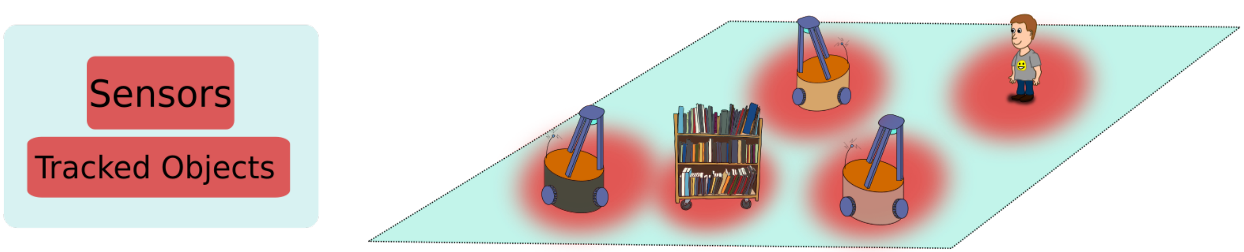


Multirobot cooperative object tracking

Assumptions and problem classes in CP

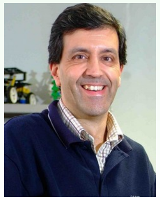


Multirobot cooperative robot localization



Unified cooperative robot localization and target tracking





Particle Filter-based Unified Cooperative Robot Localization and Object Tracking (PF-UCLT)

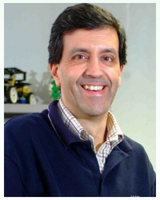
Given a team of robots with:

- On-board sensors (camera, odometry, etc.).
- Moving object(s) to be tracked.
- Known environment.

PF-UCLT: An ONLINE unified approach for

- Tracking object(s) cooperatively.
- Localizing robots cooperatively
  - using the cooperatively tracked object
- Using a particle filter-based estimator.





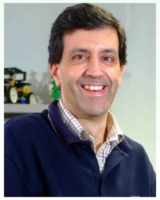
Why a Standard PF Solution is Not Suitable?

Let **M** = number of particles required for a single robot localization or a single object tracking situation.

Disadvantage of the standard PF solution:

- Required number of particles grow exponentially with the dimensions of the estimated state.
- If state consists of **N** robots' poses and **O** objects' positions, required number of particles is in the order of **M^(N+O)**.





PF-UCLT: What can our method do?

PF-UCLT is an online approach

- Recursive Bayesian estimation: estimates the most recent robots' poses and objects' positions.
- Computationally decentralized: each robot runs its own instance of PF-UCLT.
- Fast execution speed is crucial.

Core idea:

* A novel sub-particle concept and rearranging technique.

* Exploiting the properties of conditional and/or mutual independence of the involved variables to speed up the PF.



Bayesian Sensor Fusion & Optimal Control

An Online Scalable Approach to Unified Multirobot Cooperative Localization and Object Tracking

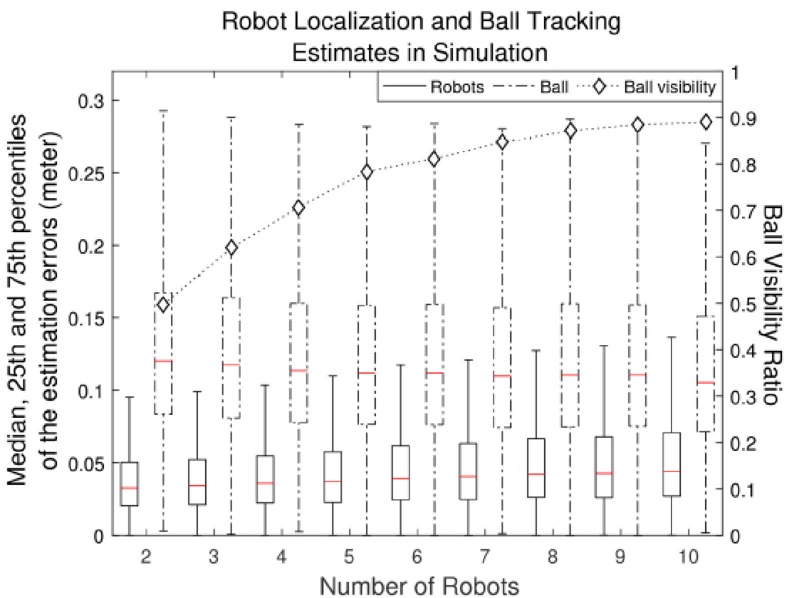
Aamir Ahmad[1,2], Guilherme Lawless[2] and Pedro Lima[2]

1. Max Planck Institute for Intelligent Systems, Tübingen, Germany.
2. Institute for Systems and Robotics, Instituto Superior Técnico, Universidade de Lisboa, Portugal.

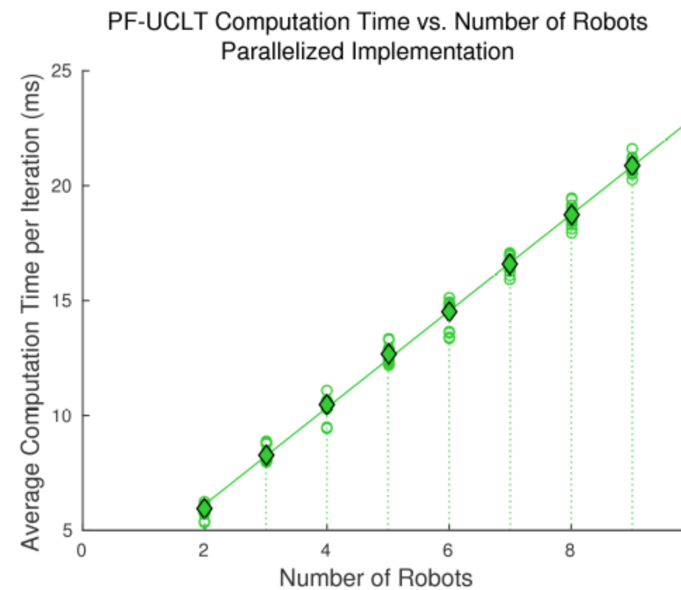
Published in: IEEE Transactions on Robotics
Volume: 33, Issue: 5, Oct. 2017

PF-UCLT Results

PF-UCLT : Experiments and Results



State estimation accuracy w.r.t.
the number of robots

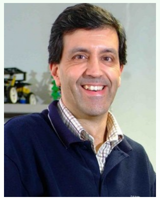


Computation time w.r.t.
the number of robots

PF-UCLT - Publications

- A. Ahmad, G. Lawless and P. Lima, "**An Online Scalable Approach to Unified Multirobot Cooperative Localization and Object Tracking,**" *in IEEE Transactions on Robotics*, vol. 33, no. 5, pp. 1184-1199, **Oct. 2017**.

- Additionally **accepted** as a journal track paper at **ICRA 2018, Brisbane.**



Model Predictive Control-based Formation Controller

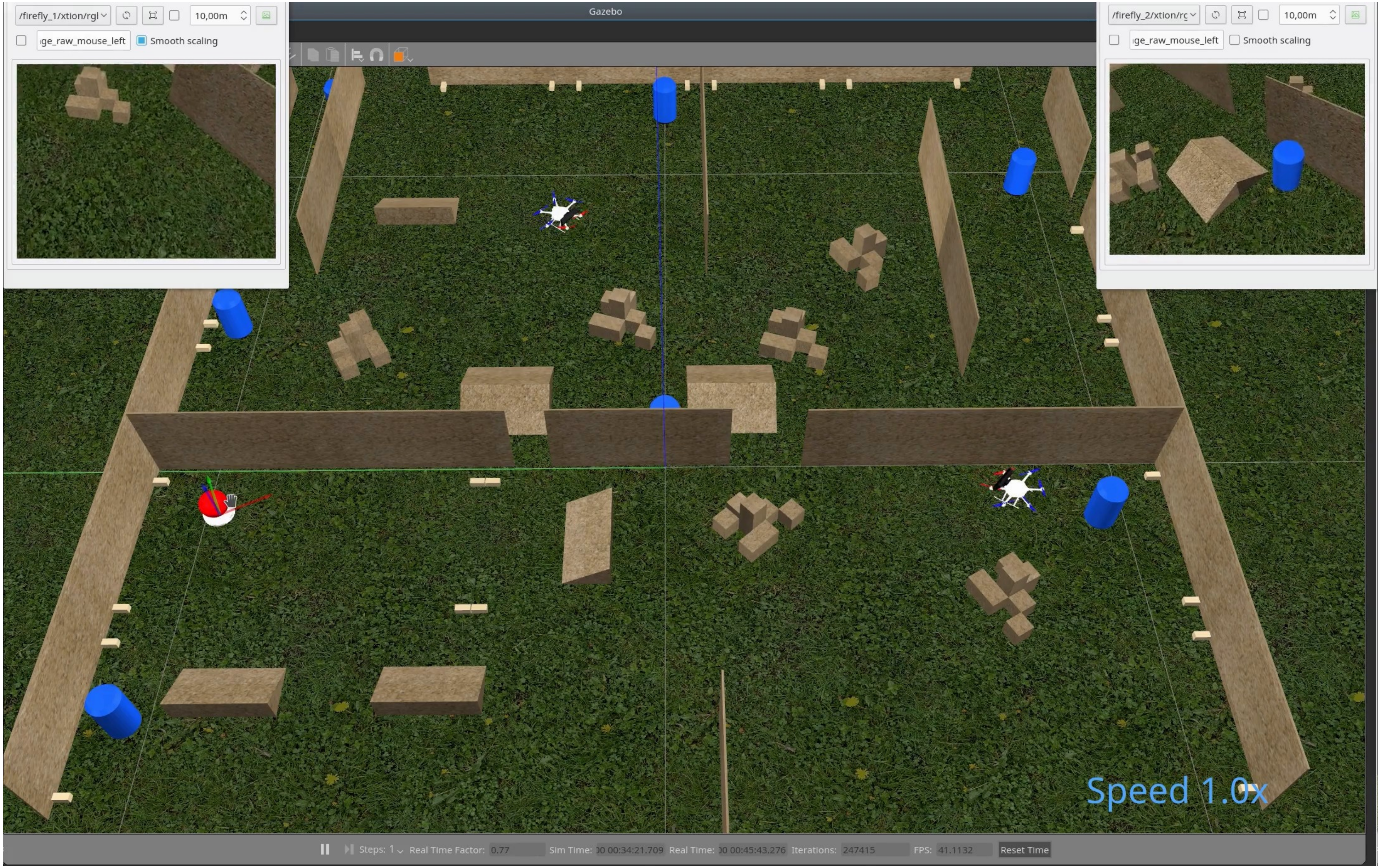
- High-level (and high order) trajectory generation for multiple aerial robots.
- Multiple convex and non-convex constraints:
  - Outer bounding box.
  - Safety distance to the target.
  - Safety distance to teammates (collision avoidance).
- **Fused target estimate with associated uncertainty.**
- Current solution: MPC for convex constraints [1] + potential field for collision avoidance.
- Ongoing work: Integrating collision avoidance within the MPC.

[1] P. Lima, A. Ahmad, A. Dias, A. Conceição, A. Moreira, E. Silva, L. Almeida, L. Oliveira, T. Nascimento, **Formation control driven by cooperative object tracking**, *Robotics and Autonomous Systems*, Volume 63, Part 1, Pages 68-79, Jan 2015.

Max Planck Institute for Intelligent Systems
Perceiving Systems Department

MAX-PLANCK-GESELLSCHAFT

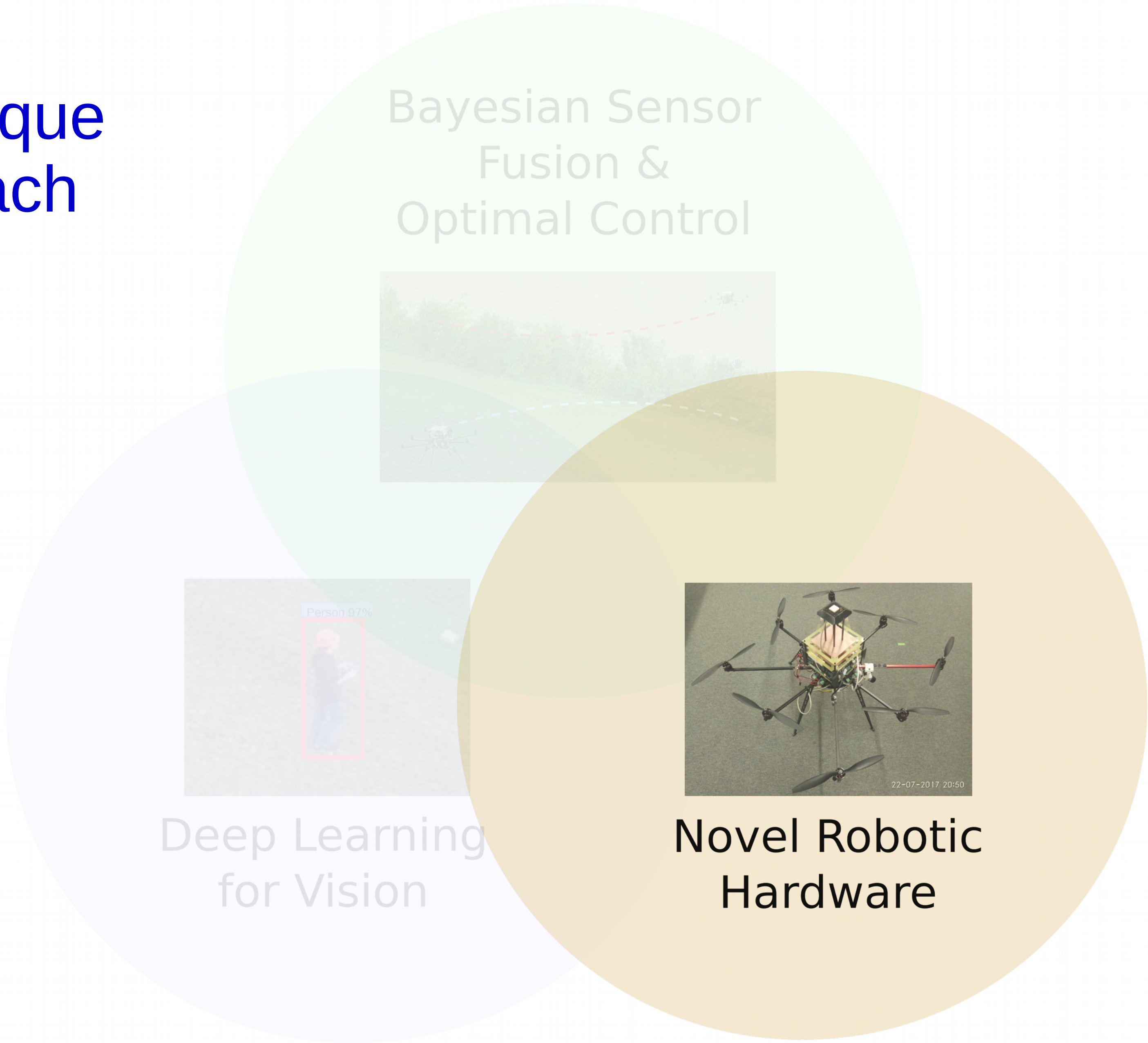
Bayesian Sensor Fusion & Optimal Control



MPC + Potential Field in Simulation

Novel Robotic Hardware

Our Unique Approach



Novel Robotic Hardware

Why our own hardware?

- Commercially available systems are too limited.
- Open Source design allows easy reproduction of results.

Max Planck Institute for Intelligent Systems
Perceiving Systems Department

Novel Robotic Hardware

Design criteria:

- On board computer vision.
- Full HD, 60 fps video recording.
- GPU.
- Reliability - for safe outdoor flight near people.

Max Planck Institute for Intelligent Systems
Perceiving Systems Department

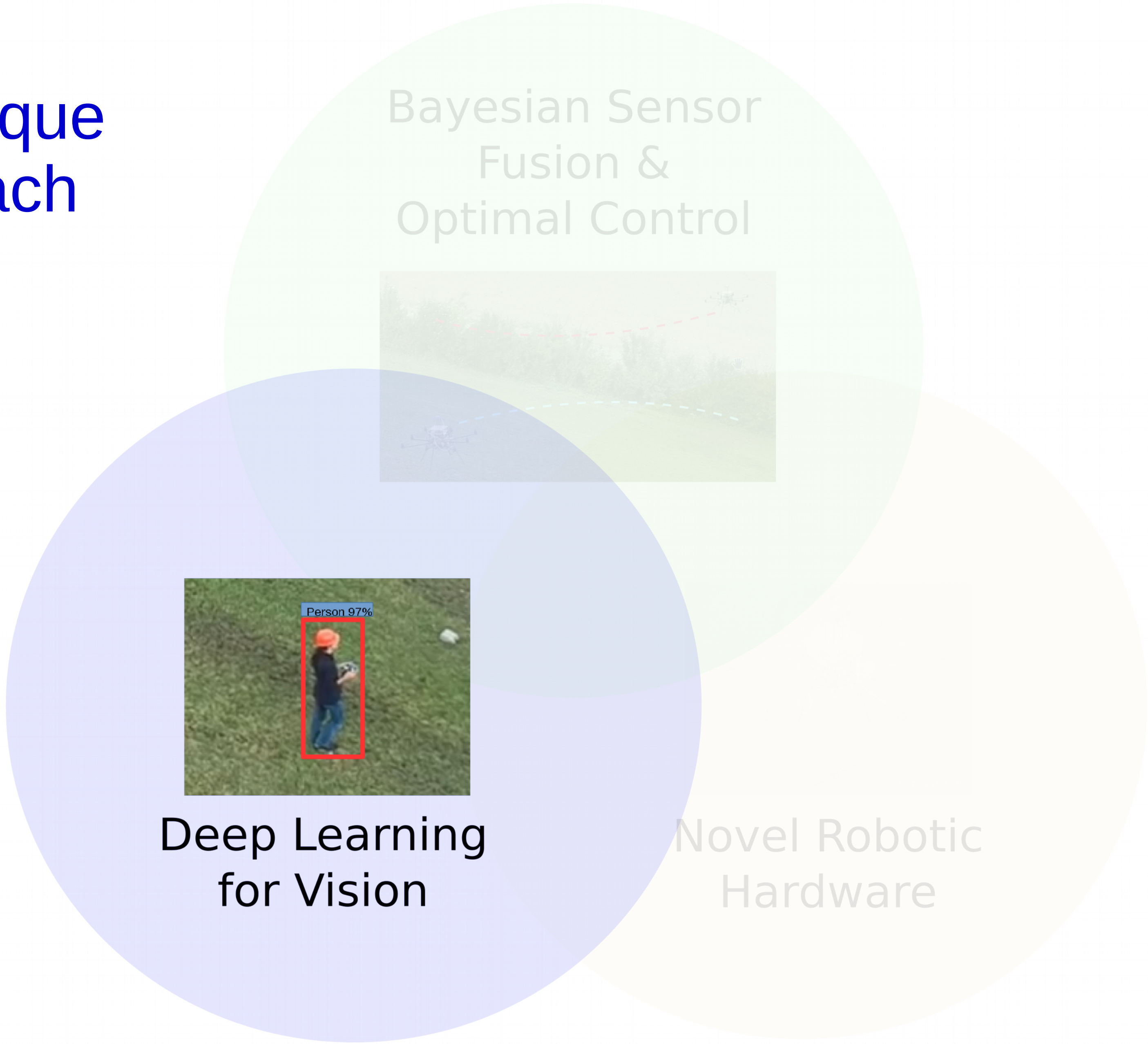
Novel Robotic Hardware

- Commercial frame.
- Open Source flight-controller.
- NVIDIA Jetson TX1 GPU.
- High BW storage HW.

Deep Learning for Vision

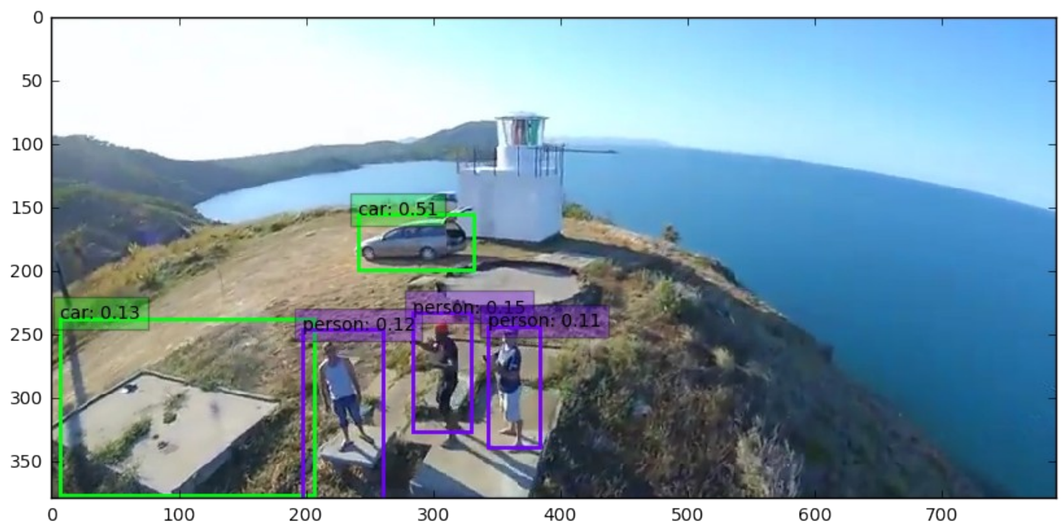
Our Unique Approach



Deep Learning for Vision

Detection and Tracking: What detector to use?



- State of the art: Wei Liu’s SSD Multibox detector.
- Very reliable detection and localization on aerial images out of the box.
- 300x300 pixel ~4 fps in sequential mode on our hardware.

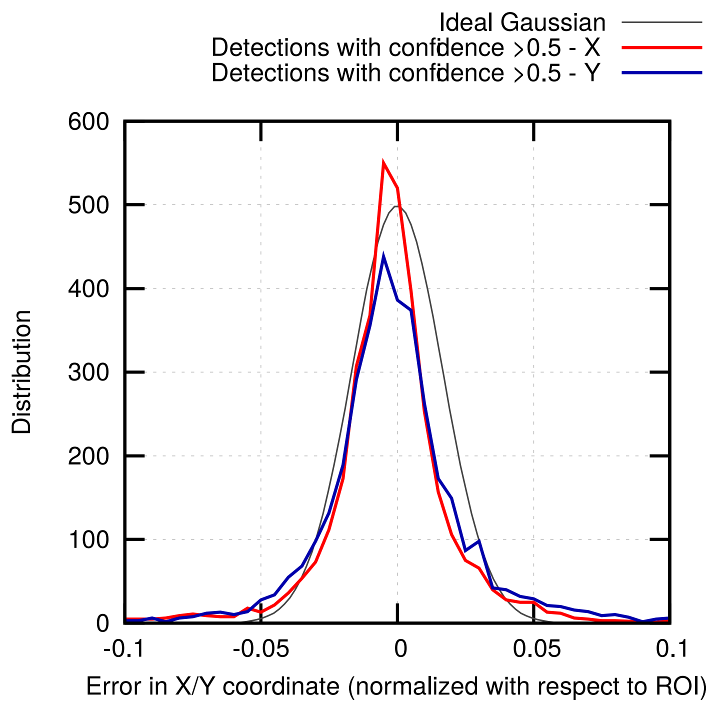
Accuracy of detection

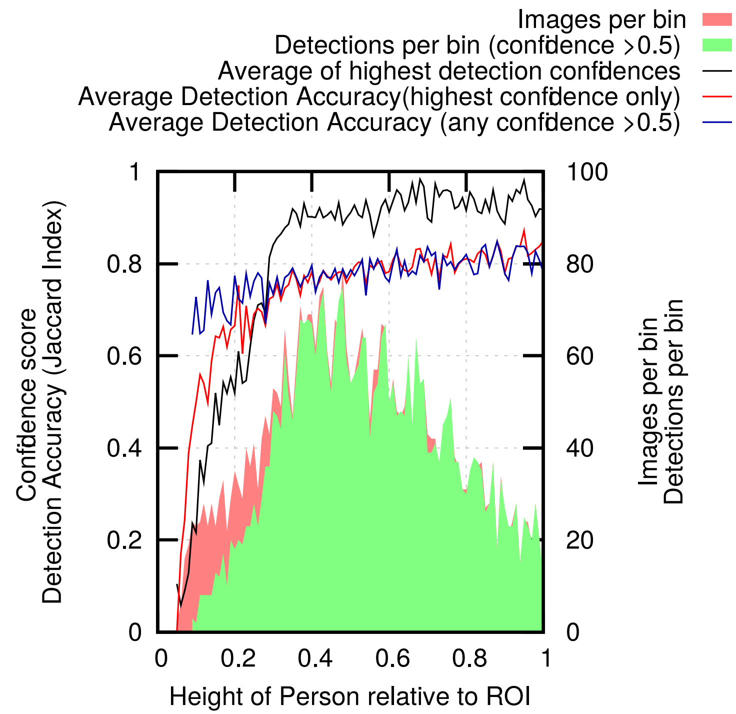
Accuracy of detection

- Test with thousands of images.
- Make detections of same content with different sizes.
- Analyze statistics.

Deep Learning for Vision

Accuracy of detection





- Person needs to be at least 10% of the image height to be detected.
- Accuracy has Gaussian error with known variance.
- Can be treated like any other noisy sensor!

Deep Learning for Vision

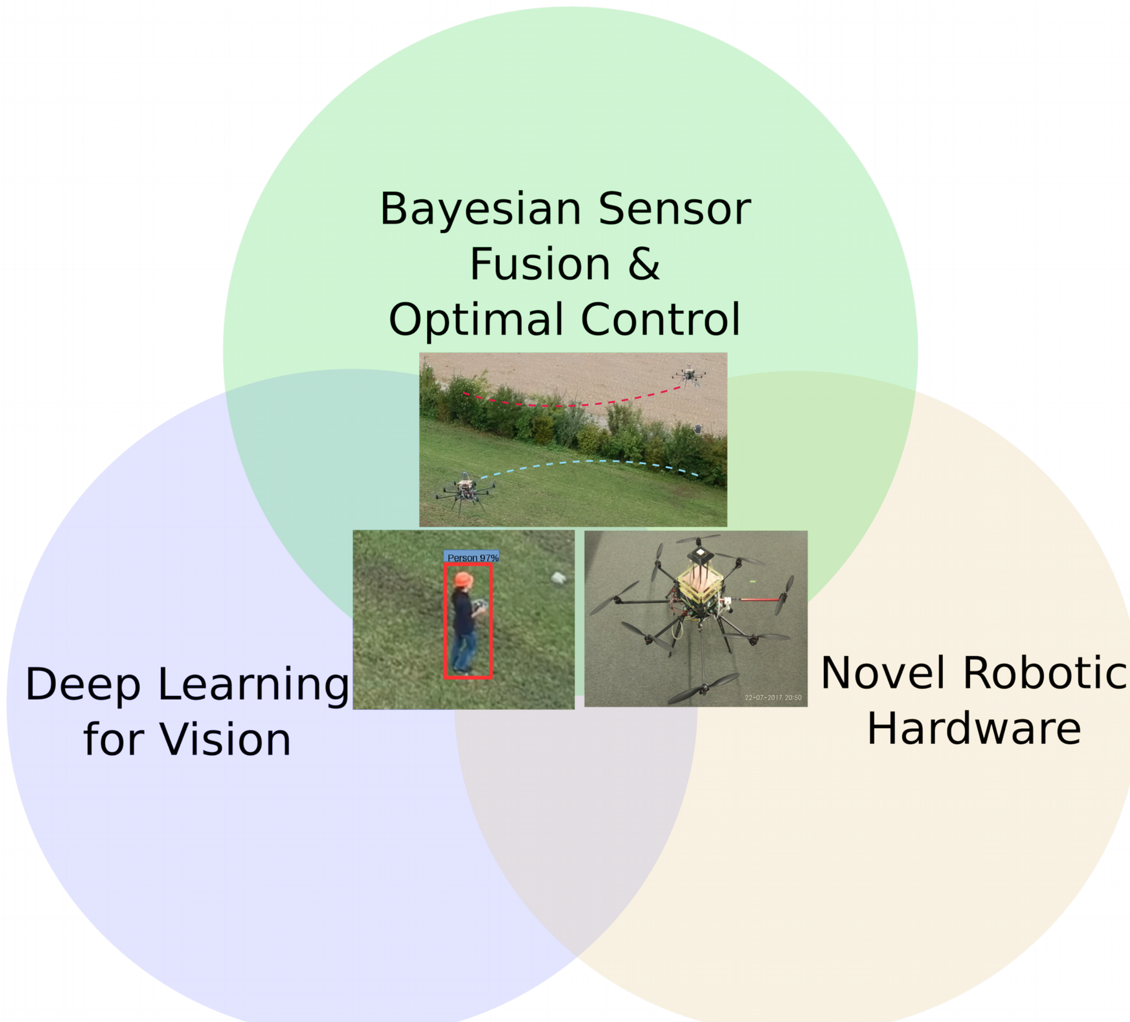
Can we detect and track

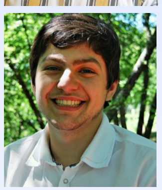
far away people

with multiple cooperating flying robots?

Deep Learning for Vision

Yes, we can…
...through our unique approach

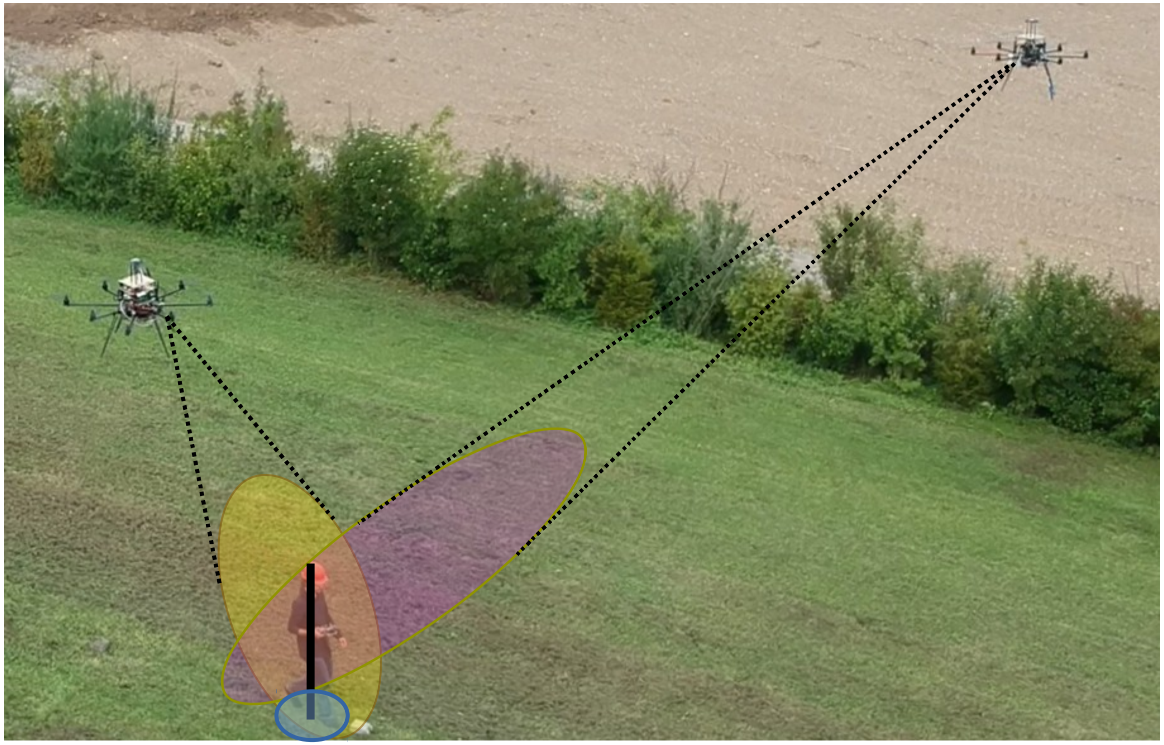




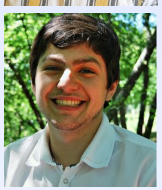

1. Detect the person

2. Fuse detections

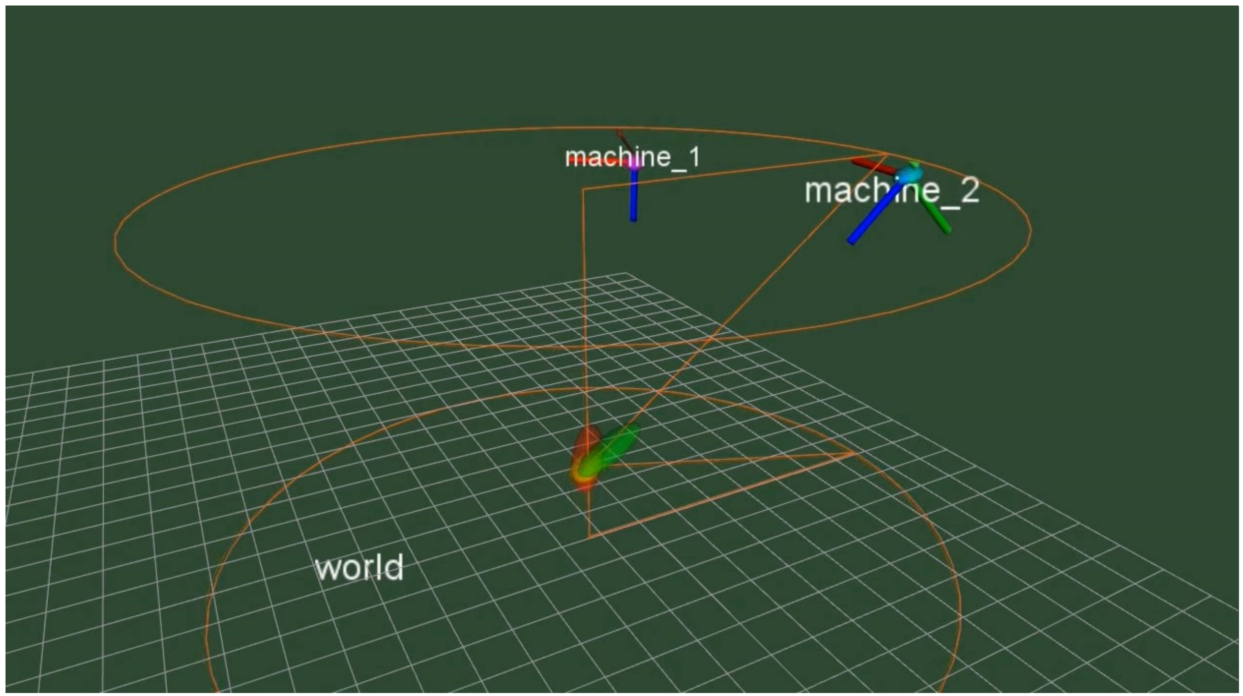


- We know how large a person can be.
- As such, we can estimate where the person could be in 3D space.
- Every vehicle can do that separately.
- The vehicles can easily share that knowledge:
  - It’s only a point with uncertainty ellipse

Deep Learning for Vision

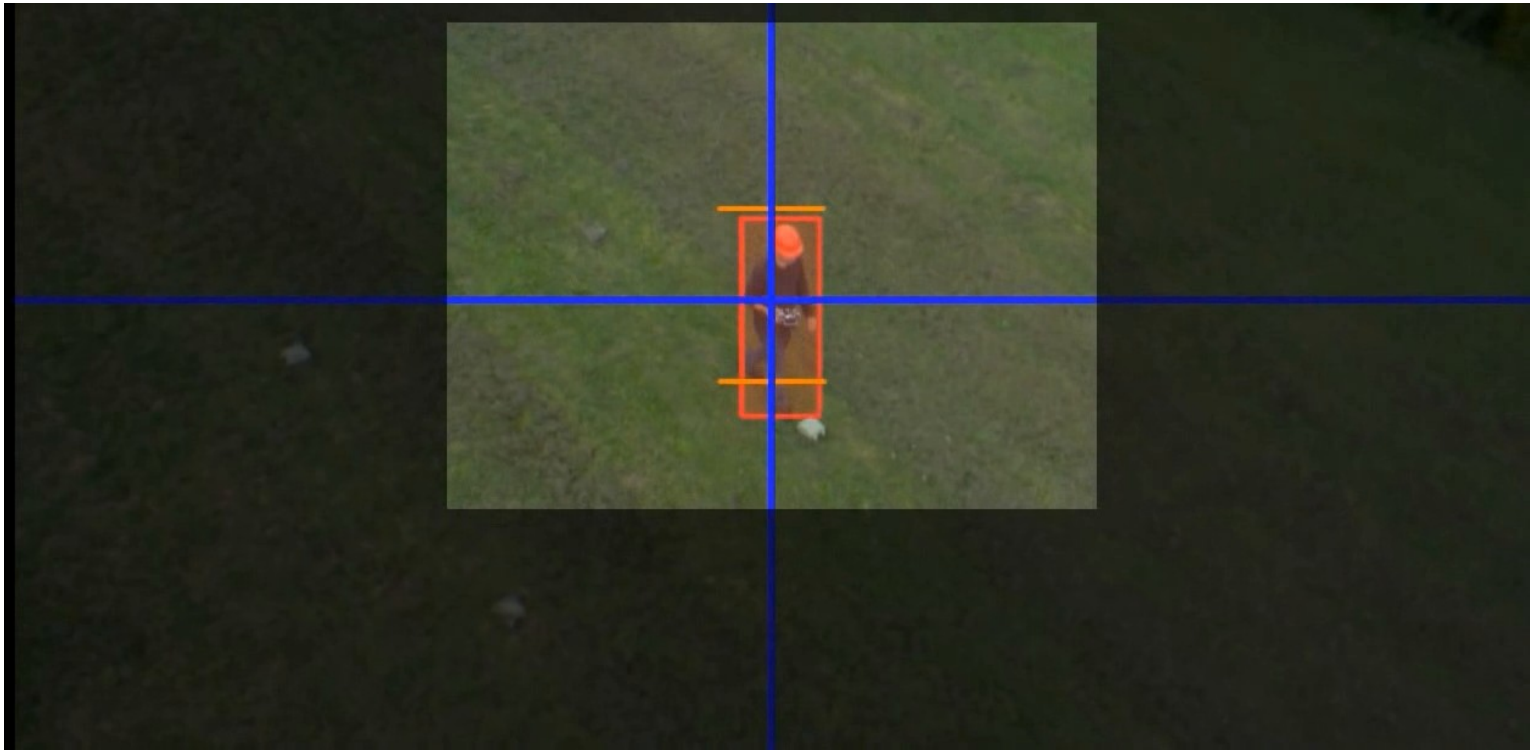


2. Fuse detections



- A Kalman filter can fuse all detections.
- We only need a very simple motion model.
- We can predict where the person is going to be!

3. Predict person position

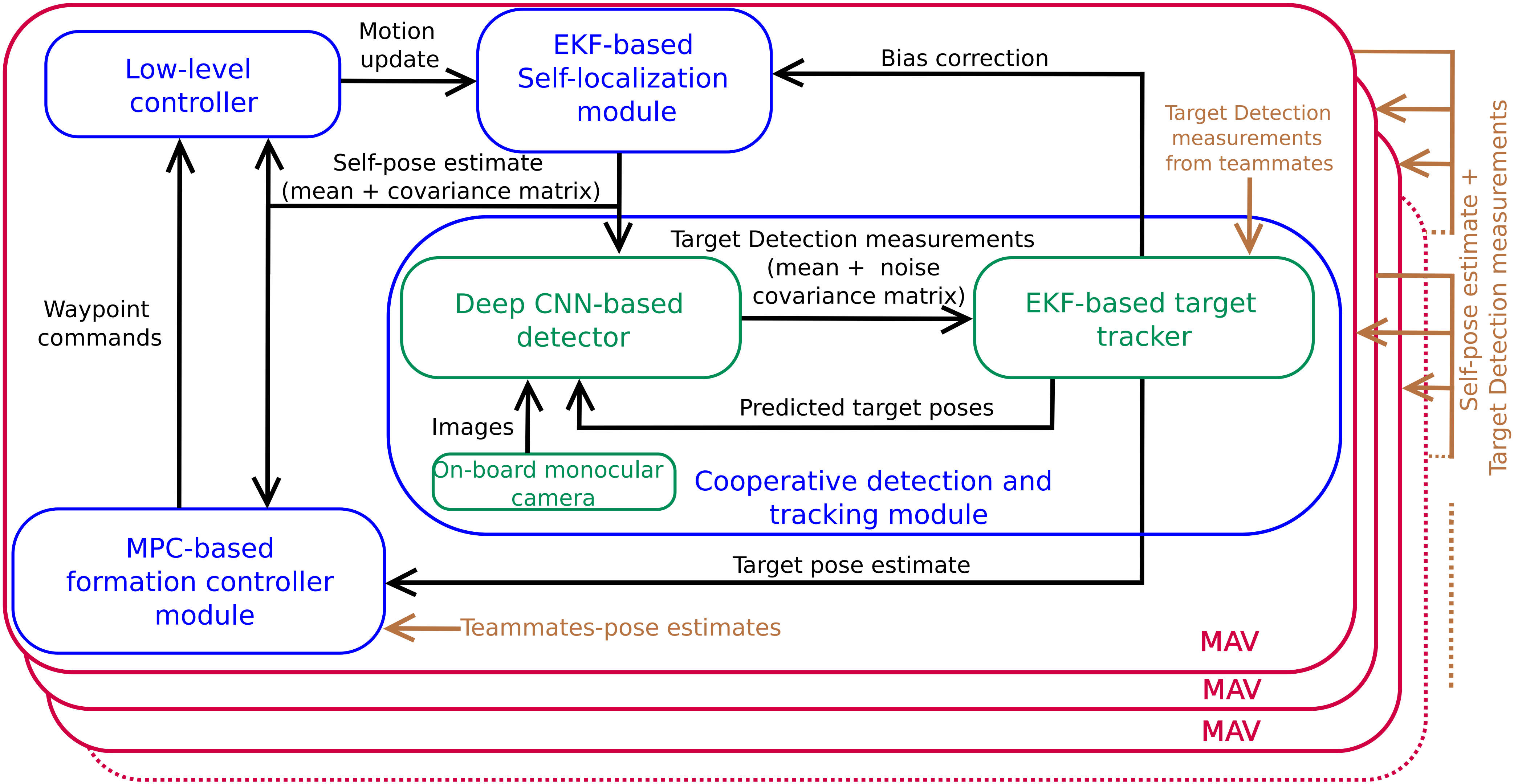


- We know where the person can and should be.

  → We know where to look.
- We can make sure the person will be large enough for successful detection!

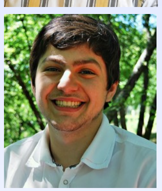
  (larger than 10%)

Max Planck Institute for Intelligent Systems
Perceiving Systems Department



System Architecture

Deep Learning for Vision

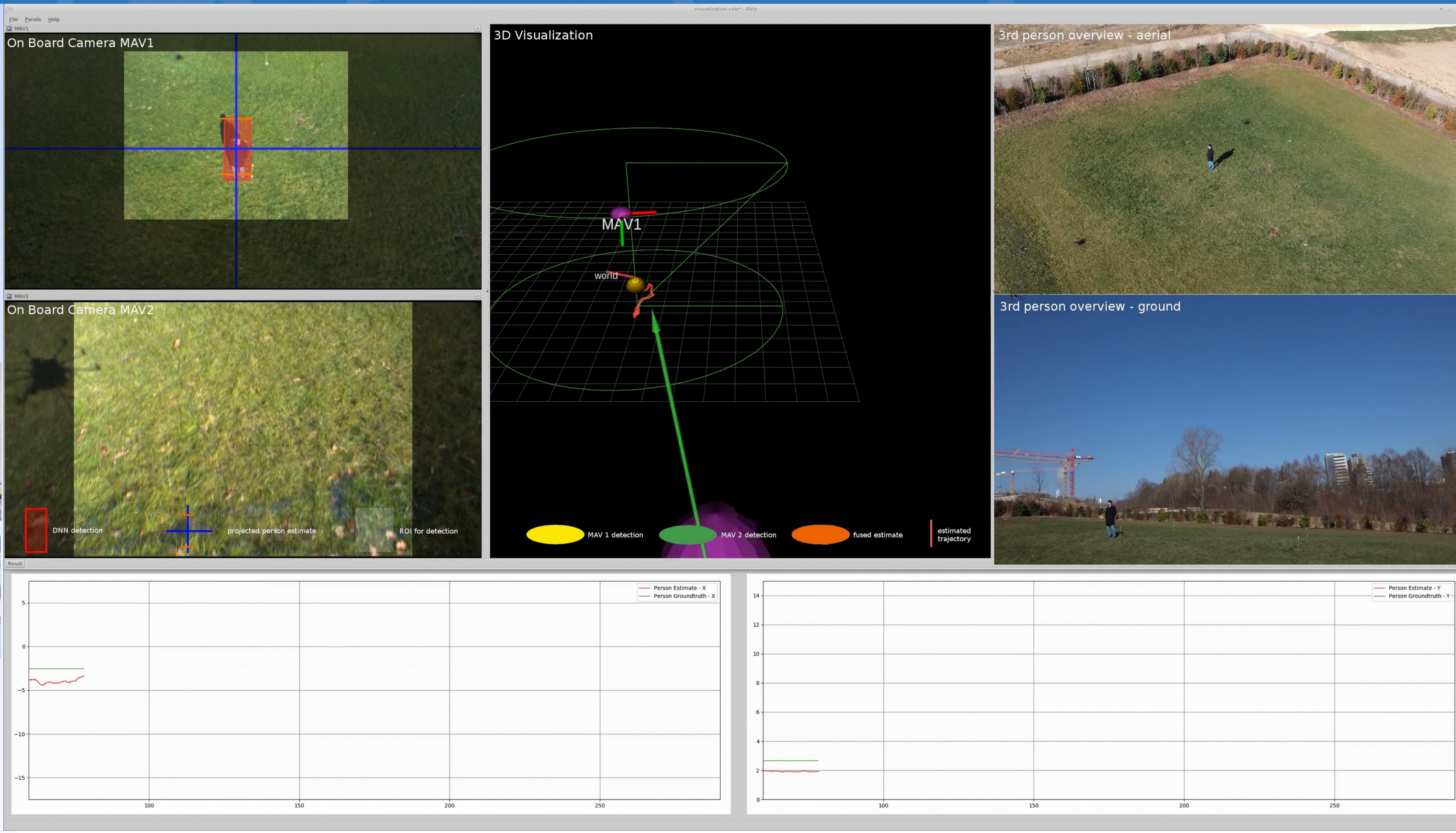


Perception-driven Formation
Real Robot Experiments

Max Planck Institute for Intelligent Systems
Perceiving Systems Department

Deep Learning for Vision

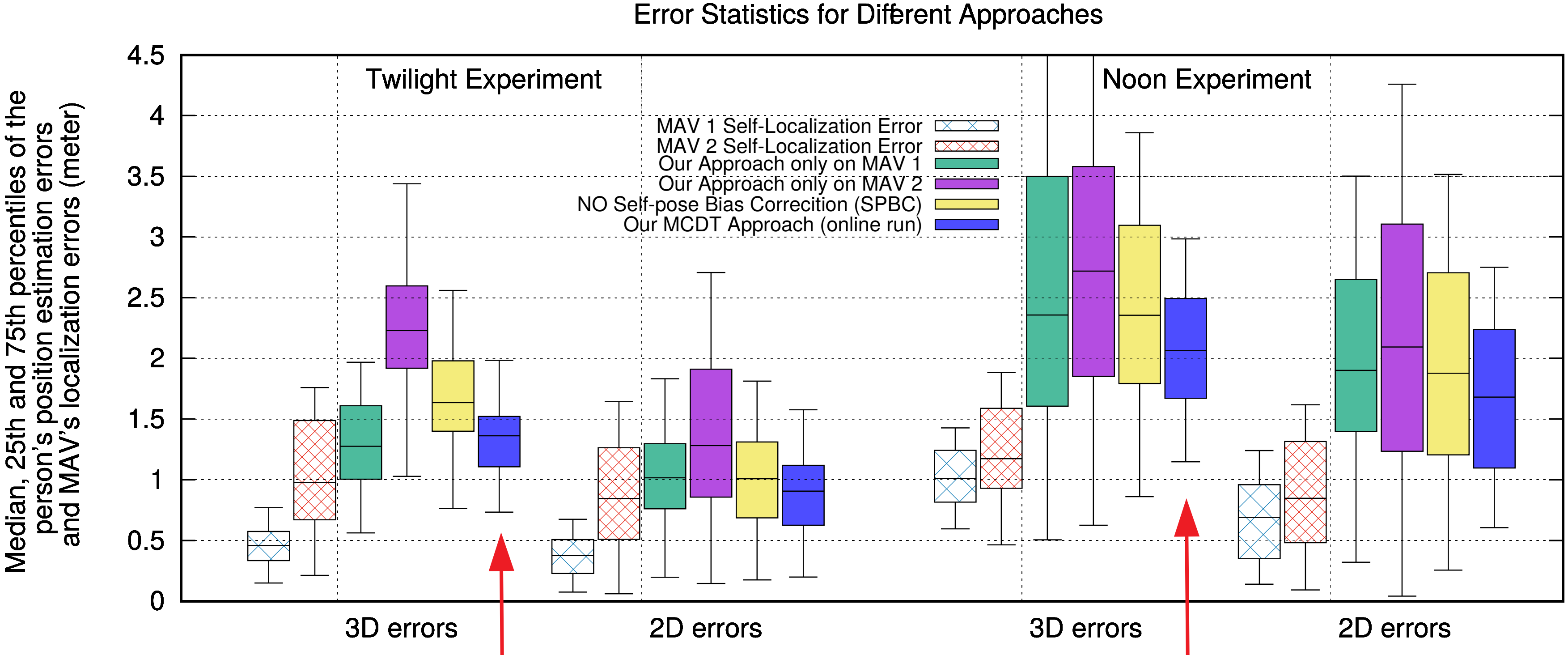


Perception-driven Formation
Real Robot Experiments

Deep Learning for Vision

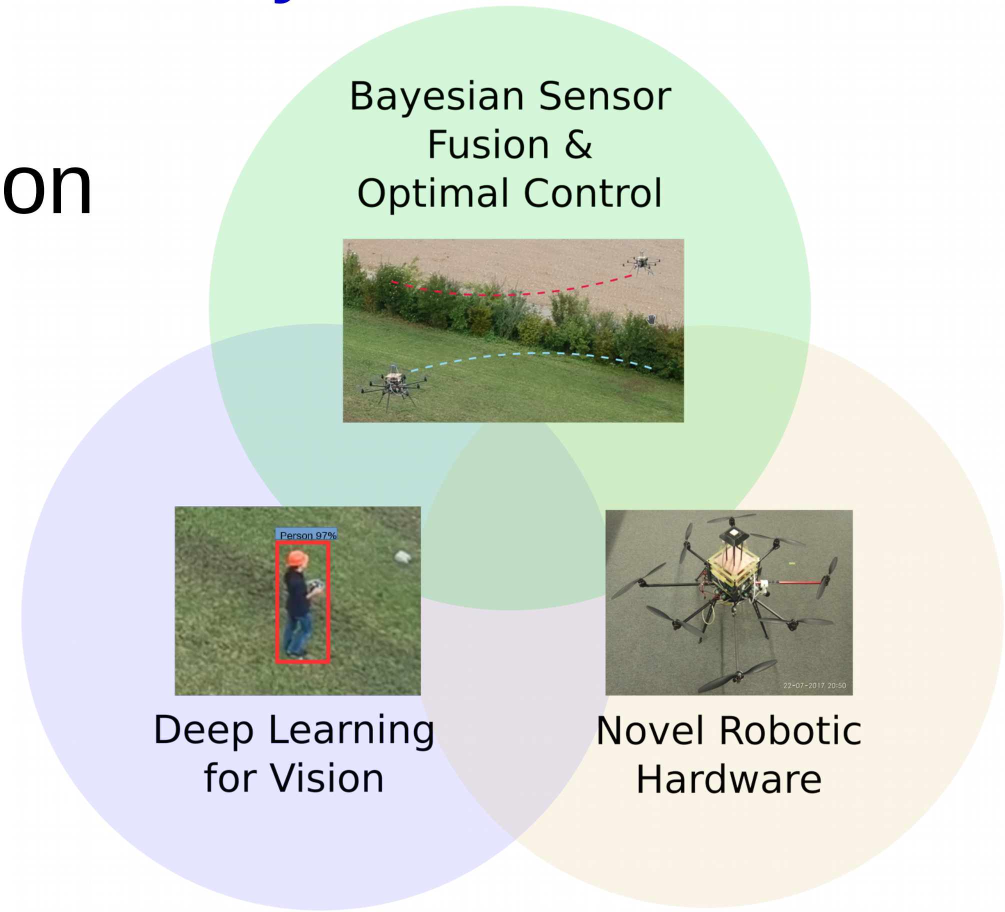
Overall accuracy



Perception-driven Formation Real Robot Experiments

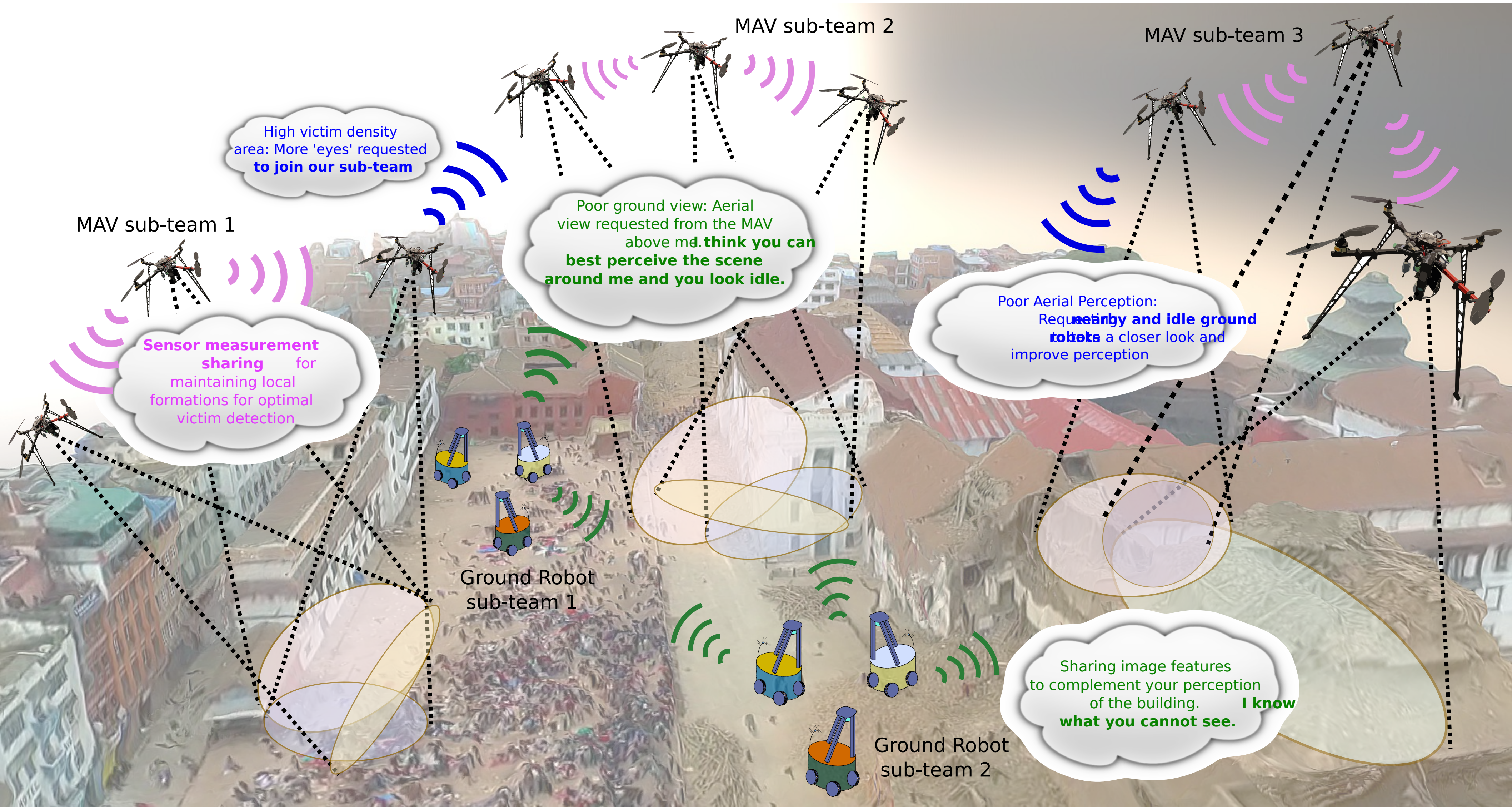
Summary

- Deep Vision
- Cooperative perception
- Coordinated multi vehicle motion
- Flying Hardware
- Enabled by Embedded GPU
- Fully Open Source: http://aircap.is.tuebingen.mpg.de



MAX-PLANCK-GESELLSCHAFT

Future Outlook




Thank you!