Drone Vision and Deep Learning for Infrastructure Inspection

V. Mygdalis, Prof. Ioannis Pitas
Aristotle University of Thessaloniki
pitas@csd.auth.gr
www.aiia.csd.auth.gr
Version 3.3.2

Artificial Intelligence &
Information Analysis Lab

Infrastructure inspection applications

- Aerial robots with different characteristics must be integrated for:
  1. Long range and local very accurate inspection of the infrastructure.
  2. Maintenance activities based on aerial manipulation involving force interactions.
  3. Aerial co-working safely and efficiently helping human workers in inspection and maintenance.

UAV Infrastructure Inspection

- **Overview**
- Sensors
- Visual analysis
- Drone operations

Technical objectives

- Cognitive functionalities for aerial robots including ***perception based on novel sensors*** such as event cameras and data fusion techniques, learning, reactivity, fast on-line planning, and teaming.
- Cognitive safe aerial robotic co-workers capable of ***physical interaction with people***.
- ***Cognitive aerial manipulation*** capabilities, including manipulation while flying, while holding with one limb, and while hanging or perching to improve accuracy and develop greater forces.
- Aerial platforms with ***morphing capabilities***, including morphing between flight configurations, and between flying and ground locomotion, to save energy and perform a very accurate inspection.

Long range inspection of power lines

Helicopter inspection of power lines

Helicopter inspection of power lines

- Complete manned helicopter flight:
  - The helicopter has on-board a pilot and a camera operator.
  - Manned helicopter is flying at low altitude and stopping at each electrical tower.
  - High quality visual, thermography and LIDAR data are obtained at the same time.
  - LIDAR is disconnected in each electrical tower since it gets bad results when it is a long time in the same spot.

Types of flights with manned helicopter

- Fast manned helicopter flight:

  - Thermography and LIDAR acquisition at the same time.
  - Helicopter does not stop at each electrical tower, but the flight is at low altitude (due to the thermography camera resolution).

  - ***Speed limited to 50-60 km/h because of the thermography.***

Artificial Intelligence & Information Analysis Lab

Disadvantages of current approach

Main disadvantages of current inspections with manned helicopters:

- Costs: 40,5 €/km.
- Difficulties to detect some devices, like connecting cable from the tower to ground.
- Safety.

- ***200 km report is ready in two weeks.***

Safe local manipulation interventions

- Examples:
  - Installing anti-birds systems.
  - Cleaning isolator in power lines.

Installing anti-birds systems

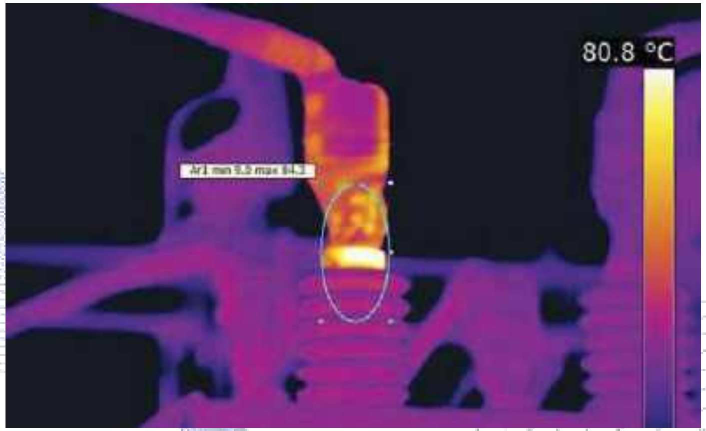
- National regulation (a few years ago) enforces their installation every 5-10 m.
- (De-)installation is performed by work at height on a basket.
- Dangerous, slow and costly.
- The electrical lines has to be without voltage, resulting in money loss.

Co-working activities

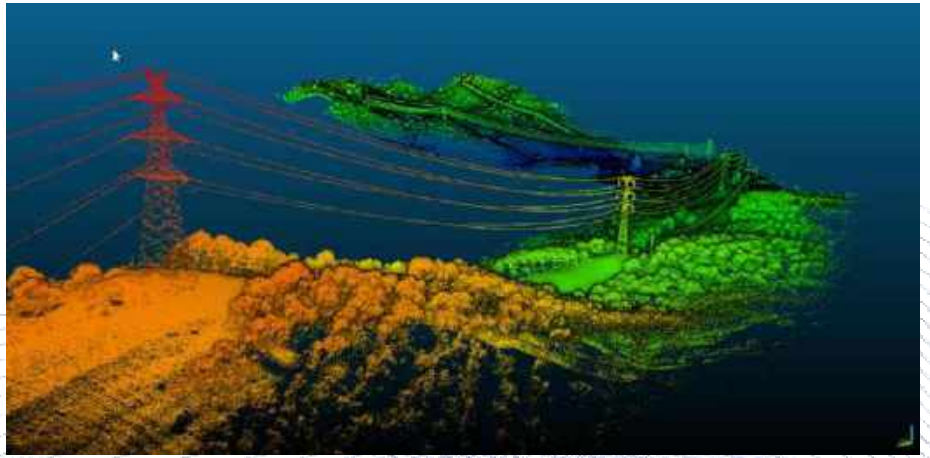
Infrastructure Inspection

- Overview
- **Sensors**
- Visual analysis
- Drone operations

Types of inspection



Thermography



3D mapping (LIDAR)

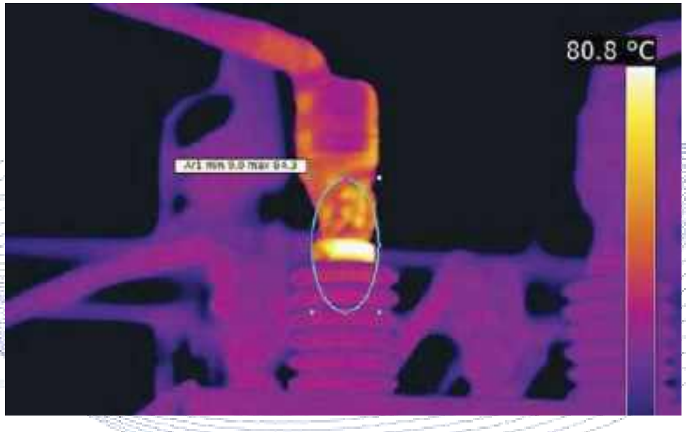
Camera & video

Inspection using camera/video

- High quality images and videos
- Detailed images of the complete electrical tower
- Requires 2 mm GSD, i.e., 1 pixel per 2 mm to be able to identify all the required details.
- For example:
  - check that the bolt on a screw is there.
- Requires that the UAV moves very slowly around the electrical tower.

Thermography

- Detection of hot spots in the electrical tower: cramps and connections
- To perform thermography, the speed of a fixed wing UAV is limited to 50-60 km/h.



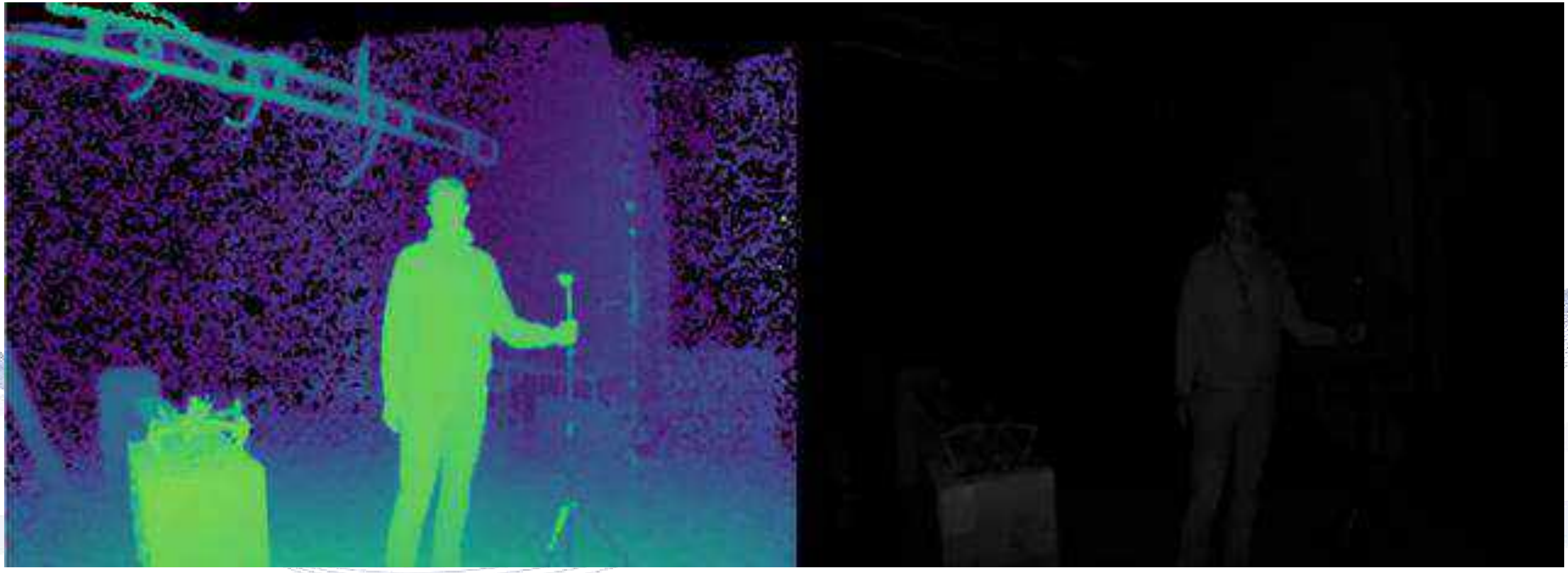
Artificial Intelligence & Information Analysis Lab

3D LIDAR

- Precise 3D mapping (with cm level accuracy and precision)
- No speed limitation on the manned helicopter
- A 3D map is constructed to:
  - Detection of obstacles close to power lines.
  - Measurement of vegetation around power lines.
  - Checking distance when crossing power lines.
  - Once the 3D map is obtained, a classifier algorithm (and also checked and adjusted by a technician) is used.
  - Afterwards, distances and other measurements are performed to develop the inspection report.

3D VGA Time-of-Flight camera

- A camera for human gesture recognition, object avoidance in close distance, landing and taking-off.



Indoor Tests, February 2021, Terabee facilities.

Event cameras - motivation

Latency & Motion blur.

Dynamic Range.

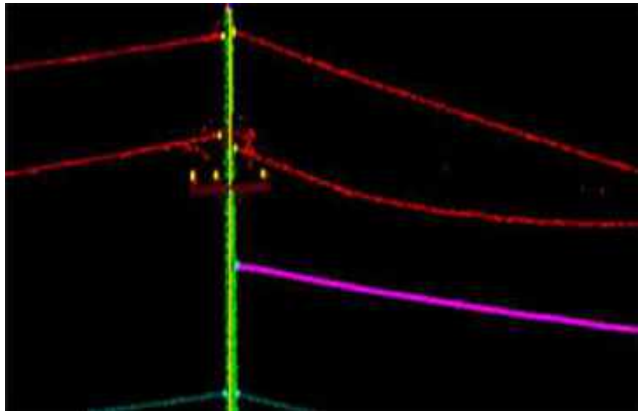
Event cameras

- Novel sensor that measures only motion in the scene.
- Low-latency (~ 1 μs).
- No motion blur.
- High dynamic range (140 dB instead of 60 dB).
- Ultra-low power (1 mW vs 1W).
- Traditional vision algorithms do not work all the time!

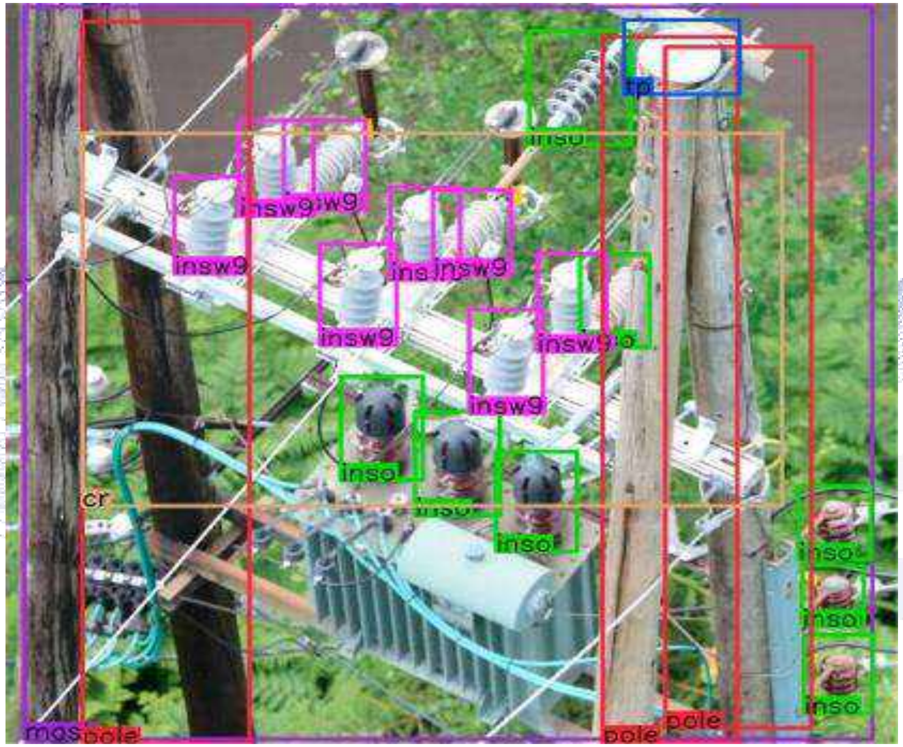
Infrastructure Inspection

- Overview
- Sensors
- **Visual analysis**
- Drone operations

Research tasks



- Semantic 3D world mapping.
- Learning methods for object detection/tracking of electric lines, rods, etc.
- Human-drone interaction:
  - Gesture drone control.
  - Body posture estimation.
  - Human action recognition.
  - Facial pose estimation.



Learning methods for aerial inspection

- Visual detection.
- Semantic segmentation of power lines to enhance robot behavior.
- Object detection for manipulation tasks.
- Focus in lightweight nets for online computing.
- Generative adversarial networks (GAN) to improve detection quality from previous learned experiences.

Semantic visual cognition

- Deep Neural Networks (DNNs) are the algorithm of choice for 2D visual object detection/tracking tasks.
- They require powerful GPU-equipped hardware platforms for real-time execution.
- E.g.: Nvidia Xavier computing board for embedded/robotics applications.
- Software execution environment: Linux + Python.

Fast 2D Convolutions

- State-of-the-art neural network architectures for visual data use convolutional layers.
- The convolution operation takes up most of the total inference and training time.
- Reducing the computational complexity of convolutions would render networks for e.g., target detection or target tracking much more efficient for deployment on embedded GPUs.

- We developed a fast convolution algorithm which splits cyclic convolution into smaller products.
- In this algorithm, cyclic convolution takes the following form:

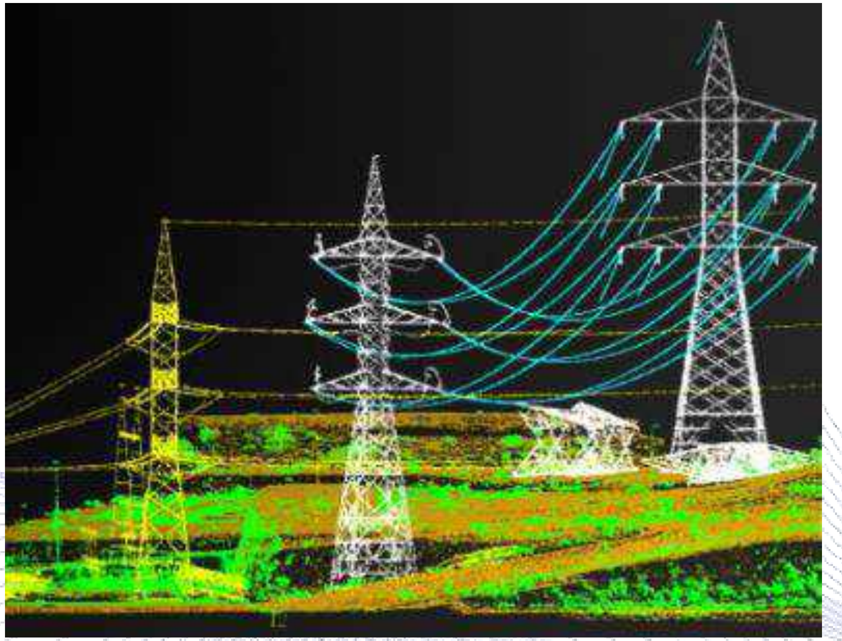
$$\mathbf{y} = \mathbf{C}(\mathbf{Ax} \otimes \mathbf{Bh}).$$

- Thus, the problem is reduced to finding matrices $\mathbf{A}, \mathbf{B}$ and $\mathbf{C}$.

Experimental Results

Algorithm	Computation time (ms)
Winograd-6 (cuDNN Winograd linear convolution)	0.9165
GEMM-0 (fastest cuDNN convolution)	0.3858
Ours	**0.0809**

Semantic 3D World Mapping



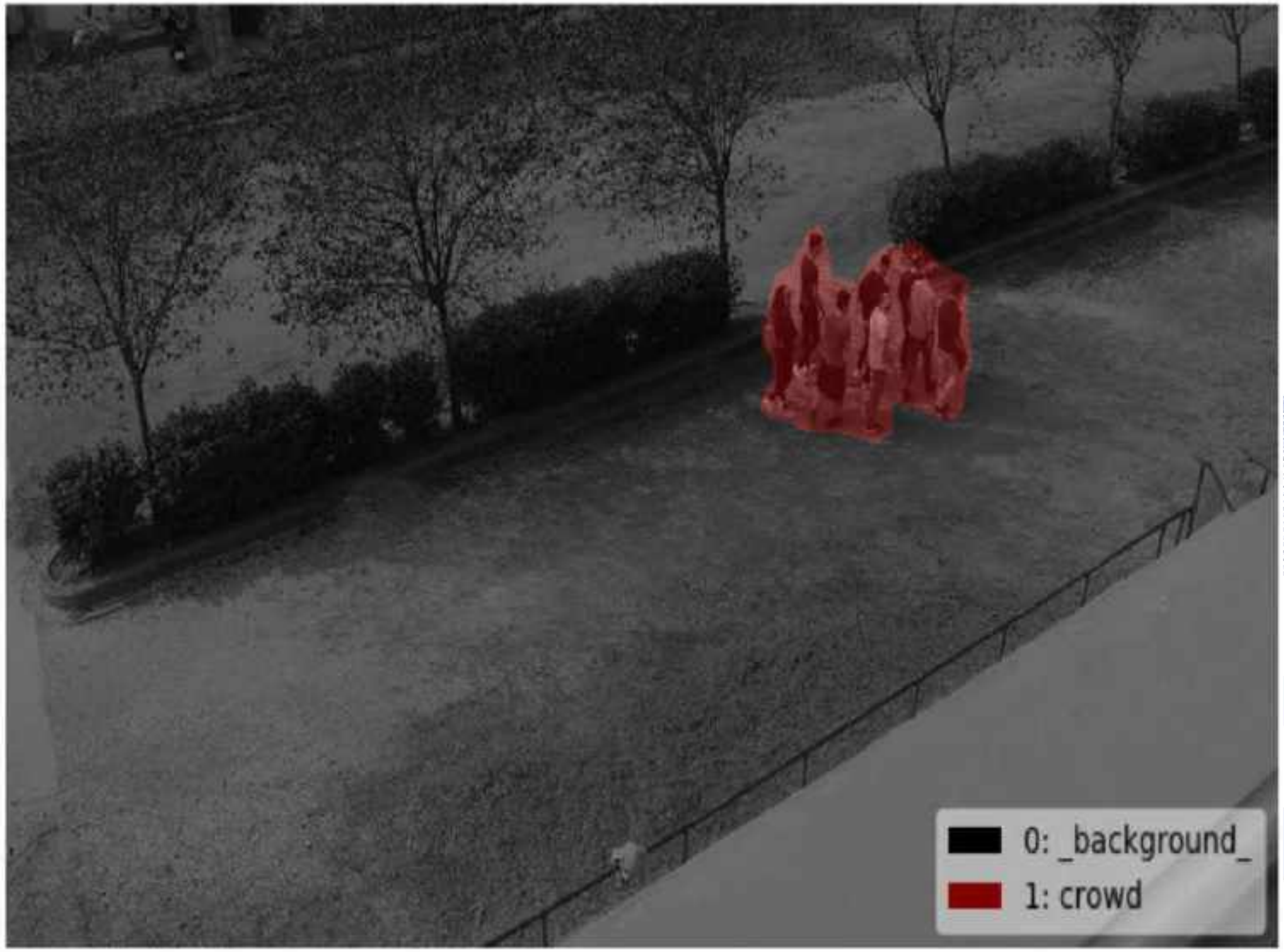
Geometric modeling of the 3D world.

Semantic 3D World Mapping

- ***Semantic image segmentation***:
  - Segment low/high vegetation regions, roads.

Semantic 3D World Mapping

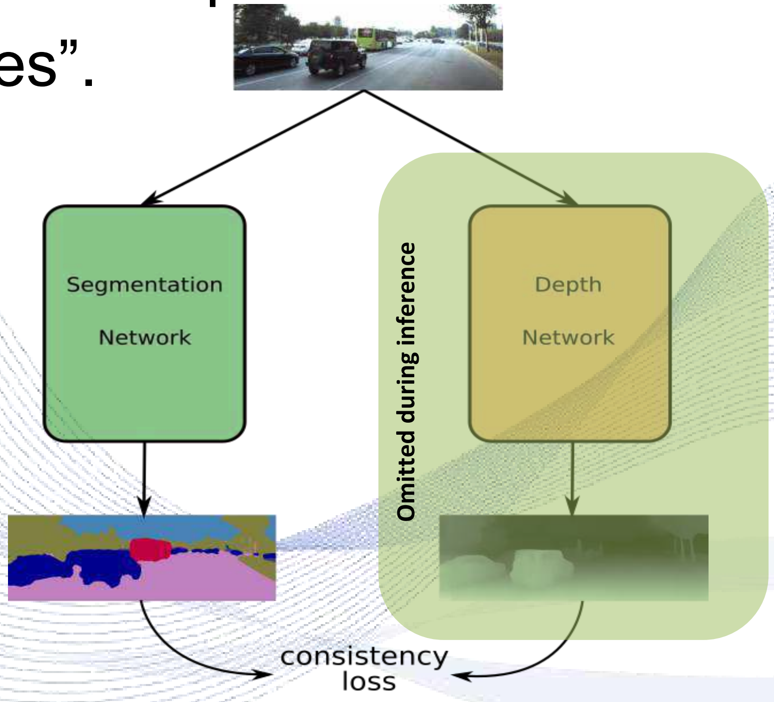
- ***Semantic image segmentation:***
  - Crowd detection and localization.



Semantic Segmentation

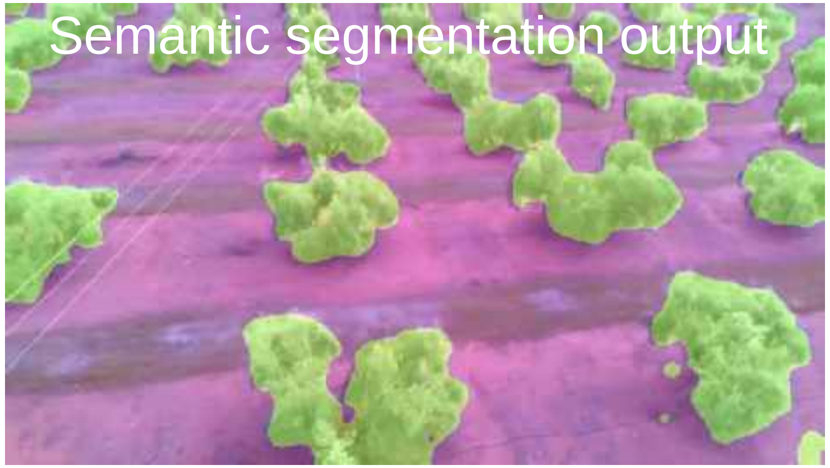
- Multitask CNN for semantic segmentation and self-supervised depth estimation.
- Novel consistency loss function to regularize segmentation output.
- “Do not form semantic edges, if there are no depth edges”.

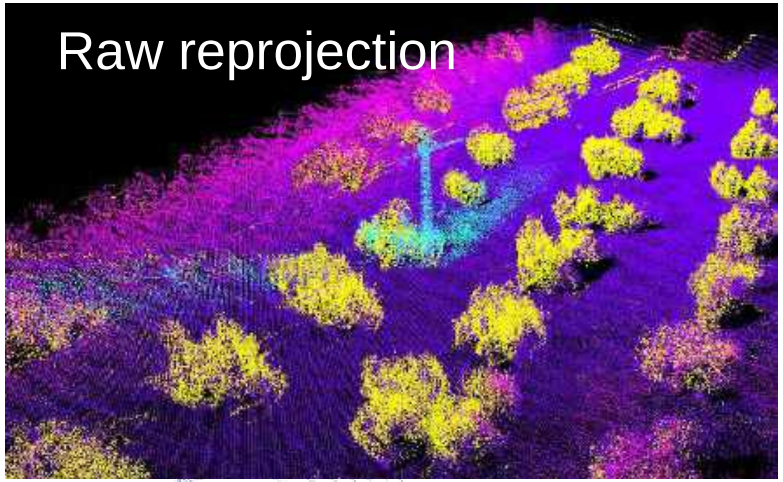
Method	Mean IoU	Inference (ms)
Yu et al.	39.557%	**6.2**
Klingner et al.	34.318%	6.4
Novosel et al.	37.683%	8.3
Chen et al. (pretrained)	39.610%	**6.2**
Chen et al. (multitask)	38.153%	9
Ours	**40.597%**	**6.2**

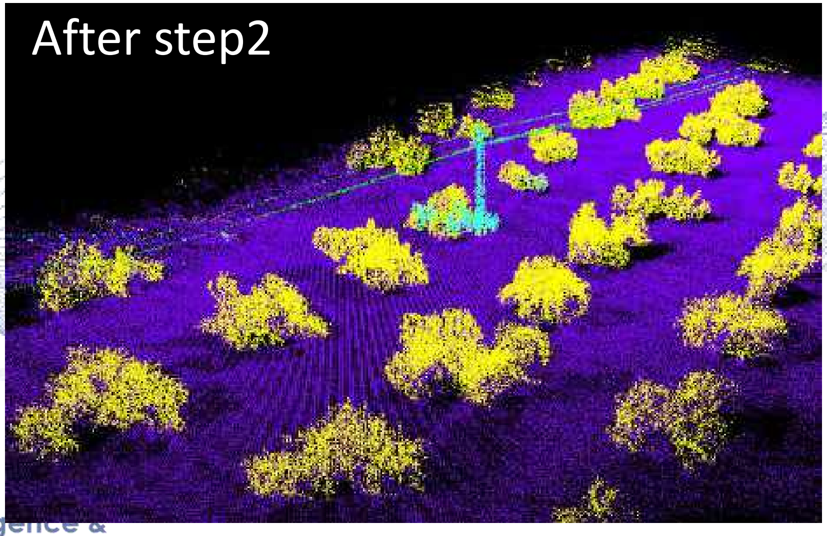


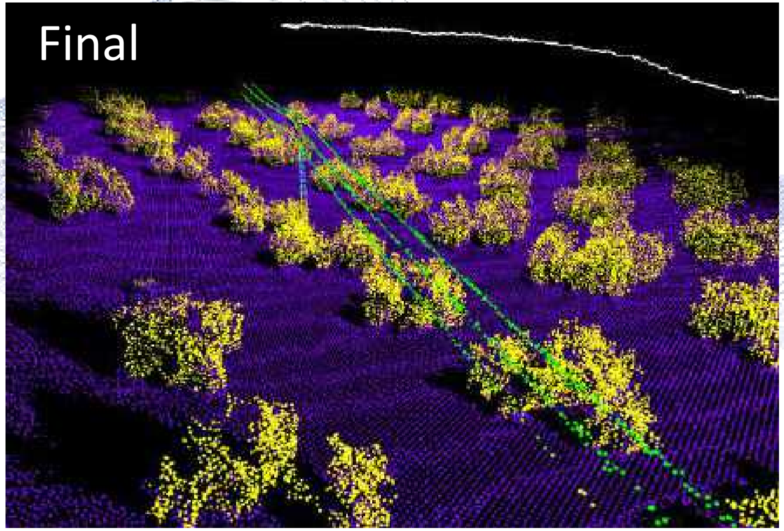
Semantic Image Segmentation Guided by Scene Geometry [PAPAD2021].

Semantic 3D World Mapping



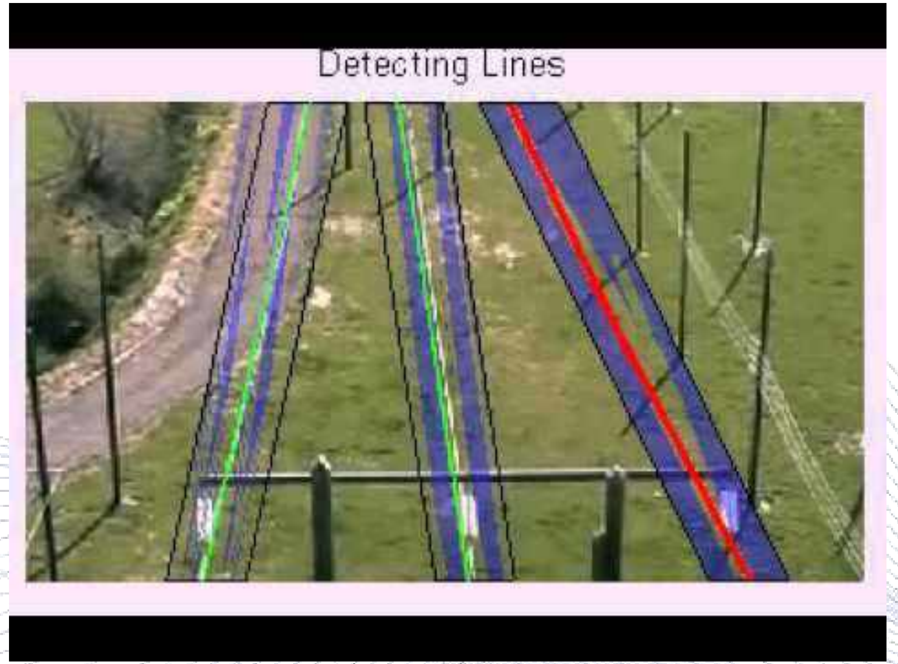






Artificial Intelligence & Information Analysis Lab

Object detection and tracking



Deep learning for power line detection and tracking.

Object detection and tracking

Event-based Powerline tracker.

Object detection and tracking

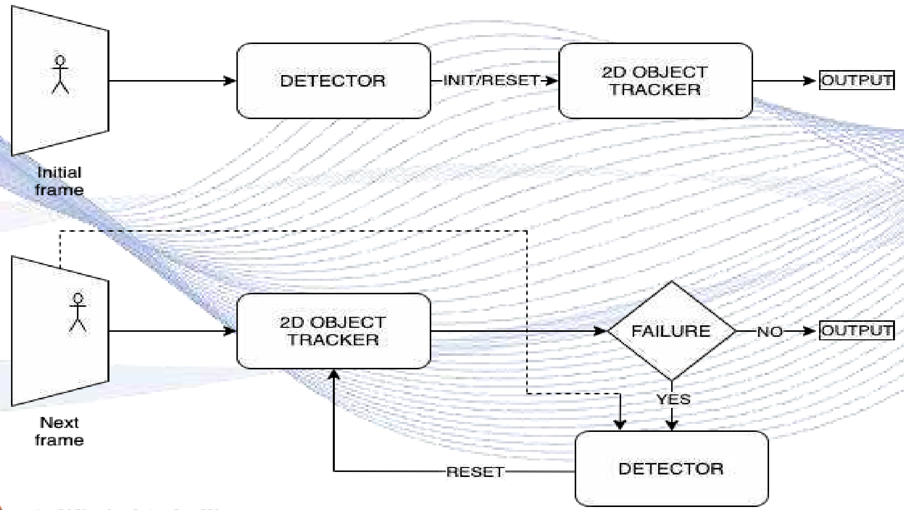
- ENDESA dataset (17K images, insulators, dumpers, towers).
- SoA detector evaluation (Single-Shot-MultiBox-Detector (SSD), You-Only-Look-Once v4 (YOLOv4), Detection-Transformer (DETR).
- Proposed approach: Content-specific image queries (based on DETR).

Model	FPS 2080 / Jetson	AP	AP_{50}
YOLO v4 CSPDarknet53	96/26	41.6	83.5
SSD Mobilenet v2	126/17	50.1	82.1
SSD Inception v2	84/13	48.7	80.0
SSD Resnet50	40/9	52.3	79.8
DETR Resnet50	35/8	52.4	83.1
Ours Resnet50	35/8	**53.9**	**83.9**

Artificial Intelligence & Information Analysis Lab

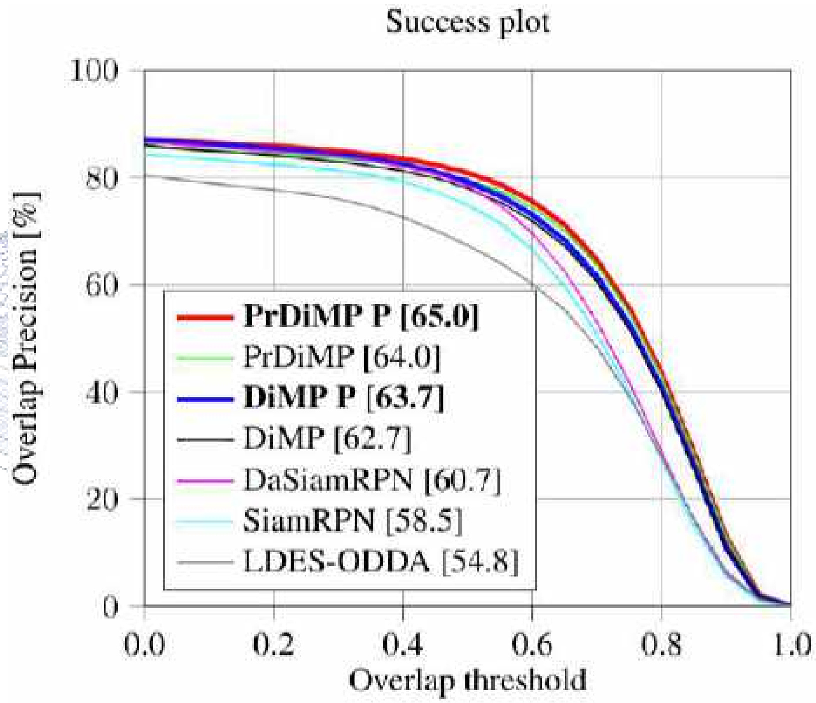
Object detection and tracking

- Combination of object detection/tracking methods.
- Object detector periodically re-initiates the tracker.



Online tracking model adaptation

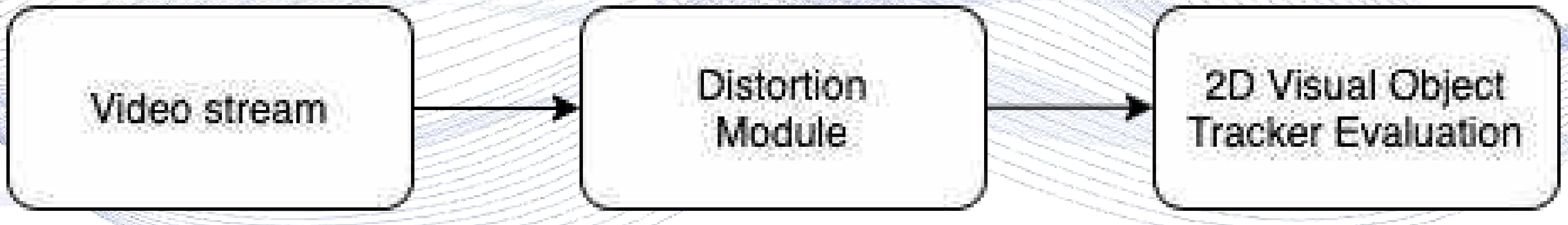
- Online tracking model updating is typically addressed as a regression problem.
- An **adversarial optimization scheme**
- **Generator** is assigned to the tracking model producing response maps.
- **Discriminator** network is trained to identify if the tracker response maps produced by the generator belong to the target distribution, or not.



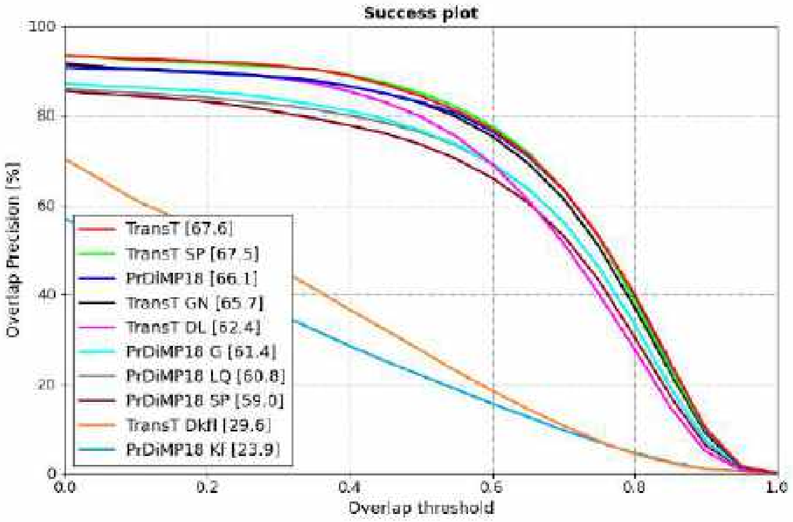
Artificial Intelligence & Information Analysis Lab

Robustness 2D Visual Object Tracking

- VOT-RT - A toolkit that allows evaluation against:
- Image acquisition: Gaussian, Salt and pepper, etc,
- Image transmission: Low Quality image, Key-frame loss.
- We evaluated many state-of-the-art tracking methods, and all suffer from performance loss in every case.



Robustness 2D Visual Object Tracking

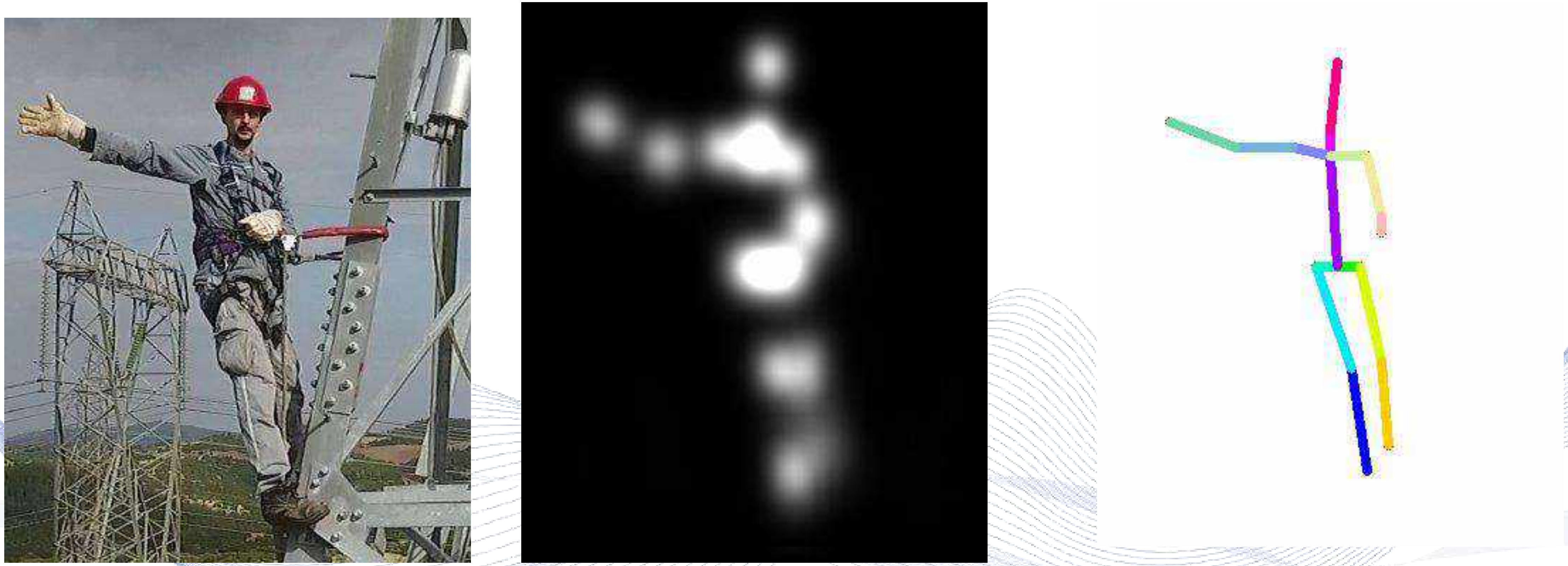


Artificial Intelligence &
Information Analysis Lab

Object detection and tracking

- Requirements similar to 2D visual detection/tracking:
- Method: Embedded DNNs.
- Hardware: GP-GPU equipped computing boards (e.g., Xavier).
- Software: Linux + Python.
- Training: Massive, annotated, domain-specific datasets.

Human posture estimation

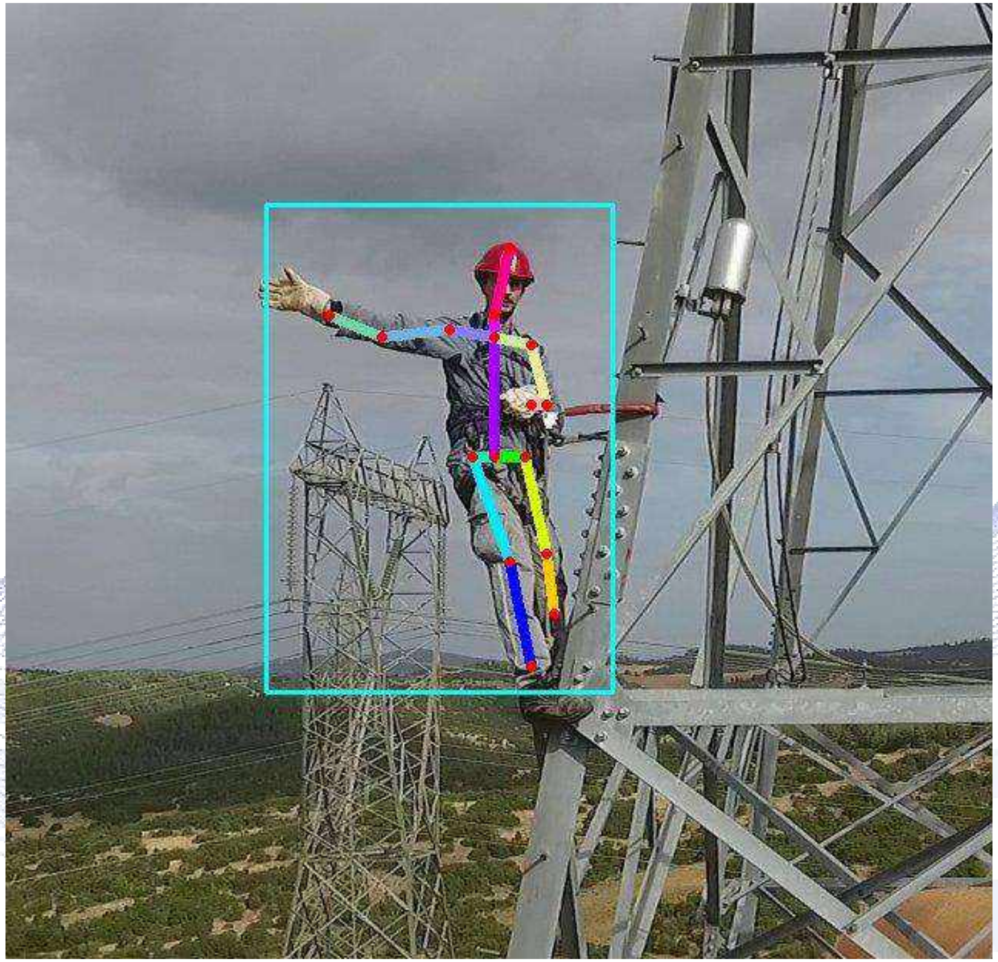


a) Original image; b) Body joints heatmap; c) Human posture estimation.

Human-drone interaction

- Goals: The UAV/Aerial Co-Worker:
  - Can verify that the technician follows pre-set safety rules at all times.
  - May perceive the technician's current activity (e.g., climbing a pole) in order to get into suitable position for assisting him.
  - Is able to interact visually with the technician by interpreting pre-defined communication hand gestures.
  - AUTH may also potentially employ semantic image/instance segmentation for assisting in the above tasks and for augmenting algorithm performance.

Human posture estimation

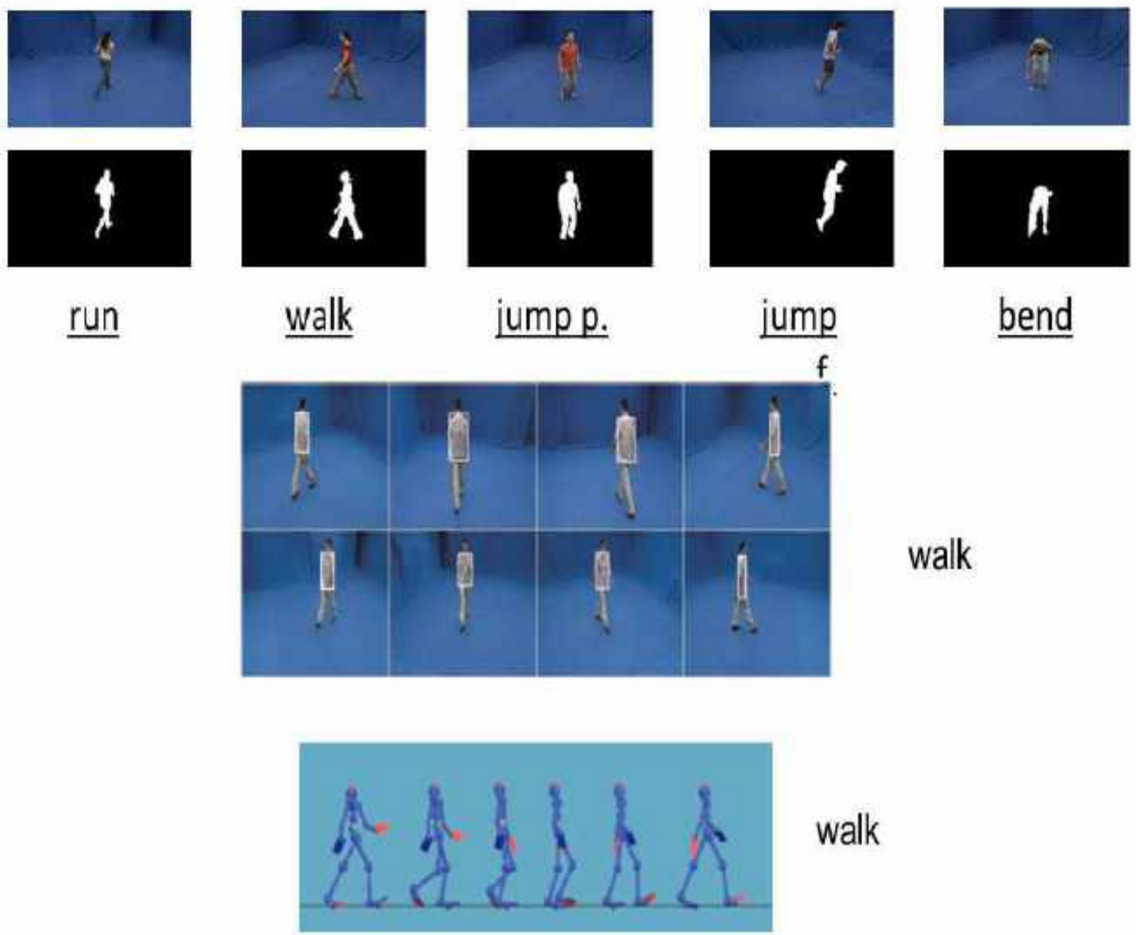


Human posture estimation.

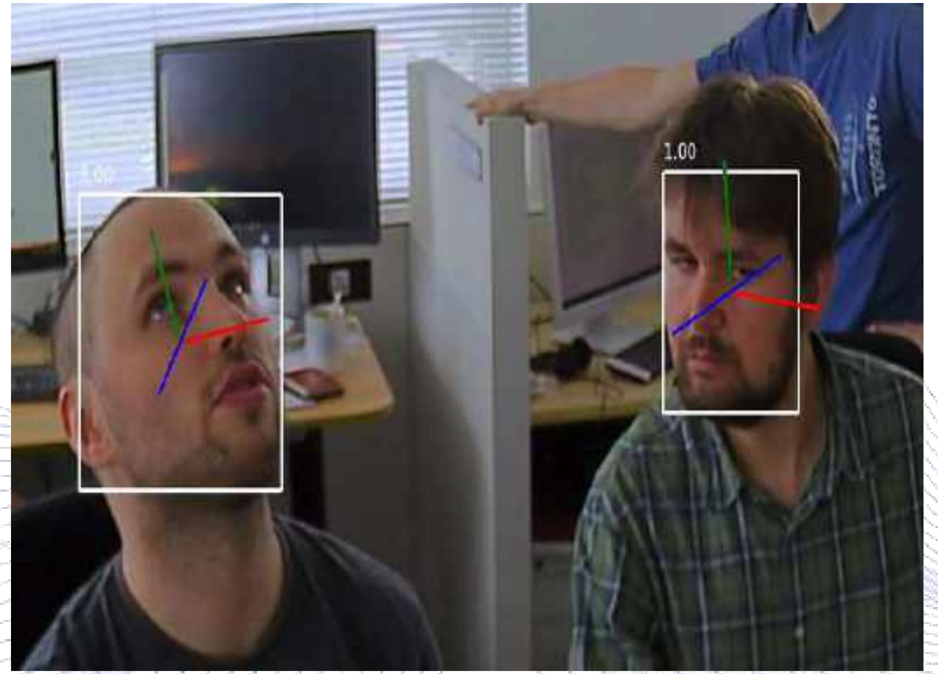
Human posture estimation

a) Original image; b) Body joints heatmap; c) Human posture estimation.

Human action recognition



Human pose estimation



Facial pose estimation.

Gesture recognition

Language of visual gestures for drone control:

- Extend one arm to the side
- Cross arms (form X with forearms)
- Raise one arm upwards
- Palms together (namaste gesture)
- Victory sign
- Ok sign (thumbs up)

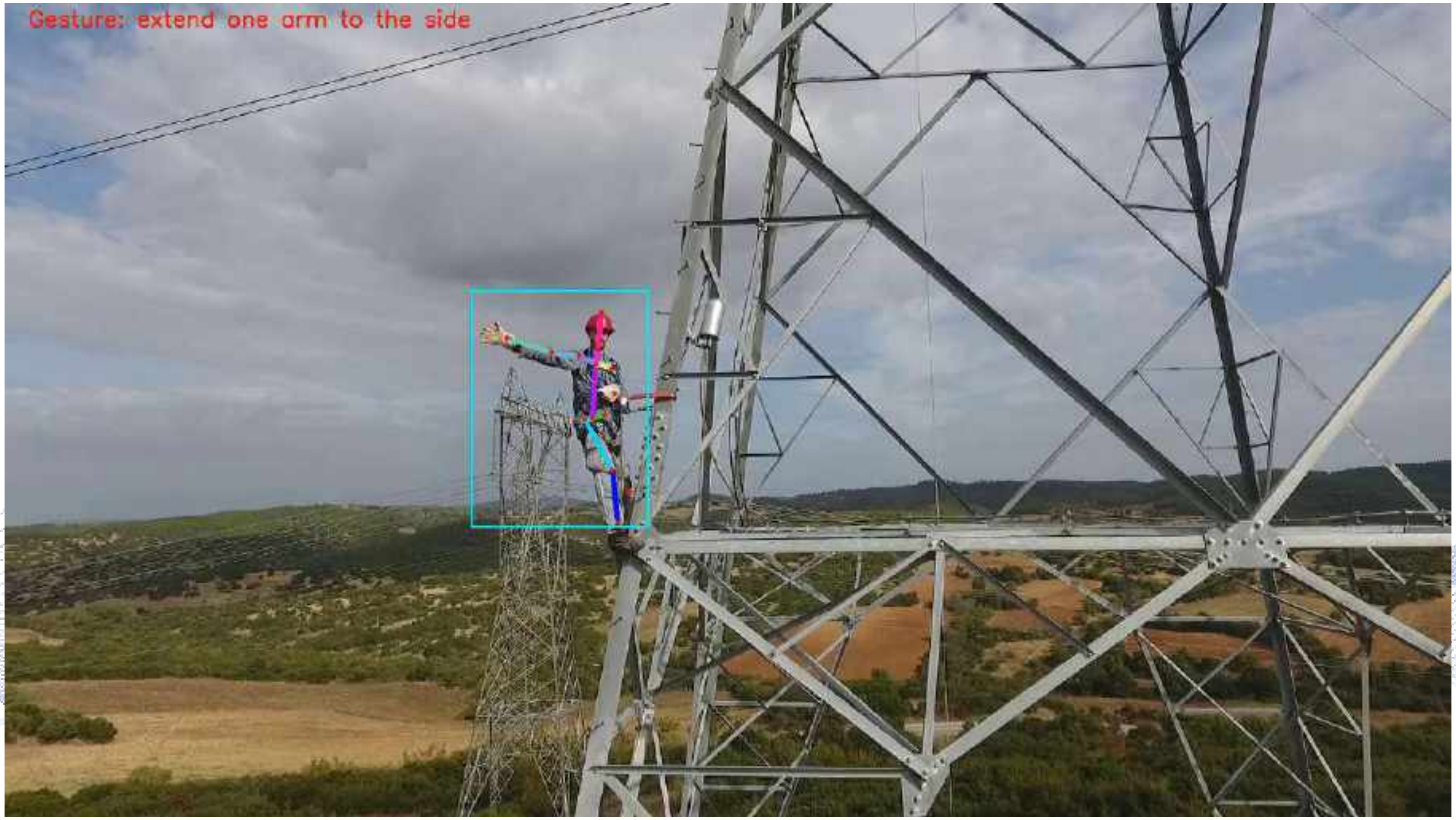
Gesture recognition

- A gesture dataset was created for training, using three data sources:
  - UAV gestures dataset (thumbs up, cross arms, victory, palms together) [PER2018].
  - NTU dataset (thumbs up, cross arms, raise one arm upwards) [SHA2019].
  - Video acquisition performed by AUTH.
- A novel gesture recognition method was developed, relying ***on CNNs and LSTM networks***, yielding a maximum test set ***classification accuracy of 89.22%***.

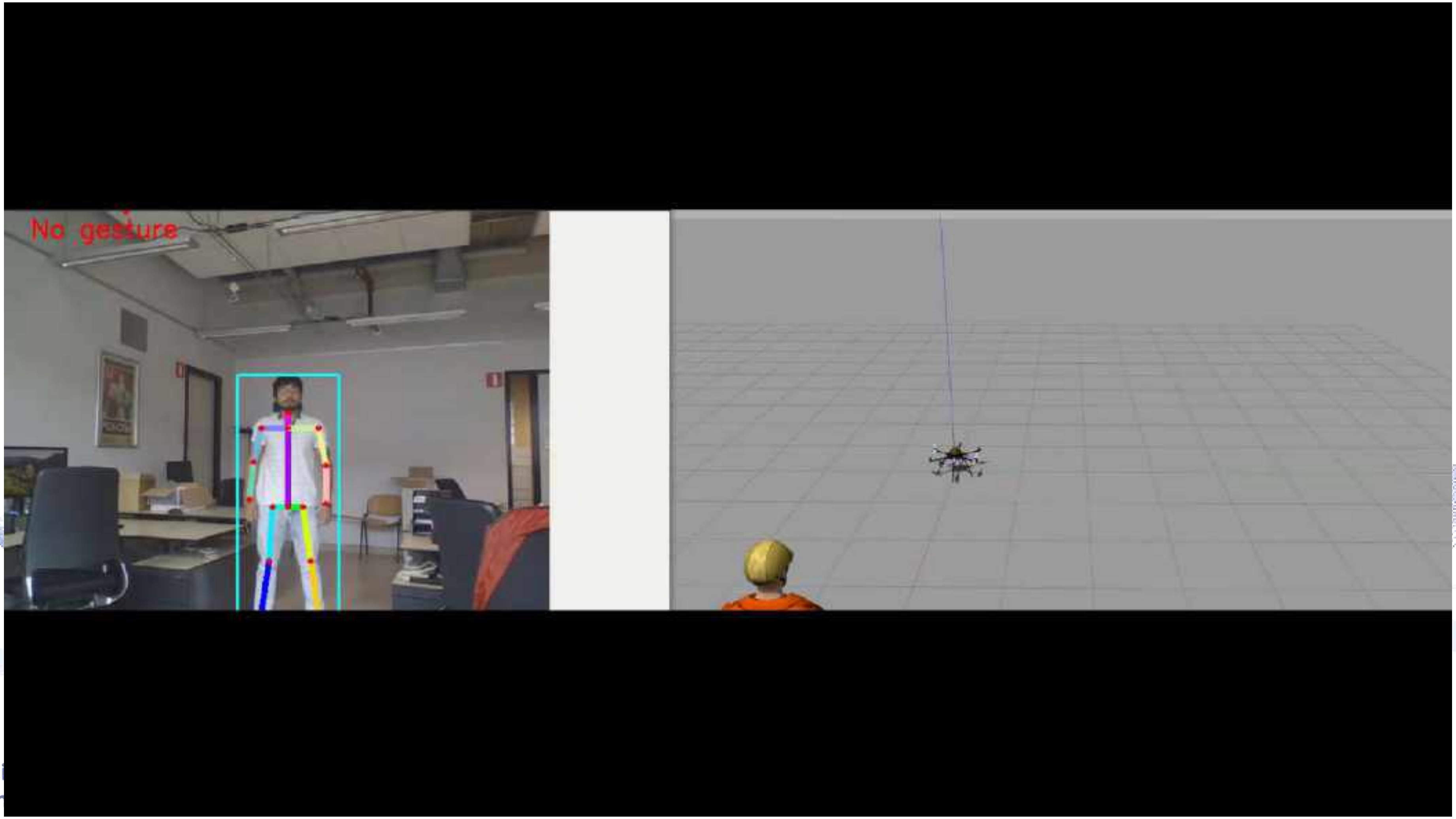
Human posture estimation

- AUTH developed a novel 2D human posture/body joint/skeleton estimation method based on deep CNNs using an image segmentation approach, utilizing a multi-task segmentation + I2I (GAN) network architecture.
- It receives an image of a localized person as input and predicts a dense heatmap for each body joint in a predefined joints set (skeleton).
- The final 2D pixel coordinates of each joint are obtained by post-processing the body joint heatmaps.

Human posture – gesture recognition

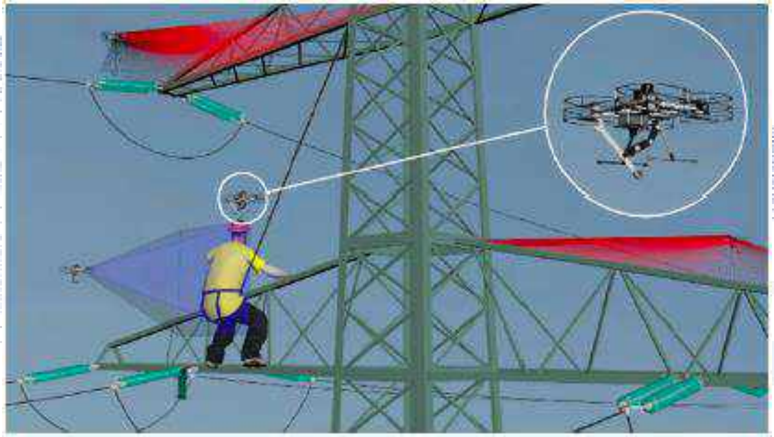


Gesture-based control



Coordination of a Heterogeneous Team of ACWs

- 3 main ACW activities:
- Safety-ACW - equipped with a surveillance camera (blue).
- Inspection-ACW – inspection sensor (red).
- Physical-ACW - equipped with a manipulator to provide tools required by workers



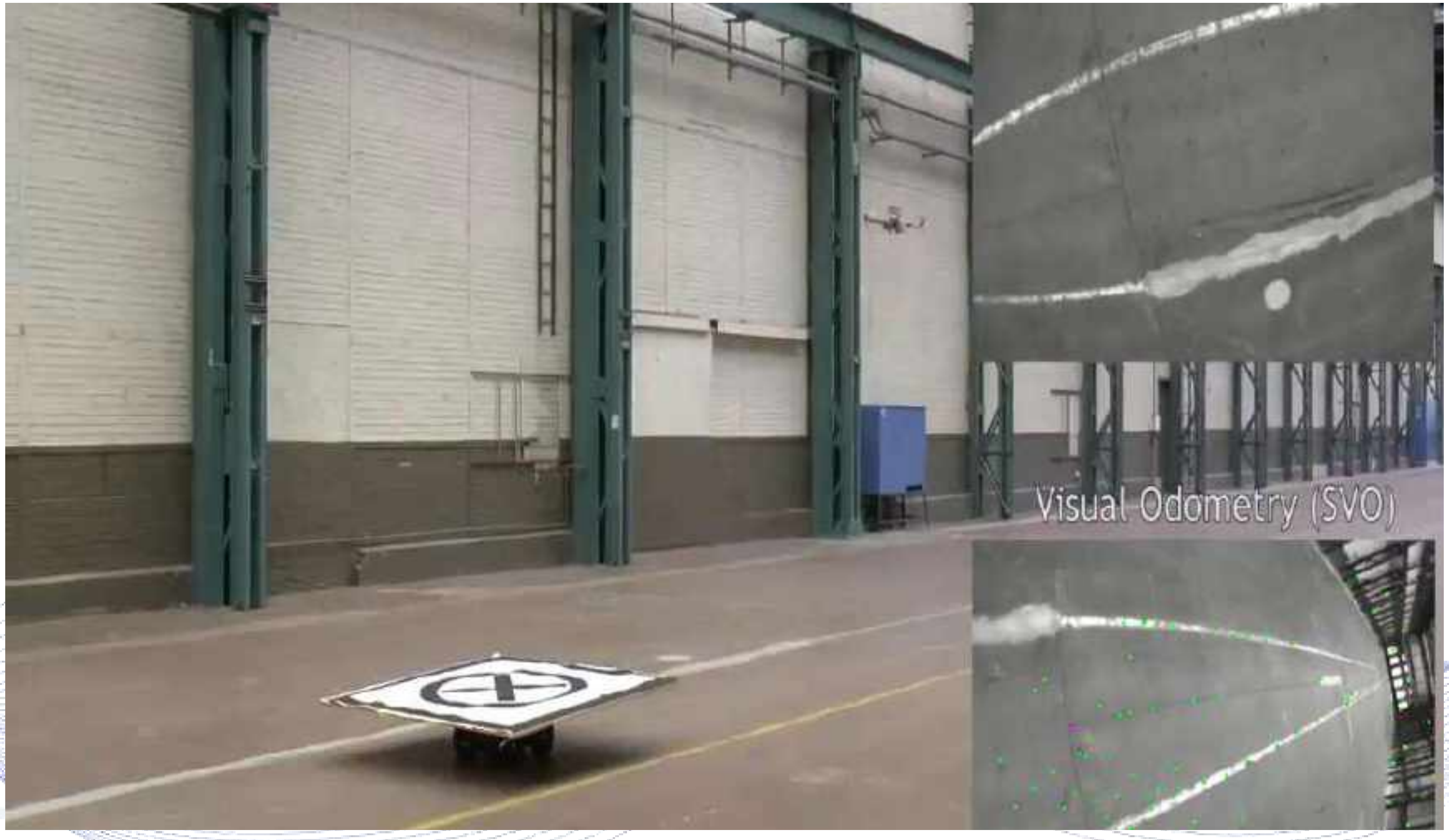
Infrastructure Inspection

- Overview
- Sensors
- Visual analysis
- **Drone operations**

Autonomous landing/perching

- Develop an autonomous landing and perching scheme (i.e., planning and control) that allows different flying platforms to land in confined spaces and perch on complex surfaces, such as, e.g., tower structures or electrical power lines.
- The system will be able to evaluate different landing positions for their feasibility and plan landing paths in real time that guide the aerial robots safely to the desired landing or perching spot while avoiding any obstacles.

Autonomous landing



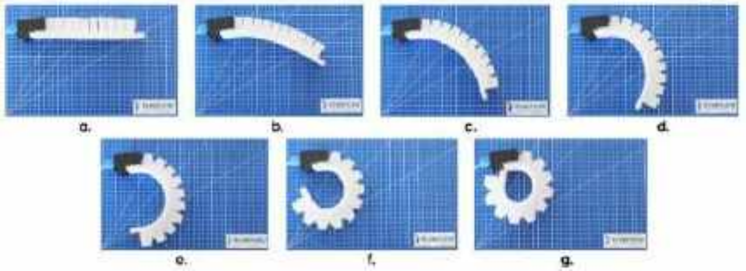
Autonomous perching

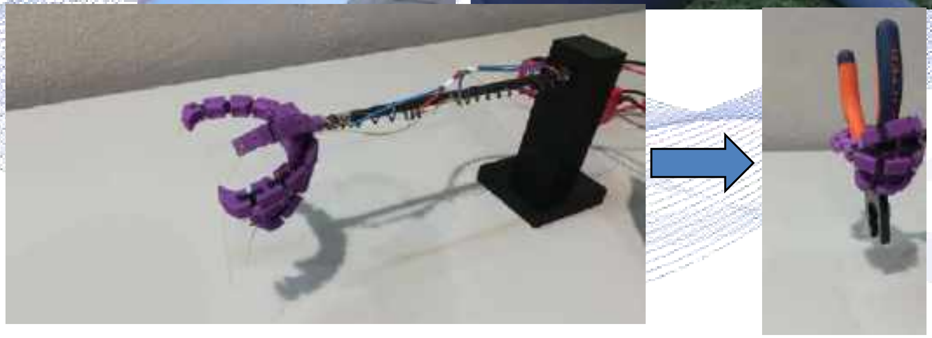
- Sensor fusion to exploit synergies:
- Perching steps:
- Preparation
  - Multi-sensor detection & tracking of perching candidates
    - LIDAR
- Fast approach to perching zone
  - Multi-sensor Visual Servoing:
    - event cameras
- Short distance approach & perching
- Multi-sensor Visual Servoing.

End-effectors for holding/grabbing

- Bio-inspired actuators for compliant co-working and close range inspection.

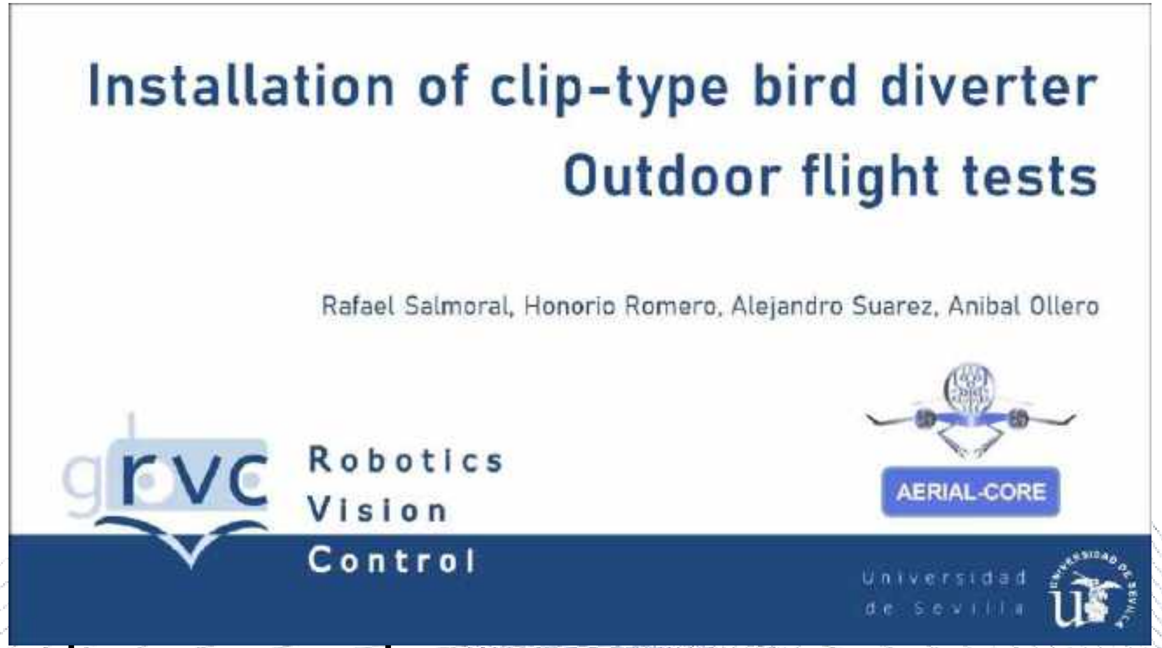






Manipulation while holding/perching

Manipulation while flying, holding and perching



Main challenges outdoor scenario:

- Physical interaction on flight during installation.
- Motion constraints during the installation phase.
- Positioning accuracy, dependent on GPS visibility.

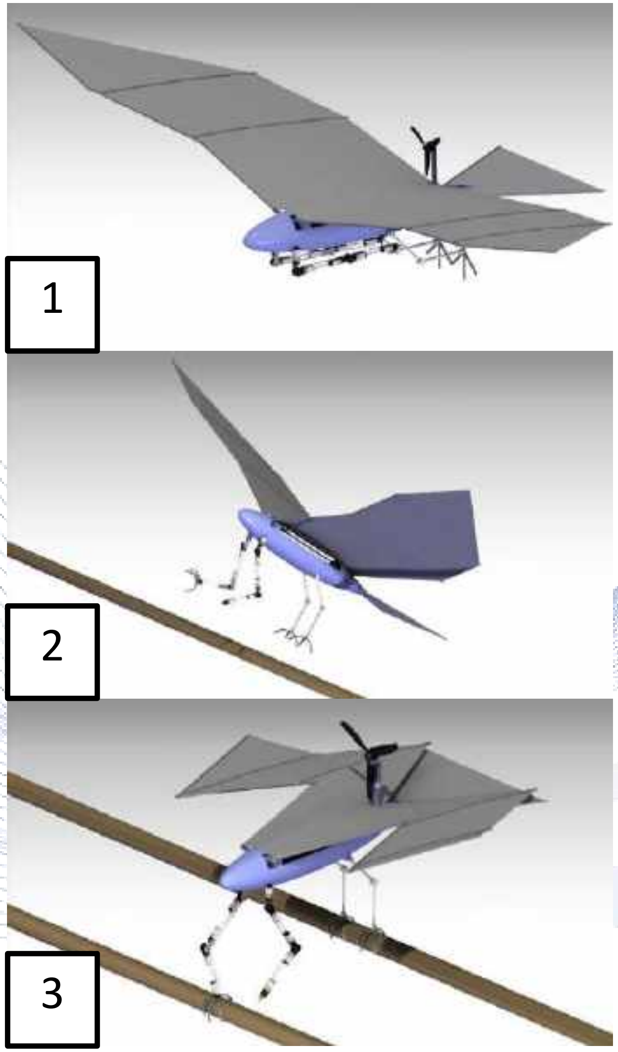
Manipulation while flying, holding and perching

Main challenges realistic scenario:

- Loss of depth perception for the human pilot.
- Risk of entrapment or collision with cables.
- Wind gusts at high altitude.

Morphing

- ***Flapped wing*** to fixed wing.
- Fixed to rotary.
- ***Ornithopters*** can potentially achieve better efficiency, maneuverability and safety.



Simulations

Bibliography

[PIT2021] I. Pitas, “Computer vision”, Createspace/Amazon, in press.
[PIT2017] I. Pitas, “Digital video processing and analysis ” , China Machine Press, 2017 (in Chinese).
[PIT2013] I. Pitas, “Digital Video and Television ” , Createspace/Amazon, 2013.
[NIK2000] N. Nikolaidis and I. Pitas, 3D Image Processing Algorithms, J. Wiley, 2000.
[PIT2000] I. Pitas, “Digital Image Processing Algorithms and Applications”, J. Wiley, 2000.

Bibliography

[PIT2021] I. Pitas, “Optimal multidimensional cyclic convolution algorithms for deep learning and computer vision applications”, in Proceedings of the International Conference on Autonomous Systems (ICAS), 2021

[KAR2021] I. Karakostas, V. Mygdalis, and I. Pitas,“Adversarial optimization scheme for online tracking model adaptation in autonomous systems”, ICIP (special session on Autonomous Vehicle Vision), 2021

[KAR2020] I. Karakostas, I. Mademlis, N.Nikolaidis and I.Pitas, "Shot Type Constraints in UAV Cinematography for Autonomous Target Tracking", Elsevier Information Sciences, vol. 506, pp. 273-294, 2020

[KAR2019] I. Karakostas, V. Mygdalis, A.Tefas and I.Pitas, "On Detecting and Handling Target Occlusions in Correlation-filter-based 2D Tracking" in Proceedings of the 27th European Signal Processing Conference (EUSIPCO), A Coruna, Spain, September 2-6, 2019

[NOU2019a] P. Nousi, A.Tefas and I.Pitas, "Deep Convolutional Feature Histograms for Visual Object Tracking" in Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK, 2019

Bibliography

[NOT2019b] P. Nousi, D. Triantafyllidou, A.Tefas and I.Pitas, "Joint Lightweight Object Tracking and Detection for Unmanned Vehicles" in Proceedings of the IEEE International Conference on Image Processing (ICIP), Taipei, Taiwan, September 22-25, 2019.

[PAP2019] C. Papaioannidis and I.Pitas, "3D Object Pose Estimation using Multi-Objective Quaternion Learning", IEEE Transactions on Circuits and Systems for Video Technology, 2019.

[PAPAD2021] Papadopoulos, Sotirios, I. Mademlis and I. Pitas, “Semantic Image Segmentation Guided by Scene Geometry”, in Proceedings of the International Conference on Autonomous Systems (ICAS)*, 2021*.

[PAP2021] C. Papaioannidis, I. Mademlis and I. Pitas, “Autonomous UAV Safety by Visual Human Crowd Detection Using Multi-Task Deep Neural Networks”, ICRA, 2021.

[PIT2021] I. Pitas, “Optimal multidimensional cyclic convolution algorithms for deep learning and computer vision applications”, in Proceedings of the International Conference on Autonomous Systems (ICAS), 2021

[PER2018] Perera, Asanka G., Yee Wei Law, and Javaan Chahl. "UAV-GESTURE: A dataset for UAV control and gesture recognition." *Proceedings of the European Conference on Computer Vision (ECCV)*. 2018.

Bibliography

[NOM2019] P. Nousi, I. Mademlis, I. Karakostas, A.Tefas and I.Pitas, "Embedded UAV Real-time Visual Object Detection and Tracking" in Proceedings of the IEEE International Conference on Real-time Computing and Robotics 2019 (RCAR2019), Irkutsk, Russia, 2019

[BOC2020] A. Bochkovskiy, CY Wang and HY M. Liao. “YOLOv4: Optimal Speed and Accuracy of Object Detection”. arXiv, 2020.

[DIET2021] A.Dietsche, G.Cioffi, J.Hidalgo-Carrió, D. Scaramuzza Autonomous Persistent Power line Tracking using Events, IROS 2021

[LIU2016] Liu, Wei, et al. "SSD: Single Shot Multibox Detector." European Conference on Computer Vision. 2016.

[NIC2020] C. Nicolas, et al. "End-to-End Object Detection with Transformers." arXiv preprint arXiv:2005.12872 (2020).

[SHA2019] A. Shahroudy, J. Liu, T. Ng and G. Wang, "NTU RGB+D: A Large Scale Dataset for 3D Human Activity Analysis," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, 2016, pp. 1010-1019.

Q & A

Thank you very much for your attention!

More material in
http://icarus.csd.auth.gr/cvml-web-lecture-series/

Contact: Prof. I. Pitas
pitas@csd.auth.gr