UAV human teleoperation using event-based and frame-based cameras

J.P. Rodríguez-Gómez, **R. Tapia**, A. Gómez Eguíluz, J.R. Martínez de-Dios, A. Ollero

Robotics
Vision
Control

GRIFFIN Meeting
October 28, 2021

Universidad de Sevilla

Outline

INTRODUCTION

PIPELINE DESCRIPTION

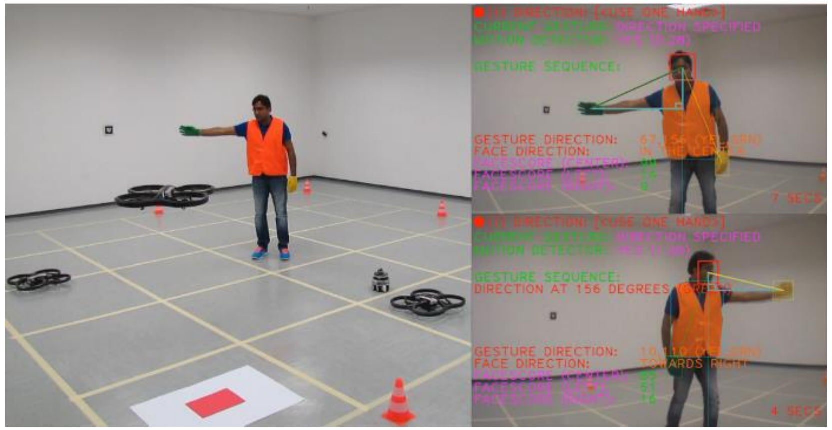
GESTURE DETECTION METHOD

EXPERIMENTS

CONCLUSIONS

Introduction

- Human-Drone Interaction (HDI) relying on user gestures detected from visual information.
- HDI requires perception systems capable of detecting unexpected events during the human-robot interaction.
- Event cameras:
  - Provide low-latency response.
  - Provide robustness against different illumination conditions.
  - Do not suffer from motion blur.



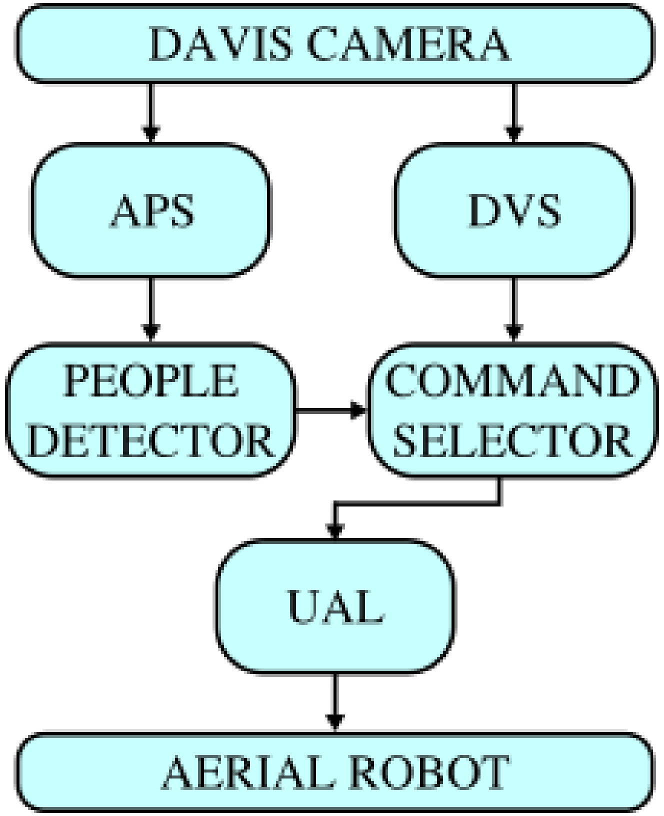
A human operator commanding a flying UAV [1].

Published in **AIRPHARO2021** Workshop on Aerial Robotic Systems Physically Interacting with the Environment

[1] HagaNagi, J., Giusti, A., Di Caro, G. A., & Gambardella, L. M. (2014, March). Human control of UAVs using face pose estimates and hand gestures. In *2014 9th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 1-2). IEEE.

Pipeline

- Grayscale images + events from a DAVIS346 camera.
- **Images**: Detect the operator using YOLO object classifier [2].
- **Events**: Detect operator gestures at high rate.
- The UAV receives the motion commands from the user gestures to move in the environment.



[2] Adarsh, P., Rathi, P., & Kumar, M. (2020, March). YOLO v3-Tiny: Object Detection and Recognition using one stage improved model. In *2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS)* (pp. 687-694). IEEE.

Gesture Detection Method

Detect gestures from events triggered by the movement of the user limbs in specific regions of the image.

- The set of detection regions are defined by: $\mathbf{Z} = [Z_0, ...Z_i]$
- Each region represents an area in the image: $Z_i \in \mathbb{R}^2$

Gesture Detection Method

- An event belongs to a region if it lies inside its boundaries.

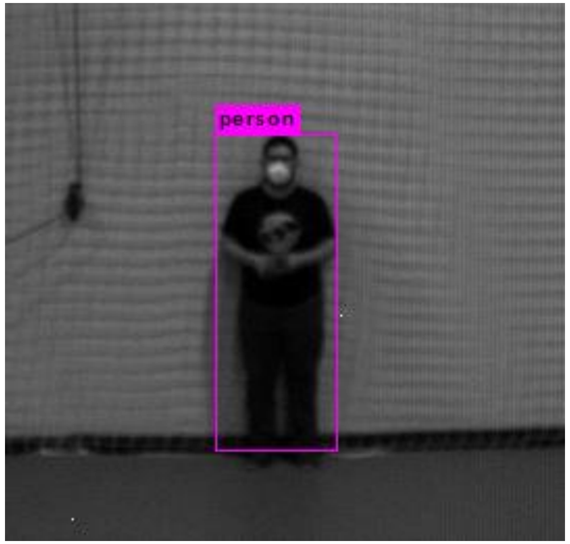
$$Z_i \Leftrightarrow u_{min} \leq u \leq u_{max}, \ v_{min} \leq v \leq v_{max}$$

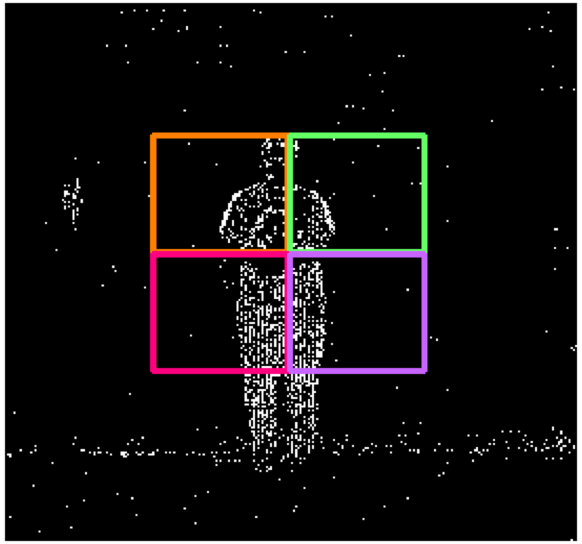
- Events are accumulated in each region using their pixel coordinates as a reference.

$$\eta_i = \sum_{k=0}^{N} \sum_{\mathbf{x} \in Z_i} \delta(\mathbf{x}_k - \mathbf{x})$$

Gesture Detection Method

- Four detection regions around the user body.





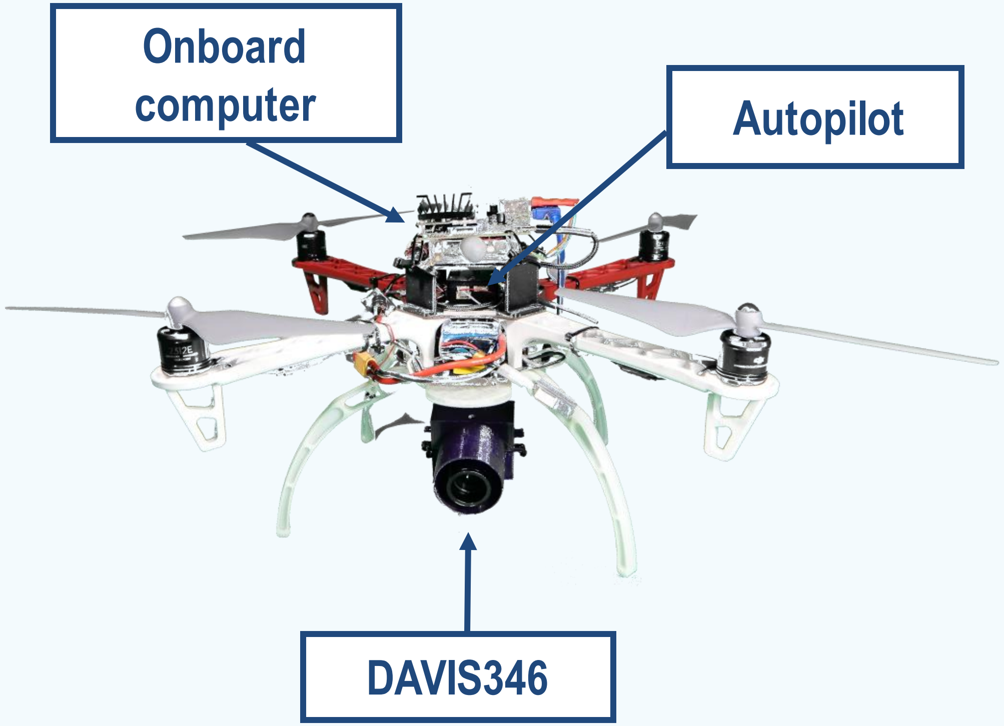
- Gesture commands are defined using the event occurrence at each region.

- Four types of commands are defined:

$$
\begin{array}{lll}
Right & if & \hat{\eta_0} \geq \kappa, \eta_0 \geq \lambda \\
Left & if & \hat{\eta_1} \geq \kappa, \eta_1 \geq \lambda \\
Up & if & \hat{\eta_0} + \hat{\eta_1} \geq \kappa, \eta_0 + \eta_1 \geq \lambda \\
Down & if & \hat{\eta_2} + \hat{\eta_3} \geq \kappa, \eta_2 + \eta_3 \geq \lambda
\end{array}
$$

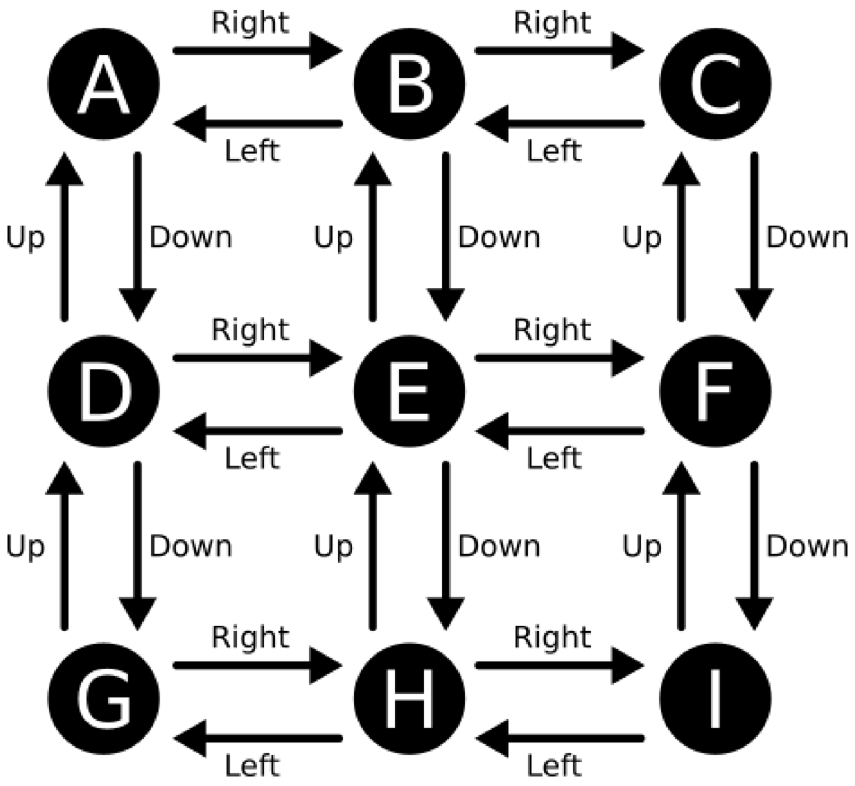
Experimental Platform

- DJI Flamewheel F450 frame.
- PixRacer autopilot.
- DAVIS346 event camera.
- NVIDIA Jetson Nano.
- Experiments performed in an indoor flight with a Motion Capture System.



Experiments

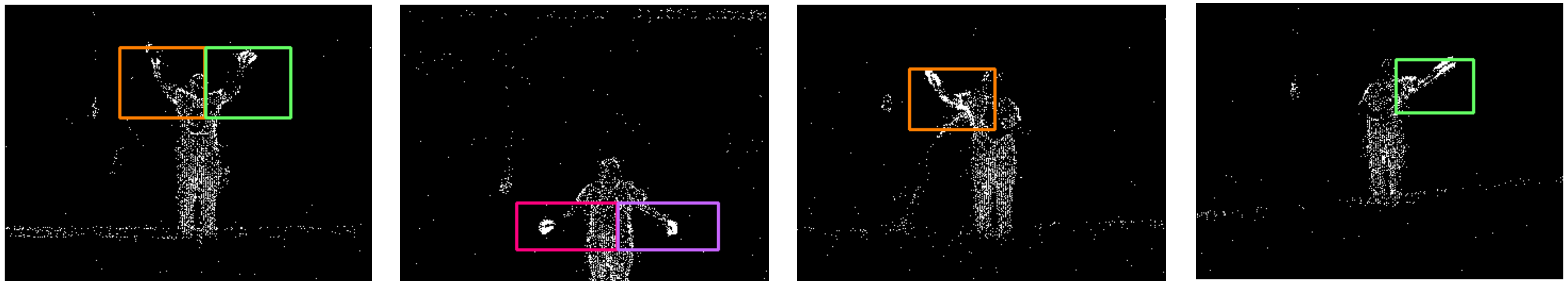
- A state machine handles the robot actions triggered by the gestures of the operator.
- Each location corresponds to a state.
- Nine locations were defined to move the robot in the flight arena.
- The robot stops at each location waiting for new commands.



State machine handling the robot actions.

Experiments

- The multirotor receives four types of commands: **up**, **down**, **left**, **right**.



Events accumulated each 25 ms just for gesture visualzation.

- The operator location is updated each ~15 Hz using YOLO.
- Gesture commands update at 200 Hz.
  - 7 times greater than many frame-based methods.

Conclusions & Future Work

- A human gesture method for multirotor teleoperation using event vision.
- Our hybrid approach combines event and frames.
- Event-based vision provides fast response at gesture recognition.
- A first step towars the integration of event cameras in UAV-human teleopertaion.

- Future work focuses on:
  - Developing a purely event-based solution.
  - Extend the method to perform human-robot collaboration activities at complex tasks.