Multimodal Interaction

Lesson 5bis
Gesture Interaction
An example application

Maria De Marsico
demarsico@di.uniroma1.it

CHITALY19

SEPTEMBER 23-25 | UNIVERSITY OF PADOVA | ITALY

SIGCHI ITALY

USING HANDS AS AN EASY UAV JOYSTICK FOR ENTERTAINMENT APPLICATIONS

Maria De Marsico

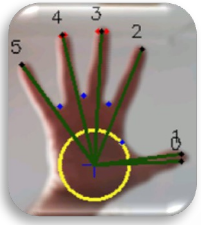
Alessandro Spagnoli

SAPIENZA UNIVERSITÀ DI ROMA

Natural Human-Computer Interacton

- Commands sent to a pc/mobile device as among humans
- Examples:«natural» gestures and spoken utterances

- Command decoding



Computer Vision for gestures

Signal processing for speech

Focused application

- Use of «**natural interaction**» to drive an Unmanned Aerial Vehicle (UAV), popularly known as **drone**

- Target context: **amusement and hobby**

Present context

- Most studies in literature (accuracy, effectiveness, efficiency, operator stress …) are related to **military** applications:
  - **Time to complete the mission**
  - **precision of flight trajectory and collisions**
  - **precision of targeting**
  - …

- **Hard operator training more than usability-oriented design**
- **Secondary goal: to avoid «collateral damage»**
- **when the use of weapons is entailed, a hybrid solution is of course to be preferred → (wise) suitable balance between automation and human control**

Differences

- Few studies in literature

- **No or little operator training; usability-oriented design is more than welcome**

- **Primary goal: to avoid «collateral damage», including … drone destruction**

Hand-tracking for drone driving vs. traditional modalities

- Goal: intuitive interface, able to track the operator's hand movements, and to transmit them as commands to a drone to drive it without the aid of any other device, sensor or glove

- Most popular traditional modalities:
  - Joystick

  - Virtal knobs

  - Instrumented love

- Difficult to map width and speed of gestures

Maria De Marsico - demarsico@di.uniroma1.it

Compared app: DJI GO

- At the time of the study, the only one with gesture control
- Main phases: take-off, hand recognition and tracking, away and follow

Drone DJI Spark

Infrared sensors

Compared app: DJI GO

Main phases: take-off, hand recognition and tracking, away and follow

Compared app: DJI GO

Main phases: take-off, hand recognition and tracking, away and follow

Compared app: DJI GO

Main phases: take-off, hand recognition and tracking, away and follow

Compared app: DJI GO

Main phases: take-off, hand recognition and tracking, away and follow

Limitations of DJI GO

- Somehow implicit in the announcement as "cameras in the sky“[1]
- Designed for a single specific operation
- Further will be discussed in comparison with our proposal

[1]https://www.newswire.ca/news-releases/dji-launches-spark-the-easy-and-funcamera-drone-for-everyone-624086813.html

Parrot AR. Drone 2.0

Drone components

High Definition 720p Camera + low resolution camera positioned under the drone

Ultrasonic sensors

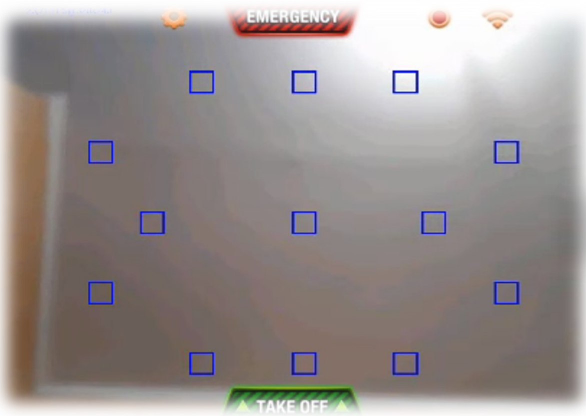
No hand tracking facility (not even in Parrot more advanced models)

Core design strategy

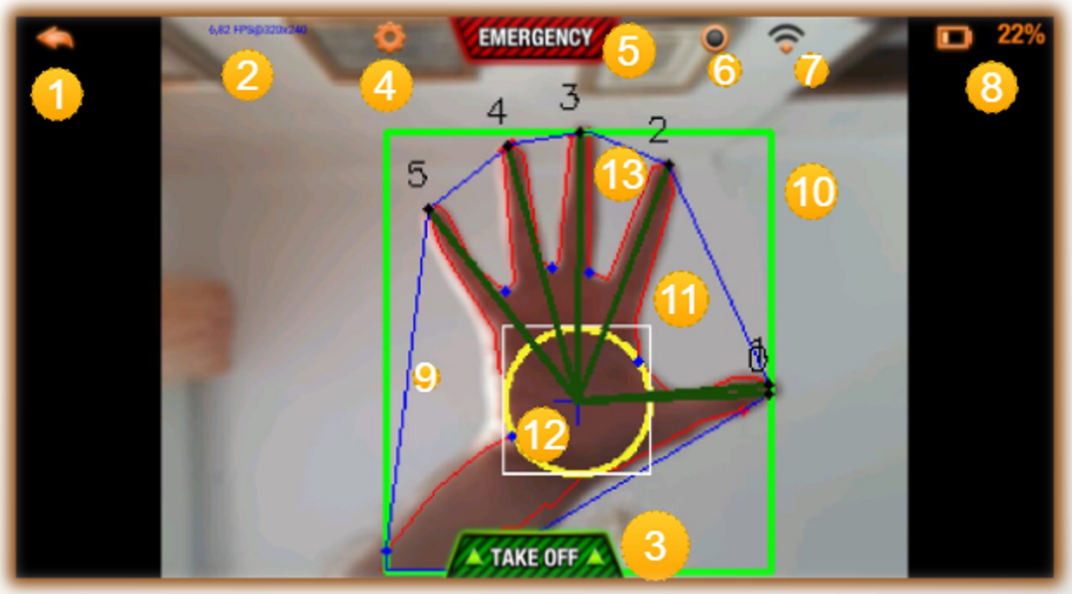
- WRAPPING!
- Mapping hand gestures onto drone commands using the AR. Drone 2,0 Open Source Library provided by Parrot

Hand detection and tracking

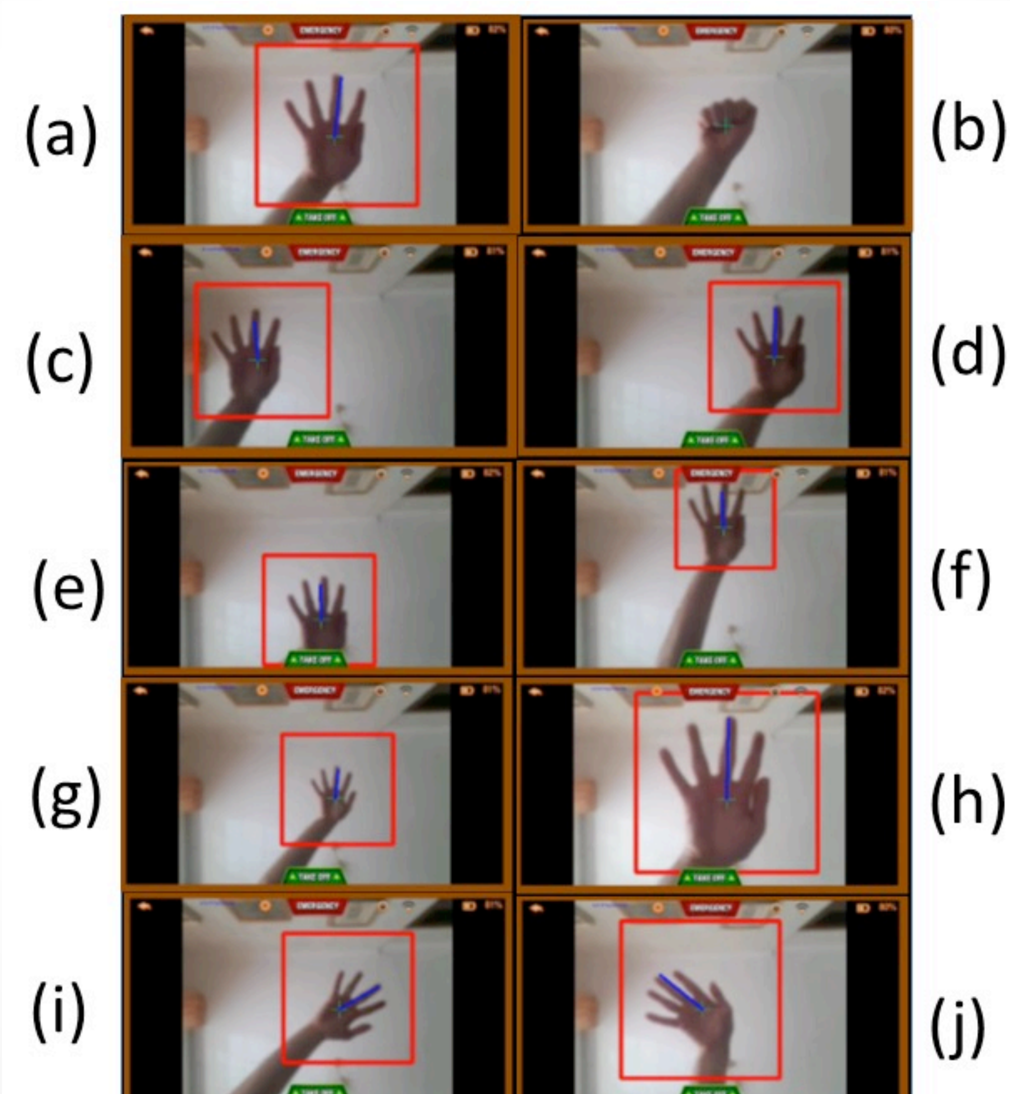
Build color model for both background and hand



Hand detection and location of relevant features (center, fingertips, size of the convex contour)



Gesture vocabulary



Comparison with DJI GO: Sight constraints

DJI GO

- direct capture of the hand image from the drone camera → the hand must be continuously in sight of the drone

Our App

- the front camera of any mobile phone captures the hand gesture and the phone processes it to map it onto a known command and then transmit it to the drone via WiFi → no need for the hand to be always insight

Comparison with DJI GO: Acquisition and extraction of hand information, and accuracy of results

DJI GO

- user must be very close to it to perform not only the extraction of the hand information, but also the acquisition of the face in order to recognize the relative position of the pilot → a phase of training is required that does not always end with correct information (as reported by users)

Our App

- after the piloting option is pressed, in little more than a few seconds the interface guides the operator to correctly execute the steps to extract the hand characteristics
- this very intuitive action is also started by the open hand

Comparison with DJI GO: Drone take off in hand-controlled mode

DJI GO

- the SPARK app allows (but actually, requires) the drone to take off and land directly from the user's hand

Our App

- the same operations is possible by using any flat surface, that does not need to be too close to the pilot

Comparison with DJI GO: Operator's freedom during drone's flight

DJI GO

- the SPARK requires the operator to never move more than 1 meter away from the drone, due to the limited range of action of its sensors → limited use of the drone, when piloted by hand

Our App

- the complete piloting of the drone at a distance is only limited by the power of the WiFi (in tests with a normal antenna was about 50 meters), and does not require the drone to know the user's position

Comparison with DJI GO: Active tracking system

DJI GO

- the SPARK active tracking system allows the drone, after the recognition of the corresponding gesture, to follow the operator during walking

Our App

- future work

Comparison with DJI GO: Video recording

DJI GO

- during DJI GO hand-gesture control, video recording is limited, since the camera is busy

Our App

- the app presented here exploits phone camera for gesture tracking, → the drone camera is fully available
- it is possible to send the video to a USB stick positioned on the drone, but also the direct transmission of the file to the mobile phone, during the flight of the drone

Comparison with DJI GO: Cost

DJI GO

- DJI SPARK with the features discussed here = about 799 euros

Our App

- Parrot AR. Drone 2.0 = about 150 euros (basic version)

Comparison with DJI GO: Further considerations

DJI GO

- not all the commands (in particular the one for tracking mode and return command) are always interpreted correctly
- the farther the hand from the camera, the higher the influence of negative factors like self occlusion of parts of the hand/arm, shadows or over-illumination

Our App

- the image capture carried out at short distance by the smartphone camera mostly avoids these problems

Tests: users

- 10 participants were involved in the tests, all adult males, with age mostly in the range 20-25
- no participant except two had driven a drone before
- each participant performed 40 gestures (4 repetitions of the set of 10, in no predefined order) for a total o 400 processed gestures
- before testing the application, the participants carried out a very short training regarding the basic supported gestures

Tests: equipment

- the application was tested on a Samsung S7 edge with 8-core processor and 2 GHZ frequency
- the mobile camera also supports full HD video, but 320 × 240 pixel resolution is sufficient

Tests: tasks

- all participants tested both the original application, controlling the AR.Drone 2.0 by the screen joystick, and the gesture recognition control (both exploit the smartphone screen)
- within-subjects evaluation: the same subject tested all the conditions
- to avoid any learning effect half of them tested the knobs first, and the other half the gestural commands first
- very simple tests to reproduce normal amusement
- requested tasks = free flight in a delimited space, producing the series of 40 gestures per participant + take-off and landing on a square region 1m × 1m (useful to test the precision achieved by the operator)

Tests: evaluated conditions

- Indoor
- Outdoor

In partcular

- Tracking accuracy
- Gesture recognition
- User experience

Tests: indoor

- closed ambient: absence of external climatic conditions that compromise the flight
- phone placed either on a table or on the other hand with the camera turned towards the ceiling
- ceiling had usually a uniform color

Tests: indoor

- ✓ The recognition of the hand did not present any problem in either condition.
- ✓ The gestures were recognized correctly and the hand was captured in its exact dimensions
- ✓ The response to the controls was precise and timely

→ → drone driving can be considered safe.

Tests: outdoor

- possible weather conditions in an external environment could cause instability and difficulty in controlling the device (common to **all** control modes, does not depend on the gesture modality)

- phone placed either on a surface or on the other hand with the camera turned towards the sky

- sky may have inhomogeneous color distribution

Tests: outdoor

- ✓ Hand recognition = successful even with clouds or unstable sky → correct operation
- ✓ A possible negative condition: objects or colors close to those of the hand in the field of view of the camera → increased workload for hand detection and delay in the transmission of commands to the drone
- ✓ Solution: a glove with sharp color that would be automatically modeled at the beginning of the piloting session → decrease in the "naturalness" of the interaction.

Tests: tracking accuracy

Starting from the first frame acquired by the camera, frame-based metrics were computed frame by frame (binary classification hand/no-hand)

- True Negative, TN: the number of frames where the hand is not present and the system does not detect it
- True Positive, TP: the number of frames where the hand is present and the system correctly identifies it
- False Negative, FN: the number of frames where the hand is present but the system does not correctly detect it
- False Positive, FP: the number of frames where the system identifies the false presence of the hand

Tests: tracking accuracy

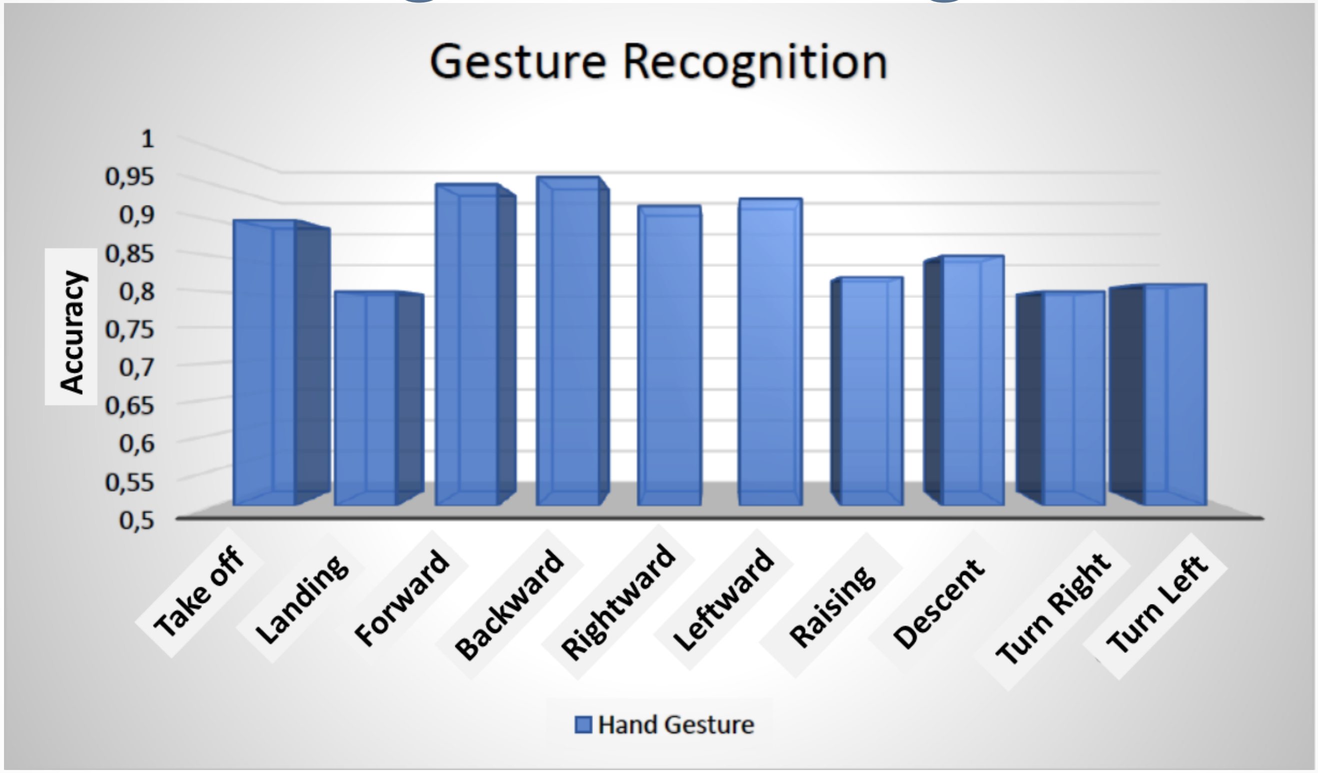
- TG = the number of frames where the hand is present
- TN = the number of frames where the hand is not present
- T F = the total number of frames from the camera

Starting from the 400 gesture clips recorded (10 participants performing 40 gestures each) the following final metrics were computed:

- Tracker Detection Rate (TRDR: TP/TG)
- False Alarm Rate or FalseAcceptance Rate (FAR): FP/TN

The prototype achieved a TRDR of about 0.8 and a FAR of 0.15

Tests: gesture recognition



- the recognition of the operator's gesture works in 90% of cases
- no significant difference in accuracy was found between the indoor and outdoor performances
- the data refer to a mix of them

Tests: user experience

A simple questionnaire was carried out after the participants completed both test rounds:

1) comparison of the ease of use of the two control modalities
2) comparison of the naturalness of using the control elements, virtual knobs vs. hand gestures
3) comparison of the learning effort to calibrate the gestures to the desired drone movements
4) comparison of the expected performance vs.the achieved one for both modalities
5) an open-ended question about the problems found.

Tests: user experience

- All participants (100%) found the gesture-based interaction much easier to use and natural even without a significant training (questions 1 and 2)
- In particular, they found more natural the scaling from the amplitude of the gesture captured by the phone camera to the drone movement, than the scaling needed by the virtual knobs (question 3)
- The above caused a lower correspondence with expected performance when using virtual knobs (question 4)
- Problems found: gestures corresponding to the turn commands need wrist rotation and were considered a little less comfortable (these are also the gestures achieving the lower recognition performance, though above 80%)

Future work

- New gestures to implement more functions (e.g., follow me) or …
- … add speech recognition with a small intuitive vocabulary to avoid the need to learn counterintutive gestures
- Improve robustness in adverse condition (background, illumination)

Videos

Readings

- De Marsico, M., & Alessandro Spagnoli, A. (2019). Using hands as an easy UAV joystick for entertainment applications. In CHItaly 2019, Proceedings of the 13th Biannual Conference of the Italian SIGCHI' Chapter: Designing the next interaction. Article n. 9. ACM.