

The Researcher's challenge

Main problems concerning the task

- The success of the system will highly depend on the limitation of RGB camera (meaning the quality of the video), the light conditions and the variation of the background;
- Data acquisition: training the deep learning software solution will require a large set of videos that should vary in light, background, etc;
- Data transformation and storage

Development challenges

- Performance : during the implementation, the solution's performance may vary depending on the host device's specs; performance should be monitored during development so that relevant feedback could be passed to the management as part of the decision of the embodiment of the technology

Step-by-step plan

1. Data acquisition:

Several options should be taken into consideration:

- data crawling (may be affected by uniformity issues, and extra effort invested into 'cleaning' and validating the input obtained in a such a manner: there may be videos too long, videos too short, shot from afar, multiple people in the footage, etc).
- 'in-house' generation of footage (the main problem will be the amount of time invested into filming the set of videos)
- some kind of 'public campaign' (either online or not) that would provide videos of people performing a set of given gestures

The assigned engineers should look into these main approaches and possible alternatives (or maybe a combination of any of them) and gather the data accordingly. A first analysis dismisses the first and second options as inefficient and not robust enough.

2. Data validation and filtering

This will depend on the acquired data as a result of the first step. Simple validations may be instantiated (for example a video length's) so that highly atypical samples can be filtered out

3. Data transformation

The filtered data should pass through a series of transformations so that it can be fed to the network in a convenient and uniform format. These transformations could include but are not limited to:

- edge detection for noise elimination (recognizing the hand from the environment): multiple approaches should be investigated (Sobel filtering, Perwit filtering, etc);
- contour completion: multiple approaches should be investigated;
- any other transformation that can improve the consistency of the data set or minimize the input size (for example image cropping for frames that contain a lot of “empty background”);
- convenient data representation (for example matrix of pixels)

It should be taken into consideration that all of these transformation will be applied for the data obtained live, which may cause performance issues!

4. Divide data into training/test/validation sets

Research optimal training/test/validation split and divide data accordingly.

5. Train & Tune

6. Interface, integration and deployment

More development tasks could arise such as designing a mock-up interface, integrating the final fully-trained classifier, deployment of the software solution to the host device. These are highly dependent on the requirements of the prototype, may be voluntary neglected (as higher priority tasks need attention) or developed in parallel to the research and implementation of the classifier.