hugomcp@di.ubi.pt, 2024/25

Human Recognition in Surveillance Settings

Recognizing humans in surveillance settings is among the major challenges in artificial intelligence, due to the wide range of potential applications (e.g., security, safety and forensics). By assuming that subjects are not aware of the data acquisition process, it is expected that the collected data has very poor quality, not only in terms of resolution and lighting, but also in terms of pose and occlusions.

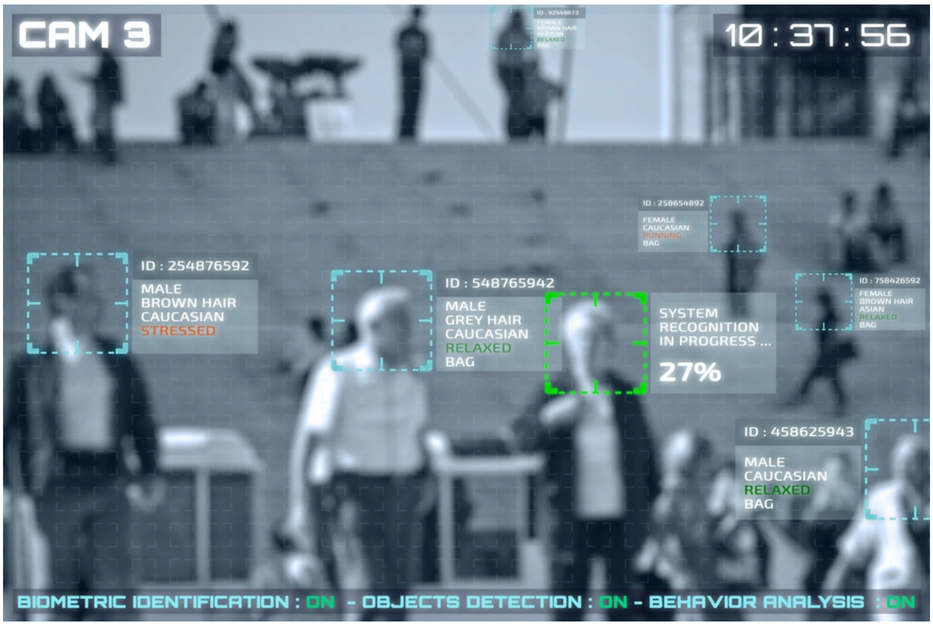

Figure 1: Typical conditions in surveillance settings, where recognition faces severe problems due to poor data quality (source: http://verdict.co.uk).

Hence, there are numerous efforts being developed to develop automata able to recognize human beings in such type of conditions, i.e., using extremely poor-quality data. Among the many difficulties that arise in this setting, one of the problems is the inexistence of solid information about the actual variations in the data collected with respect to each data covariate (e.g., distance, pose, resolution, and lighting).

A video-based dataset for biometric recognition in surveillance settings is available at: https://socia-datasets.di.ubi.pt:50004/hugomcp/CV_24_25_data.zip

This practical work is divided into 2 main parts:

1) From Image to Video Recognition. The goal is to perceive the advantages in performance that can be obtained in biometrics/object recognition, when moving

from using a single frame as input, to using "n" frames (i.e., from "image" to "video" recognition).

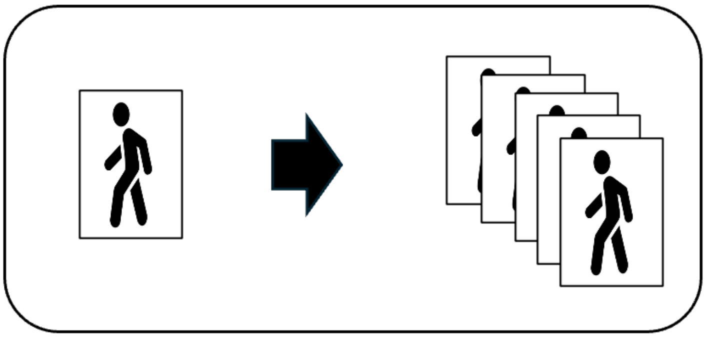

Figure 2: Illustration of the input data planned to be used in the practical project. At first a single frame (n=1) should be used as input. Next, the value of "n" should be increased and the variations in performance measured.

Each student is required to develop, using Python + Keras or Pytorch libraries, an **expert model** to perform pairwise image/video classification. This model should distinguish between gallery + probe pairs that:

a. Regard the same trait values (i.e., "**genuine**" comparisons)

b. Regard different trait values (i.e., "**impostor**" comparisons)

Note: This analysis of "**genuine**" / "**impostor**" can be made (after discussion and assignment of the tutor's course), in terms of:

- ID;
- Soft biometric labels (age, gender, ethnicity);
- Clothing colours and style.

2) Using an open access LVLM, design and implement an interaction strategy with the LVLM, in order to increase the interpretability of the responses provided by the expert model. The idea is to use the feature maps provided by the expert model, together with the final response and ask the LVLM to provide human-understandable justifications (in textual or visual format) for the responses obtained.

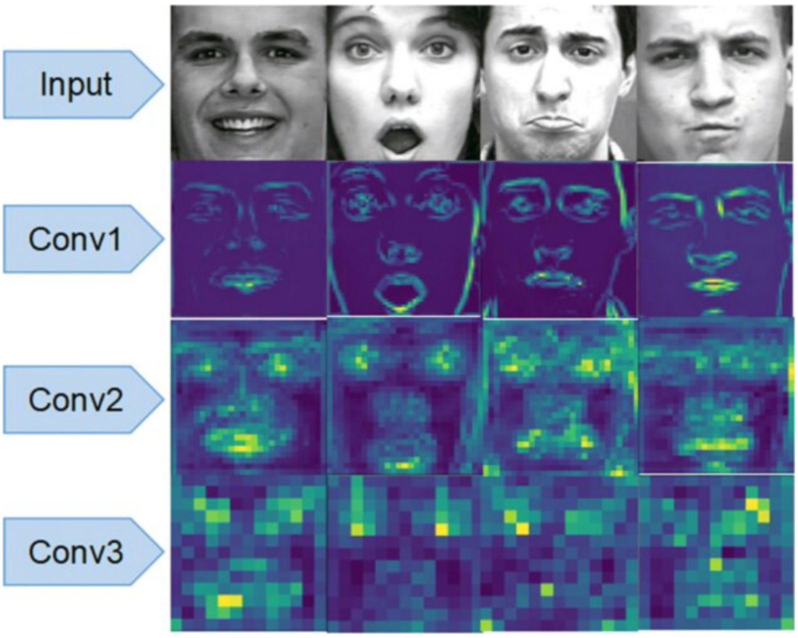

Figure 2: Example of feature maps from different convolutional layers of a face recognizer expert system (source: https://www.researchgate.net/publication/333379379).