COMPUTER VISION COURSE

September project 2022

Human gaze estimation

Introduction

Eye direction estimation techniques can provide useful clues about users' attention or can improve the lives of people with disabilities by allowing them to interact with computers. Until a few years ago, eye tracking technologies were limited to laboratory environments, but recent advancements in the computer vision field allow dealing with this problem in more challenging environments using standard cameras.

It looks relatively easy to estimate the eye direction from an image of a face, but limited resolution, motion blur, variable head poses, eyeglasses and illumination conditions can cause ambiguities in determining the eyeball shape of the iris.

The task of this project is to:

- localize the eyes in color images;
- inside the eye, detect the iris (described as a circular region enclosing the iris);
- estimate the eye direction.

For this project a simple classification of the gaze direction into 3 categories is needed – namely:

- looking straight,
- looking right,
- looking left.

You can use both deterministic approaches like template matching or the Hough transform for circles or machine learning techniques, e.g., use Viola and Jones to get the face and then try to get the eyes by knowing their position in the face. Deep learning approaches can also be used – **the same rules already discussed on the forum apply to this project**.

Test dataset

To assess the performance and robustness of your system a benchmark dataset is provided on moodle (eyes_direction_cv) - the images are divided into three directories that identify the three categories mentioned above.

Project delivery

The project must be developed in C++ based on the OpenCV library. The only allowed exception is the usage of Python code for developing the networks if you decide to exploit this family of techniques. However, **the project must compile (using CMake) on the T.2020 virtual machine with the current software and library installation**. The trained deep network shall be provided as a file.

You need to deliver your project including:

- All the source code (both C++ and Python);
- CMake configuration files (the use of CMake is mandatory);
- A report (no page limit) presenting your approach and the performance measurement on a set of images detailed below.

Working in a group, you should **clearly identify** the contribution of each member in terms of ideas, implementation, tests and performance measurement. You can organize the work as you prefer: you are not forced to assign one specific step to each group member. Please also include **the number of working hours** per person in the report. This is needed for a monitoring on our side of the effort requested – the evaluation will not depend at all on the number of working hours, but on the quality of the result.

Performance measurement

The numeric evaluation of the performance of your approach based solely on the gaze direction classification detailed above. You should provide:

- the confusion matrix
- the classification accuracy

related to the test dataset.

Your report **must include** the results of the eye and iris localization superimposed on the original image (no numeric values are needed for such two outputs of the algorithms). Moreover, the result of the gaze estimation should also be provided as a written element superimposed on the image.

The test dataset provided shall be used to assess the performance of your system. If your solution relies on machine learning/deep learning (which is not needed – just your choice), you will need the ground truth of a sensibly larger dataset. Collecting additional ground truth is a time-consuming task, but you are free to share the output of the ground truth with other colleagues belonging to other groups. In other words, you are allowed to cooperate with everyone to share the ground truth collection. The organization of the ground truth collection is completely free; if you wish, you can use the project section in the moodle forum.

If you use additional dataset(s) and related ground truth, **you must describe this in the report**, **providing a link** to an online resource (e.g. UniPD google drive) where such additional dataset is available.

If you use machine learning/deep learning in your project, please consider that **you must not use any image coming from the test dataset** in your training/validation dataset.