Computer Vision – Measuring Crowd Size and Flow from Video Streams

MobiLab Team - [redacted] 2019

Project Description

The aim of this project is to research the possibility of using existing and new video streams of squares (and other public spaces) in the municipality to **count crowd sizes** and to **get insight in crowd flows**. Currently, the municipality is working with third parties to do crowd analysis on video material. However, working with third parties means that the municipality does not have full insight in (and ownership of) the process. Hence, the municipality currently cannot make low-level decisions about the used approach, algorithm(s), trade-offs and involved ethics (privacy concerns etc.). Given the growing importance of these types of decisions, the municipality wants to do its part, and actively conducting research towards developing its own innovative solutions. Ideally, this project works towards an **open-source alternative** which makes an impact on the transparency of crowd monitoring methods used
by any governmental institution, today and in the future.

Currently, preliminary research has been performed on a small set of camera's using computer vision techniques such as YOLO (2016, Redmond et. al.) in order to check the feasibility of this approach. This has lead us to a positive outlook for this endeavor, but has raised a lot of questions on the specifics for an ideal approach.

Possible research questions

- What approach is most suitable to **consistently get reliable counting and tracking results** in a variety of situations?
    - Some possible methods: body detection, foot detection, head detection, umbrella detection, gait recognition for flow tracking, a combination of techniques.
- What **computer vision algorithms** could we build on/extend to implement such an approach?
- How do we efficiently and quickly **create a ground truth**? Could we optimize this process? (e.g. using human assisted annotation?)
- What **domain factors** should we account for? (e.g. differences in camera distance, angle and focal length; differences in square size; differences in weather and light conditions; possible lens flare; etc.).
- How can we optimize our approach in a **scalable fashion**, so it works efficiently for large sets of camera's (1000+)?

What will the work environment look like for the student?

You will work in the MobiLab team, which has working spaces at the **Datalab** of the Municipality of Amsterdam (Weesperstraat 113) and the **Amsterdam Institute for Advanced Metropolitan Solutions** (Marineterrein Amsterdam). MobiLab is a team of experienced software developers and business analysts, which is knowledgeable regarding machine learning and computer vision methods and has hands-on
experience with the video data.

What we offer:

- A pleasant working environment
- Collaboration with a team of experienced people
- Responsibility over your project and progress
- A possibility to contribute to (your own?) smart city development
- A monthly internship compensation

From students, we expect the following:

- An independent work ethic
- Machine learning knowledge
- A keen interest in computer vision
- Python programming skills
- And most importantly: a passion for their field, an ambition to learn, and a drive to make Amsterdam even more beautiful!