Optical Character Recognition

By: Marc Zoomers

How to application works:

JPG

A .JPG file is chosen

Natural Image Mode

License Plate Mode

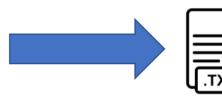**Scanned Document Mode**

PyOCR

With the use of an SWT-algorithm, it will make a high-contrast mask of the original image, that will improve the performance of the OCR-engine by more than 150%, compared to the original photo.

OPEN-ALPR

OPEN-ALPR uses the Wolf-Julien-method and the Sauvola-method to transform the image and uses a Local Binary Pattern algorithm to find license plates and read them out.

PyTesseract

The Tesseract developed by Google uses Otsu's algorithms and Otsu's method to find text from scanned documents. Thanks to Otsu's algorithm (also known as the Firefly-algorithm) it's very capable in finding re-occurring patterns in a photo in a very efficient way. With this ability it can outperform a lot of other algorithms that are used in OCR-engines.