Beeldverwerking Assignment 3

Amber Elferink - a.s.elferink@students.uu.nl - 5491525

Thijs Ratsma - t.ratsma@students.uu.nl - 6315887

11-11-18

# Introduction

For this assignment, we wanted to create a practical algorithm that could be applied and be useful in a real-life situation. Inspiration came from FAIR2Media, an acquainted design studio which is working on a table with a beamer projecting images on top of it. In the current iteration, the projection is always the same map. Through interaction with the table, the projection would update, allowing for interaction with the software. Since hands are a natural way to interact, our goal would be to develop an algorithm that could detect the position of hands above a projected table.

Because the beamer is fixed at a certain height above the table, there is not a lot of variation in size of hands, aside from the real size of course. However, this scale is near negligible. For our collection of photos, there was no variation in light conditions. Many different photos of hands were taken, with variations in rotation in relation to the table, partial occlusions caused by the projection, arms with or without sleeves and hands either pointing or spread out. We also assume only one hand at a time can be detected.

This resulted in the following context table:

|  |  |
| --- | --- |
| **Criterium** | **Possible values** |
| Minimum/maximum size | Near-same size always |
| Lightning variations | Constant |
| Rotation variations | Multiple different rotations |
| Occlusion | Often partially occluded by projection |
| Other | Arms with/without sleeves, hands pointing/spread |

The desired output is a bounding box showing the position of the hand, as well as a way to differentiate between a pointing hand and a spread hand. In case there are multiple hands in a picture, only one bounding box depicting one hand will be shown.

A picture containing mosquito net, object, indoor, table

Description automatically generated

Figure 1: Example of an input photo

# Pipeline

## Pre-processing

All input images show parts of the table which don’t have any part of the map projected onto them. Our first step was to resize the image to remove these areas and then converting the remaining image to a greyscale image. After experimenting with threshold values on multiple images and by comparing histograms, we concluded that, in most cases, the greyscale value of the hands lies between 70 and 90. However, because of the projection on the hands, this wasn’t constant, so a fixed threshold wouldn’t work. An automated threshold wasn’t the solution either, as the values for the hands in the histogram were valleys compared to the peaks of the rest of the projected map. As such, the following pre-processing pipeline was decided on:

1. Resize the image and convert to greyscale.
2. Threshold the greyscale image, starting at a threshold value of 70.
3. Apply an opening filter to get rid of any noisy pixels (3x3 filter).
4. Apply a region labeling algorithm, count the number of regions and return the bounding box of the largest region.
5. If the number of regions increased compared to the previous iteration, don’t update the bounding box.
6. Repeat steps 1-4 on the image derived from step 0 with an incrementing threshold value until the difference between the smallest number of regions and the current number of regions is larger than 3.
7. Return the thresholded opened cutout of the bounding box of the largest labeled region.

This pre-processing pipeline ensures the best threshold of the hand with the least amount of noise interfering from the projected map in the background. Steps 1-4 generate a bounding box containing (presumably) the hand, while step 5 checks if the result is more suitable than the last iteration. The difference in regions of 3 was chosen experimentally, as we observed that 1 or 2 new regions are sometimes introduced without hampering the quality of the threshold too much, while 4 to 6 new regions allowed too much leeway and drastically lowered the quality of the overall threshold.

A close up of a mans face

Description automatically generatedA close up of a mans face

Description automatically generatedA close up of a mans face

Description automatically generated

Figure 2: Opened thresholded bounding box cut-outs for thresholds of 80, 88 and 92 respectively. 88 was found to be the optimal threshold, while 92 was the point where the algorithm stopped running because more than 3 new regions were introduced.

## Object recognition

# Reflection

# Results