Ready for detection

Stair-detecting in depth images using spatial features and Adaboosting

F.D. Andriessen
1517619
Space Engineering

Final Draft
MSc. Thesis

Delft University of Technology

Aerospace Engineering

January 15, 2018

Ready for detection
A coming of age story

Frerik Andriessen
University supervisor: Prem Sundaramoorthy
Company supervisor: Anne van Rossum

January 15, 2018

Abstract

Space exploration could be significantly aided by combining the disciplines of machine learning and computer vision, but these disciplines need to be developed further for specific space-related applications to have merit. One of the applications for space exploration is the detection of certain structures designating areas of interest. This thesis demonstrates a method of structure-detecting that is applied to staircases. In addition to incorporating certain physical features, like other algorithms have done, the proposed algorithm (Step-1) also takes into account the spatial relation between these features, in order to increase its robustness. Looking at a staircase from the front, the distances between each step become warped, as they are further away from the observer. This exponential spatial distortion is known as a 'chirp'. Step-1 tries to match a chirp-waveform to every edge along a straight line randomly drawn through an image, and based on that match classify the image as containing a staircase or not. The random lines are then weighted based on their effectiveness using Adaboost, which are finally combined to obtain a final classification. The results show potential but there are still some issues to be addressed. However, once the algorithm has been upgraded it could aid space exploration by being applied to satellite images and autonomous rovers.

Contents

1 Introduction **1**
- 1.1 Thesis topic . . . 2
- 1.2 Research questions . . . 3
  - 1.2.1 Algorithm requirements . . . 3
- 1.3 Current state of computer vision in space exploration . . . 4
- 1.4 Outline of thesis . . . 4
  - 1.4.1 Methodology . . . 4

2 Literature review **6**
- 2.1 Surface segmentation . . . 6
- 2.2 Object detection . . . 7
  - 2.2.1 Stair-detection . . . 7
- 2.3 Conclusion . . . 10

3 Approach **12**
- 3.1 Algorithm . . . 12
  - 3.1.1 Dataset . . . 12
  - 3.1.2 Tools used . . . 13
  - 3.1.3 Flow diagram of proposed algorithm . . . 13

4 Stair-case RGB-D preprocessing **14**
- 4.1 Extracting surfaces . . . 14
- 4.2 Gaussian Blur . . . 16
- 4.3 Clustering pixels . . . 18
  - 4.3.1 Histogram or bins . . . 18
- 4.4 Floodfill algorithm . . . 19

5 Stair-detection algorithm **21**
- 5.1 Drawing random lines . . . 21
  - 5.1.1 Generating random coordinate-pairs . . . 21
  - 5.1.2 Bresenham's line drawing algorithm . . . 22
- 5.2 Classifying as chirp . . . 25
  - 5.2.1 Hilbert space . . . 26
- 5.3 AdaBoosting for final classification . . . 29

6 Different phases of algorithm development **33**
- 6.1 A simple Stepcounter algorithm, without chirpfitting or boosting . . . 33
- 6.2 Chirpfitting without boosting . . . 34
- 6.3 Chirpfitting using line-pairs . . . 36
- 6.4 Chirpfitting with penalties . . . 36
- 6.5 Chirpfitting with normalized error . . . 38
- 6.6 Chirpfitting with boosting . . . 39

7 Results **40**

8 Conclusion & recommendations **43**
- 8.1 Applicability to space exploration . . . 44

A C++ implementations **46**
A.1 Bitmask for segmenting . . . 46
A.2 Floodfill algorithm . . . 46
A.3 Updated Bresenham line drawing algorithm . . . 47

B Article in the AD about stairclimbing robot **49**

List of Figures

1.1 Birthrate of people in the Netherlands between 1850 and 2010 [1]. 2
1.2 Age distribution of the population in the Netherlands in 1950 [2]. 2
1.3 Age distribution of the population in the Netherlands in 2017 [2]. 3

2.1 Segmented surfaces in the point cloud using plane-fitting through spatial neighbourhoods of a 3d point cloud by Holz et al. [3]. The different orientations of the surfaces can be easily distinguished based on their color. 6
2.2 Examples of the Haar-like features used by Viola and Jones [4]. 7
2.3 Example of Haar Wavelet [5]. Note the resemblance to the Haar-like features from Fig. 2.2. . . 7
2.4 Parametric representation of line in (x,y) space [6]. 8
2.5 Example of demarcation done by a Radial Basis Function[7]. Any new datapoint falling within the circle will be designated white, while any point falling outside the circle will be designated black. Note that higher dimensions are possible. 9
2.6 Example of depth-graph of crosswalk and staircases from Wang et al.[8]. The crosswalk decreases in intensity as the line moves further away from the camera. The upstairs and downstairs graphs also show clear characteristics. The downstairs graph is more steep and has straight corners, while the upstairs graph slowly rotates as the steps move up. 9
2.7 Process of staircase detection by Lee et al. [9] 10
2.8 Start and finish of filtering stair line-segments [10] 11

3.1 The main steps of the algorithm. In color coding is differentiated what was already existing, and what was novel contribution 13

4.1 The steps taken in pre-processing part. Outputs after each step can be seen next to the step. The transformation of a noisy multi-channel image to a smooth one-channel image should be clear. 15
4.2 Loaded raw input, before doing any processing on it. Already some noise can be seen at the edges of some steps. 16
4.3 Resulting image after surface normals were calculated for depth image. There is noise due to the discrete (values of 0-255 in steps of 1) nature of the pixel values as well as the lesser quality of the input data. 16
4.4 Surface normals after Gaussian Blur 18
4.5 Flow Diagram for segmenting pixels based on their individual channel value. 19
4.6 Clustered staircase using 8 different bins 19
4.7 Noisy areas removed with Floodfill algorithm 20

5.1 Overview of stair-detection algorithm 21
5.2 Example of Bresenham's line drawing algorithm applied to clustered input 22
5.3 Flow Diagram of the original Bresenham line drawing algorithm. Note that it does not allow for negative dx or dy, or for dy to be bigger than dx. 23
5.4 The eight possible orientations for drawing a random line with (x_1, y_1) as starting point and (x_2, y_1) as ending point[11] 23
5.5 Flow Diagram of updated Bresenham Line Drawing algorithm. It is now able to deal with any value of dy and dx. 24
5.6 Showing all 8 possibilities for drawing a line using the updated Bresenham's line drawing algorithm 25
5.7 Example of a chirp-like repeatability in a real world setting[12] 26
5.8 Plot of possible combinations of a and b to satisfy $a640^b = 40\pi$. Based on this plot a range of $(0 \leq a <= 0.2)$ and $(0.5 \leq b \leq 2.0)$ was chosen for the iteration. 27

5.9 Example of a chirpfitting with a very large amount of periods, erroneously determined to be the best fit. 28
5.10 Example of a bad fit being classified as good, due to bad spacing between each period. 28
5.11 Hilbert visualization for one random line 29
5.12 Example of a chirp function fitted to the transition points of a staircase 30
5.13 Flow Diagram of the AdaBoost part of the final algorithm. 32

6.1 The effect of varying the detection threshold on the total score for 20 lines. 33
6.2 The effect of varying the detection threshold on the total score for 100 lines. 33
6.3 The 3d graph of the effect of the step and detection threshold on the stepCounter algorithm. Best score is 0.62 which is found at minimum detection threshold = 10 and minimum transition threshold = 15. 34
6.4 Heatmap of the effect of the step and detection threshold on the stepCounter algorithm. Best score is 0.62 which is found at minimum detection threshold = 10 and minimum transition threshold = 15. 34
6.5 Plot of possible combinations of a and b to satisfy $a640^b = 40\pi$. Based on this plot a range of a = (0, 0.2) and b = (0.5,2.0) was chosen for the iteration. 35
6.6 Heatmap of the effect of the step and detection threshold on the TP score of the chirpNoBoost algorithm. Best score is 85 which is found at minimum detection threshold = 5 and maximum error threshold = 9. 35
6.7 Heatmap of the effect of the step and detection threshold on the TN score of the chirpNoBoost algorithm. Best score is 100 which is found at minimum detection threshold = 10 and maximum error threshold = 1. 35
6.8 Heatmap of the effect of the step and detection threshold on the total score of the chirpNoBoost algorithm. Best score is 0.525 which is found at minimum detection threshold = 18 and maximum error threshold = 16. The corresponding values at that point are TP = 0.09 and TN = 0.96 35
6.9 Heatmap of the effect of the step and detection threshold on the CP score of the chirpNoBoost algorithm. Best score is 85 which is found at minimum detection threshold = 5 and minimum transition threshold = 9. 37
6.10 Heatmap of the effect of the step and detection threshold on the CN score of the chirpNoBoost algorithm. Best score is 100 which is found at minimum detection threshold = 10 and minimum transition threshold = 1. 37
6.11 Heatmap of the effect of the step and detection threshold on the total score of the chirpNoBoost algorithm. Best score is 0.525 which is found at minimum detection threshold = 5 and maximum error threshold = 16. The corresponding values at that point are CP = 0.09 and CN = 0.96 37
6.12 Example of a chirpfitting with a very large amount of periods, erroneously determined to be the best fit. 37
6.13 Example of a bad fit being classified as good, due to bad spacing between each period. 38

7.1 Precision recall plot for Step-1 algorithm. 41
7.2 Precision recall plot for algorithm Lee [9]. 41

1 Introduction

Space exploration knows many faces; it ranges from being physically at a location to retrieve samples, to looking at images from an orbiter, to piecing small chunks of information together about a planetary-like body you did not know existed but that should exist based on the data you have found. Now, the most concrete form of getting data, by physically visiting extraterrestrial bodies to do research, is often not feasible. Landing a piece of equipment in such a manner that it is still able to do research after landing is extremely difficult [13]. A more feasible solution is then to use satellites with imaging capabilities to form a picture of the world under investigation.

These images need to be scanned for interesting parts. A computer vision algorithm that is able to detect certain structures is one of the ways the data can be filtered. However, gathering knowledge is not the only criteria for space exploration. The other criterion is a necessary evil, keeping to a budget. Space travel is expensive. For instance, getting a payload to Low Earth Orbit from Europe costs around 10000€per kg [14], although SpaceX is working hard on bringing this number down[15]. Because of these large costs and tight budget, development for certain technologies becomes unfeasible. A computer vision algorithm that needs to detect the one feature of interest in thousands upon thousands of images will be very complex and therefore costly to develop and maintain.

An often seen solution to this problem is the practice of spin-in [16]. In this practice, an already existing technique for terrestrial application is adapted for usage in extra-terrestrial situations. This drastically reduces the costs required to get the TIL-level to the required level. This could be an interesting route for object detection. By maturing the technology on Earth, it can afterwards be implemented on space exploration missions for a fraction of the cost. This is of course only possible if there is a business-case on Earth for the technology, to propel the development. A case for the further development of object detection will therefore be made in the next paragraph. The plan is to start of with detection of simple structures, which can gradually be expanded to increasingly more complex objects and structures.

Case for object detection The aging of the population in Western countries is a serious problem. This problem finds its roots at the Second World War. More specifically, at the end of the Second World War. Due to the euphoria and relief of the war being over, our ancestors could not contain their joy, amongst other things. This resulted in a significant spike in birthrate for several years, see Fig. 1.1. Since the birthrate returned to a more normal rate after some years, there is now a generation (called the babyboomers) that is significantly larger than other generations, and it is slowly moving through life. Since the babyboom is more than 60 years ago, these large group of babies have now turned into a large group of elderly people, see Fig. 1.2 and 1.3.

These elderly people are now in need of care, and this will only increase as in the following years their age group shifts slowly towards 70-80 years. Because of this, there is a lot of pressure on the medical sector, who have too many patients and not enough people in the work-force to take good care of every patient or give them the attention, medical and social, that they require [17].

With not enough people in the medical work-force to treat the ill, a logical idea would be to look in the direction of prevention. It has been shown that keeping up a healthy diet and exercising regularly, both mentally and physically, aid in the process of staying healthy longer [18]. Indeed for an elderly person it is not often easy to keep doing gymnastics up to their retirement age, nor is that necessary. It is shown that only small amounts of exercise done during a normal day can show significant increases for a seniors health [19]. One of those activities is taking the stairs.

Unfortunately, at some point this becomes hard to combine with their everyday activity of taking a long walk towards the local supermarket and buying their groceries. Because they can with some effort climb the stairs, but it becomes very difficult when also lugging around a big bag of groceries. They are fine climbing stairs or carrying a bag of groceries, but not both at the same time. This is these days solved with elevators, stair-lifts or even worse, delivery of groceries at home. This robs the elderly person of a good daily exercise in the first two scenarios, but in the third scenario it also robs him of the social interaction with the outside world. Something that elderly people these days are in desperate need of [20].

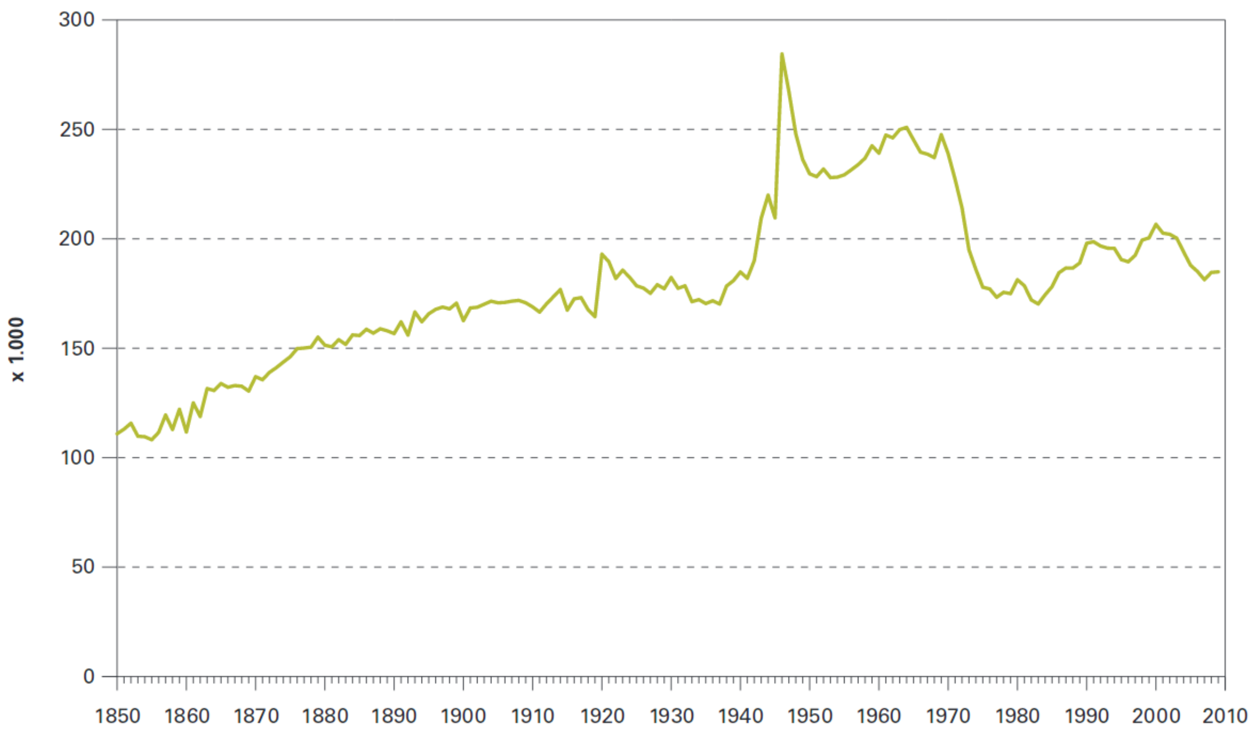


Figure 1.1: Birthrate of people in the Netherlands between 1850 and 2010 [1].

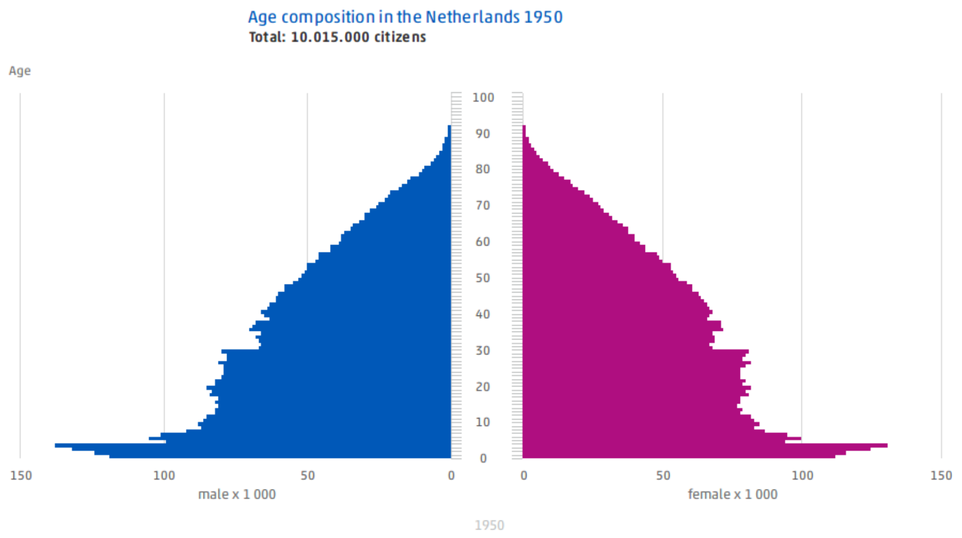


Figure 1.2: Age distribution of the population in the Netherlands in 1950 [2].

1.1 Thesis topic

This thesis was done under the supervision of DoBots. Dobots is a robotics company in Rotterdam that has experience in working with all sizes of robots. They worked with large robots in the form of a robot vaccuum cleaner, and with robot swarms like FireSwarms, a project they did in cooperation with

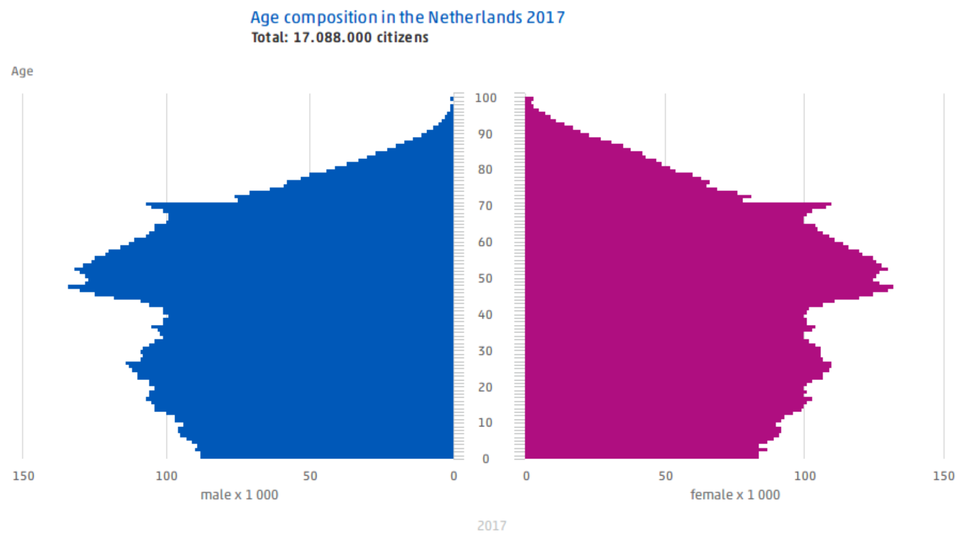


Figure 1.3: Age distribution of the population in the Netherlands in 2017 [2].

(amongst others) TU Delft. They are currently working on making smart homes an affordable reality, visit http://crownstone.rocks for more information. What DoBots proposes is a robot to help with the act of climbing stairs and carrying groceries. In an internship under DoBots a prototype of a stair-climbing robot was built by Frerik Andriessen[21]. The robot, built out of technical Lego, managed to successfully climb parts of the Groothandelstrap. There was an article about it on AD.nl which can be read in Appendix B. Meanwhile, another intern, Reka Hajnovic, is working on a robot that will carry the groceries for the elderly. A very nice touch of her robot is that it also allows the elderly person to use it for support. Combining that robot with stair-climbing capabilities will aid massively for elderly people that just need a bit of support. One of the things a stair-climbing robot needs is the ability to sense if there are stairs or not, and thus if it should climb the staircase or not. That is what this thesis aims to provide: a starting point for staircase navigation, in the form of a stair-detection algorithm.

The algorithm should be presented with a depth image from one of its own camera modules, and the algorithm should then decide that it sees a staircase or not, and send a signal accordingly.

1.2 Research questions

Developing a stair-detection algorithm is a long and difficult process. Carefully crafted research questions aid in keeping the right focus and give something to verify you are going in the right direction. For the stair-detection algorithm these are as follows:

- What are the recurring physical traits for a staircase?
- How can these traits be recognized?
- How can these traits be combined to accurately classify a staircase?
- Can an algorithm made up of simple rules accurately detect a staircase?

1.2.1 Algorithm requirements

To be able to verify if the algorithm design meets the intended performance, it is common practice to formulate requirements that steer the design of the algorithm. The following requirements were formulated for this algorithm:

1. It should consist of simple rules, based on the stairs physical traits

2. It should have at least comparable performance in accuracy as other stair-detecting algorithms
3. It should be able to perform in low-light conditions
4. Its performance should not be dependent on the angle at which the images are supplied

In a real-life application there is not always control over the lighting conditions. The algorithm might have to perform in very dark environments. Similarly, the angle at which the stairs might be encountered could vary widely, a robot has no use for stair-detection if it needs to stand in one very specific spot or otherwise it will not work. Therefor the algorithm should perform regardless of which angle the staircase image is supplied. The simple rules based on physical traits are discussed in more detail in Chapter 3. The comparison with the current state of the art stair-detecting algorithms is found in Chapter 7.

1.3 Current state of computer vision in space exploration

The current state of using computer vision in space is somewhat lacking behind as compared to terrestrial applications. Most of the computer vision developments are being done for autonomous rovers, either concerning landing, like the MER-DIMES (Descent Image Motion Estimation System) [22] or VAIN (Vison Aided Inertial Navigation) [23], or terrain mapping and navigating, like SPARTAN [24]. Another computer vision project being developed by NASA is VERTIGO on the SPHERES platform [25]. SPHERES is a testing platform consisting of 3 dice-like satellites, VERTIGO is an algorithm that uses SURF (Speeded Up Robust Features) and SLAM (Simultaneous Localization And Mapping) to create a 3D-map of its environment. Computer vision is often used in combination with machine learning, because even seemingly simple tasks for a human require a lot of intelligence from the computer vision system. In its paper about Self Supervised Learning (SSL) for potential space exploration robots, G. de Croon states that the current state of machine learning in the space sector is lacking behind because there is so little room for error on the missions [26], while machine learning has certain inherent risks. That is, learning is often done by making mistakes and receiving feedback on the results. In a situation where a tiny mistake can cause a mission-failure it is understandable that the space sector is very careful in approaching the topic of machine learning. Using machine learning and computer vision on satellite images has virtually none of these risks. However, it is mostly applied on terrestrial applications. This is probably due to the larger economic incentive for terrestrial applications combined with the naturally much larger datasets available for Earth as compared to other planets, and the so-far slow adaption of machine learning in the space sector.

1.4 Outline of thesis

1.4.1 Methodology

The following methodology will be adapted:

- Think about how we as humans are able to determine something to be a staircase. Do not just focus on the most obvious ways, but also look into subtle cues that subconsciously alert us that something might be a staircase.
- Find and read papers on staircase detection, and what features they use as indications of a staircase. See how they correspond with own ideas, what did they not address; what did they address that was missed in the previous step?
- Synthesize papers with own ideas into a set of rules/features to detect a staircase.
- Read papers to see what is currently used to collect data on the aforementioned features.
- Read papers on combining the set of rules and features into a tool for staircase or object detection.
- Develop classification algorithm based on these rules and features.
- Combine datasets from earlier mentioned papers, and other sources into a big test data-set.

- Implement algorithm in a programming language and test on data-set.
- Compare results with papers, most importantly compare the FP/TN and FN/TP rate. What was different in the results and what might have caused this.
- Update algorithm and compare again, until good results have been met, or a reasonable amount of iterations have taken place.

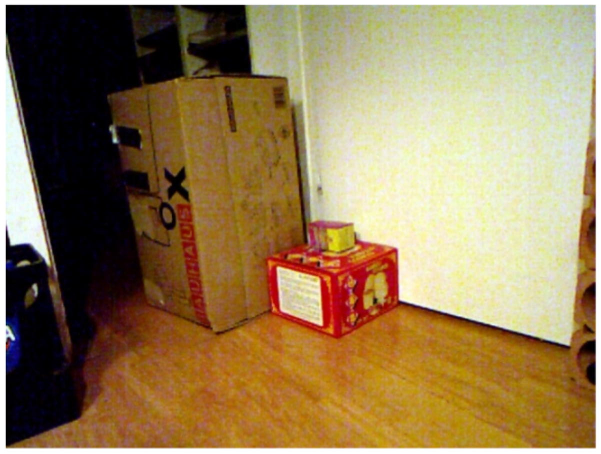
In chapter 2 the current state of the art in object and stair-case recognition is discussed. It shows which methods are seldom used and which are overly used and why. Chapter 3 will walk through the preprocessing algorithm step by step, and show the effects on an example RGB-D input image. It starts off with using the depth-info of the RGB-D image to create surfaces. It then differentiates between the surfaces and segments them. Chapter 5 then continues with the detection algorithm. First 20 random line coordinate pairs are chosen, and used to draw lines through the image. The underlying pixel values on these lines are stored in a vector, and checked for changes in pixel value (”transitions”). A chirp-signal is fitted to these transitions. Finally the results of all lines are weighted and summed using the AdaBoosting method. Chapter 7 shows the results of the algorithm on several datasets. Finally chapter 8 gives a short summary, draws conclusions, and gives several recommendations for future work.

2 Literature review

This chapter discusses the state of the art of staircase detection. It also reviews techniques relevant for staircase detection, like surface segmentation and the detection of objects in general. Section 2.1 discusses the method of surface segmentation, which allows uniform sections to be extracted from images. Section 2.2 discusses object detection using simple features, where Sect. 2.2.1 in particular discusses the current state of stair-detection.

2.1 Surface segmentation

Stairs consists mostly of a collection of flat surfaces (steps, risers, walls). A core concept of detecting a staircase therefor consists of first distilling the different surfaces from the depth image. These manipulations need to be done in real-time for a robot to have practical usage. The algorithm of S. Holz et al. allows for manipulation while dealing with images on a rate of 30Hz [3]. It allows detection of objects and obstacles while also grasping the geometry of the scene. Surfaces are detected using surface normals. There are two ways of doing so. One way is to fit a plane through the spatial neighbourhoods of a 3D point cloud, using either k-neighbours or a radius of r to delimit the neighbourhood, see Fig. 2.1. The second, computationally cheaper, way is to take the pixel structure of the input image and use those as pixel-based neighbourhood. The surface normal is obtained by calculating two tangential vectors using the left and right neighbour and top and bottom neighbour for the x and y vector respectively. If there is a large amount of noise (small spots that do not correspond to the rest of the surface) in the data, an average of multiple neighbours is used. Both techniques have to deal with blurring/smoothing when the neighbourhood is chosen too large, while they have to deal with the significant effect of noise in the depth image input when the neighbourhood is chosen to be too small. However, this effect is lesser for the pixel-based neighbourhood [3]. Objects and obstacles can be detected based on the fact that these should be situated on a groundplane or horizontal surface like a table. This simple algorithm had a 93% detection rate for objects, and 100% detection rate for obstacles.


(a) Input camera image

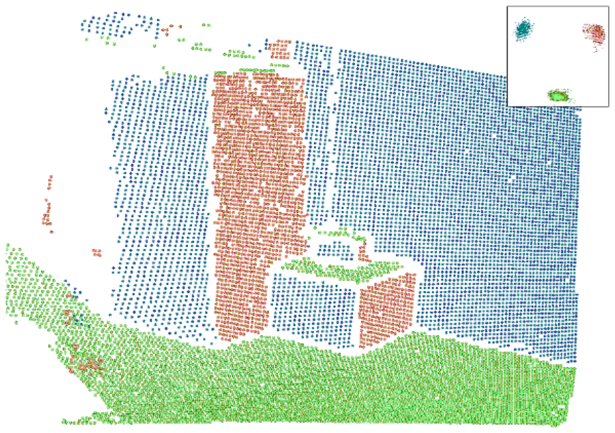

(b) Segmented cloud

Figure 2.1: Segmented surfaces in the point cloud using plane-fitting through spatial neighbourhoods of a 3d point cloud by Holz et al. [3]. The different orientations of the surfaces can be easily distinguished based on their color.

2.2 Object detection

When detecting objects it is very difficult using just one or two characteristic features to accurately determine whether or not it is the object you are looking for. Sometimes the lighting is off or the angle is such that an important trait is covered up. To deal with these problems and get a more robust detection algorithm, P. Viola and M. Jones [4] used an Adaboosted set of Haar-like features to detect faces, see Fig. 2.2. These features are called Haar-like because of their resemblance to Haar Wavelets, as can be seen in Fig. 2.3. These Haar Wavelets are used, amongst other applications, in image compression [27]. These Haar-like features are combined into a cascaded classifier of 38 stages and applied to 24x24 resolution images to detect whether or not a face was shown in the picture. The structuring of the classifier level was chosen in such a way that a large number of false images were discarded in the first levels, speeding up the process drastically. With each added level, the false positive rate decreases, while the detection rate decreases as well. While the goal is to have the false positive rate drop below a certain threshold, the detection rate should stay above a certain threshold. Therefore, it will add levels until the false positive rate has dropped to the required value, while the detection rate is still at an acceptable value.

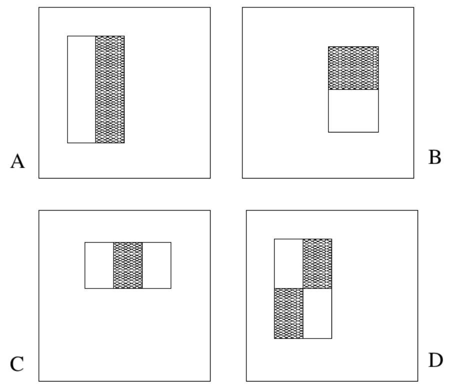


Figure 2.2: Examples of the Haar-like features used by Viola and Jones [4].

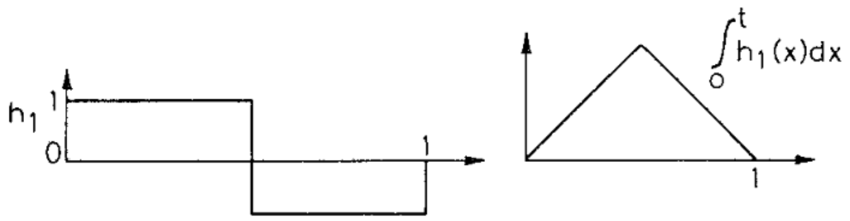


Figure 2.3: Example of Haar Wavelet [5]. Note the resemblance to the Haar-like features from Fig. 2.2.

2.2.1 Stair-detection

Although there have been many object and scene detection algorithms studied before, staircases are often not in the standard lists of everyday objects or scenes. See for instance the Microsoft COCO database[28]. While incredibly extensive, stairs are not included. Therefore, many research papers focusing on that dataset do not work on detecting staircases, and not many papers have dealt outright with detecting staircases. This is possibly because a staircase does not really have a distinctive shape as far as objects go, and is a bit of a mix between an object and an environment. That said, there have been some papers that looked into stairdetection.

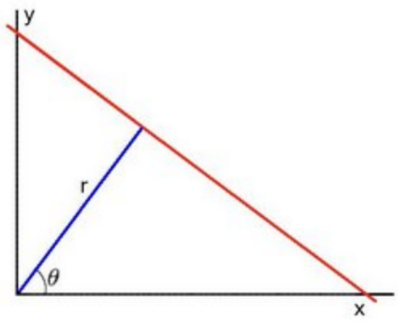


Figure 2.4: Parametric representation of line in (x,y) space [6].

S. Wang et al. worked on both stair-detection and crosswalk-detection using a combination of normal RGB images and depth images [8]. First, they used a Sobel edge detector[29] to retreive all the edges in the image. A sobel edge detector works on the fact that a relative large difference in pixel intensity often accompanies an edge. It takes an approximate derivative and searches for local maxima; this is implemented using two kernels in the following form (as an example, using a kernel of 3), which are afterwards convolved with the image:

$$G_x = \begin{pmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{pmatrix}, G_y = \begin{pmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{pmatrix}, G = \sqrt{G_x^2 + G_y^2} \tag{2.1}$$

This returned an image with a collection of pixels where an edge was detected. They then converted the loose pixels into lines using the Hough Transform. The Hough transform takes a point that was determined to be an edge, and fits a line through it. This line could be represented using $y = ax + b$, but there is a major downside to this representation. The value of the slope a goes to infinity as the line becomes vertical. This makes it impractical for storing and plotting of the (a,b) space. That is why these lines are converted to polar coordinates, using

$$y = \left(-\frac{\cos\theta}{\sin\theta}\right)x + \left(\frac{r}{\sin\theta}\right) \tag{2.2}$$

where r is the length of the perpendicular line from the origin to the line being converted to polar coordinates, see Fig. 2.4, and θ is the angle in degrees, which can range from (0-180).

This equation can be rearranged to

$$r = y\sin\theta + x\cos\theta \tag{2.3}$$

Now, for every edge-point (x,y) there are many combinations of (r,θ), and thus many lines, that go through this point. Plotting these combinations of r and θ would give a sinusoid, where each point on the sinusoid would represent a line that goes through the edge-point. Doing this for all edge-points in the image gives a different sinusoid for each point. The points where some of these sinusoids intersect signify the lines that go through multiple edge-points. The more sinusoids intersect, the more points can be found on that line. The Hough Transform has a certain threshold of these intersections. If the amount of intersections is above the threshold it will draw the line through these points. Of course, the lower the threshold, the more lines are detected, but it also increases the amount of noise and bad line-fits.

Wang et al. filtered out all lines that were not at least somewhat horizontal. They used the same representation for their lines as was used in the Hough Transform, e.g. $r = y\sin\theta + x\cos\theta$, where the origin is the starting point of the line. Now, the lines that are found need to be parallel, which is the case of their θ is equal. If this is the case the algorithm checks if the lines have a certain minimum length, and if there is a large enough amount of individual lines. In their case they chose a length of 60 pixels and a minimum of 5 lines. If the image scores low on either one of them it is discarded as a negative image. Otherwise, it is classified as either a staircase or a crosswalk. Once this is determined, it will check its RGB-D counterpart and they will create a depth-graph by drawing a line through the stairs/crosswalk and collecting the values along this line, see Fig. 2.6. They then use a hierarchical Support Vector Machine based Radial Basis Function to classify either a crosswalk, an ascending staircase or a descending staircase [7]. This Radial

Basis Function uses the training data to demarcate areas for each classification group, see Fig. 2.5. The area that every new data point falls in, is then designated as the data-point's classification group. Note that the example in the figure is 2-D for visual purposes but the RBF supports higher dimensionality.

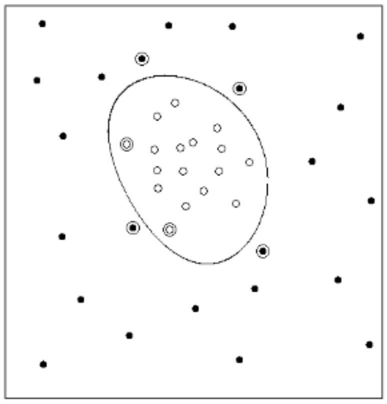


Figure 2.5: Example of demarcation done by a Radial Basis Function[7]. Any new datapoint falling within the circle will be designated white, while any point falling outside the circle will be designated black. Note that higher dimensions are possible.

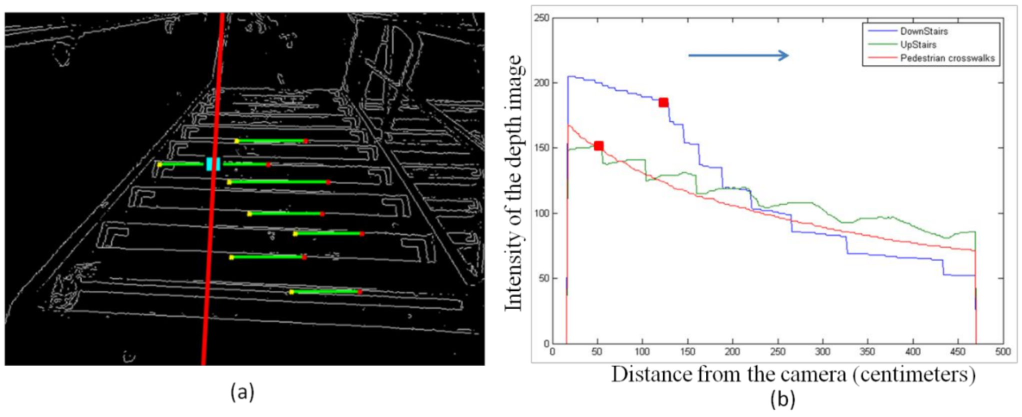


Figure 2.6: Example of depth-graph of crosswalk and staircases from Wang et al.[8]. The crosswalk decreases in intensity as the line moves further away from the camera. The upstairs and downstairs graphs also show clear characteristics. The downstairs graph is more steep and has straight corners, while the upstairs graph slowly rotates as the steps move up.

Using a training data-set of 60 stair images, 30 crosswalk images and 30 negative images, they tested on a test data-set of 106 staircase images, 70 negative images and 52 crosswalk images. A True Positive rate of 97.2% and a True Negative rate of 100% was obtained.

Lee et al. devised a portable stair-detection vest for the visually impaired [9]. It uses a combination of Haar-like features with adaBoosting, ground plane measurements and temporal consistency, see Fig. 2.7. For the first part of the algorithm, the classifier is trained with Discrete AdaBoosting for 18 rounds on 210 positive images with 25 distorted variants each and 7000 negative images all with a resolution of 40x40. At the end of this training, the True Positive rate on the training data-set was 0.983. The classifier scans the image and divides it up in 40x40 subwindows. To weed out False Positives, multiple detection subwindows need to overlap for the classifier to consider it. Using stereo vision it determines the groundplane and gives a weight to detected staircases. As the staircase is closer to the groundplane it gets a higher weight and vice versa. Finally, the affinity between two consecutive frames with detection needs to be above a certain threshold for it to be ultimately determined to be a staircase. It is shown that a combination of the three

(Cascaded Classifier, Ground Plane, Temporal Consistency) has the best result, but also that Ground Plane has very little effect on the precision of the algorithm. They attribute this sub-par performance to the not-yet matured ground-plane detector technique. They are able to detect stairs under a varying number of conditions, like differences in sunlight and angle. This gives them an advantage over the technique used by S. Wang et al [8] that is only able to detect stairs when directly facing them.

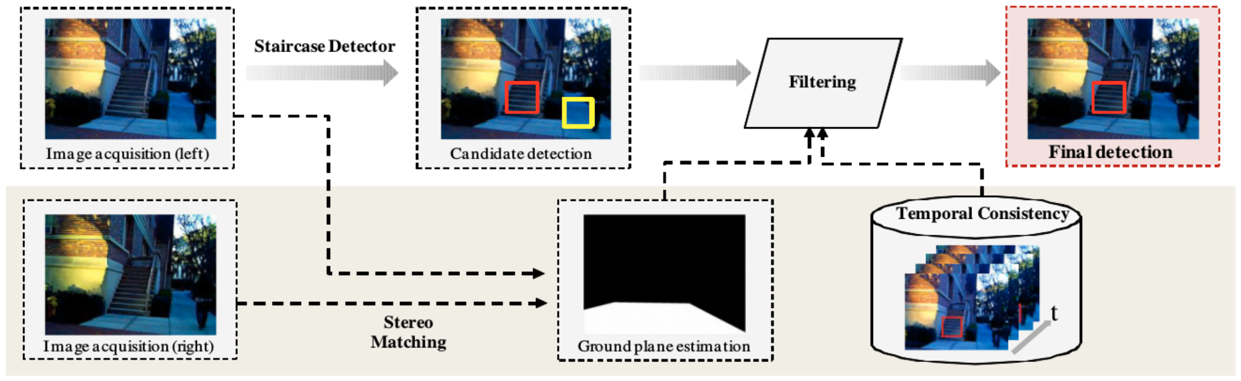


Figure 2.7: Process of staircase detection by Lee et al. [9]

Stair-navigation

Y. Xiong & L. Matthies used low-threshold edge detection to determine their relative position and attitude on the stairs. [10]. Due to its low threshold it will register a lot of edges that are not staircase edges. They incorporated the following rules to filter the large amount of edges obtained:

- Stairs edges are straight lines
- When looking straight at a staircase, the edges are relatively horizontal and mostly parallel to each other
- If there are several small edge segments that lie on the same line, they can be merged to one
- Stair edges in general, are long

Using these rules the non-relevant edges obtained with the low-threshold technique get filtered out, namely by filtering out non-straight edges, filtering out edges whose orientation is too different from the dominant orientation, linking together edges that are close and parallel, and filtering out edges that are small, since staircase edges are long. The results can be seen in Fig. 2.8. This was then used by them to determine the robots center position and heading, but it could also be used as input for a stair-detection algorithm.

2.3 Conclusion

Object detection and scene recognition is a widely-studied subject that is well-represented in the academic world. However, the subset of stair-detection is significantly less studied. The papers that addressed stair-detection as a goal both used a combination of a simple classification algorithm and machine learning to train the classifier. They have relied on Haar-like features or parallel lines for their simple classifier. They do not take advantage of the converging repeatability of staircases which seems to be an interesting way to make the classifications more robust. The reason that this could improve the classifications is that the current algorithms do not look at where their regions of interest are situated with respect to each other. In the case of parallel lines, this could mean that randomly stacked boxes will create enough parallel lines to be classified as a staircase, even though a smarter algorithm could see that the distance between the edges is not consistent to what you would expect in a staircase.

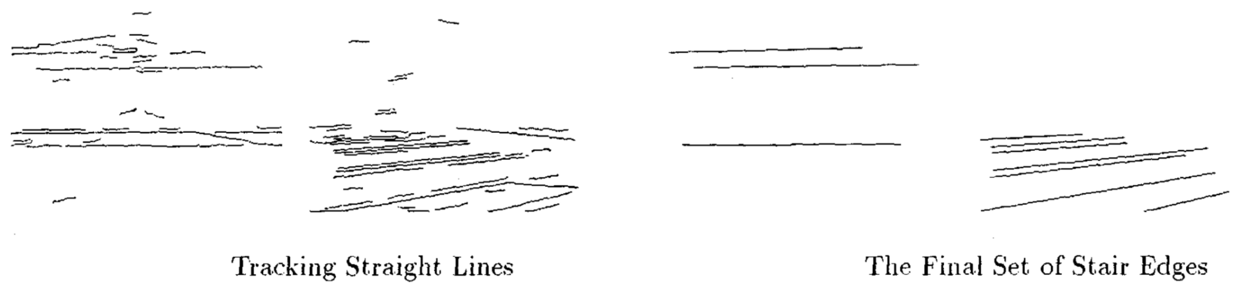


Figure 2.8: Start and finish of filtering stair line-segments [10]

Therefore the aim of this thesis is to develop an algorithm based on the relative positions of these regions of interest. The goal is to improve on the current performance and expand the options that are currently available for stair-detection.

3 Approach

As discussed in the literature review in chapter 2, the aim of this thesis is to develop an algorithm for staircase detection that takes the relative spacing of the regions of interest into account.

3.1 Algorithm

For staircases, there is one main type of feature that can be identified as a region of interest for detection, which is the steps. This could be the whole step itself, or the edges of the step. The edges of the step are more easily detected since it is just a sudden transition in value between two adjacent pixels, so it makes sense to focus on these. For any normal staircase, these edges are equidistantly spaced from an absolute point of view. Thus, the spacing between these edges is internally consistent for every staircase, which is a feature that can be relied on for classifying the staircase. However, looking at the staircase from any realistic position, the spacing of the edges will be distorted in such a way that steps that are closer seem to be larger and more distant from each other, than steps that are further away, see Fig. 5.7. This follows a phenomenon called a chirp, which is a waveform of which the frequency increases as time increases. This means the peaks of the waveform are getting spaced closer to each other as the running variable is increased, just as can be observed when looking up at an ascending staircase. See Sect. 5.2 for a more in-depth discussion on chirps. If we could show that the image follows a chirp by mapping the edges of the steps to the chirp, it would follow that the transitions are spaced in a consistent way, corresponding to how a staircase would appear. This would rule out randomly stacked boxes or random transitions scattered over the image, and thus be an improvement on the current methods of stair detection. The foundation of the chirp research did not use a chirp for classification, but rather it found images with structures following a chirp in it, as classified by human eye, and de-chirped them, leaving an undistorted view of the structure[12]. That said, using a chirp for classification is a not yet documented approach that would expand on the currently available methods of classification, with stair-detection in particular. It is therefore chosen as the core of the stair-detection algorithm in this thesis.

The step edges, or transitions, can be collected by drawing a line over the staircase and storing every transition that is encountered. Because the staircase could be looked at from different angles, it is not clear what linepath would be the most effective for traveling over enough transitions to fit this chirp. A solution for this is to draw many random lines and have their effectiveness on classifying images determine their worth, or weight. This technique is called boosting and was used successfully in [9].

3.1.1 Dataset

The input for this algorithm will be depth images, since they represent the physical distance to the camera for each individual pixel. Because of this it can clearly communicate edges, represented by a sudden change in depth value, without being reliant on good lighting. These depth images need to be processed to make them usable for the detection algorithm. The main point is that it needs to processed in such a way that the different surfaces are distinguished from each other, such that transitions can be found. There will be a clear difference in surface orientation between the steps and their adjacent risers, and therefor a representation of the surface orientation will be extracted from the depth images to be used as input for the detection algorithm, to be called Step-1 henceforth. This dataset consists of 100 16-bit depth images of ascending stairs, and 100 16-bit depth images of non-stair scenery. All staircases in this dataset are straight staircases (e.g. no spiral staircases or stairs with a 90 degrees angle after the first few steps) under a variety of angles, and all the stairs are ascending. This is because the generally used POV for descending staircases are such that only the horizontal steps are shown. This means there will only be a small line in which the surface orientation of the pixels is different. After removing the noise in each image this would then show all steps to melted together into one large segment. Therefore the data was chosen such that both the horizontal steps as the vertical risers are clearly visible; this leads to the images being a front view of ascending staircases.

3.1.2 Tools used

For the image processing software there were several libraries to choose from, the main two contenders being OpenCV and PCL (Point Cloud Library). The main difference between both libraries is that OpenCV works in 2D and PCL works with voxels in 3D. Manipulating and visualizing data in 3D is in general more complex and demanding of the hardware than 2D data manipulation. Therefore, because there was no clear advantage of choosing PCL over OpenCV, it was decided to use the C++ libraries of OpenCV[1]. There are also Python bindings for OpenCV, which is normally my language of choice. However, because it became clear quite soon that the C++ has been maintained more extensively, it was decided to use C++ for all image processing. The graphs in this report have been made using the Matplotlib library of Python.

3.1.3 Flow diagram of proposed algorithm

The final flow of the algorithm can be seen in Fig. 3.1. The color coding shows what techniques were already there, and what was novel contribution.

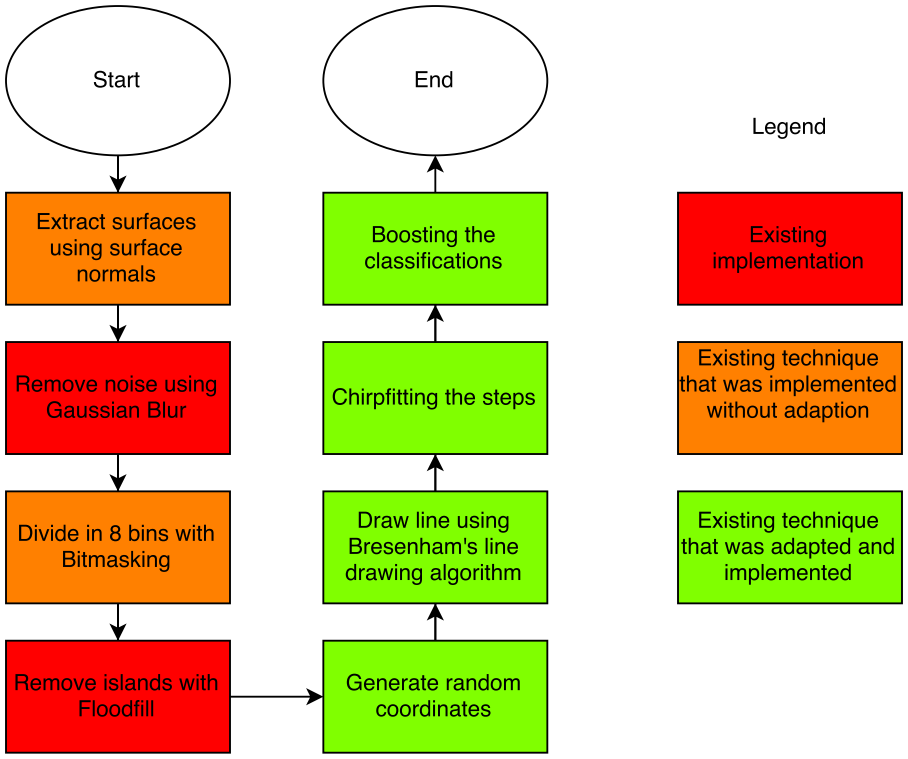

Start
End
Legend
Extract surfaces using surface normals
Boosting the classifications
Existing implementation
Remove noise using Gaussian Blur
Chirpfitting the steps
Existing technique that was implemented without adaption
Divide in 8 bins with Bitmasking
Draw line using Bresenham's line drawing algorithm
Existing technique that was adapted and implemented
Remove islands with Floodfill
Generate random coordinates

Figure 3.1: The main steps of the algorithm. In color coding is differentiated what was already existing, and what was novel contribution

It shows there are two main parts for the algorithm: dataset pre-processing and actual stair classification. The dataset preprocessing will be discussed in detail in Chapter 4, while the classification algorithm is discussed in Chapter 5.

[1] An important note: OpenCV uses BGR values instead of the commonly used RGB values; i.e. the reversed order. This has no effect on the calculations but when referring to, for instance, the blue channel, we are talking about the first element of the vector, followed by the green channel, and then the red. Similarly, pixel coordinates are given in the form of (y,x) instead of the normally seen (x,y) format.

4 Stair-case RGB-D preprocessing

The stair-detection algorithm consists of two parts. The first part consists of processing the input-images to prepare them for the stair-detection methods in the second part. An overview of the first part can be seen in Fig. 4.1. First, a depth image is loaded. Then, the surface normals are calculated using the depth gradients in Sect. 4.1. To remove most of the noise a Gaussian Blur is applied to the images in Subsect. 4.2. In Sect. 4.3.1 the different surface orientations are divided into 8 bins based on binary addition. Finally, in Subsect. 4.4 the small specks of different orientations are smoothed out using a Floodfill algorithm, that recolors sections that have an area below a certain threshold. This is all discussed in more detail in the following sections.

4.1 Extracting surfaces

With the depth information provided by the depth images, it is possible to see the image as quasi-3D, using the pixel values to visualize the position on the z-axis. The image that is loaded up in the stair-detection program is the depth image, using grayscale intensities to represent the depth of the objects in the scene. An example of the loaded input can be seen in Fig. 4.2.

Every pixel has a depth value based on its distance to the camera. We cannot directly extract steps from this information. To create coherent surfaces, we rely on the surface normals of the image, as seen in [3]. The surface normal shows us in which direction the pixel value is headed if one were to follow its neighbouring pixels. The idea is that for flat surfaces the depth-value in a certain direction changes the same amount for each pixel in the surface following that direction. Thus, the surface normal is consistent throughout the surface.

The gradient of each pixel is calculated based on its direct horizontal and vertical neighbours. Taking the cross product of the derivatives in x and y direction gives us the normal vector, see Eq. 4.1. Taking p as a vector with x, y and z coordinates as vector elements, so $p = [x, y, z]^T$, allows us to easily relate and manipulate nearby pixel values. Here x and y represent the coordinates of the pixel and z is the intensity value.

$$n = \frac{\partial p}{\partial x} \times \frac{\partial p}{\partial y} \tag{4.1}$$

where $\frac{\partial p}{\partial x} = [1, 0, \frac{\partial z}{\partial x}]$ and $\frac{\partial p}{\partial y} = [0, 1, \frac{\partial z}{\partial y}]$. In our example $\frac{\partial z}{\partial x}$ is calculated taking the difference between the depth value at (x-1) and (x+1) and dividing by 2. More neighbours could be incorporated in both directions, which would give a smoothing effect as the number of neighbours are increased. However, smoothing is not integrated in this step and rather done in the next step, to have both steps isolated from each other. This allows the effect of both surface normals and smoothing to be evaluated individually. The smoothing is discussed in Subsect. 4.2. The same principle for the surface normals is applied to $\frac{\partial z}{\partial y}$. The surface normal following from this calculation is a vector with 3 elements. These are stored in a new image container using the RGB channels. This leads to the following output, see Fig. 4.3.

As can be seen, there is quite some noise in the resulting image, as visualized by the lines and specks scattered throughout the different surfaces. This stems from the noise in the input image, where a different value means a different depth, which translates to a wildly different surface normal at those neighbourhoods. Not only that, there will also be round-off errors. Say that for each pixel moved in y-direction, z increases with 0.8, thus $\frac{\partial z}{\partial y} = 0.9$. The pixel values only work with integer values from 0-255, any decimal value is truncated. So 0.8 turns into 0. The next pixel has a value of 1.6, thus it is 1. Next has a value of 2.4, so it becomes 2. 3.2 turns into 3 and 4.0 is actually 4. Then, the next value is 4.8, but due to the truncation this is also turned into 4. The previous pixel transitions all showed $\frac{\partial z}{\partial y} = 1$ but this pixel transition is now shown as $\frac{\partial z}{\partial y} = 0$ even though the pixels are uniformly increasing in value. This shows up in the image clearly at the vertical part of the steps, the risers, where the normal value of red is sometimes interjected by a black line, which is the point where the truncation caused two succeeding pixels to be the same value.

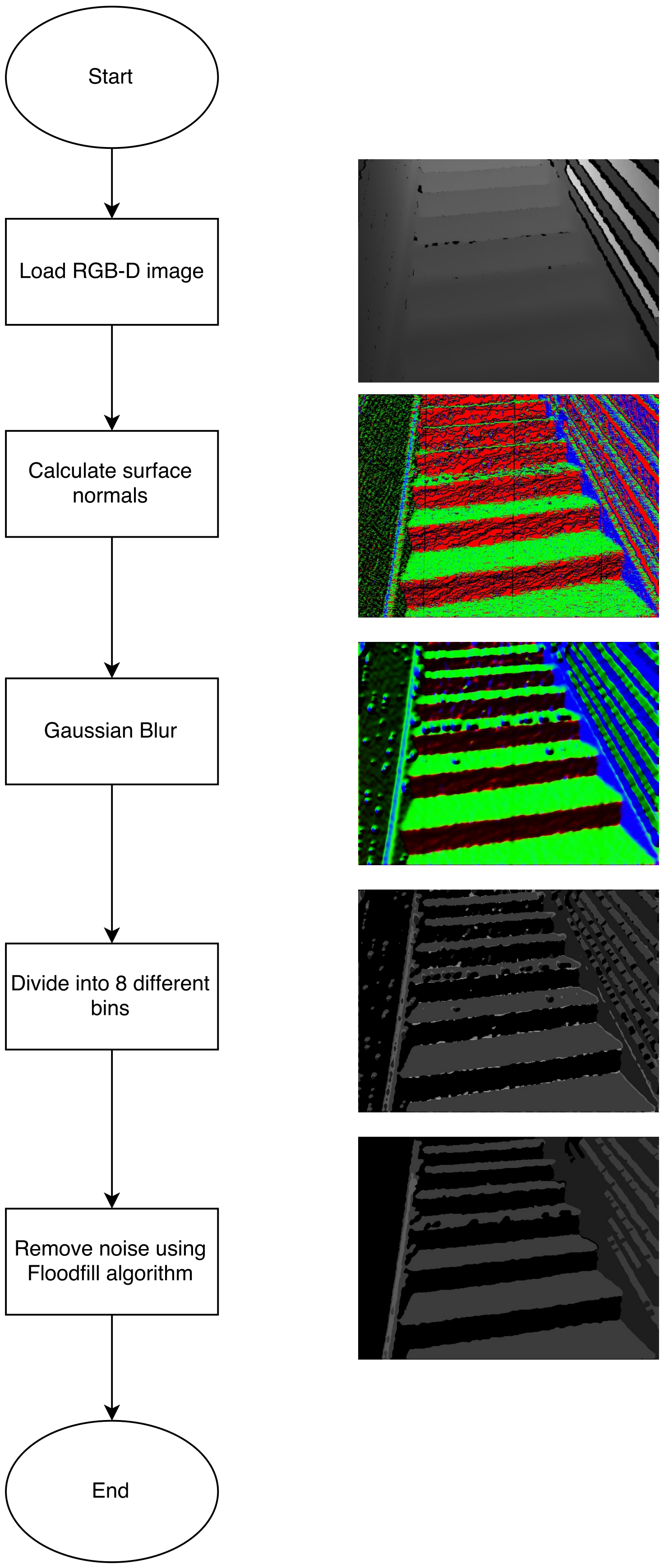

Start
Load RGB-D image
Calculate surface normals
Gaussian Blur
Divide into 8 different bins
Remove noise using Floodfill algorithm
End

Figure 4.1: The steps taken in pre-processing part. Outputs after each step can be seen next to the step. The transformation of a noisy multi-channel image to a smooth one-channel image should be clear.

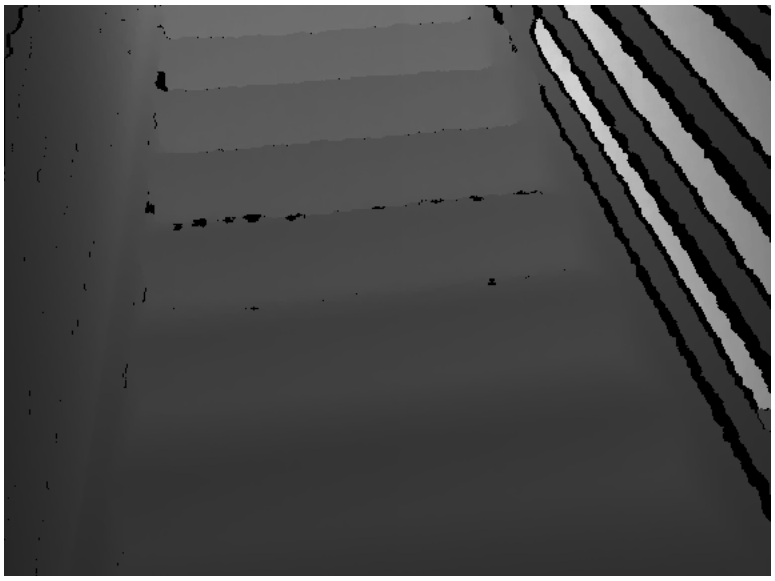


Figure 4.2: Loaded raw input, before doing any processing on it. Already some noise can be seen at the edges of some steps.

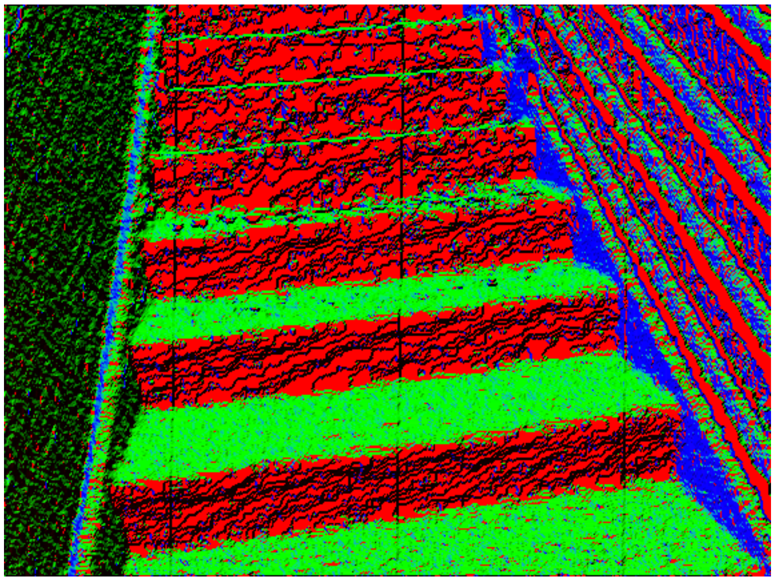


Figure 4.3: Resulting image after surface normals were calculated for depth image. There is noise due to the discrete (values of 0-255 in steps of 1) nature of the pixel values as well as the lesser quality of the input data.

4.2 Gaussian Blur

The noise in Fig. 4.3 needs to be removed, because the Step-1 algorithm is based on transitions, indicated by two adjacent pixels with a large difference in pixel value, and therefore very susceptible to noise. There are many ways to remove the noise, an often used technique is the Gaussian Blur. The reason a Gaussian Blur is suited for noise removal in this scenario is because it has a smoothing effect, averaging lone pixels with their neighbours [29]. Averaging has a stronger effect on lone pixels than it does on edges, because the edges have more pixels of a similar value in their neighbourhood. So although the edges will be somewhat less apparent, the noise will be removed to a larger degree. That said, this technique is a double-edged sword since it removes unwanted noise, but it also smooths out edges, if applied over too many neighbours per pixel. Depending on the situation, this can merge distinct sections that should not be merged. It should therefore be used sparingly. The weights of the pixels used to average the pixel that is the target of the Gaussian Blur follow a Gaussian distribution. This means the center pixels own value has the most weight, after which the neighbouring pixels have lower weight the further they are away from the center pixel, according to the 2D Gaussian distribution. The filter is implemented via

$$p^*(i,j) = \sum_{k,l} p(i+k, j+l)h(k,l) \tag{4.2}$$

where p^* is the new value of pixel p, and $h(k,l)$ is the kernel for the 2D Gaussian distribution. This kernel is a 2D matrix of [k x l] dimensions, storing the value of the weights for each cell in the kernel, according to Gaussian distribution. The kernel is calculated with

$$h(k,l) = A\exp\left(\frac{-(k-\mu_k)^2}{2\sigma_k^2} + \frac{-(l-\mu_l)^2}{2\sigma_l^2}\right) \tag{4.3}$$

where μ is the mean and σ is the standard deviation of the kernel. A is the amplitude, which is chosen such that $\sum\limits_{k,l} h(k,l) = 1$ to not add or remove any energy from the image. Now the sum of the 1-dimensional Gaussian distribution G(k) for these constraints can be calculated by taking its discrete integral:

$$\sum_k G(k) = \int_{-\infty}^{\infty} A\exp\left(\frac{-(k-\mu_k)^2}{2\sigma_k^2}\right) dx = 1 \tag{4.4}$$

Recall that for an arbitrary Gaussian integral

$$\int_{-\infty}^{\infty} \exp(-a(x+b)^2)dx = \sqrt{\frac{\pi}{a}} \tag{4.5}$$

which means that, substituting $a = \frac{1}{2\sigma_k^2}$ and adding constant A (which has no effect on the integral), the equation becomes:

$$\int_{-\infty}^{\infty} A\exp\left(\frac{-(k-\mu_k)^2}{2\sigma_k^2}\right) dx = A\sigma_k\sqrt{2\pi} = 1 \tag{4.6}$$

This equation can only be satisfied if $A = \frac{1}{\sigma_k\sqrt{2\pi}}$. So that then gives a 1D Gaussian of

$$G(k) = \frac{1}{\sigma_k\sqrt{2\pi}}\exp\left(\frac{-(k-\mu_k)^2}{2\sigma_k^2}\right) \tag{4.7}$$

Naturally, the same holds for the 1D Gaussian in the other direction, $G(l)$. To transform from a 1D Gaussian distribution to a 2D Gaussian distrbitution we simply multiply both 1D Gaussian with each other, which leads to the kernel $h(k,l)$

$$h(k,l) = G(k)G(l) = \frac{1}{2\pi\sigma_k\sigma_l}\exp\left(\frac{-(k-\mu_k)^2}{2\sigma_k^2} + \frac{-(l-\mu_l)^2}{2\sigma_l^2}\right) \tag{4.8}$$

Now, since we want the distribution to have the center pixel as the focus point, the mean is 0, thus $\mu_k = \mu_l = 0$. Furthermore, if we assume both standard deviations to be equal to each other, because the Gaussian blur should perform equally in both x and y direction, that gives $\sigma_k = \sigma_l = \sigma$. The equation can then be simplified to

$$h(k,l) = \frac{1}{2\pi\sigma^2}\exp\left(\frac{-(k^2+l^2)}{2\sigma^2}\right) \tag{4.9}$$

Sigma can be chosen freely, dependent on what spread you want the distribution to have. In OpenCV, it is normally calculated based on the size of the kernel, K, with

$$\sigma = 0.3\left(\frac{K-1}{2} - 1\right) + 0.8 \tag{4.10}$$

In this case, the kernel size was chosen to be 17 in both directions, based on empirical findings. Since the images need to focus on showing the difference in surfaces and not have a focus on having the edge at exactly the right point, this was a sacrifice that could be made, in return for a very smooth image.

The result of extracting the surface normals after Gaussian Blur has been applied can be seen in Fig. 4.4.

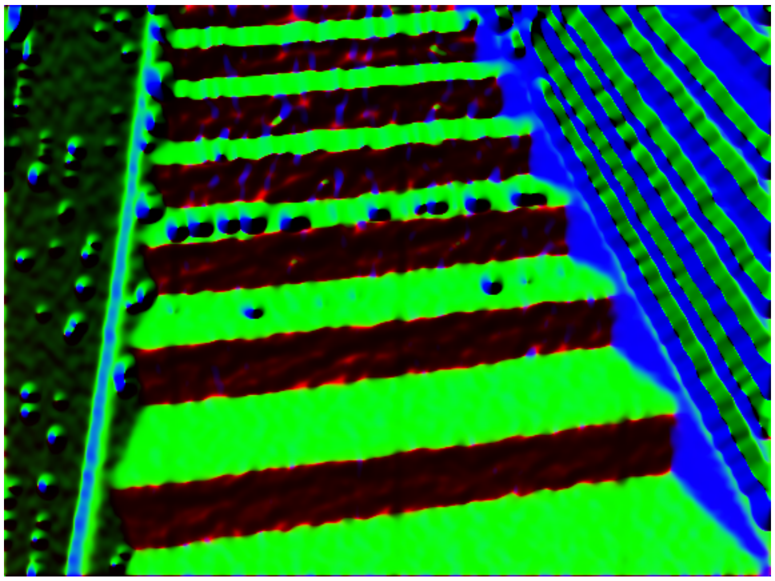


Figure 4.4: Surface normals after Gaussian Blur

4.3 Clustering pixels

We now have an image that looks very clear when looked at from a humans perspective. The steps are easy to distinguish from each other and are somewhat uniformly colored. However, for a computer the image still consists of constantly varying pixels. This is evident especially in the risers (vertical parts of the step). Furthermore, the value of each pixel is a combination of three different channels (BGR), which does not allow for easy comparison. Therefore we start off with dividing the pixels into bins.

4.3.1 Histogram or bins

The pixels will be divided in eight different bins, dependent on their combined RGB values using binary arithmetic. For each channel, the pixel is classified in either the lower half of the value spectrum, or the upper half. Zero meaning it has none of this color channel, 255 meaning it has 100% intensity for this particular color channel.

When the channel falls in the upper half of the spectrum it gets a '1', otherwise it gets a '0'. The BGR channels, represented by 2^0, 2^1, and 2^2 respectively, can form a binary number from 000 to 111. This binary value is then converted to a decimal value and applied to a new image container. For example, a pixel with BGR values of Red(230): '1', Green(150): '1' and Blue(90): '0', gets a binary value of '110' or 6 in decimals. This is done for every pixel in the image.

First, an empty 1-channel matrix of the same size is initialized with zeros. The algorithm then loops through the surface normal matrix and checks each channel. If the value of the first, second or third channel exceeds 127, a value of +1, +2, or +4 is added to the corresponding pixel in the new 1-channel matrix. See Fig. 4.5 for the flow diagram of this process. The results can be seen in Fig. 4.6. The steps have now mostly become a uniform surface, both the horizontal parts and the vertical parts.

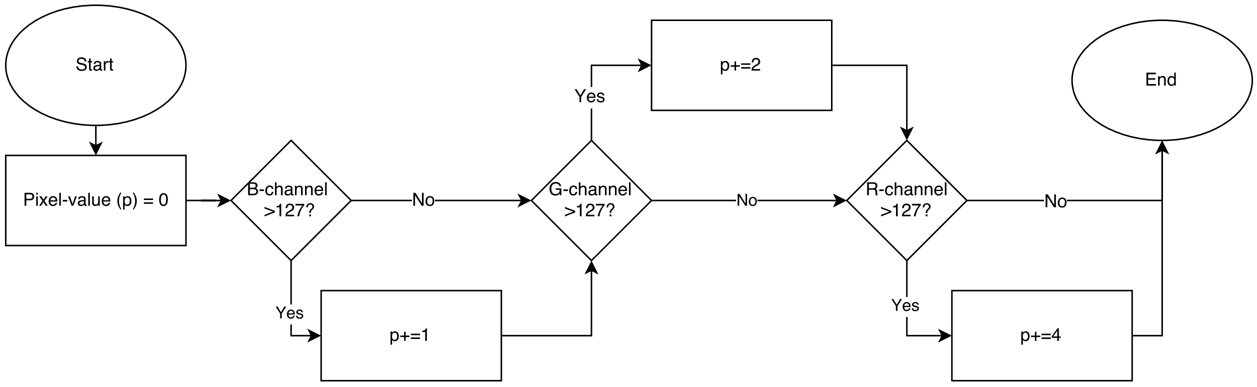


Figure 4.5: Flow Diagram for segmenting pixels based on their individual channel value.

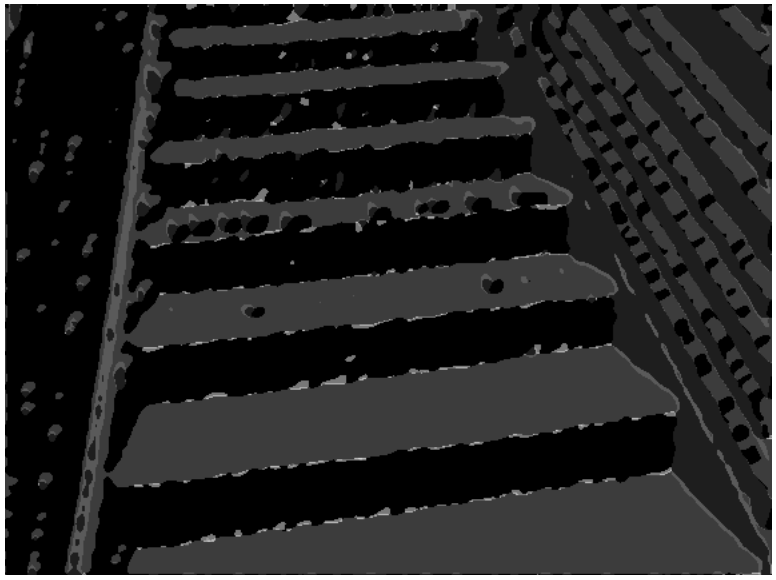


Figure 4.6: Clustered staircase using 8 different bins

4.4 Floodfill algorithm

However, especially at the edges there are a lot of noisy patches. Since these patches of pixels have different values compared to the surface they are a part of, they will erroneously signal a transition/edge to the algorithm, adding noise to the chirp signal. To prevent them from interfering with the stairdetection algorithm, these need to be removed. Preferably, they should blend in with their neighbours. Normal noise removal techniques also affect the larger segments, and do not remove the islands, they just smooth them. There are techniques like erosion and dilation[30], but they have better effect on lone pixels than on islands. To get rid of the islands, what needs to happen is that small patches of pixels of a different value should be given the same value as a neighbouring large area, while the large areas stay unaffected. To do so the number of pixels in the patch is computed, for each segment in the image. Any segment that has a size below a certain threshold needs to be given the same value as a neighbouring segment, such that it becomes one. To do this a Floodfill algorithm is used. Floodfill is comparable to the bucket-function in most Paint-like programs. It uses a recursive algorithm where it performs a function (count pixels, or change pixel value), and then checks each neighbour (using either a Von Neumann or Moore neighbourhood) for the same value; if this neighbouring pixel has the same value as the original pixel it will go to that neighbouring pixel and do the same thing, until done. In the implementation of the Step-1 algorithm a Moore neighbourhood was used.

Sections with an area below 300 pixels were deemed noise, and were floodfilled with the value of the left-neighbour, $(y, x-1)$, unless it was the first column of the matrix, then the neighbour above, $(y-1, x)$, was used. The result can be seen in Fig. 4.7.

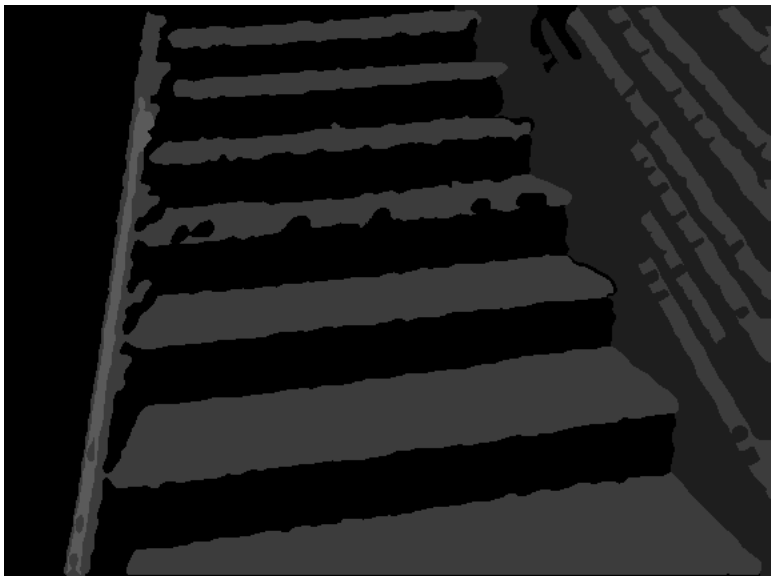


Figure 4.7: Noisy areas removed with Floodfill algorithm

Now that the input has been processed, it is ready for the stair-detection algorithm.

5 Stair-detection algorithm

The stair-detection process consists of several steps, the overview of which can be found in Fig. 5.1. First the images are loaded again and the correct labels are retreived from the filenames. Using a random number generator an n-number of random line-pair coordinates (consisting of a horizontal $[x_0 - x_{max}]$ and a vertical $[y_0 - y_{max}]$ line) are generated. These coordinates are used as input for Bresenham's line drawing algorithm [11] to collect the values of the pixels on the line between these points. This vector of values is then searched for transitions, difference in value between two adjacent pixels, which are noted with a '1'. Every pixel that is the same value as the pixel before it gets a value of '0'. There is a certain spacing between these transition points that tend to get smaller or larger over time. These gaps are then fitted with a sine-wave with a chirp function. A Hilbert space is created of these different chirp functions, and the chirp function with the smallest error is selected as the best fit. Its error is used as the basis for classification. This is done for every image and every coordinate-pair. AdaBoost is used to get a final classification based on the sum of the weighted results of each coordinate-pair.

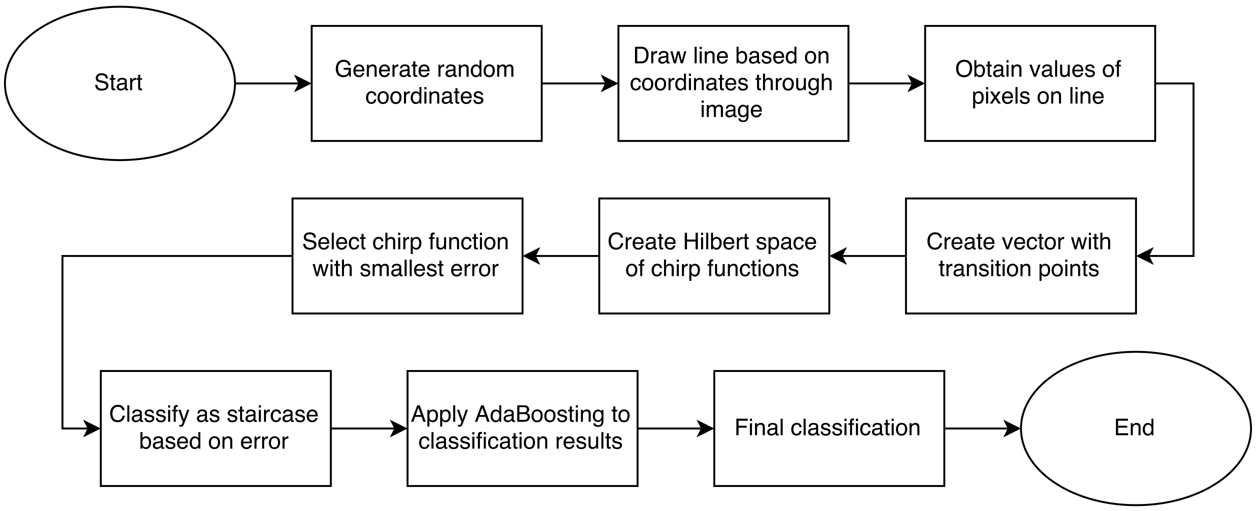

Start
Generate random coordinates
Draw line based on coordinates through image
Obtain values of pixels on line
Create vector with transition points
Create Hilbert space of chirp functions
Select chirp function with smallest error
Classify as staircase based on error
Apply AdaBoosting to classification results
Final classification
End

Figure 5.1: Overview of stair-detection algorithm

5.1 Drawing random lines

Since input images for the algorithm do not have a fixed angle, it needs to be robust against deviations in perspective or angle, on a per-picture basis. To aid in this, randomly drawn lines will be used, ensuring that the orientation of the input does not matter since the lines deviate amongst themselves anyway.

5.1.1 Generating random coordinate-pairs

A very simple algorithm for drawing random lines was used, consisting of three steps:

1. The program selects the first column as starting point for the horizontal line, and the first row as the starting point for the vertical line. Similarly, it selects the last column as endpoint for the horizontal line and the last row as endpoint for the vertical line.

2. It then chooses two random values for the other start and end coordinates of the line.

3. It draws a line using Bresenham's line drawing algorithm[31] using these coordinates.

5.1.2 Bresenham's line drawing algorithm

Bresenham's line drawing algorithm uses two points as input, the starting point (X_0, Y_0) and the end point (X_1, Y_1). The basic version of the algorithm is as follows: In the situation that $\Delta X > \Delta Y$, the line will travel over the x-values, and depending on the buildup error in y, travel one pixel in y-direction everytime the error has build up too far. The error gradient is calculated with

$$\Delta\epsilon = \frac{\Delta Y}{\Delta X} \tag{5.1}$$

This means that, for every pixel the line travels in X-direction, it builds up an error of $\Delta\epsilon$. Once the error becomes larger than one, i.e. $\epsilon > 1$, the line being drawn should shift one pixel in the Y-direction. The value of ϵ is then subtracted by 1 (since the error has now been decreased by 1 pixel) and the algorithm continues to the end point of X. An example can be seen in Fig. 5.2.

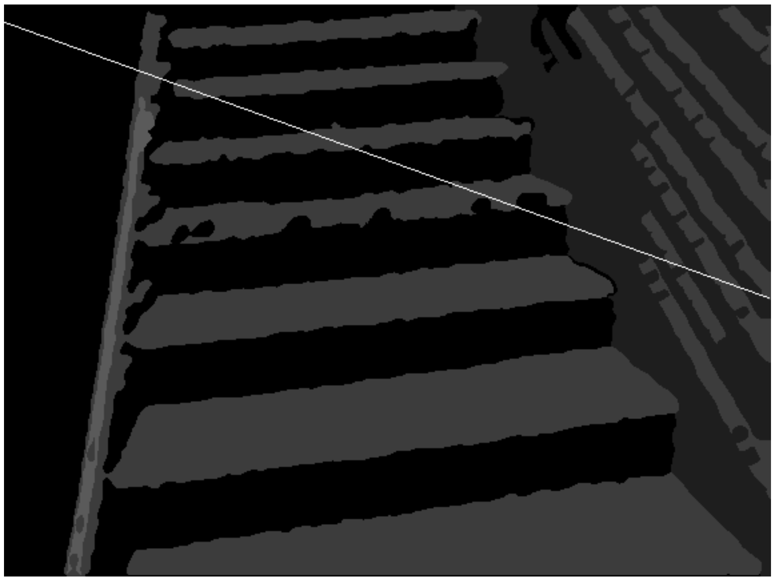


Figure 5.2: Example of Bresenham's line drawing algorithm applied to clustered input

There are some limitations to the current implementation:

- If the error increases with more than 1 for each pixel it travels, i.e. $\frac{\Delta Y}{\Delta X} > 1$, it will not travel far enough in Y-direction.
- If either ΔY or ΔX are negative, the error will be negative as well and the line will never move in Y-direction.
- If $\Delta X = 0$ (straight vertical line) there will be a division by zero error.
- It can only travel from top to bottom and from left to right.

The error gradient $\Delta\epsilon$ of this algorithm is also the slope of the intended line, so what this means is that the current implementation does not allow for slopes larger than 1, and the line can only be drawn from a top-left position to a bottom-right position. This only addresses one of the possible eight line-orientations, see Fig. 5.4. In this drawing, (x_1, y_1) is the starting point, (x_2, y_2) is the end point, and m is the slope, i.e. $m = \Delta\epsilon = \frac{\Delta Y}{\Delta X}$.

So, some adjustments need to be made. This can be done using logical statements that check for the conditions in the octants and change the code to accompany those conditions. For example, if $1 < m < \infty$ the error gradient needs to be reversed, i.e. $\Delta\epsilon = \frac{\Delta X}{\Delta Y}$. The slope now falls between 0 and 1, and the leading direction changes from an x-direction to a y-direction. Similarly, if the error is negative, then every time a

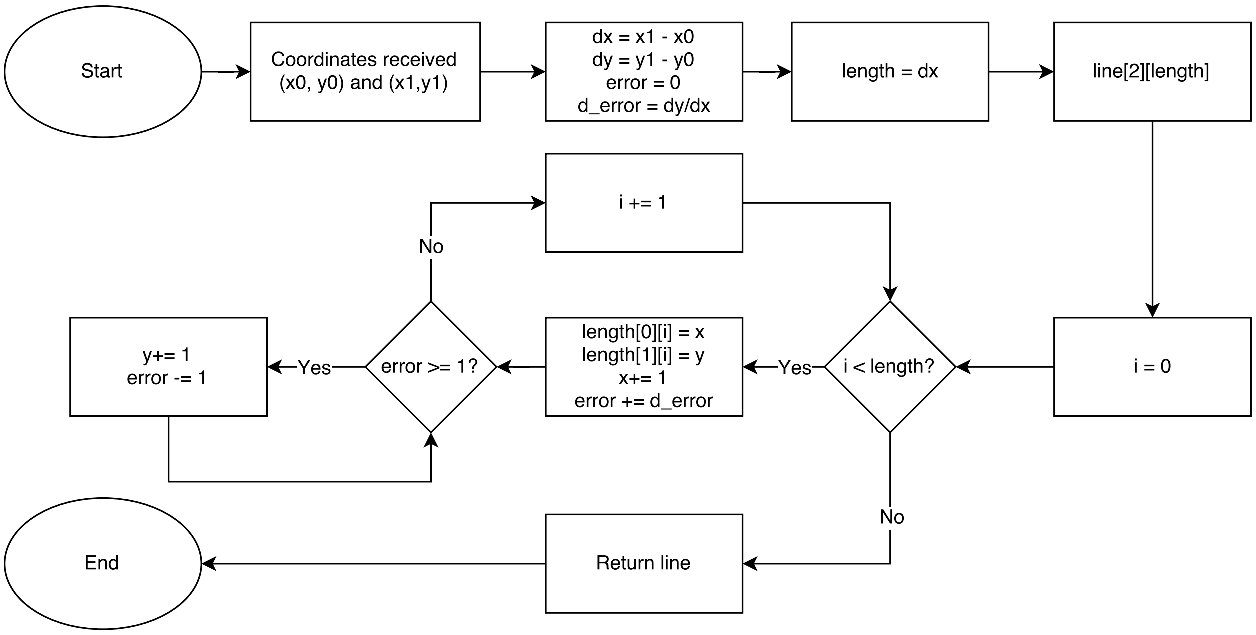


Figure 5.3: Flow Diagram of the original Bresenham line drawing algorithm. Note that it does not allow for negative dx or dy, or for dy to be bigger than dx.

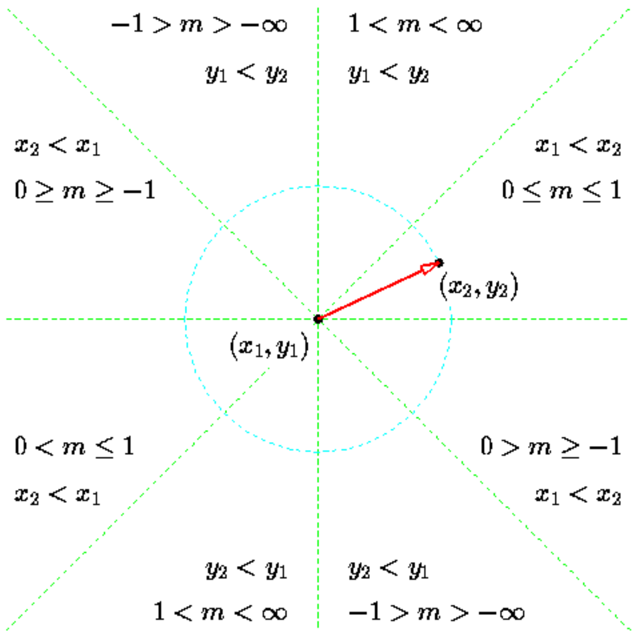


Figure 5.4: The eight possible orientations for drawing a random line with (x_1, y_1) as starting point and (x_2, y_1) as ending point[11]

pixel is jumped and the error needs to be decreased in absolute terms, you increase it by 1, i.e. from -1.5 to -0.5.

To prevent having to write out the code 8 times, a condensed version was used where the sign of ΔY and ΔX determines in what direction the next pixel moves. See Fig. 5.3.

With these adjustments made, the line-drawing algorithm is now able to draw any line on the image, see

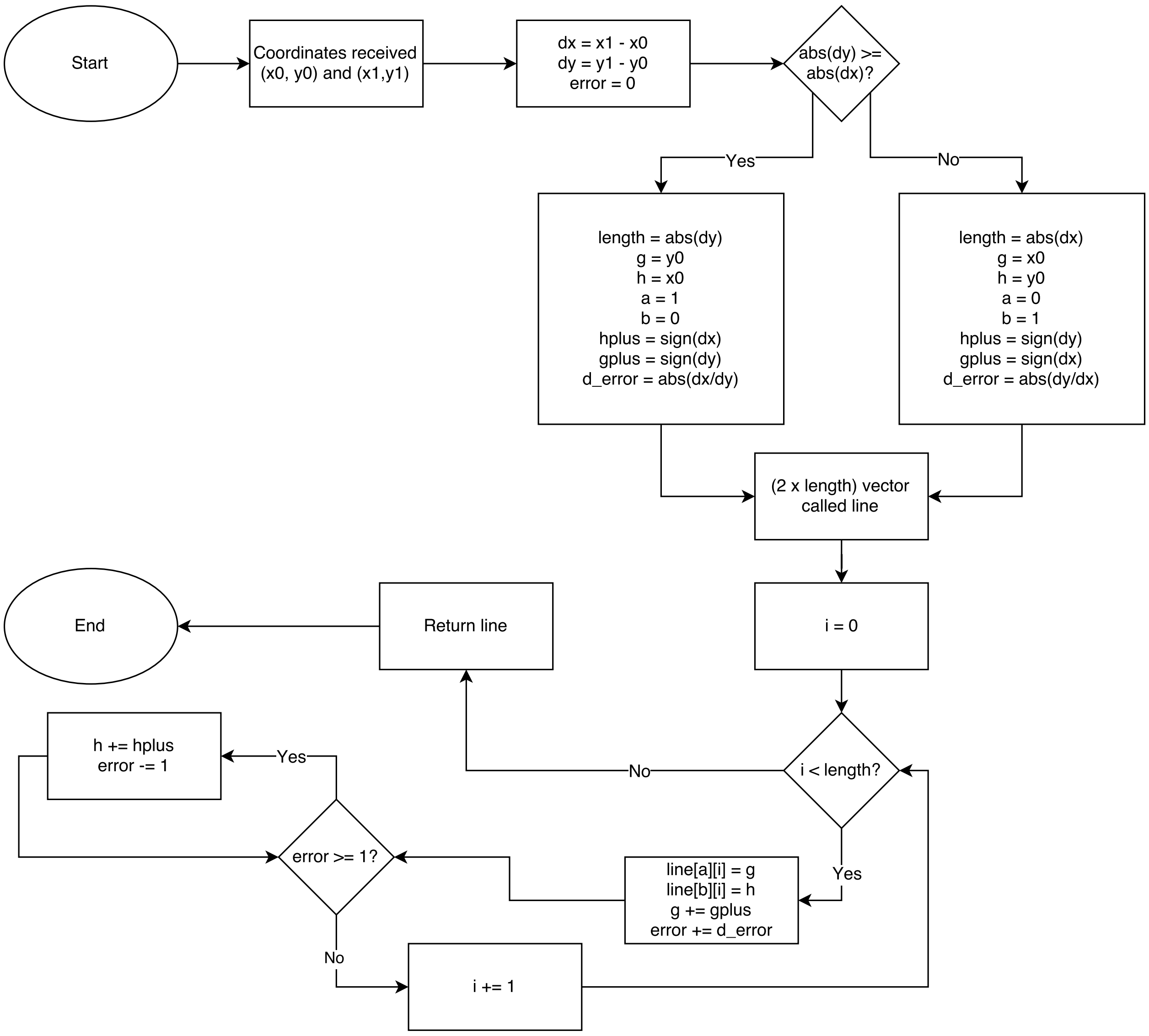


Figure 5.5: Flow Diagram of updated Bresenham Line Drawing algorithm. It is now able to deal with any value of dy and dx.

Fig. 5.6.

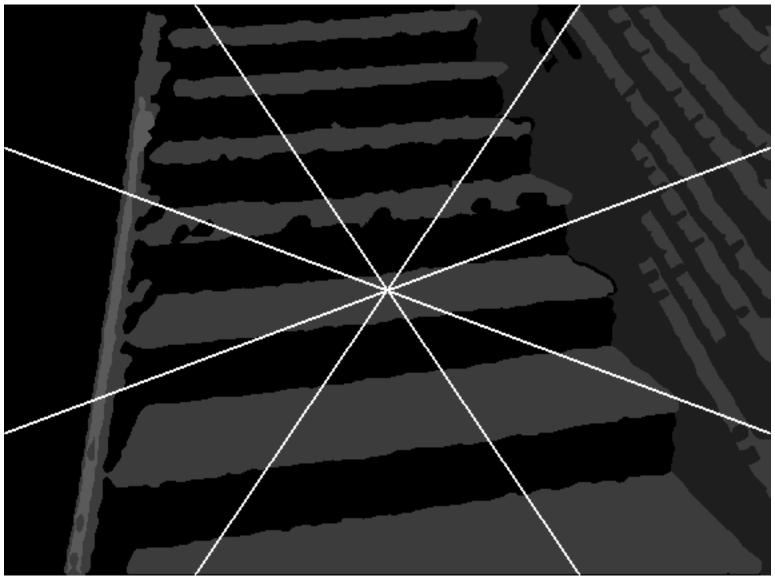


Figure 5.6: Showing all 8 possibilities for drawing a line using the updated Bresenham's line drawing algorithm

5.2 Classifying as chirp

As mentioned, a consistent feature of a staircase is the equidistant spacing between the edges of its steps. These distances get more warped as they get further away from the observer, and follow the shape of a chirp. A chirp is a waveform in which the frequency increases as its variable increases, which in the case of the staircase is the value of the position along the staircase. As can be seen in Fig. 5.7, the windows of the building are equidistantly spaced, but appear to be closer to each other as they move further away. Below that is a chirp waveform following the same pattern. Similarly, a chirp could be fitted to a staircase, since it is also a structure consisting of equidistant parts. Therefor, given the right environment (indoors), we could state that, if a chirpfunction can be plotted against a line across an image, there is a good chance this image contains a staircase. The main chirp waveform can be expressed as

$$f(x) = \sin\left(a(x+c)^b + d\right) \tag{5.2}$$

where a (frequency factor) determines the starting frequency, while b (acceleration factor) determines how fast the frequency of the waveform increases, causing the chirp effect. Both c and d translate the waveform along the x-axis. The x-axis in this equation consists of the values of the pixels found along the randomly drawn line. For every pixel along this line, x increases with 1.

If a variant of this chirpfunction can be found, such that the points of interest (edges of the steps) correspond to the peaks of the chirpfunction, the image will be concluded to be a staircase. The way to compute the fit is to calculate all possible variants of the chirpfunction and see which one fits best. The best fit is the fit with the smallest error, which will be discussed in more detail in Sect. 5.2.1. However, there are 4 variables in Eq. 5.2. Varying each of them over 100 different values will lead to 100000000 different forms of the chirpfunction per image. To eliminate two of the variables, it is assumed that the chirpfunction has to be '1' at the point of the first transition (x_f); or: the peak of the chirpfunction lines up with the first transition. This means the equation changes to

$$f(x) = \sin\left(a(x-x_f)^b + \frac{\pi}{2}\right) \tag{5.3}$$

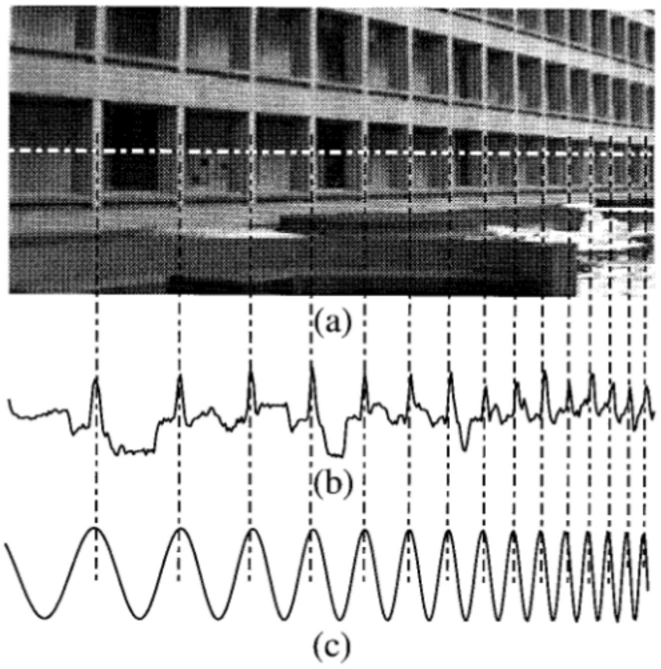


Figure 5.7: Example of a chirp-like repeatability in a real world setting[12]

if now $x = x_f$ is plugged in, the equation evaluates to $\sin \frac{\pi}{2} = 1$, satisfying our requirement to start from 1 at the first transition. This means there are only two variables left. The frequency factor (a) and the acceleration factor (b). To find out the range in which these will be varied, it assumed that there will be a maximum of 10 steps in each image. This means there will be a maximum of 20 transitions in the image, which corresponds to 19 periods but is rounded up to 20 for extra margin. Thus, the part inside the sine function should, at its highest point, be equal to the value of 20 periods. Noting that $x_{max} = 640$ because that is the amount of pixels in the image, and the equation can be rewritten as

$$a640^b = 20 \cdot 2\pi \tag{5.4}$$

plotting this equation gives Fig. 6.5. Based on this plot a range for the frequency factor of $(0 \leq a \leq 0.2)$ and for the acceleration factor of $(0.5 \leq b \leq 2.0)$ was chosen for the iteration, using steps of 0.001 and 0.01 respectively. The results have been shown to not deviate strongly when larger step-sizes were used, so taking computation time into account, these step-sizes were deemed sufficiently precise.

5.2.1 Hilbert space

The different variations of the chirpfunction in Eq. 5.3 for the ranges of a and b can be stored in a Hilbert space. A Hilbert space is in its simplest form an infinite-dimensional vector space. In our scenario we have two vectors, $\vec{a}$ and $\vec{b}$, that have finite length. For computational purposes, the domain of $\vec{a}$ is chosen to be $0 < \vec{a} <= 0.200$ in steps of 0.001, and the domain of $\vec{b}$ is chosen as $0.5 < \vec{b} <= 2.0$ in steps of 0.01. Since the vectors are not really infinitely dimensional, the Hilbert space should actually be referred to as a pseudo-Hilbert space. However, for the sake of simplification it will continue to be referenced as Hilbert space. The Hilbert space $H_{a,b}$ can be seen in Eq.5.6. It is filled using

$$H_{i,j} = \{\sin\left(0.001i(x - x_f)^{0.01j} + \frac{\pi}{2}\right) \mid 1 \leq i \leq 200, 51 \leq j \leq 200\} \tag{5.5}$$

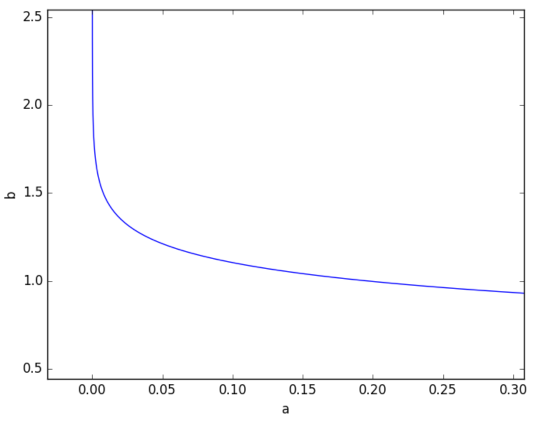


Figure 5.8: Plot of possible combinations of a and b to satisfy $a640^b = 40\pi$. Based on this plot a range of $(0 \leq a <= 0.2)$ and $(0.5 \leq b \leq 2.0)$ was chosen for the iteration.

$$H_{a,b} = \begin{pmatrix} \sin\left(0.001(x-x_f)^{0.51}+\frac{\pi}{2}\right) & \sin\left(0.001(x-x_f)^{0.52}+\frac{\pi}{2}\right) & \cdots & \sin\left(0.001(x-x_f)^{2.00}+\frac{\pi}{2}\right) \\ \sin\left(0.002(x-x_f)^{0.51}+\frac{\pi}{2}\right) & \sin\left(0.002(x-x_f)^{0.52}+\frac{\pi}{2}\right) & \cdots & \sin\left(0.002(x-x_f)^{2.00}+\frac{\pi}{2}\right) \\ \sin\left(0.003(x-x_f)^{0.51}+\frac{\pi}{2}\right) & \sin\left(0.003(x-x_f)^{0.52}+\frac{\pi}{2}\right) & \cdots & \sin\left(0.003(x-x_f)^{2.00}+\frac{\pi}{2}\right) \\ \vdots & \vdots & \ddots & \vdots \\ \sin\left(0.200(x-x_f)^{0.51}+\frac{\pi}{2}\right) & \sin\left(0.200(x-x_f)^{0.52}+\frac{\pi}{2}\right) & \cdots & \sin\left(0.200(x-x_f)^{2.00}+\frac{\pi}{2}\right) \end{pmatrix} \tag{5.6}$$

Error calculation

For every function in the Hilbert space, the error will be calculated, after which the smallest error will be selected as the best fit. Errors can be defined as a deviation between the expected value and the actual result. In this case, between the points of transition of the image, and the chirp function. There are a couple of ways in which the chirp function can deviate from the transition points:

1. The value of the chirp function at each transition should be equal to 1, and equal to -1 between the transitions. A deviation from this is an error as large as the deviation itself.

2. The number of periods should be equal to the amount of steps minus 1.

3. There should be one, and only one, period between each transition.

Now, these deviations objectively have some form of correlation, since they are all based on the periodic characteristic of the chirp function. Applying only deviation type 1. will cause the chirp to be fitted like in Fig. 6.12. A large number of periods, that are 1 and -1 at just the right points. This is not a good or even acceptable fit. Applying type 2. as well will result in Fig. 6.13, where the number of periods will be acceptable but the periods between the transition points are completely off. Type 3. will ensure that the periodicity of the fitted function will match up as much as possible to the transition points. Therefore type 3. is chosen as our error measurement method. Recall from Eq. 5.3 that the number of periods that have passed since $x = 0$ up to transition point n are calculated with

$$P_n = a(x_n - x_f)^b \tag{5.7}$$

As such, the error for each transition point can be calculated with

$$\epsilon_n = 2\pi(1 - [P_n - P_{n-1}]) \tag{5.8}$$

or

$$\epsilon_n = 2\pi(1 - [a(x_n - x_f)^b - a(x_{n-1} - x_f)^b]) \tag{5.9}$$

where ϵ_n is the deviation at transition n, x_n is the x-coordinate for transition n and x_{n-1} is the previous transition coordinate.

The total error is then calculated using the Normalized Root Mean Square (NRMS). This is because NRMS penalizes outliers more than it rewards good fitting points, meaning consistent behaviour is rewarded and luck becomes less of a factor. Due to its normalization factor images with little to no transitions do not have an advantage over staircases with as much as 10 steps in them. Without normalization, images that have more transitions in them would have more opportunities to build up errors and would then be wrongly classified, similarly for empty images with barely any transitions in them. The NRMS is calculated with

$$E_T = \sqrt{\sum_{n=1}^{t} \frac{\epsilon_n^2}{N}} \tag{5.10}$$

where E_T is total error and N is the number of steps, as represented by the number of ϵ values

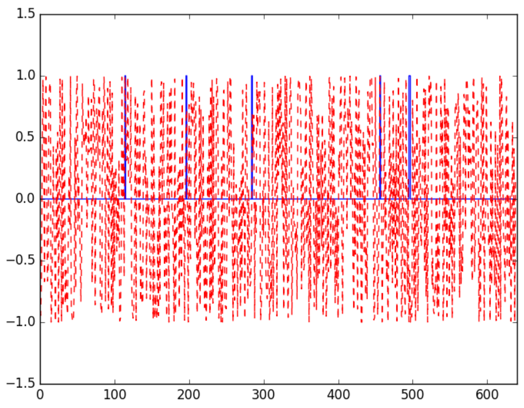


Figure 5.9: Example of a chirpfitting with a very large amount of periods, erroneously determined to be the best fit.

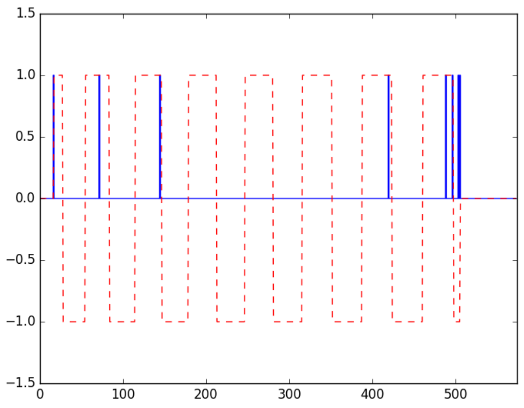


Figure 5.10: Example of a bad fit being classified as good, due to bad spacing between each period.

Using an exhaustive search on the Hilbert space, the NRMS error is computed for each point in the Hilbert space. The best fit is found for the point that shows the smallest error. If this error is below a certain threshold, it will be determined there is a chirp in the picture and thus a staircase. The fit starts at the first transition x_f, such that $f(x_f) = 1$, and ends with the last transition x_l.

Visualization Hilbert Space

The total error E_T is calculated for each element in the Hilbert Space. These errors can be visualized using the a and b index as x and y, where the errors are represented by the values of the pixels. One would expect a well-fitted function to be increasingly worse-fitted the further it gets away from its perfect-fit point. It is expected to see a gradual change from black/dark grey to light grey as the distance from the best-fitted pixel increases. If the changes in values are very random, the fit is not very solid in the first place. An example of this visualization can be seen in Fig. 5.11. It can be seen that there is a certain region through which the combination of a and b result in a small error, as displayed via the darker color of the pixels. It can also be seen that as the pixels travel further away from the correct fit, they become lighter, as expected. At some point, the errors become larger very quickly, due to the x^b relationship, which then shows in a complete white region, which means $E_T > 255$.

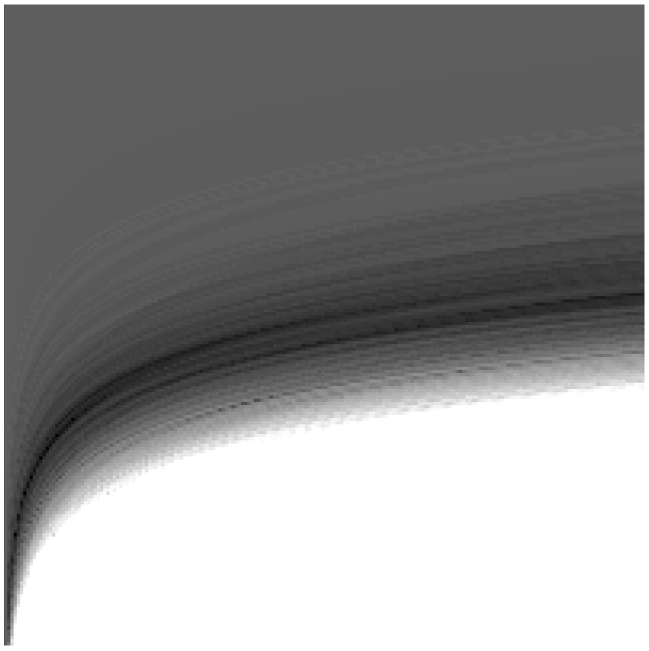


Figure 5.11: Hilbert visualization for one random line

To find the best chirp-fit, we run a minimization algorithm on the Hilbert Space, such that

$$E_T(x, y) = \min(E_T(a, b)) \tag{5.11}$$

The chirp function based on these coordinates is then deemed the best fit. An example of a fitted function can be seen in Fig. 5.12.

The error is then used to classify the image based on certain thresholds. Is the error below this threshold, the image gets classified as a staircase. Otherwise, it is classified as a non-staircase. A precision vs recall curve is created by by classifying the test images while varying the threshold value.

5.3 AdaBoosting for final classification

Now the random line that gave us the transition points might not be a good choice for the particular image. If it comes in from the side it will only see one step and will not be able to fit a stair-like chirp. This cannot be solved by hard-coding the line to travel from the top to the bottom of the image, since you do not know

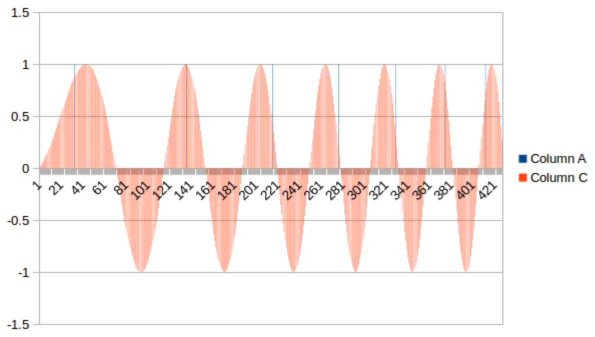


Figure 5.12: Example of a chirp function fitted to the transition points of a staircase

the orientation of the image. To account for this, we draw multiple random lines and rely on a technique called AdaBoosting. Adaboosting is a machine learning algorithm which combines weak classifiers into a strong classifier. A weak classifier is a classifier that has mediocre performance (say <60% accuracy) but is better than random guessing. If the classifiers are barely better than random guessing, a lot of them need to be combined to get adequate results. Vice versa, if the classifier already has a quite good performance on its own, it only needs to be combined with several other similar performing classifiers. Adaboosting was selected as the boosting method because it is designed to work with weak classifiers and has shown to be robust against overfitting [32]. Furthermore, 10 rounds of training have shown to already drop the generalization error (misclassification on a new data-set after training) to as low as 5%[32].

The Adaboosting algorithm gives 'rules' (simple classifiers) voting power based on their effectiveness in correctly classifying the object or scene. In this case, each random line leads to a classification and thus represents a rule. A training dataset of equal size for positive and negative examples (200 images in total) is used to train the rules and give each rule a weight. The weight for each rule gets multiplied with the classification outcome, which can be either 1 or -1, and the final classification is based on the total sum of these weighted rules. If the sum is positive it is a staircase, if it is negative it is not. The equation representing this classifier is written as:

$$H(x) = \sum_{t=1}^{T} a_t h_t(x) \tag{5.12}$$

and

$$\mathrm{I}(x) \begin{cases} 1, & \text{if } H(x) \geq 0 \\ \text{-1}, & \text{if } H(x) < 0 \end{cases} \tag{5.13}$$

where $H(x)$ is the total weighted sum for image $x \mid x = 1, .., m$. $I(x)$ is the classification of image x, where 1 means it is a staircase and -1 means it is not a staircase. a_t is the weight for rule t, $t = 1, 2, ..T$ and $h_t(x)$ is the classification for image x using rule t. The weight for each rule is inversely proportional to the error rate in classification using this rule, see Eq. 5.14.

$$a_t = \frac{1}{2} \ln\left(\frac{1-\epsilon_t}{\epsilon_t}\right) \tag{5.14}$$

where ϵ_t is the error rate. It can be seen that as the error rate reaches zero, the weight goes to a relatively large positive number. As the error rate reaches 0.5 the weight is 0. Since the results of this rule were exactly the same as random chance, it gets no weight and thus has no effect on the classification. Finally, as the error rate goes below 0.5 (worse than random chance) the weight becomes negative. The rule is then so bad at classifying things correctly you can get correct classifications doing the opposite of what it says. The error rate for the first round is obtained with

$$\epsilon_1 = \frac{1}{m} \sum_{x=1}^{m} \eta_1(x) \begin{cases} \eta_1(x) = 0, & \text{if } h_1(x) = y(x) \\ \eta_1(x) = 1, & \text{if } h_1(x) \neq y(x) \end{cases} \tag{5.15}$$

where $y(x)$ is the label (stairs/no stairs) that image x falls under. Or in more simple terms: the error rate is the number of images that were incorrectly classified, divided by the total number of classified images. Some images will be easy to classify, and maybe 1 or 2 rules already give a clear classification. Others will be more difficult. The algorithm is designed such that it will spend more time fitting the harder images. For this, a distribution is used and updated between each round. At the start, the distribution is equal for every image, which is

$$D_1(x) = \frac{1}{m} \tag{5.16}$$

for every x, so that

$$\sum_{x=1}^{m} D_t(x) = 1 \tag{5.17}$$

So the distribution D_t for any round t has a total sum of 1. The distribution is to the individual images what a_t is to the different lines (rules), it gives a larger weight to the images we want the algorithm to focus on more.

For every subsequent round, the error rate is then found by summing the distributed weight of every incorrectly classified image, see Eq. 5.18.

$$\epsilon_t = \sum_{x=1}^{m} D_t(x)\eta_t(x) \begin{cases} \eta_t(x) = 0, & \text{if } h_t(x) = y(x) \\ \eta_t(x) = 1, & \text{if } h_t(x) \neq y(x) \end{cases} \tag{5.18}$$

where $D_t(x)$ is calculated at the end of every previous round based on the error rate of this previous round

$$D_{t+1}(x) = \frac{D_t(x)\exp(-a_t y(x) h_t(x))}{Z_t} \tag{5.19}$$

It can be seen that as an image has had a correct classification, $\exp(-a_t y(x) h_t(x)) < 1$ and thus will have a lower weight in the distribution in the next round. Vice versa if the classification was incorrect this particular image will have a larger weight in the next round. Thus, the algorithm will start focusing more on correctly classifying these difficult images. Z_t is a normalization factor which ensures $\sum\limits_{x=1}^{m} D_{t+1}(x) = 1$ such that the new distribution conforms to Eq. 5.17. Thus

$$Z_t = \sum_{x=1}^{m} D_t(x)\exp(-a_t y(x) h_t(x)) \tag{5.20}$$

so that $D_{t+1}(x)$ becomes

$$D_{t+1}(x) = \frac{D_t(x)\exp(-a_t y(x) h_t(x))}{\sum\limits_{x=1}^{m} D_t(x)\exp(-a_t y(x) h_t(x))} \tag{5.21}$$

Now that all equations have been defined, the algorithm needs to go through the training process. It starts off setting a simple distribution according to Eq. 5.16, and calculates the error rate based on the classifications of the first rule and the actual labels. It then calculates the weight for this specific rule, and calculates the new distribution. It repeats this for every rule and at the end uses all these rule weights to calculate the final classification in Eq. 5.12. The AdaBoosting algorithm is thus as follows:

- Set distribution of first round $D_1(x) = \frac{1}{m}$
- Use simple rule to classify images
- Calculate error rate ϵ_t using distribution, classification and labels
- Calculate rule-weight a_t using error rate
- Calculate new distribution $D_{t+1}(x)$ using rule-weight, classification and labels

- Repeat for each rule
- Finally, come to a total classification using the sum of the weighted rules

These steps are visualized in the Flow Diagram in Fig. 5.13. Note that one could also do the classification per rule beforehand, and then just input the classification results into the training. This would give the exact same final results. Advantage of this approach is that it keeps the classification and adaboosting algorithms separated.

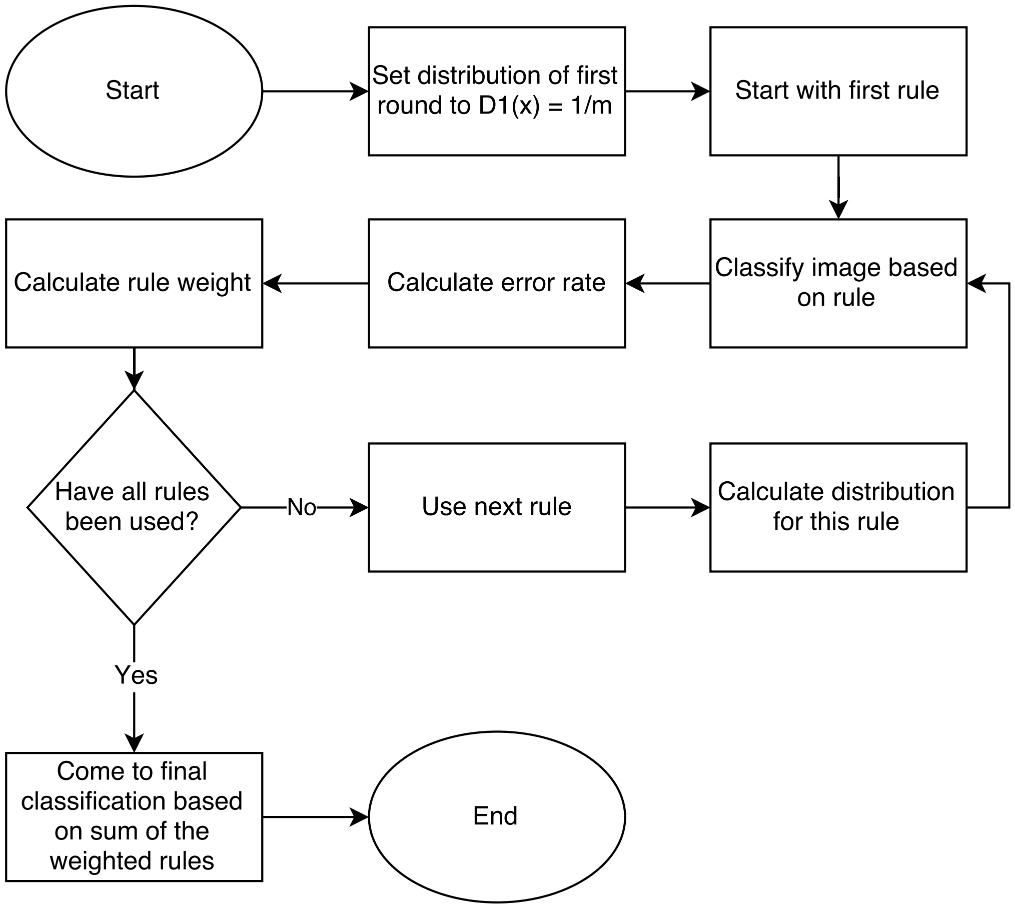

Start
Set distribution of first round to D1(x) = 1/m
Start with first rule
Classify image based on rule
Calculate error rate
Calculate rule weight
Have all rules been used?
No
Use next rule
Calculate distribution for this rule
Yes
Come to final classification based on sum of the weighted rules
End

Figure 5.13: Flow Diagram of the AdaBoost part of the final algorithm.

6 Different phases of algorithm development

This section discusses the changes the algorithm went through, from the start to the last update. It illustrates how the algorithm has logically progressed over the course of the project, and shows what is still to be done.

6.1 A simple Stepcounter algorithm, without chirpfitting or boosting

The first version of the algorithm did not incorporate chirpfitting or a boosting method. It would draw a random line through the image and read out the values. Whenever a change in value occurs, either positive or negative, it counts as a transition. Once more than a certain number of transitions have been counted, the algorithm declares that it has found a staircase. Let's call this number of transitions t. This happens for a number of lines. Naturally, from an accuracy point of view this should be done with as many lines as possible. But from a time-perspective the lowest amount of lines possible are preferable. To find the most optimal balance for the number of lines, a test was done using 20 lines and 100 lines, varying the number of detections. Since it followed from the results in Fig. 6.1 and Fig. 6.2 that the best score is found at a relatively similar detection threshold for both 20 and 100 lines, it was decided that 20 lines should be a large enough sample-base for a correct detection. Calculation time was not yet an issue at this stage, but it was projected that it could become a bottleneck later on as the algorithm development progressed.

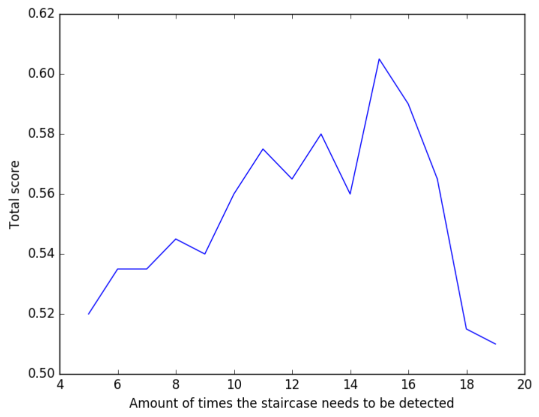


Figure 6.1: The effect of varying the detection threshold on the total score for 20 lines.

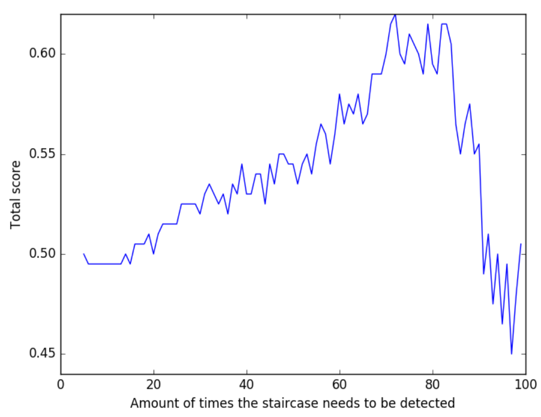


Figure 6.2: The effect of varying the detection threshold on the total score for 100 lines.

From these 20 lines, there need to be at least a certain number of detections to show a staircase is in the image for the classifier to classify the image as a staircase. Let us call this number of detections d. To find out what the best combination of transitions and detections thresholds was, the code was run while varying the number for both the required transitions and detections between 5 and 20. The result can be seen plotted as a 3D-plot in Fig. 6.3 and as a heatmap in Fig. 6.4. The best total score can be found at (10,15), so a threshold of 10 for d and 15 for t. Using these values as input for the algorithm, the following results are obtained, see Table 6.1.

This was just slightly better than random guessing, with a total score of 0.62. There were too many negative images that incorporated random transitions for the algorithm to accurately predict a staircase based on just the number of transitions. A more robust solution would incorporate the spatial aspect of the

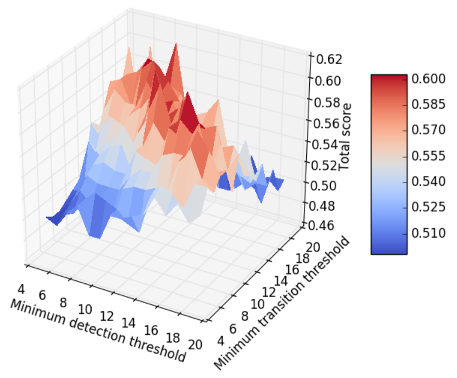


Figure 6.3: The 3d graph of the effect of the step and detection threshold on the stepCounter algorithm. Best score is 0.62 which is found at minimum detection threshold = 10 and minimum transition threshold = 15.

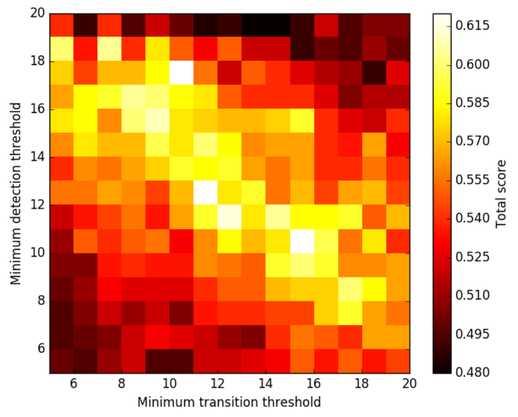


Figure 6.4: Heatmap of the effect of the step and detection threshold on the stepCounter algorithm. Best score is 0.62 which is found at minimum detection threshold = 10 and minimum transition threshold = 15.

Table 6.1: Results of using the stepCounter algorithm, without adjustments

FP	0.5
FN	0.26
TP	0.74
TN	0.5
time spent	0m47s

transitions in the algorithm. That is why a chirp was fitted for more accurate results.

6.2 Chirpfitting without boosting

The fitted chirpfunction should incorporate both the number of transitions as well as the spatial information of these transitions. See Sect. 5.2 for a more detailed overview of what chirpfunctions are and why it is hypothesized to work in this scenario. The chirpfunction that needed to be fit is of the form

$$f(x) = \sin\left(a(x-c)^b + d\right) \tag{6.1}$$

The best fit will be found by iterating over the a, b, and c values. To save on computation time, it was assumed that the best starting point for the fitting would be at the first peak, and thus at the first transition point. This means that f(x) should be equal to 1 at the first transition point x_f. This is ensured by adapting the equation to

$$f(x) = \sin\left(a\left(x - x_f\right)^b + \frac{\pi}{2}\right) \tag{6.2}$$

Now iteration only needs to be done for a and b. It is assumed that there will be a maximum of 21 transitions that need to be fitted, since that translates to around 10 steps because every step should have two transitions, at the front edge and the rear edge. This leads to 20 periods that need to be covered. The maximum value of x is around 640 (pixels). Thus the range of a and b should be able to satisfy at least

$$a640^b = 40\pi \tag{6.3}$$

Rewriting to $a = \frac{40\pi}{640^b}$ and plotting the result gives Fig. 6.5. Based on this plot a range of a = (0, 0.2) and b = (0.5, 2.0) was chosen for the iteration, using steps of 0.001 and 0.01 respectively.

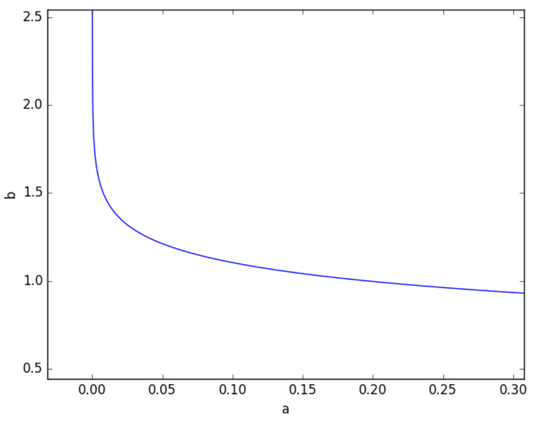


Figure 6.5: Plot of possible combinations of a and b to satisfy $a640^b = 40\pi$. Based on this plot a range of a = (0, 0.2) and b = (0.5,2.0) was chosen for the iteration.

After running the iteration the following figures were obtained, see Figs. 6.9, 6.10, and 6.11.

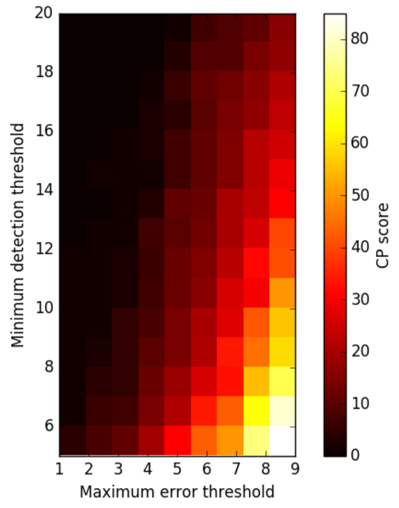


Figure 6.6: Heatmap of the effect of the step and detection threshold on the TP score of the chirpNoBoost algorithm. Best score is 85 which is found at minimum detection threshold = 5 and maximum error threshold = 9.

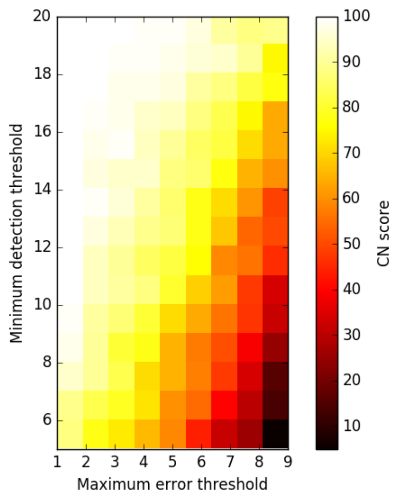


Figure 6.7: Heatmap of the effect of the step and detection threshold on the TN score of the chirpNoBoost algorithm. Best score is 100 which is found at minimum detection threshold = 10 and maximum error threshold = 1.

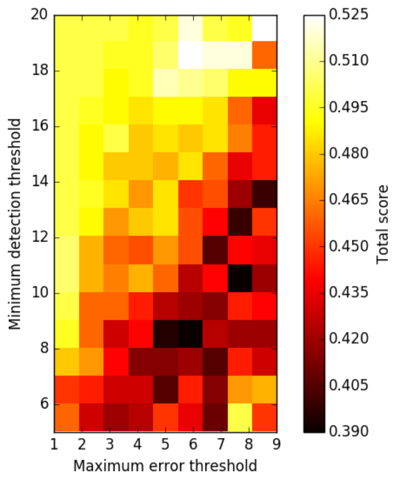


Figure 6.8: Heatmap of the effect of the step and detection threshold on the total score of the chirpNoBoost algorithm. Best score is 0.525 which is found at minimum detection threshold = 18 and maximum error threshold = 16. The corresponding values at that point are TP = 0.09 and TN = 0.96

As can be seen, the best total score is found at a point where the amount of correctly identified positive images is only 0.09 and the amount of correctly identified negative images are 0.96. This is not a well balanced result and it is clear the algorithm needs more tweaking. As it currently is, it is overfitted for negative images. The total results can be seen in Table 6.2.

Now, the first thing that needs to be addressed is the large computation time. Although a considerable amount of time is spent looking for the right combination of the error and detection threshold, it should be looked into if this can be improved. A proposed first improvement is to change the stepsize for the fitting of

Table 6.2: Results of using the chirpNoBoost algorithm, without adjustments

FP	0.04
FN	0.91
TP	0.09
TN	0.97
time spent	856m41s

a from 0.001 to 0.01, and the stepsize of b from 0.01 to 0.1.

As can be seen by comparing the Tables. 6.2 and 6.3, using larger steps for the fitting process gives a significant improvement in running time (100 times faster), while having a negligible effect on the performance.

Table 6.3: Results of using the sped-up chirpNoBoost algorithm, without further adjustments

FP	0.03
FN	0.93
TP	0.07
TN	0.97
time spent	8m58s

6.3 Chirpfitting using line-pairs

In the next version of the algorithm, a combination of line-pairs is used. This means, a random horizontal and a random vertical oriented line is drawn across the image. Naturally a good chirpfit should be found in only one of these lines, since they are perpendicular. Thus, they will be combined into line-pairs, such that if one of the lines finds a staircase, it counts as a staircase detection. This means that 40 random lines will be drawn resulting in 20 line-pairs. A further change in the algorithm is that the lines will be read out both ways. In the previous iterations of the algorithm, if the orientation of the image was such that the chirp would be found from the top going down, it would not be able to be fitted if the line was only read from the bottom going up. That is now prevented by reading the lines out both ways. The result is found in Figs. 6.9, 6.10 and 6.11, and Table 6.4.

Table 6.4: Results of using the sped-up chirpNoBoost algorithm, using horizontal and vertical line-pairs, reading the line out both ways

FP	0.84
FN	0.11
TP	0.89
TN	0.16
time spent	32m20s

Unfortunately the results are again not great. The improvements on the previous iteration did not seem to have a great effect on the results, even though it logically should at least have some impact. Apparently, the way the algorithm is implemented, the factors that are thought of to have effect on the stairdetection might not actually be as relevant in its current form. The next section will look deeper into what is actually fitted based on the algorithm so far.

6.4 Chirpfitting with penalties

The algorithm was clearly not performing as intended. A look into the the graphs of the chosen fitted function shows what went wrong, see Fig. 6.12. It shows a very large amount of periods in the plot; there

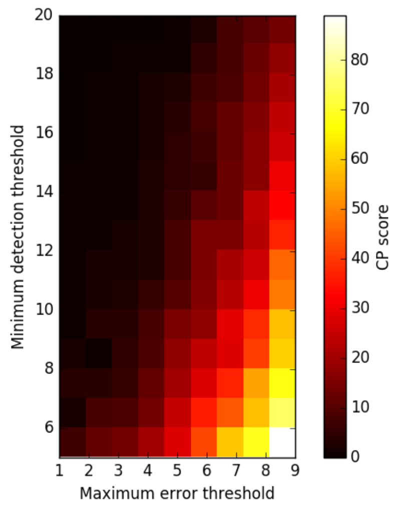


Figure 6.9: Heatmap of the effect of the step and detection threshold on the CP score of the chirpNoBoost algorithm. Best score is 85 which is found at minimum detection threshold = 5 and minimum transition threshold = 9.

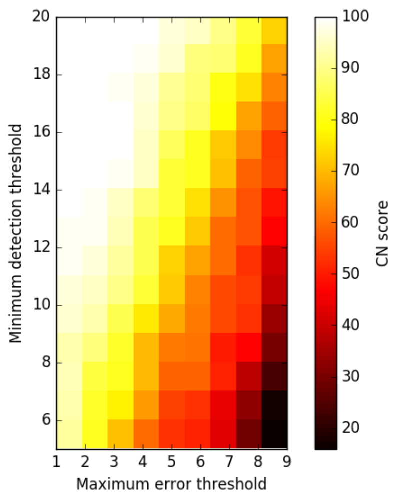


Figure 6.10: Heatmap of the effect of the step and detection threshold on the CN score of the chirpNoBoost algorithm. Best score is 100 which is found at minimum detection threshold = 10 and minimum transition threshold = 1.

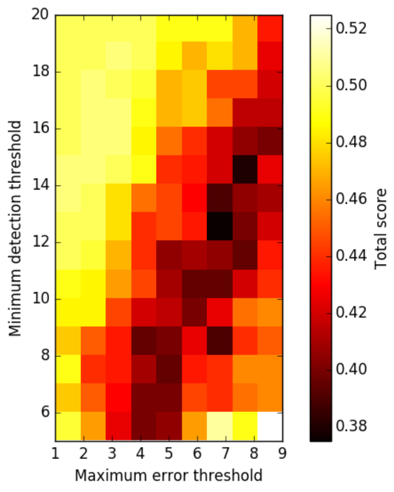


Figure 6.11: Heatmap of the effect of the step and detection threshold on the total score of the chirpNoBoost algorithm. Best score is 0.525 which is found at minimum detection threshold = 5 and maximum error threshold = 16. The corresponding values at that point are CP = 0.09 and CN = 0.96

are in fact so many periods that the plot has difficulty displaying them all. This is a reason to give penalties for large differences between the supposed number of periods (the amount of transitions minus one) and the actual total number of periods, as found by

$$P = \frac{a(x_l - x_f)^b}{2\pi} \tag{6.4}$$

where P is the number of periods, x_l is the x-coordinate of the last peak and x_f is the x-coordinate of the first peak.

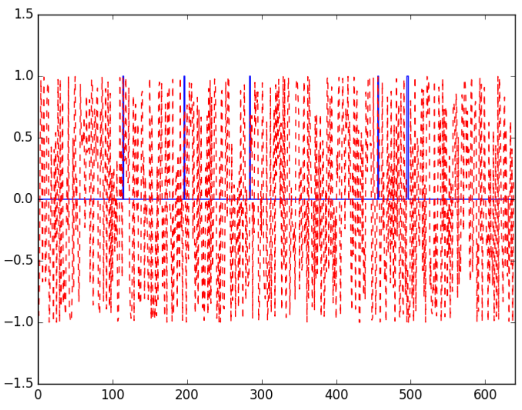


Figure 6.12: Example of a chirpfitting with a very large amount of periods, erroneously determined to be the best fit.

However, only changing the algorithm to account for this is not enough, because there is a second phenomenon that can occur, which can be seen in Fig. 6.13.

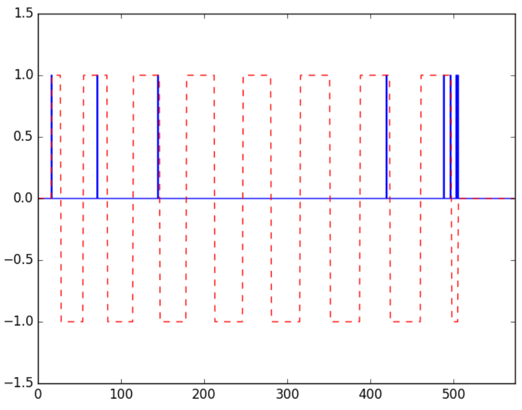


Figure 6.13: Example of a bad fit being classified as good, due to bad spacing between each period.

As can be seen from the figure, the fit is off in this example. However, the algorithm in its current form sees nothing wrong with this. It looks at the peaks, and the valleys in between the peaks, and sees that they correspond to the fitted wave. It counts the number of periods and compares it to the number of peaks and sees that they are very close. Thus, the error is small and this is deemed a good fit. To prevent this, there should also be a penalty for the difference in actual period advancement between each peak and the projected period advancement (which is 1). Recall from Eq. 6.4 what the equation for calculating the number of periods since the start for a certain point is. The difference in periods between two peaks can then be calculated with

$$E_T = \sum_{n=1}^{t} \left(2\pi - [a(x_n - x_f)^b - a(x_{n-1} - x_f)^b]\right)^2 \tag{6.5}$$

where x_n is one of the transition points on the line.

Implementing this in the algorithm gives the following results, see Table 6.5.

Table 6.5: Results of using the sped-up chirpNoBoost algorithm, using horizontal and vertical line-pairs, reading the line out both ways, adjusting for the periods.

FP	0.02
FN	0.92
TP	0.08
TN	0.98
time spent	73m19s

The results are still not up to standards. As can be seen from the results, the algorithm is currently way too strict. It rejects most of the images, resulting in many True Negatives, but many False Negatives as well. At least it rejects more Negatives than Positives, so perhaps if it is taught to be more lenient it allows more positives to pass while still catching the negatives.

6.5 Chirpfitting with normalized error

As the algorithm is currently functioning, it favors images with smaller number of transitions. This is because every transition is a possibility for an error to accumulate. However, in reality you would want many transitions since that is quite typical for a staircase. To therefore remove the penalties that many transitions give, the error is normalized by dividing the error over the number of transitions, see Eq. 6.6.

$$E_T = \sqrt{\sum_{n=1}^{t} \frac{(2\pi - [a(x_n - x_f)^b - a(x_{n-1} - x_f)^b])^2}{S}} \tag{6.6}$$

where S is the number of transitions.

See Table 6.6 for the results. There is a clear improvement but it still has difficulty correctly identifying positive images.

Table 6.6: Results of using the sped-up chirpNoBoost algorithm, using horizontal and vertical line-pairs, reading the line out both ways, adjusting for the periods and normalizing the error.

FP	0.03
FN	0.83
TP	0.17
TN	0.97
time spent	76m02s

6.6 Chirpfitting with boosting

The results mentioned in Sect. 6.5 indicate a weak classifier for staircase detection. Luckily, this weak classifier can be boosted and trained to perform better. Using the AdaBoosting method discussed in Sect. 5.3 every line-pair used functioned as a classification rule, totalling to 20 rules to be boosted. The results can be seen in Table 6.7.

Table 6.7: Results of using the chirpBoosting algorithm, using horizontal and vertical line-pairs, reading the line out both ways, adjusting for the periods and normalizing the error.

FP	0.45
FN	0.00
TP	1.00
TN	0.55
time spent	9m52s

The results are better than they were before, but they are still far from acceptable when it comes to accurately detecting staircases. It seems the algorithm does not have an issue with detecting positive images anymore, in fact it looks like it is now slightly skewed to being too positive when classifying images.

7 Results

The algorithm as discussed in Sect. 5.3 outputs a classification for each image. This is compared with the actual classification using the labels of the images. The weights for each rule has been trained using a training data-set of 100 positive images (stairs) and 100 negative images (no stairs). After the training is completed, the weighted rules are then used to classify a testing data-set of 20 positive and 20 negative images. The Type I and Type II errors for the different phases of the stair-detection algorithm discussed in Chapter 6 can be seen in Table 7.1. False Positive rate (FP) is calculated by summing all images that were wrongly classified to be a positive (type I error) and dividing it by the total number of negative images in the dataset. This FP-rate thus tells you how many of the negative images have been incorrectly classified as positive. It is inversely correlated with the True Negative (TN) rate. The same goes for the False Negative (FN) rate, which is the number of images incorrectly classified as negative (type II error) divided by the number of positive images, and the True Positive (TP) rate, such that

$$FP + TN = FN + TP = 1 \tag{7.1}$$

Naturally, any algorithm aims to get high scores for the 'True' rates and thus low scores for the 'False' rates. It depends on the goal of the algorithm whether it is preferable to have a higher FP over FN rate or vice versa. In the case of the Step-1 algorithm, they are both deemed equally important and thus no weighting is required.

Table 7.1: Results of the different phases of the stair-detection algorithm compared with other stair-detection algorithms. The algorithm acronyms are explained in Table. 7.2.

	sC	cF	cFLP	cFLP3	cFLP3N	Step-1	Wang[8]	Lee[9]
FP	0.50	0.03	0.84	0.02	0.03	0.45	0.00	0.23*
FN	0.26	0.93	0.11	0.92	0.83	0.00	0.03	0.30*
TP	0.74	0.07	0.89	0.08	0.17	1.00	0.97	0.70*
TN	0.50	0.97	0.16	0.98	0.97	0.55	1.00	0.77*

* Approximated from recall vs precision plot

Table 7.2: Abbreviations of algorithms explained.

sC	stepCounter
cF	chirpFit
cFLP	chirpFit with line-pairs
cFLP3	chirpFit with line-pairs and period penalties
cFLP3N	chirpFit with line-pairs and period penalties, with a normalized error
Step-1	Final algorithm: boosted chirpFit with line-pairs and period penalties, with a normalized error

To give the results more meaning they are often translated into precision and recall values. Precision is the number of images correctly classified as stairs divided by the total number of images classified as stairs. Thus

$$\text{precision} = \frac{TP}{TP + FP} \tag{7.2}$$

In other words, precision is an indication of how accurate your classification algorithm is. If it has a large TP rate, but its FP rate is also large, precision will be not as high as when an algorithm has high TP rate and low FP rate; which is what you'd want your measures to reflect. Recall is the number of images correctly classified as a staircase divided by the total number of staircase images. Thus it indicates how well the algorithm did on catching every staircase in the dataset. It is calculated with

$$\text{recall} = \frac{TP}{TP + FN} \tag{7.3}$$

An algorithm that is more strict will have a small recall but high precision. Vice versa an algorithm that is more accepting will be more quick to classify as a staircase, which causes its recall to be high and its precision to be low. These acceptance thresholds thus influence the precision and recall values. The Step-1 algorithm has been run for different error thresholds, and the result can be seen in the precision vs recall plot in Fig. 7.1. It is compared with the precision vs recall plot from Lee in Fig. 7.2; there was no precision vs recall plot given by Wang so it cannot be compared with their results.

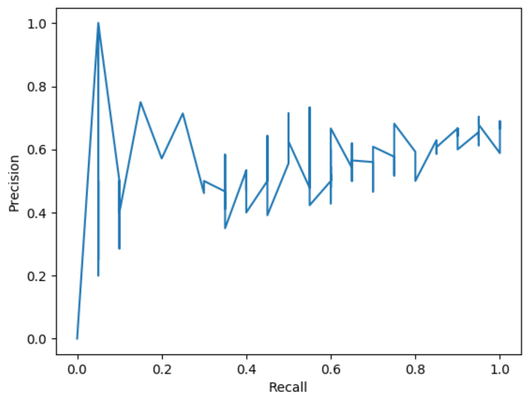


Figure 7.1: Precision recall plot for Step-1 algorithm.

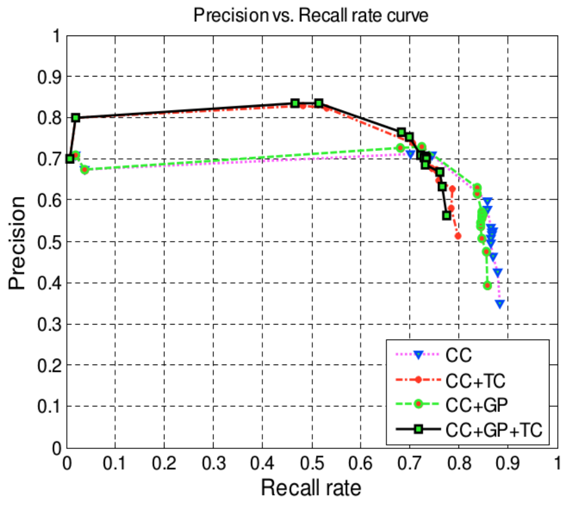


Figure 7.2: Precision recall plot for algorithm Lee [9].

The precision vs recall plot for the Step-1 algorithm is more jumpy than the plot of Lee. One of the reasons for this is that the Step-1 plot has about 6 times as many data points. However, it is strange; one would expect the precision to decrease and the recall to increase, as the error threshold is increased. However the plot shows this is different for each individual point, and if anything, it shows an upward trend for precision, as recall is increased. This behaviour can be somewhat explained by the boosting approach. A change in error threshold means a change in results for each rounds. Since consecutive rounds build on the rounds before them, this could have a large effect on the final outcome, which could slightly deviate in either way. This explains why it is not a smooth curve. Still, it does not explain that the recall manages to become 1 while still having a precision rate of more than 0.5. A similar outcome could be obtained saying everything is a staircase. The precision will be 0.5 (because the guesses where good 50%) of the time, and the recall will be 1 because you correctly identified every positive image.

The F-measure is a method to compare different precision vs recall values as one value, incorporating different weights for either of them dependent on what is deemed a more important measure for the algorithm. A more specific variant, the F1-score, assumes equal weights for the importance of both precision and recall. It is calculated with:

$$F_1 = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \tag{7.4}$$

This score is taken along every datapoint available and the maximum score is taken for each algorithm. See Table 7.3 for the F_1 scores of Step-1, Lee, Wang, and the dumb approach of Always Positive (classify everything positively as a staircase).

It can be seen here that Step-1 performs quite well compared to Lee. However, because the results do not show the expected behaviour (precision decreasing as recall increases), I fear there might be some error behind these results. A peculiar thing was for instance that recall starts off at almost zero, as the error threshold is way too large, but then suddenly jumps to 1.0 and slowly decreases again as the error threshold becomes higher. This is very counter-intuitive and shows there might still be some bugs in the code.

Lee tested on 852 frames consisting of staircases. It is unclear if they meant 852 frames that included staircases but also included scenes without staircase, and if so, how many of each were there.

Table 7.3: F_1 score for Step-1, Lee, Wang and Always Positive, and the value of Precision and Recall for which these are obtained.

Algorithm	F_1 score	Precision	Recall
Step-1	0.82	0.67	1.00
Lee	0.73	0.77	0.70
Wang*	0.98	1.00	0.97
Always Positive	0.66	0.50	1.00

* Approximated from Type I and Type II errors from Table 7.1

As can be seen the algorithm slightly outperforms Lee et al, but is dwarfed by the results of Wang et al. It is interesting to see that the first version of the Step-1 algorithm, which was not reliant on chirpfitting and boosting, is not that far off in performance compared to the final algorithm. This probably has to do with the fact that the first algorithm is a much simpler algorithm which has already been optimized, while the final algorithm still requires more development to obtain good results.

8 Conclusion & recommendations

In this thesis we looked into detecting staircases using recurring physical traits in depth images. Using depth images as input, these images were preprocessed by calculating surface normals to extract surfaces. These surfaces where blurred to remove most of the noise and divided in 8 bins based on their RGB values. They were subsequently floodfilled to remove the remaining 'lone spots' in the image. Using the processed images, 20 random lines were drawn, using the same 20 coordinate-pairs for every image. A vector of transition points was collected along the pixels of the line, and a chirp function based on a sine-wave was fitted to this vector. If this was possible with a small error, a staircase would be declared detected. Using adaBoost, every line was evaluated for its effectiveness on the stair-detection, which was finally added together for a final classification.

Doing so, it was able to get an F_1 score of 0.82, which is quite good. However, the configuration for which this score is obtained is cause to believe it might be the consequence of a bug in the implementation. That said, it was still outperformed by the algorithm of Wang[8]. Intuitively, the algorithm seems solid. A staircase is in essence just a collection of simple horizontal lines, and adaBoosting is very effective in taking simple rules and creating a well-performing algorithm. There are a number of possibilities that could explain the lesser performance:

1. AdaBoost is not actually that good in boosting simple rules to a high performance classification
2. Staircases are not as easily defined by simple rules as initially thought
3. When implementing the algorithm, some vital details were not handled adequately.

Possibility 1 is not very likely. The following papers show excellent results using adaBoosting to increase the performance of simple classification rules, Schapire [32], Schapire et al [33], Schwenk and Bengio [34]. Error rates converge to near-zero after only a few rounds.

Possibility 2 could have some merit. A staircase is in fact more than just a collection of horizontal lines. It also contains the relative position and orientation of those lines, a certain number of those lines, but also contextual clues that help us humans recognize a staircase as a staircase. However, Wang et. al [8] and Lee et. al [9] did manage to obtain reasonable results using simple classification rules; Wang used a line-detector, filtering out lines that were not parallel or that were not long enough. Lee used Haar-like features in a cascaded adaBoosting training, and combined that with a temporal consistency-check to filter out False Positives.

The algorithm discussed in this thesis consists of finding transition points, or edges, and drawing a random line through them. The transition points on this line should have a positional spacing similar to that seen in a staircase, where the distance between the steps decreases as you look further away from your own point of view. These lines are then boosted with the adaBoost algorithm.

The algorithm actually performs quite well on the True Positives part: 89% versus 97.2% from Wang [8] and $\tilde{7}0\%$ from Lee [9]. However, it has a problem with False Positives, scoring 62% vs 0% from Wang and 23% from Lee. As discussed before, both their algorithms incorporated a form of filtering out False Positives (minimal line length, temporal consistency) which seems to be the main reason for them performing so well. The random line drawing technique relies only on non-related points, which means there is room for noise to affect the outcome. The idea was that by adaBoosting the multitude of lines, the randomness would be filtered out or rendered ineffective, but that does not seem to be the case.

Possibility 3 is also likely to be a factor. There are a couple of aspects that could be handled better by a future researcher:

- In the algorithm, a certain threshold is used for the error. Below this threshold the image is determined to be a staircase, and above the threshold it is determined to not be a staircase. However, the whole algorithm is based on robustness against small deviations. This is why many random lines are drawn and why these lines are boosted, such that a small change in angle or spots on the image should not matter. However, the error threshold reduces the outcome at the lower level back to a boolean True/False result. This could lead to round-off errors in the total result. It would be better to use the errors to indicate a likelihood of being a staircase or not. The lower the error, the more likely it is to be a staircase and vice versa.

- As mentioned in Possibility 2, the method as it currently is might also be more susceptible to noise. The training data was difficult to obtain and there was not a wide variety of choice, both with respect to quality and sources. There were a large amount of images that preferably would not be used but that would leave a data-set that is too small for training anything effectively. A possibility could be for a future researcher to collect his own depth images such that he can guarantee its quality and diversity.
- The algorithm currently uses every transition-point to fit a chirp. However, every step has two transition-points, at the front and at the back. This makes it difficult for the chirp to be fitted since there are actually two chirps taking place at the same time. These should be separated from each other for better results.

Ultimately, it seems to be a combination of both 2 and 3 that cause the algorithm to under-perform on the False Positive aspect.

The main question of this thesis was: is it possible to detect a staircase using simple rules? This question was guided by the following sub-questions, which have been answered during this thesis.

- **What are the recurring physical traits for a staircase?** Edges, represented by long lines or points that are related in position and orientation.
- **How can these traits be recognized?** There are a multitude of options, surface segmentation, Sobel Edge detector, Haar-like features, etc.
- **How can these traits be combined to accurately classify a staircase?** Using a boosting algorithm that favors strong traits and penalizes weak traits.

Which leaves the main question: **is it possible to detect a staircase using simple rules?**
As it currently stands, yes. The current F_1 score of 0.81 gives hope for the future, because although the score is not great, there is clear room for improvement. The main reasons for this have been outlined above, being a combination of not filtering the False Positives adequately and forcing a boolean classification on a low level, instead of a floating classification. The False Positives could be filtered out by first detecting lines along the edges, and then fit the chirp. As it currently stands only 1 lone pixel is enough to function as a transition, making the current algorithm more susceptible to noise and random small edges.

Furthermore, a more clear and larger data-set should be obtained, or created yourself. There was a lot of noise on the data-set used, which meant there were a lot of transitions not accounted for in the correct chirp-fits.

It would also need some improvements to make it robust against similar structures, like bookcases, since the detection makes use of the repeatability in the structure of staircases. It therefore might also have problems when confronted with outside structures like fences, large flats and train tracks or cross walks.

8.1 Applicability to space exploration

Space exploration deals with discovery of the unknown. Discovering the unknown means a system needs to be able to deal with the unexpected. To do so it requires a way of gathering information about its current situation (e.g. using a stereo camera) and a way of dealing with it in a non-preset way. For this the system requires artificial intelligence. As can be seen in Sect. 1.3 the space sector is in fact working on landing using video navigation and other applications. Aside from using the current algorithm for detecting stairs, this method can also be used for detecting other objects based on simple rules and a certain repeatability of structure, like the similar structures mentioned before. It could also be used to detect non-naturally-made structures from satellite images of uninhabited places, since non-naturally-made structures will likely adhere to a certain repeatability due to its efficient nature. The Step-1 algorithm can therefor be seen as one of many stepping stones that the space sector needs to climb up in order to reap the rewards of current AI developments.

A C++ implementations

In this chapter, the various implementations of algorithms in C++ are shown, for completeness.

A.1 Bitmask for segmenting

```cpp
Mat clust(nor.size());
int bin = 127;
for (int x = 0; x < clust.cols; x++){
        for (int y = 0; y < clust.rows; y++){
                clust.at(y,x) = 0;

                if (nor.at(y,x)[0] > bin) {
                        clust.at(y,x) += 1;
                }
                if (nor.at(y,x)[1] > bin) {
                        clust.at(y,x) += 2;
                }
                if (nor.at(y,x)[2] > bin) {
                    clust.at(y,x) += 4;
                }
        }
}
```

A.2 Floodfill algorithm

```cpp
Floodfill (segment, oldColor, newColor, x, y){

    if (segment[y,x] == oldColor){
        segment[y,x] = newColor;
        for (int j = -1; j < 2; j++){
            for (int k = -1; k < 2; k++){
                if (j != 0 || k != 0){
                    Floodfill(segment, oldColor, newColor, x+j, y+k);
                    }
                }
            }
        }
    return;
}
```

A similar algorithm is used for calculating the area of these sections but instead of coloring the pixel it adds 1 to a counting variable, such as

```
Floodfill (segment, color, x, y, i){

    if (segment[y,x] == oldColor){
        i++;
        for (int j = -1; j < 2; j++){
            for (int k = -1; k < 2; k++){
                if (j != 0 || k != 0){
                    Floodfill(segment, oldColor, newColor, x+j, y+k);
                    }
                }
            }
        }
    return i;
}
```

A.3 Updated Bresenham line drawing algorithm

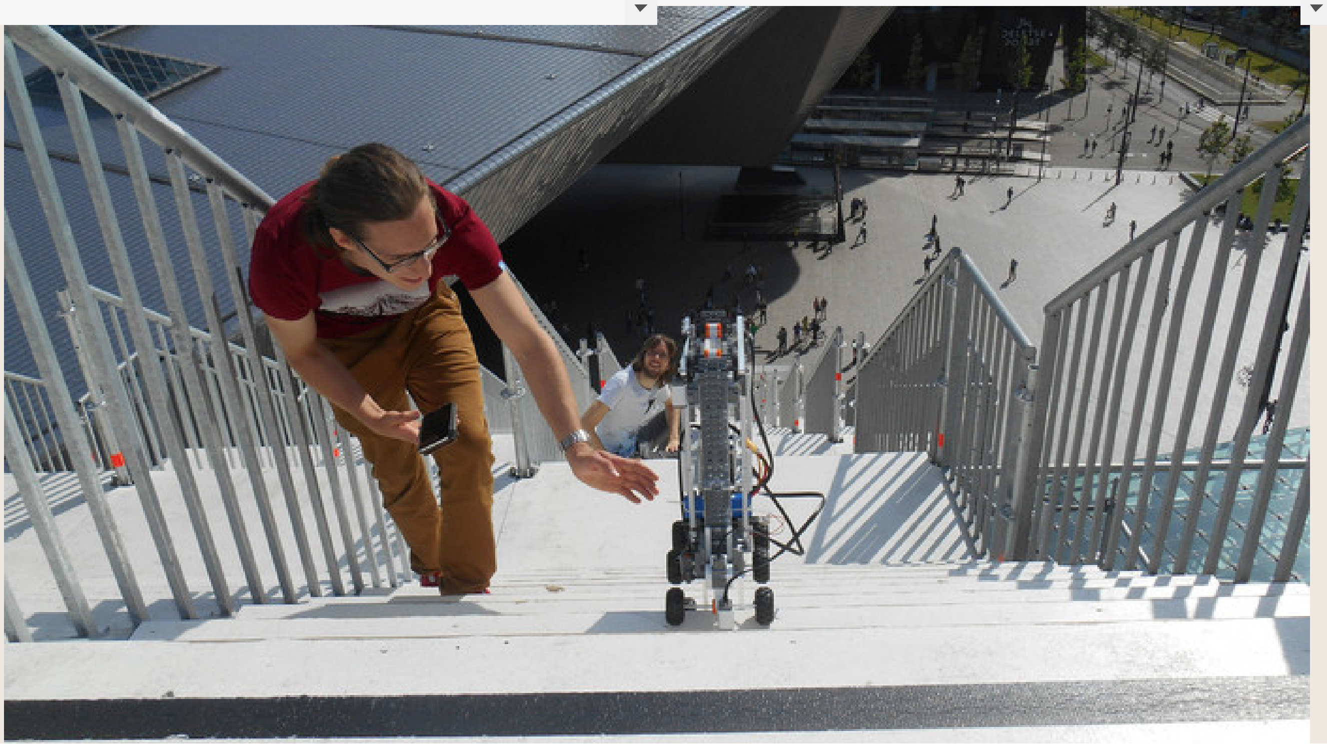
```
vector<vector<int> > random_line_all_points (Mat matrix,
int x0, int x1, int y0, int y1){
        comment("Running␣random␣line␣all␣points␣function..");
        cout<<"x0␣=␣"<<x0<<"␣|␣x1␣=␣"<<x1<<"␣|␣y0␣=␣"<<y0<<"␣|␣y1␣=␣"<<y1<<endl;

        // catching any coordinate that is too large for the image
        y0 = min(y0, matrix.rows - 1);
        y1 = min(y1, matrix.rows - 1);
        x0 = min(x0, matrix.cols - 1);
        x1 = min(x1, matrix.cols - 1);

        int dx = x1 - x0;
        int dy = y1 - y0;
        float error = 0.0;
        float d_error = 0.0;
        int length, g, h, a, b, hplusser, gplusser;
        if ( abs(dy) >= abs(dx) ){
                length = abs(dy);
                g = y0;
                h = x0;
                a = 1;
                b = 0;
                hplusser = sign(dx);
                gplusser = sign(dy);
                d_error = abs((float)dx / (float)dy);
        }
        else {
                length = abs(dx);
                g = x0;
                h = y0;
                a = 0;
                b = 1;
                hplusser = sign(dy);
                gplusser = sign(dx);
                d_error = abs((float)dy / (float)dx);
        }
        //vector that will be returned
        vector<vector<int> > line(2, vector<int>(length) );

        //Actual line plotting
        for (int i=0; i < length; i++){
                line[a][i] = g;
                line[b][i] = h;
                g = g + gplusser;
                error += d_error;
                while (error >= 1.0){
                        h = h + hplusser;
                        error--;
                }
        }
        return line;
}
```

B Article in the AD about stairclimbing robot



▲ Student Frerik leidt de robot naar de top van de trap. © AD

Robot steelt de show op Rotterdamse supertrap

Robot Freddy heeft geen haast. De dag is nog jong, en de supertrap lang. Heel erg lang. Onder toeziend oog van zijn 'baasje' Frerik (24), student ruimtevaarttechniek aan de TU Delft, gaat het wieltje voor wieltje naar boven, waarbij het uit Lego opgebouwde apparaatje sfeervolle, snorrende geluidjes maakt. Eenmaal boven aangekomen, klinkt luid applaus. ,,Hij heeft het gehaald", roept iemand.

Marcel Potters 10-06-16, 11:29 **Laatste update:** 11-06-16, 09:02

22 13 2

Eerst heb ik 'm thuis uitgeprobeerd, maar toen ik hoorde van de megatrap in Rotterdam, heb ik

Feedback

meteen contact opgenomen

▲ © AD

Frerik is trots. Alsof hij zojuist een langharige teckel 'n boeiend kunstje heeft geleerd, showt hij zijn geavanceerde machientje aan het publiek. ,,Dit is mijn afstudeerproject", legt hij uit. ,,Hier ben ik al een jaar mee bezig. Ik heb Freddy in ongeveer twee maanden gebouwd. Eerst heb ik 'm thuis uitgeprobeerd, maar toen ik hoorde van de megatrap in Rotterdam, heb ik meteen contact opgenomen. Deze kans werd me op een presenteerblaadje aangeboden."

Ingewikkeld

De werking is niet zo ingewikkeld, benadrukt de techneut in spe. ,,De robot is zo geconstrueerd, dat hij zichzelf als het ware optrekt. Net zoals een rups zich voortbeweegt", klinkt het. En dit is nog niet het einde van het project. ,,Nu help ik 'm nog een handje, bestuur ik hem. Maar op den duur moet 'ie zélf naar boven kunnen."

Zijn droom? ,,Dat Freddy ouderen kan helpen, bijvoorbeeld met boodschappen doen. Hij moet dan wel een stuk groter worden en zeker 10 kilo aankunnen. Dát concept ga ik de komende tijd verder uitwerken."

Lees ook

toon alles(4) +

Lingeriemodellen schitteren op de Rotterdamse Trap

Lees meer

Feedback

Bibliography

[1] GIJS BEETS. De geboortepiek van 1946. *Demos*, 27(2):6–8, 2011.

[2] CBS. Bevolkingspiramide.

[3] Dirk Holz, Stefan Holzer, Radu Bogdan Rusu, and Sven Behnke. Real-time plane segmentation using RGB-D cameras. In *Robot Soccer World Cup*, pages 306–317. Springer, 2011.

[4] Paul Viola and Michael Jones. Rapid object detection using a boosted cascade of simple features. In *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, volume 1, pages I–511. IEEE, 2001.

[5] C. F. Chen and C. H. Hsiao. Haar wavelet method for solving lumped and distributed-parameter systems. *IEE Proceedings-Control Theory and Applications*, 144(1):87–94, 1997.

[6] Hough Line Transform — OpenCV 2.4.13.2 documentation.

[7] Bernhard Scholkopf, Kah-Kay Sung, Christopher JC Burges, Federico Girosi, Partha Niyogi, Tomaso Poggio, and Vladimir Vapnik. Comparing support vector machines with Gaussian kernels to radial basis function classifiers. *IEEE transactions on Signal Processing*, 45(11):2758–2765, 1997.

[8] Shuihua Wang, Hangrong Pan, Chenyang Zhang, and Yingli Tian. RGB-D image-based detection of stairs, pedestrian crosswalks and traffic signs. *Journal of Visual Communication and Image Representation*, 25(2):263–272, 2014.

[9] Young Hoon Lee, Tung-Sing Leung, and Gérard Medioni. Real-time staircase detection from a wearable stereo system. In *Pattern Recognition (ICPR), 2012 21st International Conference On*, pages 3770–3773. IEEE, 2012.

[10] Yalin Xiong and Larry Matthies. Vision-guided autonomous stair climbing. In *Robotics and Automation, 2000. Proceedings. ICRA'00. IEEE International Conference on*, volume 2, pages 1842–1847. IEEE, 2000.

[11] The Bresenham Line-Drawing Algorithm.

[12] Steve Mann and Rosalind W. Picard. Video orbits of the projective group a simple approach to featureless estimation of parameters. *IEEE Transactions on Image Processing*, 6(9):1281–1295, 1997.

[13] Robert M. Manning and Mark Adler. Landing on Mars. 2005.

[14] Wiley J. Larson and James Richard Wertz. Space mission analysis and design. Technical report, Microcosm, Inc., Torrance, CA (US), 1992.

[15] S. Alan Stern. The low-cost ticket to space. *Scientific American*, 308(4):68–73, 2013.

[16] Michael Simpson. Spin-Out and Spin-In in the Newest Space Age. 2010.

[17] Jeanie Kayser-Jones, Ellen Schell, William Lyons, Alison E. Kris, Joyce Chan, and Renee L. Beard. Factors that influence end-of-life care in nursing homes: the physical environment, inadequate staffing, and lack of supervision. *The Gerontologist*, 43(suppl_2):76–84, 2003.

[18] Tiia Ngandu, Jenni Lehtisalo, Alina Solomon, Esko Levälahti, Satu Ahtiluoto, Riitta Antikainen, Lars Bäckman, Tuomo Hänninen, Antti Jula, Tiina Laatikainen, and others. A 2 year multidomain intervention of diet, exercise, cognitive training, and vascular risk monitoring versus control to prevent cognitive decline in at-risk elderly people (FINGER): a randomised controlled trial. *The Lancet*, 385(9984):2255–2263, 2015.

[19] Miriam E. Nelson, Jennifer E. Layne, Melissa J. Bernstein, Andrea Nuernberger, Carmen Castaneda, David Kaliton, Jeffrey Hausdorff, James O. Judge, David M. Buchner, Ronenn Roubenoff, and others. The effects of multidimensional home-based exercise on functional performance in elderly people. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, 59(2):M154–M160, 2004.

[20] Shu-Chuan Jennifer Yeh and Sing Kai Lo. Living alone, social support, and feeling lonely among the elderly. *Social Behavior and Personality: an international journal*, 32(2):129–138, 2004.

[21] Frerik Andriessen. Internship Report Stairclimbing Robot. Technical report, Delft University of Technology, Delft, 2016.

[22] Yang Cheng, Andrew Johnson, and Larry Matthies. MER-DIMES: a planetary landing application of computer vision. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1, pages 806–813. IEEE, 2005.

[23] Shuang Li, Pingyuan Cui, and Hutao Cui. Vision-aided inertial navigation for pinpoint planetary landing. *Aerospace Science and Technology*, 11(6):499–506, 2007.

[24] Ioannis Kostavelis, Lazaros Nalpantidis, Evangelos Boukas, Marcos Aviles Rodrigalvarez, Ioannis Stamoulias, George Lentaris, Dionysios Diamantopoulos, Kostas Siozios, Dimitrios Soudris, and Antonios Gasteratos. Spartan: Developing a vision system for future autonomous space exploration robots. *Journal of Field Robotics*, 31(1):107–140, 2014.

[25] Brent E. Tweddle, E. Muggler, Alvar Saenz-Otero, and David W. Miller. The SPHERES VERTIGO Goggles: Vision based mapping and localization onboard the International Space Station. In *Proc. iSAIRAS*, 2012.

[26] Kevin van Hecke, Guido CHE de Croon, Daniel Hennesb, Timothy P. Setterfieldc, Alvar Saenz-Oteroc, and Dario Izzob. Self-supervised learning as an enabling technology for future space exploration robots: ISS experiments. 2016.

[27] Yung-Kai Lai and C.-C. Jay Kuo. A Haar wavelet approach to compressed image quality measurement. *Journal of Visual Communication and Image Representation*, 11(1):17–40, 2000.

[28] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. Microsoft coco: Common objects in context. In *European Conference on Computer Vision*, pages 740–755. Springer, 2014.

[29] G. T. Shrivakshan, C. Chandrasekar, and others. A comparison of various edge detection techniques used in image processing. *IJCSI International Journal of Computer Science Issues*, 9(5):272–276, 2012.

[30] J. I. Liang, Jim Piper, and Jing-Yan Tang. Erosion and dilation of binary images by arbitrary structuring elements using interval coding. *Pattern Recognition Letters*, 9(3):201–209, 1989.

[31] Jack E. Bresenham. Algorithm for computer control of a digital plotter. *IBM Systems journal*, 4(1):25–30, 1965.

[32] Robert E. Schapire. The boosting approach to machine learning: An overview. In *Nonlinear estimation and classification*, pages 149–171. Springer, 2003.

[33] Robert E. Schapire and Yoram Singer. Improved boosting algorithms using confidence-rated predictions. In *Proceedings of the eleventh annual conference on Computational learning theory*, pages 80–91. ACM, 1998.

[34] Holger Schwenk and Yoshua Bengio. AdaBoosting neural networks: Application to on-line character recognition. In *International Conference on Artificial Neural Networks*, pages 967–972. Springer, 1997.