People Tracking in Outdoor Environments

Evaluating the Kinect 2 Performance in Different Lighting Conditions

People Tracking in Outdoor Environments

Evaluating the Kinect 2 Performance in Different Lighting Conditions

Ruben Seggers
6088473

Bachelor thesis
Credits: 18 EC

Bachelor Opleiding Kunstmatige Intelligentie

University of Amsterdam
Faculty of Science
Science Park 904
1098 XH Amsterdam

Supervisors
dhr. R. Bakker MSc
dhr. dr. A. Visser

Informatics Institute
Faculty of Science
University of Amsterdam
Science Park 904
1098 XH Amsterdam

June 26th, 2015

Abstract

Robots have proven to be useful for a wide range of industrial and commercial applications, principally in controlled, indoor environments. However, there are many outdoor scenarios where the use of robots could be beneficial, for instance by reducing the cost of labour-intensive tasks. The outdoor environment brings its own challenges and complications. For instance, when operating in an outdoor environment, lighting intensity is more dynamic, reducing the robustness of visual perception. People tracking and following is considered a valuable feature in robotics. This can be achieved with a HOG descriptor, which is at the moment considered one of the most robust solutions to detect people. Another recent development is the upward availability of affordable depth cameras, most notable the Kinect for Windows, of which the second version was released in September 2014. This research evaluates the performance of the HOG descriptor on both the colour and infrared images produced by the Kinect 2 for Windows in various lighting conditions. It was shown that the performance was higher in all conditions when using the infrared images. Furthermore, the algorithm was extended with a depth filter using the depth data of the Kinect 2. This improved the algorithm significantly by reducing the number of false positives, resulting in a feasible solution for outdoor person tracking.

Contents

1 Introduction

Robots have proven to be useful for a wide range of industrial and commercial applications, principally in controlled, indoor environments.[10] However, there are many outdoor scenarios where the use of robots could be beneficial, for instance by reducing the cost of labour-intensive tasks.[6] Some robots intended for outdoor use are indeed becoming popular, e.g. drones for recreational use and video capturing.[16] The outdoor environment brings its own challenges and complications. For instance, when operating in an outdoor environment, lighting intensity is more dynamic, reducing the robustness of visual perception. Another potential issue is uneven terrain, which results in movement of the sensors and therefore less stable image data. Furthermore, occlusions are more likely and of more complex form: since plants and bushes are often shaped complexly compared to objects encountered in indoor environment, e.g. tables and chairs. This calls for robust techniques performance adequately under these challenging circumstances.

1.1 Hortimotion

The topic of this research arose as a use-case of the Hortimotion robot. The start-up company Hortimotion produces this robot of the same name intended to reduce the labour cost by automating labour-intensive tasks on small- to middle-sized greeneries. This enables horticulturists to produce high-quality products at a lower cost, without the use of harmful pesticides. The Hortimotion is still in the prototype stage, and possible applications are now being explored. One of these applications is using the Hortimotion as an assistant, i.e. assisting horticulturists in tasks the robot cannot perform autonomously. For instance, the Hortimotion could follow a person and supply that person with tools or plants, eliminating the cumbersome task of drawing a heavy cart through the greenery.

1.2 People tracking

Detecting people in images is valuable in a wide variety of areas, such as visual surveillance, autonomous and auxiliary vehicle driving, human interaction, domestic service applications and image understanding. While image processing techniques to track people have been developing rapidly and with great success over the past years, there are still scenarios imaginable where the performance of these methods could perform substandard. This field of research will be described in more detail in section 2.

1.3 Research Goals

The main goal of this thesis is to explore the different available methods for the tracking of people in outdoor environments, with the aim to use this knowledge to propose a method that is optimal for this task. The secondary objective is to examine how the results obtained from testing the proposed method relate to the Hortimotion use-case.

1.4 Outline

This thesis is organised as follows. In section 2 related work is reviewed, exploring the methods currently available, in order to create a selection of algorithms that will be used for further experiments. The research method and its implementation are described in section 3. Section 4 lists the experiments conducted, and is where the produced results are evaluated. In section 5 conclusions are drawn based on this evaluation, both in respect to people tracking in general and to the Hortimotion use-case.

2 Related Work

The past years have seen increasingly rapid progress in the field of object recognition in general, and specifically vision-based person tracking. These advances were achieved with many different approaches that differ primarily in sensor type, feature extraction and the process of modelling motion.[12]

However, the tracking of objects can also be accomplished without people recognition by the use of beacons[15]. The most notable beacons in this context are: GPS trackers, RFID tags and visual markers. An example of a visual marker is a pair of coloured circles with a defined radius and distance between them. These methods require an extra item that has to be carried by the user at all times to enable tracking, which is both inconvenient and inflexible. Furthermore, these methods have additional complications: RFID performs poorly in tracking moving objects [11]; GPS is either very expensive with precise licensed sensors, or performs insufficiently; visual marker systems are extremely prone to occlusion or require numerous markers to ensure visibility from different angles.[2]

Another means of tracking people is by the use of static cameras. This is extensively implemented in various applications, most importantly surveillance camera systems. The methods used with static cameras almost all depend on background subtraction to define a region of interest; background subtraction detects displacement of objects. Sequentially, the found regions are examined for features of relevant objects. The use of multiple cameras ensures robust detection outdoors, since occlusions and problematic lighting in one field of view are likely to be absent from another perspective.[18] However, robots generally hold the benefits of autonomy, while the use of static cameras limits the area in which they are able to operate. Background subtraction is a challenging method for mobile platforms, since the movement of the camera results in a moving background.

Systems with (omnidirectional) colour cameras mounted on mobile platforms make use of different feature detection algorithms to determine the position of the person being tracked. These features can be based on (skin) colour, shape or gradients. The performance of these approaches differs in application and environment. The use of colour requires clothing of predefined hue or the presence of visible skin. This is cumbersome, although it can be combined with other features, e.g. by saving a histogram of colours after the person is detected by said features in an initialisation phase.[13] Since the shape of a person differs significantly between individuals as well as resulting from difference in angle, this method appears inferior. Presently, one of the most robust solutions for tracking people is a histogram of oriented gradients descriptor

(HOG descriptor) due to its invariance to geometric and photometric transformations.[8][5]

When coping with difficult lighting, as is often the case in outdoor environments, the performance of the HOG descriptor decreases significantly. Low light conditions in particular are a great complication for the use of any algorithm based on a colour camera. Methods that are based on depth mapping are more potent in this scenario. A powerful sensor to create such a depth map is a 3D laser range finder. 3D laser range finders are, however, prohibitively expensive and are often not available in mobile robot platforms, except for the Google car.[1] A more procurable sensor is the Kinect 2 for Windows. The Kinect 2 creates a depth-map by analysing infrared images of objects on which it projects a dense non-uniform array of infrared dots. This method has a downside: The presence of sunlight saturates the infrared image to such an extent that the projected depth map is near impossible to reconstruct, which results in the low performance in outdoor environments during daytime, at least for the Kinect 2's predecessor, the Kinect 1.[17][1]

Since the Kinect 2 also includes a colour camera, it can be used for people tracking using a HOG descriptor in bright lighting conditions. This can be combined with the depth-map analysis for the tracking task in low light conditions. The aim of this research is to evaluate if these algorithms prove to be robust in different lighting conditions encountered in an outdoor environment, possibly by combining them by the use of a fusion algorithm.

3 Method

This section lists which hardware, software tools and libraries were used to conduct the experiments. Firstly, the specifications of the Kinect 2 for Windows are described. Secondly, the different software utilities used to both capture and modify the sensor data are explained and exemplified. Lastly, the means to control and evaluate the experiments are specified.

3.1 Kinect 2 for Windows

In November 2010 Microsoft introduced the first version of the Kinect RBG-D camera, the Kinect 1 for the Xbox 360 video game console.[3] Designed to be positioned below or above the display, it enabled the users to interact with the system through body and hand motion without holding or wearing sensors. Both the RGB and depth images created had a resolution of 640x480 pixels. The first software release to use the Kinect 1 with the computer, Kinect 1 for Windows, followed in December of the same year. This was the fist time computer vision had played a such a great role in a mass-market product intended for consumer use.[9]

Besides its original purpose, the Kinects also enabled scientific research in many fields, including computer vision an robotics. For example by using the Kinect for autonomous navigation, to reconstruct detailed models of indoor environments, and to allow surgeons to examine patient's CT and MRI scans while performing surgery without the need to disinfect peripherals.[9]

The Kinect 2 for Windows is the second major version of the Kinect product line, which was released in September 2014. Both in hardware and software various improvements were made

[1] `http://googlesautonomousvehicle.weebly.com/technology-and-costs.html`

to increase performance in several aspects and thus enhance the user experience; and most likely increase performance in other applications as well. The Kinect for Windows 3 sensor contains both a colour camera and an infrared camera. The colour camera has a resolution of 1920x1080; a field of view of 84.1°and 53.8°; and focal distances of 1036.32±3.72 and 1030.4±3.87. The infrared camera has a resolution of 512x424, a field of view of 70.6°and 60.0°; and focal distances of 364.15±1.45 and 362.40±1.45. Al values in the preceding enumeration were respectively the horizontal and vertical directions. Because of these differences and other distortion coefficients, the camera has to be calibrated to obtain appropriate transformations to be applied in order to align the colour, infrared and depth images.[7]

The Kinect infrared camera detects infrared light with a wavelength between approximately 827 and 850 nm.[2] This is a much smaller range than used by conventional infrared cameras used for night vision. This makes sense since the Kinect uses the camera to record the projected dots of a certain wavelength to construct the depth map, while night visions cameras aim to capture as much light as possible to capture the scene in low-light conditions. The Kinect 2 uses several algorithms to create a depth map, most notable a time-of-flight algorithm. This results in much hihger resolution depth data compared to the structured light method of the Kinect 1, which is very similar to triangulation used by in stereo-vision cameras.
Note that from this point on 'Kinect' will refer to the Kinect 2 for Windows.

3.2 Libfreenect2

To access the data generated by the Kinect, the open source driver libreenect2 for Kinect for Windows 2 devices was used. this driver supports[3] colour image, infrared image and depth image transfer. The libreenect2 driver is developed and maintained by OpenKinect[4], and open source community with the main purpose of enabling the use of the Xbox Kinect 1 and 2 on Linux, Mac and Windows.

3.3 ROS

The Robot Operating System[5] (ROS) is a framework that enables rapid development due to the numerous functionalities it offers in the form of tools, libraries, and conventions. Through simplifying various solutions to problems that occur in robotics applications frequently, it helps to create robust software for many purposes. ROS was designed specifically for groups of different expertise to collaborate and build upon each other's work.

Since the architecture of ROS is highly modular, it grants the advantage of straightforward including and excluding functionalities from ROS, keeping the system as lightweight as possible. Furthermore, it facilitates easy implementation of functionalities written by other ROS programmers. Another considerable advantage is that the modules can operate on different systems and communicate on the standard Internet Protocol. This enables parts of the program that

[2] https://social.msdn.microsoft.com/Forums/en-US/e92e6f9b-4800-4b48-8ae7-5c8b1353d661/infrared-wavelength?forum=kinectv2sdk

[3] https://github.com/OpenKinect/libfreenect2

[4] http://openkinect.org/

[5] http://www.ros.org/

require high computational power, such as image processing, to run on computers power while simultaneously other modules can run on a robot with the means of for instance perception and actuation.

In this thesis the ROS Indigo Igloo version was used, because this is the currently supported distribution release that is currently better documented and for which considerably more modules are available than the younger supported distributions. This release is also targeted at the Ubuntu 14.04 LTS distribution, the system used for this project.

3.4 IAI Kinect2

IAI Kinect2[6] is a collection of tools and libraries for the ROS Interface to the Kinect. It contains a calibration tool for calibrating the IR sensor of the Kinect to the RGB sensor and the depth measurements. This tool enables was designed specifically for the Kinect 2 using OpenCV, which is described in the next subsection. It enables almost effortless calibration, since the process is highly automated. It makes use of an included chess or circle board image, of which the latter was used for this research. The image has to be printed in the right size and subsequently placed in the view of the Kinect. While ensuring neither the Kinect nor the printed image change position, the calibration program is run. This process is repeated, varying the orientation and distance of the calibration image in respect to the Kinect.

Furthermore, the IAI Kinect package provides communication between the libfreenect2 library and ROS, a viewer for images and point clouds and a library for depth registration with OpenCL support. The use of OpenCL enables computation using the GPU and therefore granting a higher frame rate, from approximately 5 to 25 frames per second on the machine used in this research.

3.5 OpenCV

OpenCV[7] is an open source computer vision library, which centralises computational efficiency to enable use in real-time applications. This is achieved through its implementation in optimised C/C++ and the ability to take advantage of multi-core processing. Furthermore, it is enabled with OpenCL, through which it can take advantage of hardware acceleration of the underlying platform. OpenCV offers plentiful computer vision and machine learning algorithms including to recognise, identify, classify and track objects and persons. In addition, the library offers tools to extract, modify and process 3D models and is able to process 3D point clouds as produced by the Kinect.

In this research, OpenCV was used for various minor functionalities, such as reading, writing, resizing and cropping images. It was also used to measure the relative illuminance, which is further explained in subsection. However, the particular function that was of great importance in this research, is its fast implementation of the HOG descriptor, which is described in the next subsection. The version of OpenCV that is used for this thesis, is 2.4.8, since OpenCV 3 is not yet compatible with the used ROS and IAI Kinect2 libraries.

[6] `https://github.com/code-iai/iai_kinect2`
[7] `http://opencv.org/`

3.6 Histogram of Oriented Gradients

The Histogram of Oriented Gradients (HOG) descriptor is a feature descriptor proposed by Navneet Dalal and Bill Triggs in 2005.[8] It is an algorithm that extract features from an image, which can be used to train object detection algorithms. It generalises properties of objects by rendering various conditions invariant in describing the feature. In the original research, a linear support vector machine (SVM) was used for human detection as a test case.

The HOG descriptor returns global rather than local features, in contrast to for instance the SIFT or SURF feature detectors.[14][4] This means objects are represented by a single, global feature vector, in stead of a collection of smaller, local features representing different parts of the object.

In short the algorithm is as follows: In the first stage colour and gamma are normalised. Subsequently the descriptor first slides a window of size 64x128 pixels over the image. At each position the window is divided in cells of size 8x8 pixels. Of each pixel in this block, the gradient vector is calculated and stored in a 9-bin histogram. Since the gradient is unsigned, each bin is 20 degrees wide. This stage of the algorithm serves two purposes: 64 values are stored in nine values enabling faster training, and small changes in gradient values will contribute less to the overall histogram value, generalising the information of the block. These histograms are normalised in magnitude, which makes them robust to variation in illumination. In the next stage, groups of 2x2 cells are combined to form blocks. This means the windows is divided in 15 blocks in the vertical direction and 7 in the horizontal direction. The histograms of the cells in these blocks are concatenated, which results in vector of 36 values. These blocks overlap each vertically and horizontally surrounding block by half its size, and are normalised in respect to the blocks they overlap. This means cells are contributing to the feature vector several times with different weightings, which increases invariance to illumination and contrast with respect to the background.

The final descriptor of the detection window contains 105 of vectors of 36 values, resulting in a single vector of 3,780 values for every window. Every vector is passed to the SVM detector, which returns whether the image contains a person or not. After the image is analysed, the image size is reduced by a scale parameter and the process is repeated. This results in a relatively larger window, detecting taller persons as well as persons closer to the camera.

3.7 Depth filter

To improve the performance, the depth data was used to verify whether detections of the HOG descriptor were justified with the following algorithm: In the first stage, the area where the HOG descriptor detected a person is projected on the depth image provided by the IAI Kinect tool. Subsequently, the areas is reduced in size by a factor 10. This area is translated upwards by 30 percent of the original height of the area in order to be positioned on the torso of the detected person, considering this is the part of the human body most likely to provide correct depth information. The average pixel value of this area is now considered to be the distance from the detected person. In the next stage, the focal length f of the camera and the distance are to calculate the angular extent α:

$\alpha = 2 * arctan(\frac{ds}{2f});\ d = distance$

The angular extend relates linearly with the actual size of the person. When the the size of the object is 10% smaller or larger than a predefined value, the detection is regarded incorrect and thus rejected.

3.8 Infrared & lighting conditions

As described earlier, sunlight differs from light in indoor environments in several aspects. The electromagnetic spectrum is normally divided into ultraviolet, visible, and infrared light. Different light sources produce different distributions over this spectrum. Furthermore, the spectrum of sunlight changes over time, since the amount of atmosphere it traverses changes with the angle of incidence and different layers of the atmosphere absorb different ranges of the electromagnetic spectrum. The overall intensity of the light varies as well: light in indoor environments is of relatively moderate and stable intensity, whereas sunlight may vary due to overcast and again the angle of incidence.

To simulate sunlight in a controlled procedure, experiments need to be conducted in an environment obtained from light sources. To produce light with a similar spectrum of sunlight, floodlights with 300W tungsten halogen lamps were used accounting for 4000 lumen each[8]. As can be seen in figure 1, halogen lamps create light that is continuous in spectrum and, relative to other lamps, similar to sunlight in distribution over the spectrum.

The light intensity was measured in the quantity illuminance is measured in lux (lx) and describes the total luminous flux incident on a surface per unit area; incident light is simply light reflecting from a surface. It is determined using the luminosity function to correlate with human brightness perception. This is appropriate for colour cameras, since they are constructed correspondingly to resemble the human vision as closely as possible. To measure the illuminance, a Samsung Galaxy S6 with the Light Meter [9] app was placed next to the Kinect, which has a maximum error of 5%.

3.9 Labelling

In order to evaluate the performance of the algorithm, the results of the HOG descriptor were to be quantised. The frames were labelled in terms of true positives: the actual person was detected; false positives: something other than the person was classified a a person; and false negatives: the person was not found on the frame, while it was completely visible on the image.

With these results, the ratio of correctly labelled frames is calculated: These are the frames where solely the person was detected and no false detections occurred, and additionally the frames where no person was detected while there was no person visible on the images. If the subject was partially on screen, it should not be classified, since the HOG descriptor is trained on images where the person is entirely visible. Lastly the F1 score, the harmonic mean of precision and recall, is calculated from these results as well, since this yields a better representation of the algorithms performance, as will be clarified in the evaluation.

[8] Philips Halogen Double Ended Linear RS7 300W T3 CL 2BC (3222 640 56171)

[9] `https://play.google.com/store/apps/details?id=com.bti.lightMeter&hl=en`

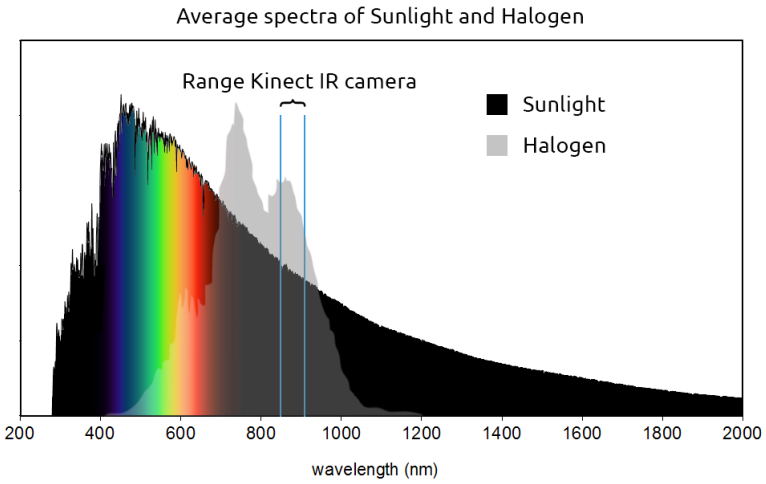

Figure 1: Average spectra of sunlight and halogen

The labelling was performed manually, observing one colour and one depth frame at a time. Since the infrared images were recorded at a higher frame rate than the colour images, i.e. approximately 10 and 20 frames per second, the colour frames were opened in order of recording, subsequently finding the infrared image with the closest created at time in milliseconds.

4 Experiments

4.1 Set-up

The experiments were conducted with the aim to exclude variance in other determinants as much as possible. To accomplish this, a track was plotted on the floor using markers and duct tape, in order to ensure that the filmed subject would traverse an equal path through the room in all recordings, consequently producing similar images to be analysed. This path was chosen in such a manner, that the subject is recorded from all perspectives: namely the front, side and back view. Similarly, the track ensures images with the subject on varying distances from the Kinect, varying from 0 to 5 meters.

Furthermore, the Kinect and the floodlights were placed at fixed positions, which can be seen in figure 2. The location where the experiments were conducted is the Game Studies lab, situated at Amsterdam Science Park. This location is exceedingly suitable, since all windows are equipped with darkening curtains and there are relatively few reflecting surfaces. This allows almost full control of the lighting intensity using the floodlights. The ceiling is also provided with vertical panels, which also contributes to eliminating reflections.

The floodlights were placed at four positions, as shown in figure 2. Two groups of two floodlights each were directed to the ceiling, which provided diffuse light increasing the overall lighting of the environment. Furthermore, two floodlights in the back and one floodlight in the front provided direct light into the camera, simulating direct sunlight as it would be perceived

outdoors in certain conditions.

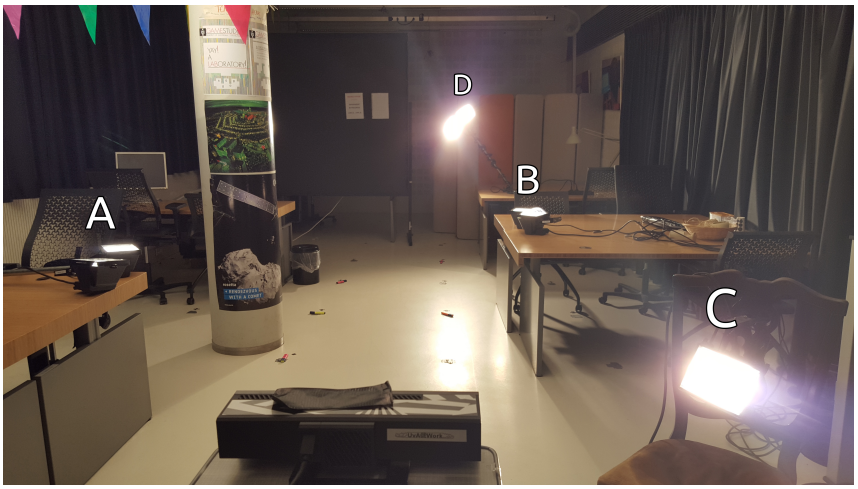

Figure 2: Overview of the Game Studies lab at the University of Amsterdam. The four groups of flashlights are labelled A-D.

Four recordings were made in different lighting conditions, producing a set of images from the Kinects colour and infrared camera. Additionally, a depth cloud was recorded and used to created depth images using the IAI Kinect2 library. These are grey scale images where the pixel values represent the distance of every corresponding pixel of the colour image and infrared images, i.e. the pixels on the same x and y location of the image. Each recording was about 20 seconds in length and the frame rate 10 frames per second, providing approximately 200 images per recording. In total 1677 frames were labelled. The first recording (i) was in almost total darkness; in the second recording (ii) the lights providing diffuse light were turned on; in the third recording with both diffuse light and direct light into the camera; and the fourth recording was shot with all the lights turned on.


fps: 14.5349 (68.8 ms)


fps: 16.1905 (61.7647 ms)


fps: 16.7818 (59.5882 ms)


fps: 21.093 (47.4091 ms)

4.2 Results

The results listed in this section are divided in two sets. In the first set, table 1 and figure 1, the performance of the HOG descriptor is described. The results listed in the second set, table 2 and figure , are of the HOG descriptors performance after filtering the results using the depth filter.

Table 1: Results of HOG descriptor before depth filtering

			Ratio Properly Labeled Frames		F1 score	
Experiment	Floodlights	Illuminace (lx)	Colour	Infrared	Colour	IR
i	none	0	0.338	0.662	0.027	0.737
ii	AB	13	0.728	0.832	0.761	0.851
iii	ABC	26	0.209	0.874	0.441	0.881
iv	ABCD	44	0.488	0.813	0.480	0.832

Table 2: Results of HOG descriptor after depth filtering

			Ratio Properly Labeled Frames		F1 score	
Experiment	Floodlights	Illuminace (lx)	Colour	Infrared	Colour	IR
i	none	0	0.374	0	0.813	0.845
ii	AB	13	0.759	0.793	0.834	0.881
iii	ABC	26	0.791	0.775	0.935	0.945
iv	ABCD	44	0.704	0.673	0.846	0.856

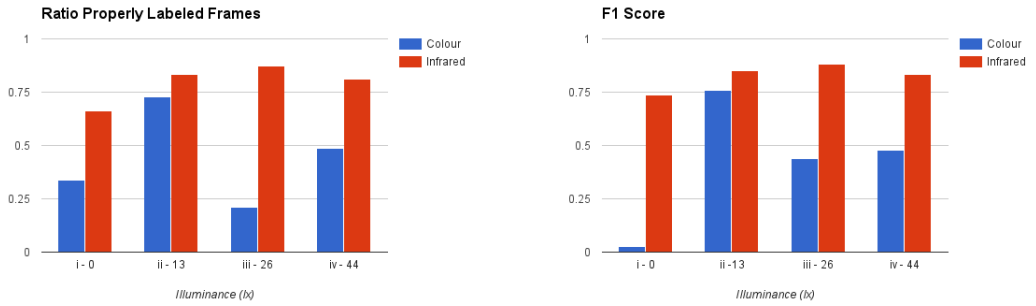

Figure 3: Results of HOG descript before depth filtering

4.3 Evaluation

In the first set of results, where no depth filtering was applied, it is shown that in low-light conditions the HOG descriptor was able to correctly classify over one quarter of the analysed images.

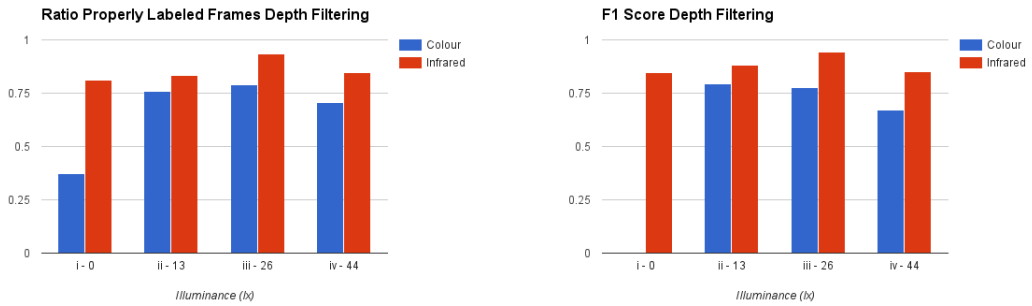

Figure 4: Results of HOG descriptor after depth filtering

However, these are the frames where no person was in the image, and the HOG descriptor therefore correctly classified it as a true negative. Hence the F1 Score provides a much better metric for evaluation, since the recall is nearly zero in this situation.

Not surprisingly, the performance of the HOG descriptor was much higher with the use of infrared images. It should be noted that this high performance is also achieved because of to the infrared lighting the Kinects infrared projector produces. The infrared projector of the Kinect illuminates the scene with light of the spectrum that the infrared camera records, resulting in very bright images even without other light sources.

The performance of the HOG descriptors on the colour and infrared images is nearly the same when there is enough diffuse light for the colour camera to capture the scene. However, when there is bright light aimed directly towards the camera, the performance of the HOG descriptor on the colour image decreases greatly. An intuitive explanation would be that the light distorts silhouettes and therefore decreases the chance of a correct classification. However, the number of true positives does not decline; the major cause for this decrease in performance is the increase of false positives. This seems to be a result of the many extra gradients that the light directed at the camera produces, resulting in complex shapes the HOG detector can classify as persons. The results of the HOG descriptor using the infrared images do not demonstrate such a decrease in performance. This can be explained by the fact that the Kinect records on a narrow band of the infrared spectrum, thus preventing bright light sources from overexposing the image.

When depth filtering is used to eliminate incorrect detection of the HOG descriptor, the performance of the HOG descriptor on both the colour and infrared images increases. This is due to the decrease of false positives that are eliminated since the depth data shows these detections would of persons that are unreasonably short or tall. This effect is much larger in when applied when using the colour camera, since it produced more false positive detections.

5 Conclusion

As expected, the HOG descriptor used with the Kinects infrared camera has a much higher performance compared to the HOG descriptor used on colour images in low-light environments. However, in moderate lighting conditions the performance does not differ significantly. When

light is directed directly at the camera, the colour image is distorted much more in comparison to the infrared image. Therefore, direct light results in a great decrease of performance when using the colour camera, but no significant change in performance was found when using the infrared image. Furthermore, using the Kinects depth map to eliminate false positives increases the performance for both the use of colour and infrared images, although this effect is much greater for the colour images. In all experiments the use of infrared images resulted in higher performance, and in low-light environments this difference is immense. Combined with the depth filtering algorithm, this appears robust enough to be a feasible solution for outdoor person tracking. The detections in the colour and infrared images could be combined as well, although this does not appear promising, since in none of the lighting conditions examined in this research the use of colour images resulted in a better performance.

6 Discussion & future work

To prove if this approach would hold in real world scenarios, it should be tested in actual direct sunlight in outdoor environments, since the intensity and spectrum of sunlight differs from halogen lamps. These experiments should be conducted at different times and preferably also in different weather conditions, since the spectrum and intensity of sunlight changes accordingly. Furthermore, this research did not focus on the computational power required for an implementation in a robot, i.e. method might be computational too expensive for certain commercial applications.

To implement this algorithm to enable the tracking and additionally following of people, as in the Hortimotion use-case, it should be extended to uphold performance in various scenarios. It should for instance be determined what the robot does when not a single person is detected, or what the do in the opposite case when more than one person is found. Another choice to be made is the maximum distance of a person to be followed; limiting this distance would also accelerate the detection algorithm. Furthermore, solutions have to be found to cope with situations where a person is detected, but the robot is prevented from reaching that person due to obstacles on its path. The HOG descriptor only detects persons that are completely within field of view of the camera, which, despite the great vertical field of view of the Kinect, will not be the case when the robot gets close to the followed subject. Other features, possibly defined on the go when the person is detected from a greater distance, could account for this. For instance, a colour histogram of the area returned by the HOG descriptor. However, this will only work when the person in question wears clothing of colours differing significantly from the environment.

7 References

[1] S. M. Abbas and A. Muhammad. Outdoor rgb-d slam performance in slow mine detection. In *ROBOTIK*. VDE-Verlag, 2012.

[2] M. N. A. Bakar and A. R. M. Saad. A monocular vision-based specific person detection system for mobile robot applications. *Procedia Engineering*, 41(0):22 – 31, 2012. International Symposium on Robotics and Intelligent Sensors 2012 (IRIS 2012).

[3] J. Ballester and C. Pheatt. Using the Xbox Kinect sensor for positional data acquisition. *American Journal of Physics*, 81:71–77, Jan. 2013.

[4] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool. Speeded-up robust features (surf). *Comput. Vis. Image Underst.*, 110(3):346–359, June 2008.

[5] H. Beiping and Z. Wen. Fast human detection using motion detection and histogram of oriented gradients. *Journal of Computers*, 6(8), 2011.

[6] J. Billingsley, A. Visala, and M. Dunn. Robotics in agriculture and forestry. In *Springer Handbook of Robotics*, pages 1065–1077. 2008.

[7] T. Butkiewicz. Low-cost coastal mapping using kinect v2 time-of-flight cameras. In *Oceans - St. John's, 2014*, pages 1–9, Sept 2014.

[8] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1, pages 886–893 vol. 1, June 2005.

[9] A. Fossati, J. Gall, H. Grabner, X. Ren, and K. Konolige, editors. *Consumer Depth Cameras for Computer Vision*. Springer, 2013.

[10] M. Hägele, K. Nilsson, and J. N. Pires. Industrial robotics. In *Springer Handbook of Robotics*, pages 963–986. 2008.

[11] M. Kim, N. Y. Chong, H.-S. Ahn, and W. Yu. Rfid-enabled target tracking and following with a mobile robot using direction finding antennas. In *CASE*, pages 1014–1019. IEEE, 2007.

[12] M. Kobilarov, G. Sukhatme, J. Hyams, and P. Batavia. People tracking and following with mobile robot using an omnidirectional camera and a laser. In *Robotics and Automation, 2006. ICRA 2006. Proceedings 2006 IEEE International Conference on*, pages 557–562, May 2006.

[13] M. Liem, A. Visser, and F. Groen. A hybrid algorithm for tracking and following people using a robotic dog. In *Human-Robot Interaction (HRI), 2008 3rd ACM/IEEE International Conference on*, pages 185–192, March 2008.

[14] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision*, 60(2):91–110, Nov. 2004.

[15] F.-R. Rideau and R. P. Goldman. Evolving asdf: More cooperation, less coordination. In *Proceedings of the 2010 International Conference on Lisp*, ILC '10, pages 29–42, New York, NY, USA, 2010. ACM.

[16] D. Schneider. Flying selfie bots. *Spectrum, IEEE*, 52(1):49–51, January 2015.

[17] J. Suarez and R. Murphy. Using the kinect for search and rescue robotics. In *Safety, Security, and Rescue Robotics (SSRR), 2012 IEEE International Symposium on*, pages 1–2, Nov 2012.

[18] Q. Zhou and J. Aggarwal. Object tracking in an outdoor environment using fusion of features and cameras. *Image and Vision Computing*, 24(11):1244 – 1255, 2006. Performance Evaluation of Tracking and Surveillance.