Evaluating the Azure Kinect for Monitoring Patient Movement During Brachytherapy

Part of the project: Dose and Motion Measurements in Brachytherapy

Broeckaert, F.V.A.M. (1227626)

***Abstract*— Cancer is a widely known illness with several treatment options available, one of which is brachytherapy (BT). However, during a BT treatment real time verification of the procedure is not yet available. Therefore the Dose and Motion Measurements in Brachytherapy (DETERMINE) project was set up by researchers from Maastro to use a combination of an EPID panel and a camera system to make real time treatment verification possible by monitoring the patient. This research is a sub-part of the DETERMINE project and will focus on evaluating the Azure Kinect to be used for the camera system. This camera system has to be able to detect patient movement and position relative to the panel.**

For the specific context, the unbinned narrow field-of-view was suggested and multiple experiments were performed using these settings. A depth estimation over time was conducted from which was concluded that measurement errors ranged between 2.4 and 12.5 mm, that there is a build-up time of 86 minutes before a stable output is generated and that temperature has no clear direct relation with depth estimation. Another set of experiments focused on the depth accuracy at different distances using a wall or different objects. Using the wall, a flat angle was used to determine the measurement error which resulted in a variation of $\sim$10 mm centering around zero. This suggests the earlier overestimation to be due to the angle, whereas the variation remains unaffected. For the objects where the camera was at an angle, different variations of overestimations were encountered for the different objects with the total ranging between 0 and 31 mm with standard deviations between 0.79 to 1.99 mm, except for the dark head phantom which was not detected properly by the camera. Lastly, the internal gyroscope output was evaluated and it was concluded that large drifts were present in all directions.

All in all, this study presented information on depth accuracy, behaviour over time and an evaluation of the internal gyroscope. Some marginal notes to explain observations during the depth accuracy experiments were offered, like the angle of the camera, (human) repeatability of experiments and accuracy of tools. Taking these marginal notes into account as well as including the encountered low standard deviations between frames, which seem promising for movement detection, further research into the Azure Kinect is advised to properly determine its suitability in the context of the DETERMINE project.

Study, cluster group:
Medical Engineering, Medical Image Analysis

Internship location:
Maastro, Dr. Tanslaan 12, 6229 ET Maastricht

Supervisors:
Frank Verhaegen (Prof. dr. ir., Professor)
&
Gabriel Paiva Fonseca (PhD, Assistant professor)
&
Teun van Wagenberg (MSc, PhD candidate)

Mentor:
Maureen van Eijnatten (Dr., Assistant professor)

Course:
Internship MIA (8ZM05)

Period:
03/05/2021 - 29/08/2021

Code and relating materials can be found on:
https://github.com/Francoiseb99/Maastro_internship

***Keywords*— Azure Kinect, Brachytherapy, Depth camera, Motion tracking**

I. Introduction

Cancer treatment is a widely known and discussed medical focus point for which several treatment options are available. One of these treatment options is radiotherapy and where external beam radiation therapy (EBRT) is known by most, there is another form of radiotherapy in clinical use called brachytherapy (BT). Opposed to EBRT, BT makes use of radioactive seeds. Common types of cancer for which BT is used, are for example skin, prostate or breast cancer [1]. BT can be applied in several ways differing in for example dose rate, duration and placement relative to the tumor. By using a radioactive seed, the dose fall-off is fast which means that only the area surrounding the seed will be exposed to a high dose, whereas other areas will only get a limited dose. Therefore exact placement of the seed is very important as this advantage of BT can turn into a disadvantage when the placement is inaccurate. If the seed is in the wrong place this means that the tumor will not (properly) be irradiated and healthy tissue will get an unnecessary high dose [1]. This can occur for instance due to patient movement or something going wrong with the applicator resulting in seed placement at the wrong site. Yet there is no equipment which can verify the treatment in real time and therefore such errors are mostly not seen in time by the medical personal [2] .

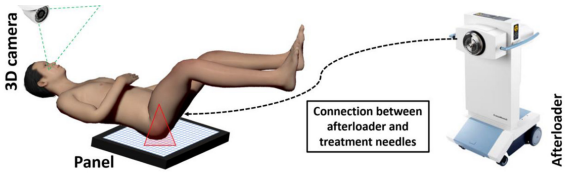


Fig. 1: This figure shows the proposed setup from the DETERMINE project. Reprinted from [3].

Researchers from Maastro have therefore taken up the challenge to use a combination of two different technologies to monitor the patient and process during BT treatments and propose this as the Iridium Imaging System (IrIS). Their project is called ”Dose and Motion Measurements in Brachytherapy (DETERMINE)” [2]. In this project, an Electronic Portal Imaging Devices (EPID) panel and a camera system will be used to detect possible errors in real time, see Fig. 1 for the proposed setup. The EPID panel is used to obtain approximate real time images of the position of the seed relative to some basic anatomical structures and the dosimetry, with the help of photons emitted by the high dose rate Iridium-192 source [3]. During EBRT treatment, a cone beam CT or lasers could be used to determine patient position relative to the panel, however, as these

are not available for BT, a camera system is chosen. The camera system will consist of one or multiple depth cameras, the camera being an Azure Kinect, and will be used to detect patient position and motion during treatment relative to the panel. IrIS is meant to detect deviations and alert the staff in case certain limits are exceeded which are likely to result in safety issues for the patient.

In this study attention will be given to the depth camera, the Azure Kinect. The Azure Kinect is able to provide RGB images, infrared (IR) images and depth images. The camera is based on a time-of-flight (TOF) principle meaning it works by illuminating the area and recapturing and analyzing the reflected IR data [4].

Prior research in the field of motion tracking is not limited to the use of the Azure Kinect and has been used in several ways. Its predecessors or other depth cameras / sensors have already been used in this field too. For example, in order to get real-time 3D respiratory motion detection (Penne *et al.*), breath hold monitoring and gating during breast cancer VMAT radiotherapy (Edmunds *et al.*) or pose tracking for gait analysis (Albert *et al.*) [5] [6] [7]. Each of these studies concluded the system to be applicable for the intended purpose, Penne *et al.* reported that available TOF cameras have a depth resolution up to 1 mm, their TOF modeling approach using best-fitting planes even resulted in a respiratory motion tracking accuracy of 0.1 mm [5]. Secondly, Edmunds *et al.* described that the Kinect V2 was able to track motion patterns having a root mean squared accuracy of roughly 1.5 mm [6]. Finally, Albert *et al.* more specifically investigated the upgrade from the Kinect V2 to the Azure Kinect and determined the Azure Kinect to have a significantly higher accuracy and similar temporal parameters [7].

Solely looking at the Azure Kinect on the other hand, the possibilities seem to exceed the healthcare field. Other areas opted by Microsoft are production, media, retail and robotics [8]. More specific examples of research being conducted would be using the Azure Kinect for urban tree diameter estimations (McGlade *et al.*) or making 3D human images to make virtual cultural heritage environments more realistic (Chalmers *et al.*) [9] [10]. McGlade *et al.* reported that for regular-shaped stems, a diameter at breast height error was found with RMSE = 3.53 cm, rRMSE = 0.11 and a bias of 1.89 cm (n =38) at an optimal distance of 2 m using the binned NFOV mode [9]. As for Chalmers *et al.*, this research was mentioned to describe the range of possibilities of the Azure Kinect, however, this research did not report relevant quantitative information on the performance of the Azure Kinect [10]. The multi-purpose proposition of the creator and the variety of research in which the Azure Kinect is used in combination with the encouraging results from multiple sources suggests this RGB-D camera to be very promising.

For the DETERMINE project, the Azure Kinect was chosen for the camera system to acquire color and depth images. In this study the Azure Kinect will be evaluated to see its suitability for IrIS. This study is a primary investigation and is meant to be a steppingstone intended for successors to build upon. For this reason, some experiments and actions performed during this study could later be determined to be either non-useful in the context of the DETERMINE project or are only intended to be useful for successors without, as of yet, being able to gain results from it. Multiple aspects will be discussed and evaluated in this study. Firstly, desirable settings will be determined based on its purpose and relating literature. Secondly there is the calibration, which can later on be used to for example overlay the color and depth image. Furthermore, the camera's qualities like depth estimation over time and accuracy will be evaluated based on experiments. Lastly, the gyroscope has been evaluated to see how stable this is. The aim of this study is to map as much relevant features and shortcomings of the Azure Kinect relating to motion detection during brachytherapy so this can be taken into account in the remaining part of the DETERMINE project and in the practical setting. In addition to this paper, code and a more elaborate manual like report can be found on the following GitHub page:
https://github.com/Francoiseb99/Maastro_internship

II. Methodology

This research mainly focused on evaluating the Azure Kinect and its settings for application for the DETERMINE project. More specifically the depth camera of the Azure Kinect has been evaluated. In this section the preferred settings for this purpose will shortly be described as they will be used in the experiments. Furthermore, the setup of the performed experiments will be explained. These experiments have focused on several relevant aspects of the camera which were relevant for IrIS to be mapped.

A. Settings

The chosen settings in this study are based on the estimated distance which will need to be measured and the ranges per setting given in the documentation of the Azure Kinect from Microsoft [11]. In this case this is the patient undergoing BT who is approximated to be 1.5-3.0 meters away from the camera, which will also be called moderate range. Translating this to the five possible settings only two had a range including the 1.5-3.0 meters, namely narrow field-of-view (NFOV) binned and NFOV unbinned. Since acquiring the unbinned version still gives the option of binning it afterwards, the NFOV unbinned was chosen to be used in the experiments.

B. Calibration

In order to calibrate the camera, a 6x8 checkerboard pattern with a size of 59.5 cm x 84.2 cm was used. 200 frames with the checkerboard pattern in different positions from the camera were taken after which a single camera calibration and a stereo camera calibration between the infrared and the color camera were performed. It should be noted that for this calibration different settings were used, namely binned wide field-of-view (WFOV), as the provided calibration code only worked for these specific settings. From this calibration the lens distortion could be obtained which in turn can be used to improve the obtained images. The calibration can also be used to overlay the depth and color images.

C. Experiment Depth Estimation Over Time

This section includes the depth estimation over time experiment which is meant to display the Azure Kinects behaviour over time. First the general experimental setup will be elaborated on, after which the determination of the advised warm-up time will be explained and a later addition of actually including the internal temperature data relative to time and depth estimation is included.

1) Depth estimation over time: In the paper of Tölgyessy *et al.* it was stated that the Azure Kinect has a warm-up time affecting its precision and this warm-up time was stated to be 40 to 60 minutes [12]. However, since accuracy and precision are very important in the context of patient tracking during BT, it was decided to redo this experiment taking into account both aspects. This was done by determining the average distance of a manually selected area of an object and ploting the estimated distance over time (the DepthOverTime function was used, see GitHub). The object used during this experiment was a box with dimensions 15.5 cm x 29 cm (x 6 cm). However, for each run, the selected area differed as it was done manually and the aim was to just take multiple pixels in

the middle of the box in order not to rely on only one pixel. Like mentioned in section II-A, the desired distance is between 1.5 and 3.0 meters, therefore the depth over time has been established for a range of distances between these two with an interval of 0.5 meters. To evaluate the repeatability of this experiment, three runs have been performed for the 1.5 m distance. To ensure accuracy of placement of the object opposed to the camera, a laser based measurement tool was used. This way not only precision over time could be evaluated, but also the accuracy. To make sure the camera was not warmed-up beforehand, the experiment was done at the beginning of the day before the camera was used for other purposes. A 1.000.000 frames were taken, translating to approximately 9 to 10 hours of footage. Such a big time window was chosen to ensure a good overview and not conclude anything based on a possible intermediate plateau. Pictures of the experimental setup are included in Appendix I-A.

2) Determination warm-up time: To determine a warm-up time for clinical use, a few steps were taken. First, for each of the measurements (minus 1500 mm run 3, as this one was not yet obtained) the difference between the actual encountered depth at each time point and the maximum and final depth were taken. Secondly, this was plotted and the average slope coefficient was determined. Lastly, the average slope coefficient was matched to the time point in the actual graph having the same slope coefficient, this will be called the inflection point. Based on the curve of the graph it can be stated that everything before this time point has a steeper slope coefficient. Therefore everything before this point is considered to be beneficial in the warm-up time whereas everything after does not give enough added value opposed to the extra time needed to be invested. To determine this optimum warm-up time a MATLAB function was used, namely the OptimumWarmUpTime function (see GitHub).

3) Temperature: For 1500 mm run 3, the internally measured temperature of the camera was taken during each iteration. This allowed for plots to be made relating time and temperature as well as temperature and depth to evaluate how this relation holds. This experiment was performed indoors at normal room temperature.

D. Experiment Depth Accuracy

This experiment was divided into two parts, namely determining the depth accuracy on a flat white wall and determining it on objects with either a flat or gradient surface. Both those experiments have been performed at moderate range, being between 1.5 m and 3.0 m with an interval of 0.5 m.

1) Depth accuracy on a flat white wall for moderate range: The first sub-experiment made use of a flat white wall and was meant to determine the measurement error for two modes, namely binned WFOV and unbinned NFOV. The second mode was included to evaluate potential differences. A white wall with landmarks to determine the center in the depth image was used, for pictures of this setup see Appendix I-B. The center and the alignment of the camera were verified by two cross line lasers. After placing the camera at the respective distances with the help of a laser measurement tool and verifying its position respective to the lasers, 25 frames were taken in each mode. Afterwards, the center in the depth image was determined manually using the landmarks. The coordinates of the center were used to determine the pixel value throughout all frames from which the mean and standard deviation could be calculated. Lastly, the difference of the average estimations by the camera and the set distances were taken to determine the measurement error.

2) Depth accuracy including flat and gradient object surfaces with different colors for moderate range: In this second sub-experiment two types of surfaces were used as well as two different colors for each. Firstly, there was the flat surface with the colors blue and red. Secondly, two 3D printed phantom heads were use, one being light of color and the other dark. The objects were placed at the moderate distances, for the head phantoms this distance was calculated from the nose, with the help of a laser measurement tool. In addition a cross line laser was used to make sure the objects, especially the heads, were placed at the correct position. For each setup, 25 frames were taken and saved for evaluation where the average estimated distance of a single pixel given by the camera could be compared to the distance measured by the laser measurement tool. For an illustration of the setup used during this experiment, see Appendix I-C.

E. Gyroscope evaluation

Lastly, the internal gyroscope was evaluated to see if it could be used for depth corrections in a later stage. In order to evaluate the stability of the gyroscope, three runs were performed. During the first two runs the camera was placed horizontally on a flat surface and during the third run the camera was placed vertically on a flat surface, see Appendix I-D. In these positions the camera was activated and the gyroscope output was taken and plotted cumulatively over time. This was done for approximately 250 minutes to see if the output would remain constant or would show a certain drift.

III. Results

In this section the results of the depth estimation over time, depth accuracy and gyroscope evaluation experiments will be displayed.

A. Experiment Depth Estimation Over Time

1) Depth estimation over time: In Fig. 2 two plots are shown, the first plot shows the measurement error over time for each of the distances which are given in the legend. The second plot shows the standard deviations of each run, as the upper plot is averaged per minute. A similar behaviour of the lines can be observed as they all present logarithmic behaviour; starting of with a somewhat steeper increase before seeming to stabilize. The measurement error differs per measurement and ranges between 2.4 and 12.5 mm. In addition, it can be seen that, with the exception of 2500 mm, the measurement error seems to decrease per distance. Furthermore, the 2500 mm run seems to not only be an exception to this, but also has a significantly smaller measurement error of 2.4-3.2 compared to the other five runs of which the values ranged between 7.9 and 12.5 mm. Lastly, it can be observed that the three runs at a distance of 1500 mm each differ slightly by a maximum of 1.4 mm.

2) Warm-up time: In Fig. 3 the average depth difference based on the maximum / final encountered depth, the average gradient line and the inflection point are shown along with the standard deviation in a second sub-plot. It can be seen that the the difference relative to the maximum and the final encountered values show the same trend. Starting at 0.8 mm, first a sharp decrease can be seen after which it slowly stabilizes around zero. In section II-C.2 it was explained how the inflection point, where the slope gradient is equal to the average gradient, can be found. This inflection point is found at 86 minutes which can be considered the recommended warm-up time.

3) Temperature: The last part of the depth estimation over time experiment shows the relation of the actual temperature respective to both time and depth estimation, this is shown in Fig. 4. Sub-figure 4a shows a sharp increase in the first half hour starting at 12.8 degrees after which it becomes more stable at 23-24 degrees, yet still showing some deviations and even some peaks. As for the second sub-figure 4b, the graph seems to increase at different rates starting at 1512.2 mm and without a clear trend going to 1512.5 mm, the standard deviation also grows larger as the temperature increases having a maximum of roughly 0.115 mm.

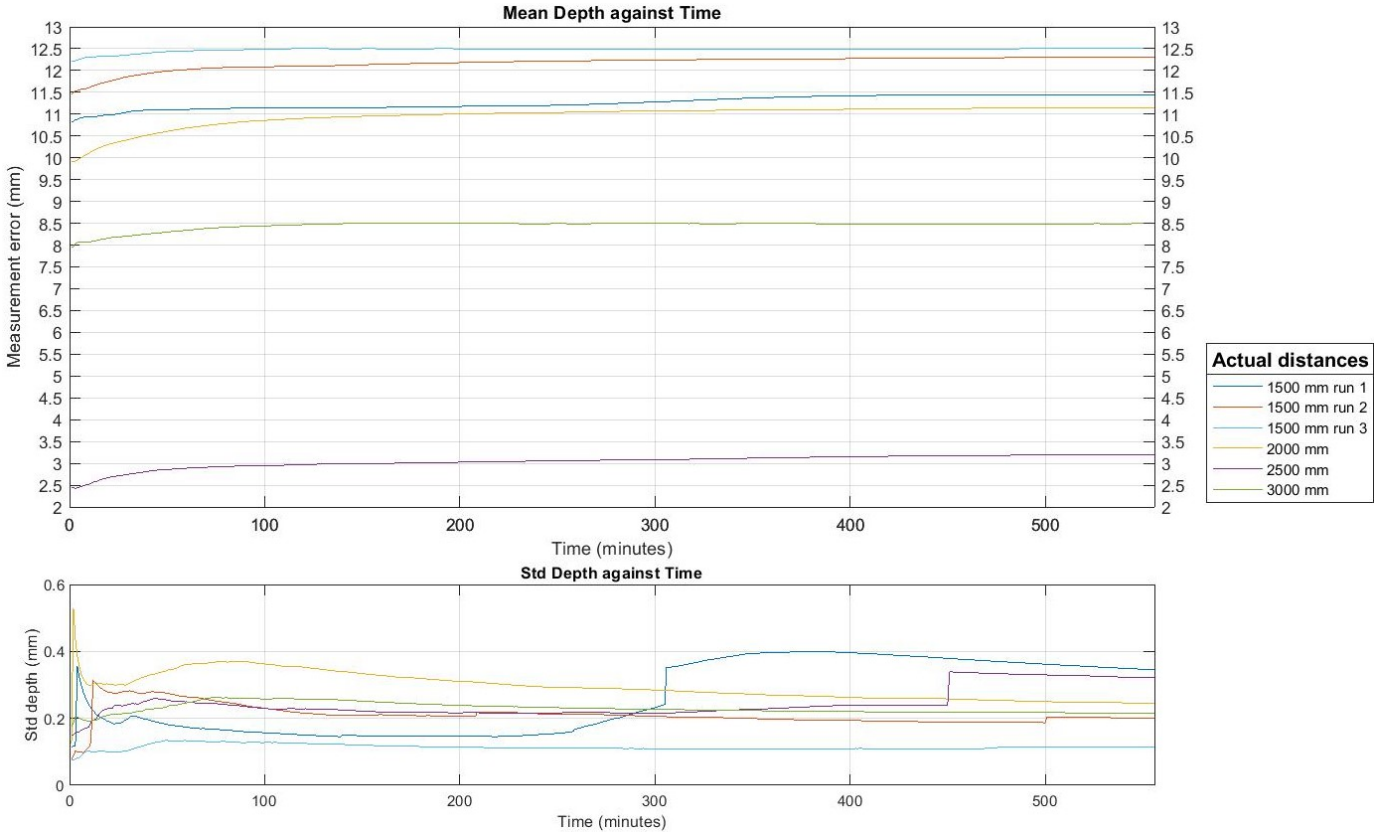


Fig. 2: This figure shows the results from the depth estimation over time experiment. The results in the first subplot are normalized by subtracting the actual distance from the measured distance with the Azure Kinect resulting in the measurement error. The actual distance per line is given in the legend.

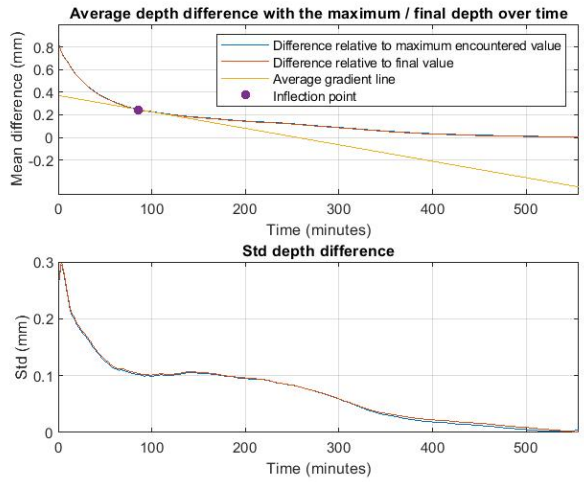


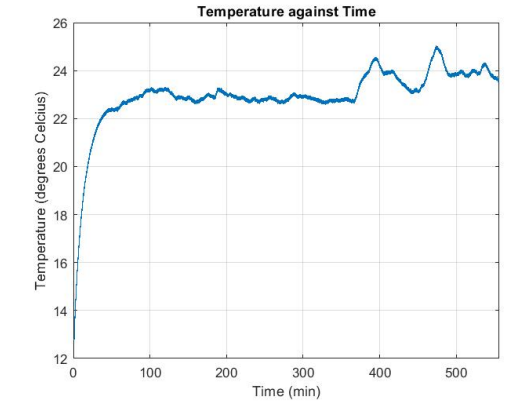
Fig. 3: This figure shows two plots. The first plot shows the average depth difference based on the maximum / final encountered depth value, an average gradient line and the point at which the gradient is equal to the average gradient (line). The second plot shows the standard deviation of the above plot.

B. Experiment Depth Accuracy

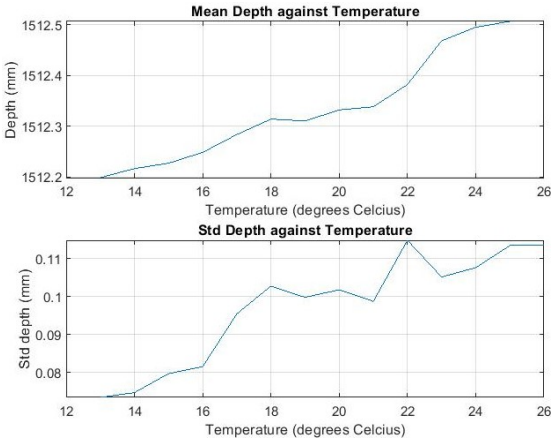
1) Depth accuracy on a flat white wall for moderate range: The first part of this depth accuracy experiment consisted of the measurement error at different distances from a flat white wall for two different modes of the Azure Kinect. The results from this experiment are shown in Fig. 5. For the binned WFOV mode it can be seen that the measurement error oscillates around zero for the different depths having its maxima at 3 and -3. On the other hand, the unbinned NFOV starts at 4 and for increasing depth the measurement error goes down to -6. In addition, whereas binned WFOV oscilates, unbinned NFOV only drops as the distance increases. For all points a standard deviation of 1 mm was encountered.

2) Depth accuracy including flat and gradient object surfaces with different colors for moderate range: Table I shows the results for the second part of the depth accuracy experiment. The first column shows the actual distance measured with the laser measurement tool. The second and third columns show the results from the depth camera on a flat surface, being blue and red respectively. For the second column deviations range between 2 and 6 millimeters relative to the actual distance with the standard deviation of the measurement itself increasing from 0.79 to 1.80 mm as the distance increases. For the third column this is a difference in the range of 0 to 22 millimeters with the standard deviation starting at 1.02 mm and increasing to 1.99 mm as the distance increases. Coming to the fourth column it can be seen that all values estimated by the camera are zero as well as their standard deviations, meaning it could not detect the nose of the dark head phantom. The last column shows the estimated distances obtained when using the light head phantom. Unlike the nose of the dark head phantom, the nose of the light head phantom does give results for all distances with deviations from the actual distance ranging between 18 and 31 millimeters. The standard deviations again increase as the distance increases, this time starting at 0.91 mm and going to 1.56 mm.

In Fig. 6 an illustration is given of how the depth is perceived by the



(a) Temperature against time



(b) Mean depth against temperature and its standard deviation

Fig. 4: These figures show the relation between temperature and time as well as the relation between temperature and depth estimation.

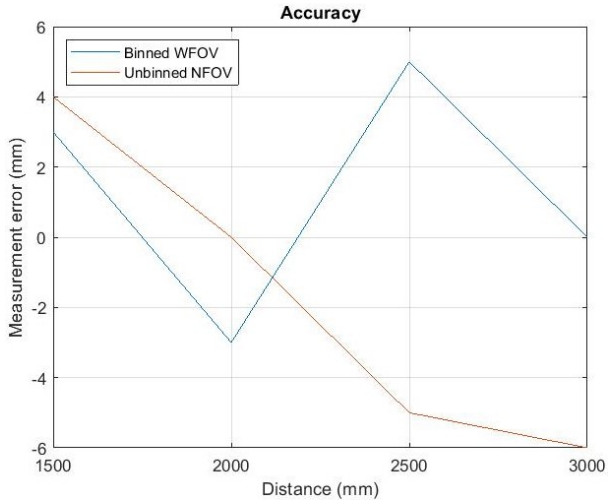


Fig. 5: This figure shows the measurement errors for two modes at different distances. The standard deviation was 1 mm for each point.

camera. It shows each of the objects presented in Table I at a distance of 1500 mm. The range of the color bar was set to range from 1450 mm to 1650 mm, to ensure nuances could be seen with the naked eye. For figures 6a and 6b it can be seen that differently colored blue spots are present on both blocks. These spots represent variation in depth estimates in the range of a few millimeters. In figures 6e and 6f it can also be seen that there is quite some uncertainty as the standard deviation map has a lot of white in it, meaning a high standard deviation is present in these pixels. For the figures containing the heads (6c and 6d), a larger difference is observed. Not only the nose, but the entire dark head is barely detected by the camera whereas the light head shows clear features of a head. In addition, it can be noted that for the dark head the standard deviation is low for most of the head where zero depth was estimated. For the light head the standard deviation map shows roughly the same uncertainty as with the two blocks.

Distance (mm)	**Estimated distance mean ± std (mm)**			
	Blue block	*Red block*	*Dark head*	*Light head*
1500	1506 ±0.79	1509 ±1.02	0 ±0.00	1518 ±0.91
2000	2004 ±1.01	2007 ±1.08	0 ±0.00	2019 ±1.21
2500	2502 ±1.68	2500 ±1.29	0 ±0.00	2518 ±1.32
3000	3005 ±1.80	3022 ±1.99	0 ±0.00	3031 ±1.56

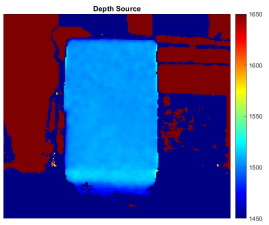
TABLE I: These are the results from the depth accuracy experiment for flat and gradient object surfaces with different colors. For the two blocks, the distance from the depth camera to the middle of the block is used. For the two heads, the distance from the depth camera to the nose are used.

C. Gyroscope evaluation

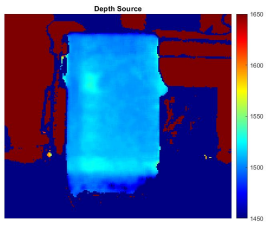
The last experiment involved evaluating the internal gyroscope of the Azure Kinect. In Fig. 7 a summation of the pitch, yaw and roll output of the gyroscope can be found for each of the three runs. Over time rotations varying between -524 and 820 degrees were observed. Most of the lines show a constant increase, however, some lines display some curvature. In addition, it can be noted that the respective lines of the pitch, yaw and roll do not overlap and therefore it can be determined that a different drift is present per run.

IV. Discussion

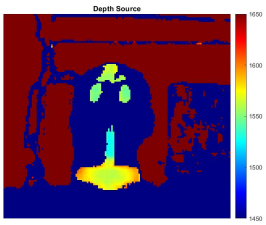
In section III-A, the results of the first set of experiments were presented. Figure 2 showed the measurement error over time where a trend was observed in which a larger distance seemingly had a lower measurement error. The only exception to this was the 2500 mm run which surpassed the 3000 mm in the mentioned trend and also showed a significantly lower measurement error compared to the other measurements. During all measurements, conditions were kept the same as much as possible. However, as the setup was placed manually it could be possible that the camera was misplaced by a few millimeters which could explain the deviating result from the 2500 mm run. Another observation made in this figure is the fact that three runs were performed at a distance of 1500 mm, each differing a few millimeters in output whereas the setup was the same. This could suggest that either the camera does not have an optimal precision or the experiment is not completely repeatable. However, since the difference is 1.4 mm at most, it is most likely that the experimental setup causes this difference. The setup relies on human repeatability and also makes use of a laser measurement tool as ground truth, the laser measurement tool having an accuracy of +/- 1.5 mm, which could very well explain the observations. Lastly, Fig. 2 gives an impression of the measurement error in general, meaning apart form the time. The measurement errors ranged between 2.4 and 12.5 mm overestimation which can be quite significant, especially in the DETERMINE context where patient movement should be monitored precisely to be effective. These measurement errors can



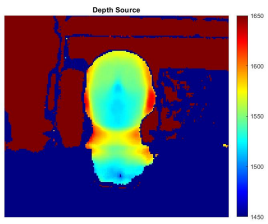
(a) Blue block depth map



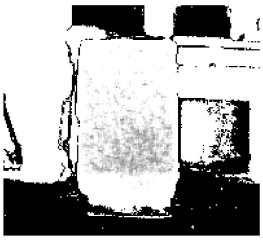
(b) Red block depth map



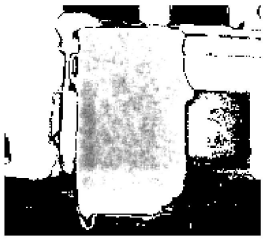
(c) Dark head depth map



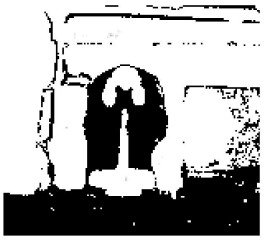
(d) Light head depth map



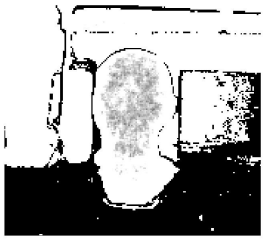
(e) Blue block std map



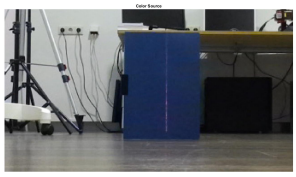
(f) Red block std map



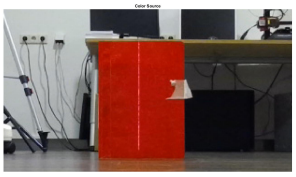
(g) Dark head std map



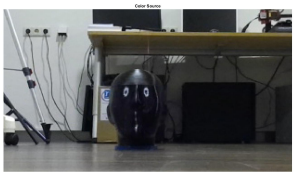
(h) Light head std map



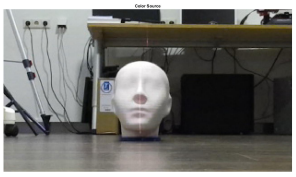
(i) Blue block color image



(j) Red block color image



(k) Dark head color image



(l) Light head color image

Fig. 6: The first row shows depth images taken during the depth accuracy experiment at a distance of 1500 mm. To be able to visually spot the nuances, the color range has been set to 1450 mm - 1650 mm. The second row visualizes a standard deviation map based on the 25 frames taken per run. The third row shows the RGB images of the objects.

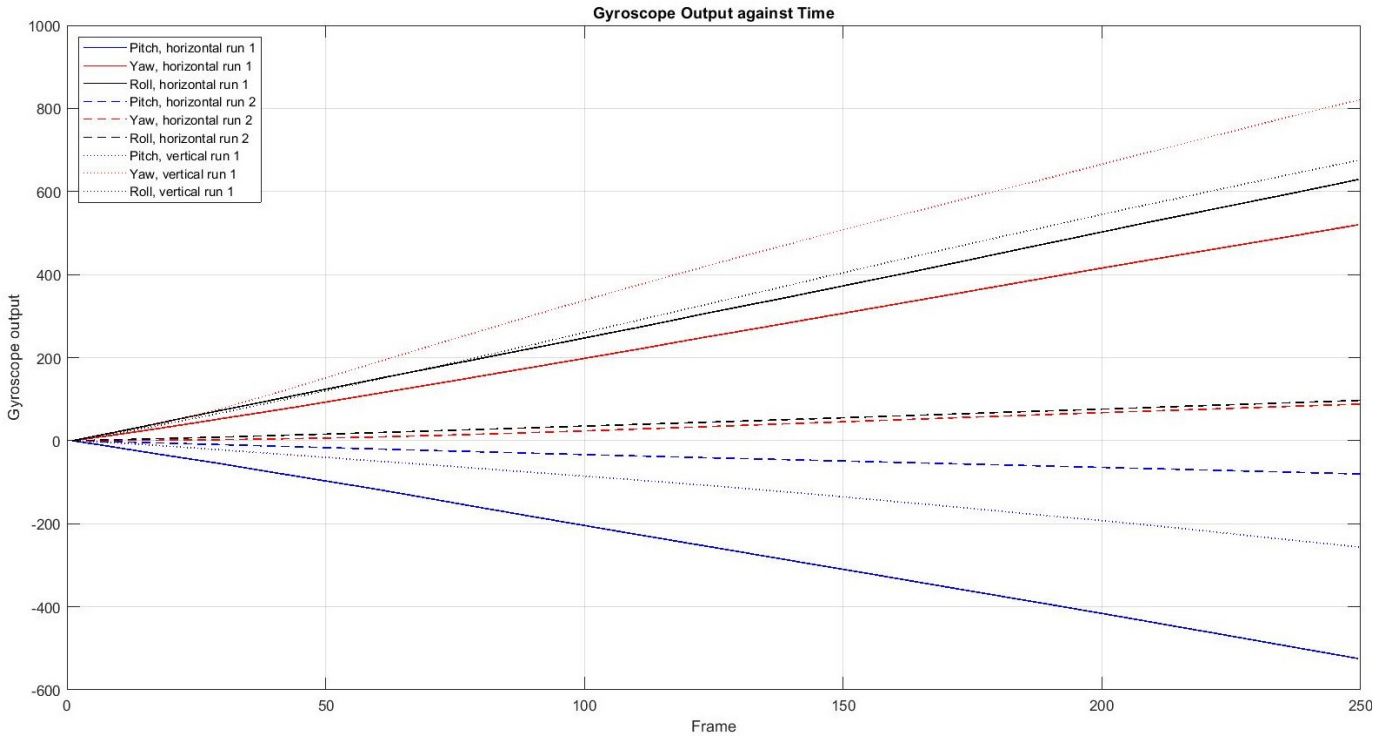


Fig. 7: This figure shows a summation of the output of the internal gyroscope for ∼250 minutes for each of the three runs.

however not fully be assigned to the Azure Kinect as there are a few factors which are possibly also playing a role in this. The earlier mentioned accuracy of the laser measurement tool as well as the fact that all experiments in this paper are subject to human error and human repeatability capability could both account for some error. Lastly and probably most importantly, the Azure Kinect is delivered with a stand which puts the camera at an angle. This was initially not considered and therefore used during the experiments, however, in retrospect this could cause a certain measurement error. Each of these aspects could have contributed to the observed measurement errors, but it is debatable whether or not it can account for everything so it is expected that some amount of the measurement error could still be assigned to the camera. Especially since the different distances report different measurement errors.

In section III-A.2 a warm-up time of 86 minutes was suggested based on multiple experiments. This warm-up time exceeds the warm-up time suggested in other literature like Tölgyessy *et al.* which suggested approximately 40-60 minutes. Their setup was however slightly different and their duration was only 80 minutes. Additionally, their definition of stabilization was not given and therefore can not be compared to this study [12].

The last part of this first set of experiments involved relating the actual temperature of the camera to both time and depth estimation. It was stated that, as expected, the camera warmed up during the run. However, the biggest increase happened during the first half hour after which it stabilized, which does not seem to correspond with the results in Fig. 2. In addition, sub-figure 4b also does not show a clear trend, suggesting that the warm-up time does not have a clear relationship with the actual temperature and other factors are involved.

The results of the depth accuracy experiment were presented in section III-B and comprised of two sub-experiments. The first sub-experiment showed the measurement error for two modes at different distances from a flat white wall. It is interesting as this is the only experiment where negative measurement errors, meaning underestimations by the depth camera, were encountered. This encourages the argument given earlier about the influence of the angle, as this is the only (sub-)experiment where depth estimations were made without making use of the stand and where the camera was positioned flat with the help of cross line lasers. The variety is still similar, as this was between 2.4 and 12.5 in Fig. 2 and in Fig. 5 for the same mode this was between -6 and 4. They both have a difference of approximately 10 mm, but the second one centers more around zero. The same experiment was performed by Tölgyessy *et al.*, however, they used the first measurement with the mode unbinned NFOV as reference and set this as 0 measurement error since they did not have a ground truth. In this experiment, the laser measurement tool was used as a ground truth. Their experiment was also more elaborate as they used all four modes, used a wider range of distances and used a smaller interval between the distances. As for comparison of behaviours between this paper and theirs for the same distance region, for binned WFOV the maximum encountered variation is similar (+/- 1 mm), unlike for unbinned NFOV where the results from Tölgyessy *et al.* only show a variation of 3 mm instead of the encountered 10 mm in this research. For the behaviour of the lines themselves, a similarity does not seem to be present between the respective results of this research and those of Tölgyessy *et al.* for either of the modes since increases and decreases do not appear between the same distances [12].

The second sub-experiment made use of four differently shaped and / or colored objects and the results were presented in both a table presenting a single pixel value in the center and in a set of images showing the depth estimate, standard deviation map and RGB image for the different objects. For the two blocks the depth overestimations were similar to the previously performed experiments and therefore the same possible explanations hold. These being the angle under which the camera was relative to the objects, human error and a possible inaccuracy of the laser measurement tool. Still, at most distances the red block seems to give a slightly higher estimate, but most of these do not seem significant enough to say that the color had an influence in this case. However, Tölgyessy *et al.* did look into different colors and materials and reported that this can indeed make a difference [12]. This difference was encountered more clearly when using the two head phantoms, where the dark colored head phantom was barely observed by the camera contrary to the light colored head which was perceived adequately. The light head did encounter a relatively high overestimation ranging between 18 and 31 mm, which could possibly be due to difficulty manually determining the ground truth distance from the nose to the camera as well as the gradient difference which might be harder for the camera to detect properly. Furthermore, the distance line between the camera and the nose was measured straight, however, it was not horizontal as the camera was on the tilted stand and was at a lower position compared to the nose. Each of the measurements of these objects showed a standard deviation ranging between 0.79 and 1.99, all of which increased as the distance from the respective objects increased. The center pixel of the dark head phantom on the other hand was not encountered by the camera throughout all frames for the dark head and therefore this and other pixels on the head were reported to have a low standard deviation. This means the event was not limited to one pixel or frame, but the camera has a consistent problem with the dark colored head phantom. This could present a problem for dark skinned people during BT if IRiS is being used, as the camera might not be able to detect the patient properly.

In section III-C, the results of the evaluation of the internal gyroscope were discussed. Initially the idea was to use the gyroscope to correct the distance estimation for the angle, however, since the output was in degrees / second this could not be used as the initial angle still needed to be precisely known. Furthermore, the results showed a significant drift over time in stable positions, from which can be concluded that it would not be wise to use this internal gyroscope, or at least without correction. However, based on the data from Fig. 7, applying a (simple) correction does not seem plausible as the lines are not all linear and the pitch, yaw and roll drifts do not respectively overlap for the different runs.

It should be clear that this research is a composition of several experiments, all with the goal to evaluate the suitability of the Azure Kinect for its part in the DETERMINE project. This paper presents the most important experiments and findings, but it should be noted that not everything investigated and produced during this research has been included. In case of interest, additional information can be found in the manual / logbook on GitHub. Another point which should be noted is the introduction of the calibration in the method section. This calibration has been performed and described, but further steps to be able to obtain actual results were out of the scope of this research as there was not enough time to include this. Nonetheless, it is of great importance for successors and for this reason it was included in this report without the displaying of results.

Since this is a primary research with the Azure Kinect in the context of the DETERMINE project, a lot more future research is advised. A few suggestions based on this research would be to look into the influence of the angle of the camera for depth estimation and a way to correct for this. Additionally, a correction for the gyroscope could prove to be useful to detect accidental camera movement so depth estimations can be corrected in real time. This research showed that a simple correction on the output of the internal gyroscope

will probably not work, however, other options might be available and could be looked into. Furthermore, as human repeatability is a factor, it could prove useful to repeat some of the experiments. Considering the code has already been created and is available, this should not be too much effort and could provide some more stable results like for example a more generally determined warm-up time. Another strongly advised research option would be to look into follow-up steps for the calibration, like overlapping the depth and color image which again could open new doors for further research. Moreover, options with deep learning could be investigated, for example delineation of the human body to properly track the patient, for this the overlapping of depth and color images could prove useful which again leads to encouraged investigation of the follow-up steps of the calibration. Lastly, this research has mainly focused on evaluating some basic aspects of the Azure Kinect. However, whereas the DETERMINE context was kept in mind, the experiments were not completely realistic. During BT the patient is asked to lay still and the camera system is supposed to track movement. During this research movement was not yet included as depth accuracy for stationary objects had to be determined first. In future research, it is advised to investigate scenarios closer to the BT treatment with for example moving objects.

V. Conclusion

Whereas literature shows quite some enthusiasm about the Azure Kinect, this research has shown some aspects to be taken into consideration. Some of these, like the warm-up time, can be addressed quite easily, especially with the suggested warm-up time of 86 minutes. On the other hand, the measurement errors could provide a bigger obstacle in the DETERMINE context as both accuracy and precision are key for good patient movement monitoring. Nonetheless, the arguments presented in section IV could prove to be (parts of) the reason for this as for example it is shown that taking away the angle of the camera can decrease the overestimation. This resulted in the estimations centering around zero while keeping the variation between the minimum and maximum encountered values the same with approximately 10 mm. Even so, this variation of 10 mm is still to be considered too high in this context and could prove to be reason for rejection of the Azure Kinect. However, the 0.79 to 1.99 mm standard deviations observed in the depth accuracy experiments still seem promising for movement tracking as this relies mostly on differences per frame and not per measurement. All in all, without further research which omits the presented explanations, the Azure Kinect should not be passed over for the DETERMINE project. Further research is therefore strongly recommended and the results and discussion points are suggested to be taken into account as a starting point for this future research.

References

[1] J. Skowronek, "Current status of brachytherapy in cancer treatment – short overview," *Journal of Contemporary Brachytherapy*, vol. 9, no. 6, pp. 581–589, 2017.

[2] G. P. Fonseca, "Dose and motion measurements in brachytherapy (DETERMINE), grant application," 2019.

[3] G. P. Fonseca, "Brachytherapy treatment verification using gamma ra-diation from the internal treatment source combined with an imaging panel – a phantom study," *Physics in Medicine & Biology*, 2021.

[4] R. Horaud, M. Hansard, G. Evangelidis, and C. Ménier, "An overview of depth cameras and range scanners based on time-of-flight technologies," *Machine Vision and Applications*, vol. 27, no. 7, pp. 1005–1020, 2016.

[5] J. Penne, C. Schaller, J. Hornegger, and T. Kuwert, "Robust real-time 3D respiratory motion detection using time-of-flight cameras," *International Journal of Computer Assisted Radiology and Surgery*, vol. 3, no. 5, pp. 427–431, 2008.

[6] D. M. Edmunds, L. Gothard, K. Khabra, A. Kirby, P. Madhale, H. McNair, D. Roberts, K. K. Tang, R. Symonds-Tayler, F. Tahavori, K. Wells, and E. Donovan, "Low-cost Kinect Version 2 imaging system for breath hold monitoring and gating: Proof of concept study for breast cancer VMAT radiotherapy," *Journal of Applied Clinical Medical Physics*, vol. 19, no. 3, pp. 71–78, 2018.

[7] J. A. Albert, V. Owolabi, A. Gebel, C. M. Brahms, U. Granacher, and B. Arnrich, "Evaluation of the pose tracking performance of the azure kinect and kinect v2 for gait analysis in comparison with a gold standard: A pilot study," *Sensors (Switzerland)*, vol. 20, no. 18, pp. 1–22, 2020.

[8] Microsoft, "Azure Kinect DK Overview," 2021. [Online]. Available: https://azure.microsoft.com/nl-nl/services/kinect-dk/#overview.

[9] J. McGlade, L. Wallace, B. Hally, A. White, K. Reinke, and S. Jones, "An early exploration of the use of the Microsoft Azure Kinect for estimation of urban tree Diameter at Breast Height," *Remote Sensing Letters*, vol. 11, no. 11, pp. 963–972, 2020.

[10] A. Chalmers, J. Parkins, M. Webb, and K. Debattista, "Realistic Humans in Virtual Cultural Heritage," *Springer Nature Switzerland*, 2021.

[11] Microsoft, *Documentatie voor Azure Kinect DK*, 2021. [Online]. Available: https://docs.microsoft.com/nl-nl/azure/kinect-dk/.

[12] M. Tölgyessy, M. Dekan, Ľ. Chovanec, and P. Hubinský, "Evaluation of the azure kinect and its comparison to kinect v1 and kinect v2," *Sensors (Switzerland)*, vol. 21, no. 2, pp. 1–25, 2021.

APPENDIX I
EXPERIMENTAL SETUPS

A. Experimental setup: Depth estimation over time

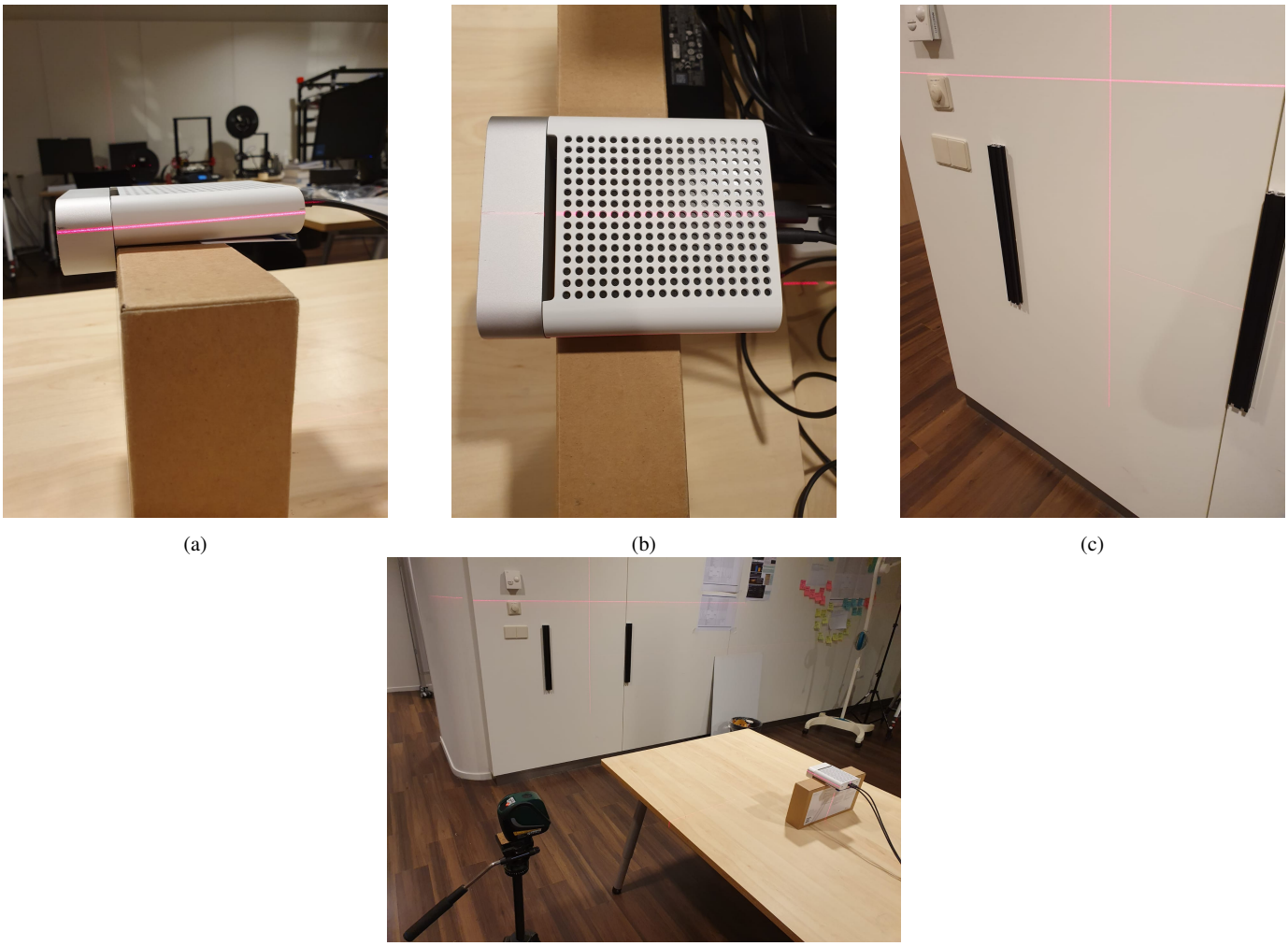
In Fig. 8 some pictures are shown of the setup during the depth estimation over time experiment. The camera was placed at the set distance, with the help of the laser measurement tool, away from the box on a flat surface. Afterwards, the laser measurement tool is removed and the camera is started using the DepthOverTime function. This is repeated for multiple distances.

(a) (b)

Fig. 8: Pictures from the setup of the depth estimation over time experiment

B. Experimental setup: Depth accuracy on a flat white wall for moderate range

Here in Fig. 9 the pictures of the setup from the depth accuracy experiment using a flat white wall are shown. The camera is placed on a platform to raise it from the table and is placed according to the lasers to make sure it has no angle and the center of the wall is in line with the center of the camera. Two bars are placed on the wall as landmarks to be able to determine the center in the depth image. The two bars were each placed 30 cm away from the laser. Once everything was put in place, the function SaveImages was used to record and save the footage afterwhich it could be processed.



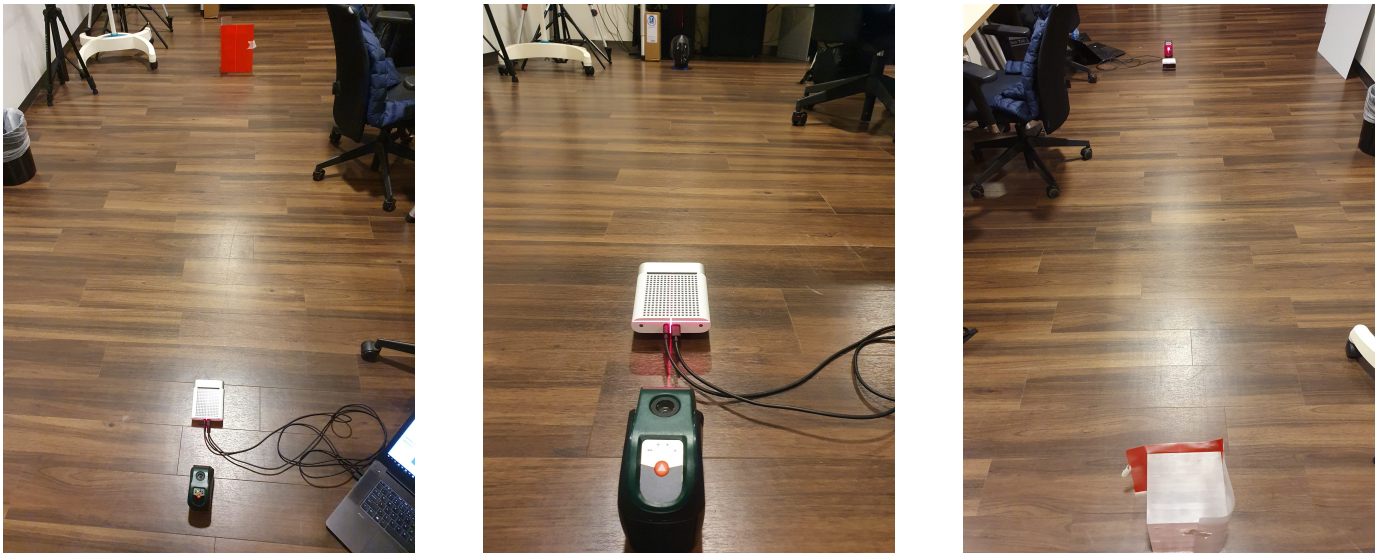
(a) (b) (c)

(d)

Fig. 9: Pictures from the setup of the depth accuracy on a flat white wall for moderate range experiment

C. Experimental setup: Depth accuracy including flat and gradient object surfaces with different colors for moderate range

In Fig. 10 it can be seen how the camera, object and cross line laser tool were placed. The cross line laser was used to align the camera with the target object. Once everything was put in the right place, the SaveImages function was used to record all data afterwhich it could be processed.

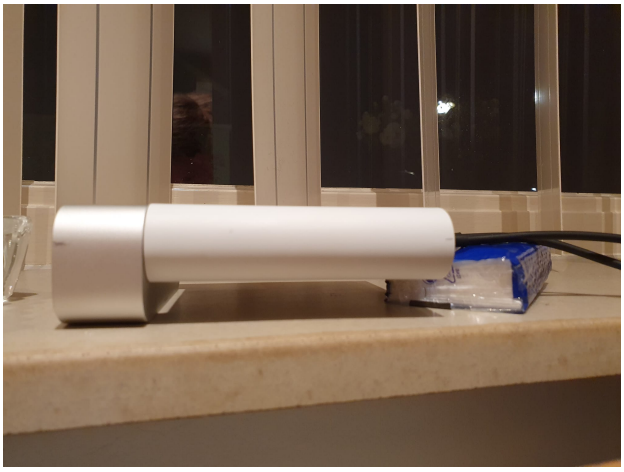


(a) (b) (c)

Fig. 10: Pictures from the setup of the depth accuracy including flat and gradient object surfaces with different colors for moderate range experiment

D. Experimental setup: Gyroscope evaluation

Fig. 11 shows the two orientations in which the camera was placed during the gyroscope evaluation experiment, namely horizontally and vertically. After placing the camera, it had to be ensured that no movement would happen which could disturb the measurement. Once everything was placed, the ShowOrientation function was run.



(a)

(b)

Fig. 11: Pictures from the setup of the gyroscope evaluation experiment