VidIt

Bachelor Thesis

Object Tracking System (VidIT)

Bachelor 2021

Authors

Frederik Mads Pil (frpi@itu.dk)

Mathias Højte Olsen (mhoo@itu.dk)

Mathias Dashtbani Mikkelsen (mdmi@itu.dk)

Supervisors

Main-Supervisor - Andres Faina (anfv@itu.dk)

Co-Supervisor - Fabricio Batista Narcizo (fabn@itu.dk)

STADS Code: BIBAPRO1PE

May 2021

Abstract

This thesis investigates, if an IT product can increase learning in an online setting. Information is included in regards of learning and the development of VidIT, which is an automated tracking system powered by a smartphone and an Arduino. The system can track people with the help of a motorized pan tilt mount. The purpose of VidIT is to enhance learning during COVID-19, by enabling students and teachers to record themselves single-handily while moving around. A survey, a user test and a performance test was conducted to gather data on the current situation of teaching in an online setting, testing of the usability and performance of VidIT. Based on the tests, it was concluded that the resulting system worked as intended. However, some improvements are needed to effectively improve learning and teaching in an online setting. These improvements includes but are not limited to, streaming functionality, movement prediction and faster computation in relation to the objection detection algorithm.

Keywords: PT-mount, VidIT, tracking system, BLE, Arduino, Object detection, Online teaching, COVID-19, pandemic, practical teaching.

Contents

1 Introduction

This report is written in connection with the development of the bachelor project on the last semester at the IT-University of Copenhagen, during spring 2021. The supervisors of this thesis are Andres Faina (Main-Supervisor) and Fabricio Batista Narcizo (Co-Supervisor). The report is based on the development of a tracking system for smartphones with coherent user tests and survey about online teaching.

1.1 Background

The impact of COVID-19 on the society is severe. Governments are putting restrictions in place in order to keep their citizens safe. This means that it is difficult, complicated and dangerous to meet in person. Never before, we have been so reliant on technology and the internet in order to communicate and work with each other to keep our daily life progressing. The impact of COVID-19 have shown some complications with learning, especially within education, where physical presence, physical activities and student interactivity plays an important role.

In 1938 an American philosopher John Dewey expressed the idea "Learning by Doing". The idea explained that learning and effective education should be socially including, where students actively participate and interact with the lesson [25]. This idea has through the years developed into an important aspect of the educational system, where attempts are made to incorporate interactive learning in a majority of educations. Eric Mazur, the Professor of Physics and Applied Physics at the Harvard School of Engineering and Applied Sciences, has experienced how the lack of interactive learning can decrease the understanding of a subject and make students exhausted. He made an experiment with his students, who expressed that his lectures was boring. He changed his way of learning to incorporate interactive learning and experienced a progressive learning gain from his students, because they got more emotionally invested in the lecture[5].

Today there can be seen a regression in interactive learning, since all teachings and presentations has to be handled online. Teachers have to sit stationary in front of a computer and teach using tools such as Zoom and PowerPoint. Students have to present their work online. This means that there is a need for a solution, so students do not lack the necessary knowledge and experience that is required of them. Furthermore, they should also be able to present work that has to be performed physically. An example would be a high school student attending a sports class, where they have to present and show different workouts in accordance with their curriculum[11]. To accommodate the issue, an understanding of the situation has to be obtained. During this bachelor project a survey has been made to investigate the students' opinion about, how the different educations handle teaching under COVID-19.

Video recording or streaming is a very common, but important tool to increase learning, if used correctly[1]. Video recording have a big range of benefits such as the option to watch it whenever, pause it, speed it up, translate it, rewatch it, etc. It is features that helps the user to understand the content better. Making a video recording or educational material is not a simple task. A lot of different factors has to be taken into consideration. Creating videos that students have to watch has to be a well thought process, in order to create effective educational videos, that impact the students understanding of a subject and improve their learning. According to the center for teaching at the University of Vanderbilt, three elements has to be taken into consideration in order to make an effective educational video. The three elements are; cognitive load, non-cognitive elements that impact engagement, and features that promote interactive learning[1].

Cognitive load theory

Within cognitive load, a theory was made by John Sweller around 1988 explaining that the memory contains three components; intrinsic load, germane load and extraneous load [18]. Intrinsic load is focused on the level of connectivity between a course and a subject. Germane load is focused on using the understanding of a particular subject to perform an analysis in order to reach the learning outcome. An example could be that in order to make a proper cost benefit analysis, you are required to have knowledge about this, and master it as well. Extraneous load is focused on elements that does not help with reaching the learning outcome, such as background music, irrelevant pictures etc. The goal is to manage intrinsic load, enhance germane load and minimize extraneous load.

Non-cognitive elements that impact engagement

Non cognitive elements that impact engagement are the performance of the presenter on the educational video. A lot of the elements are about that the language used by the presenter should be conversational instead of formal. The speaking rate should be quickly and enthusiastic, the subject should be relevant for the students in the class, and when talking about a subject, then use something visual to increase student engagement. The elements are based on the findings of Philip J. Guo[1], who examined the attention span and engagement of students, who watched different length videos. The majority of the students were only engaged within the first 9 minutes.

Features that promote interactivity

Vanderbilt highlights some important tools and elements that promote interactivity in educational videos, which they find very important. One of them is to use tools that gives students control or influence on the actions. Dongsong Zhang investigated the impact of interactivity on students when watching videos[26]. Zhang found out that those who were able to interact, control or have an influence on the actions of a video had a greater learning outcome.

The article by Vanderbilt, will be used to discuss, argue and measure if the final product can increase students' learning outcome and improve teaching in an online setting.

1.2 Related work

As the pandemic continues to cause lock downs, other solutions to online sports teaching have been suggested. In Spain, research was conducted on pre-service teachers by Valeria Varea et al.[24], where they reflected on the current problem. The soon to be physical education (PE) teachers in the elementary school were worried about the remote aspect of the teaching, and not being able to have physical contact with the students. However, the teachers were looking forward to introduce new ways to do PE without face to face communication, making it hard to access each individual pupils' skill, and evaluate the exercises after a finished lesson[24]. The teachers anticipates, that a lot of the teaching will be conducted like video tutorials of exercises, which the teachers think will make learning ineffective. The paper comments on this statement, that the teachers is underestimating their ability to use technology to interact with and teach their pupils, as the current video chatting technology still allows for the teachers to get immediate feedback and response from the students[24].

As a positive outcome, V. Varea et al. sees that pupils are becoming more independent by designing and filming their own activity circuit. The paper focuses more on how the pre-service teachers will be capable of adapting to do video tutorials. V. Varea et al. concludes that teachers are ready to use digital technology, and might also do so in the future after COVID-19. This paper took base in a Spanish audience, which have a culture with even more physical contact than in Denmark and other Scandinavian countries[24]. Therefore, the shift to digital platforms and technologies might become easier in Denmark and be relevant after the pandemic, even for PE.

Many sports teachers in Turkey thinks that distance learning is inadequate and inefficient, concludes Şule Kırbaş[12]. This was concluded based on interviews conducted on sports teaching instructors. The reason for the instructors' concerns were problems like a bad IT system, inexperience with distance learning and poor feedback from the students. The few instructors who were optimistic, about 10%, believed that the course goals could be achieved through the following:

"By creating a visual and video-assisted tracking system, they can repeat and record the skill in an appropriate environment where they live in accordance with the guidelines given."[12].

Such a solution might be viable, but is yet to be realized.

A paper by Hamid Ghasemi and Leonardo Jose Mataruna-Dos-Santos, chooses a positive angle to teaching in an online setting[7]. In this paper, various opportunities is presented as aids in Sport and Physical courses in the corona pandemic. Among the opportunities is virtual reality, augmented reality and virtual lab. These technologies is claimed to be getting most attention from experts. A Learning Management System (LMS) is also teased to be a viable suite for online learning, supporting actions such as creating teacher-student interaction online and offline. It also records actions performed by the student.

In 2019, Tan-Hsu Tan et al. proposed an "Intelligent Lecturer Tracking and Capturing System Based on Face Detection and Wireless Sensing Technology" (ILTC)[20]. The system is supposed to provide an alternative to expensive tracking solutions such as PTZ-cameras[1] and multiple camera setups. The resulting system was a stationary computer which were put a few metres in front of the lecturer. With face detection using SSD Mobilenet and IR thermal sensors, the system tracks the lecturer with an Arduino Uno which pans the camera accordingly with the help of a servo motor. The resulting product detected with a frame rate of 24 fps on a 3.2 GHz Intel i7-8700 CPU and 32 GB RAM. This system tracked a lecturer in a test about 86.95% of the time with both face detection and only 66.5% without the IR sensors. It was proved by a user survey, that the system has greater practicality than the existing solutions, which is static video recording and professional studio recording[20]. The system could not handle abrupt or rapid movement.

As presented above, there are many ideas for solutions to the problem domain of this report, but they are not very specific, apart from the ILTC system. This system is yet to be formally tested out during COVID-19. Many teachers are dealing with the shift to online sports teaching in their own way, which they are the most comfortable with. The physical limitations caused by the pandemic have many solutions, and the most optimal one may never be found before the lock downs are over. However, it is not certain that things will go back to the way they were before, as remote learning is often cheaper and easier when looking at the situation from an administrative angle[2]. However, the teachers will always prefer to have their audience face to face[24].

1.3 Related products

Different products similar to the prototype is available, most of these product only does panning, and are expensive. One of the cheapest solutions is the Pivo Pod, which costs 164$[3]. This solution is the most similar, as it uses a smartphone as computing center for AI tracking. It is possible to track just the face as well as persons. The pod consists of a single motor and is controlled over Bluetooth.

[1] Pan/Tilt/Zoom cameras
[2] Cost of education
[3] getpivo.com

The company DJI also have solutions which is comparable to the VidIT. Their already advanced technology is also supporting auto-tracking from within their app. The app controls a handheld mount, which acts as a stabilizer, yielding professional-like videos, when on the move. This solution costs 149$[4].

Another different solution is Jigabot, which uses IR-tracking (infrared tracking). This tracking is very accurate, but requires the user to carry small dots on them to get detected. The solution is very expensive starting at 895$[5].

1.4 Description of the target project

For this bachelor project, a minimal viable product (MVP) of a tracking system for smartphones has been developed containing the core features. The product is called VidIT. The requirements for VidIT is derived from two main requirements. First, the user should be able to record themselves without the need of a cameraman. Secondly, VidIT should keep track of the user and keep them in the frame regardless of them moving around. These requirements has been identified based on a informal interview with a high school student expressing her concerns with learning and teaching in an online setting. The main use case is: "As a student who has sport lectures during COVID-19, I would like to record myself with no help of others"

The product requirements were to build all functionalities adhering to the requirements as one solution.

Research Questions:

- How can learning be improved in an online setting with the use of an IT (VidIT) product?
- How can VidIT compete with already existing solutions in price and simplicity?

1.5 Methodology

The following section describes the approach of the report to the research questions as well as the used libraries in the project, to ease the development process.

1.5.1 Description of the approach to the project

Prior to the bachelor project agreement and the problem statement, an informal interview was made with a high school student. The student expressed concerns in regards to how lectures are held in an online environment. The lectures was limited in terms of interactive learning, physical experiments and sports. The

[4] dji store
[5] jigabot.com

students still had to follow the regular curriculum, which contained sport lectures where they had to develop their own exercise program[11]. According to the student, it was very hard to record herself without help from another person. Based on the informal interview, a bachelor project agreement and a problem statement were created.

The group were not well known with the technologies needed to develop the solution to the problem statement. Thorough research was needed to find the most suitable technologies to accommodate the requirements of the project.

During the development of the project, further validation of the problem statement was needed. A survey was made by the bachelor group, that should investigate multiple students' experience with teaching in an online setting. The survey also investigate their opinion to the product. The gathered information was processed and additional changes was made to the product based on the findings.

1.5.2 Used technologies and libraries

Since the target project is focused on the research questions, the developed mobile application uses libraries. This is due the focus of the project, which is not to develop object detection, camera API, etc. from the bottom, but to make a minimal viable product, that is testable and can answer the research questions.

The following is an overview of all the libraries used in the mobile application in Flutter[6]:

- camera: 0.7.0+2 [7]
- flutter hooks: 0.15.0 [8]
- flutter launcher icons: 0.9.0 [9]
- tflite: 1.1.1 [10]
- flutter blue: 0.7.3 [11]
- path provider: 1.2.0 [12]
- gallery saver: 2.0.3 [13]

[6] Flutter official
[7] camera
[8] Flutter hooks
[9] Flutter launcher icons
[10] tflite
[11] Flutter blue
[12] path provider
[13] gallery saver

- flutter ffmpeg: 0.3.1 [14]
- native device orientation: 0.4.3 [15]
- flutter audio recorder: 0.5.5 [16]
- permission handler: 5.1.0+2 [17]

1.5.3 Description of the team

The team behind this bachelor project consists of three Bsc. students in software development on ITU. The students' experience varies based on their electives and their student jobs. The students' electives are: Machine learning, Mobile App development, and Security. In terms of roles in the project, all students are equal and have all participated and contributed to the development of each artifact.

The team were assisted by a high school student, who was the origin of the problem statement and helped with understanding the application domain. This was to make sure that all members of the group understood the project's scope.

[14] Flutter ffmpeg
[15] native device orientation
[16] Flutter audio recorder
[17] permission handler

2 Identification of requirements

The high school student, who was informally interviewed, expressed a range of concerns regarding the current teaching situation at the student's high school. The student explained that she had experienced issues during her sport lectures. The sport curriculum in a high school contains different topics such as developing a warm-up program, learn how to play volleyball, athletics etc[11]. Based on the information gathered from the informal interview, a scenario can be created, showcasing the situation:

Scenario	
Scenario name:	Create a warm-up program during COVID-19
Participating actor instances	(Sophie - Real name redacted): 3rd year high school student
Flow of events	1. Sophie needs to create and record a video of her own warm-up program as a part of the curriculum. Because of COVID-19, she is forced to do it at home by herself. 2. Sophie takes her smartphone and goes out in her garden. She places the phone on a table in portrait mode, where the phone is supported by a pile of books. 3. She starts recording herself doing some warm-up exercises. She starts out with some exercises standing up and everything works well. She can clearly see her own movement and technique on the camera. 4. Sophie starts moving back and forth doing some warm-up exercises. After reviewing the video footage of her moving, she finds out that it is very hard to see her technique since she keeps going in and out of frame. She tries to record it again, but with no luck. 5. Sophie adjusts the camera and lies on the ground to do some exercises. She does different exercises. The smartphone camera captures some of the exercises perfectly and in other exercises, she needs to adjust the camera frequently to capture her technique. In some cases, it is impossible for her to capture her movements because she does some exercise on the ground and then have to jump or run immediately after.

Based on the identified scenario and the informal interview, user stories could be created that covers the concerns expressed by the student. These can be seen in the appendix [9.4]

2.1 Functional Requirements

Based on the user stories, functional requirements can be identified, that is needed in order to create a solution to the problem statement:

FR-1 The user should be able to make video recordings by using the different cameras of a smartphone.

FR-2 The system should be able to track the user.

FR-3 The app should work both in portrait and landscape mode.

FR-4 The smartphone should be able to connect wirelessly to a mount.

FR-5 If the system loses tracking, it should attempt to find the user again.

FR-6 The system should be cheap.

FR-7 The system should be easy to use.

FR-8 The system should be able to pan and tilt to keep the user in frame.

3 System Architecture and Specification

The following section describes the development, design choices and specifications of VidIT.

3.1 Arduino

The following section describes the design and implementation of the hardware developed during the project. Figure 1 shows the final prototype of the hardware used to operate the pan/tilt mount (referred to as PT-mount). This is implemented to accommodate FR-4 as defined in section 2.1.

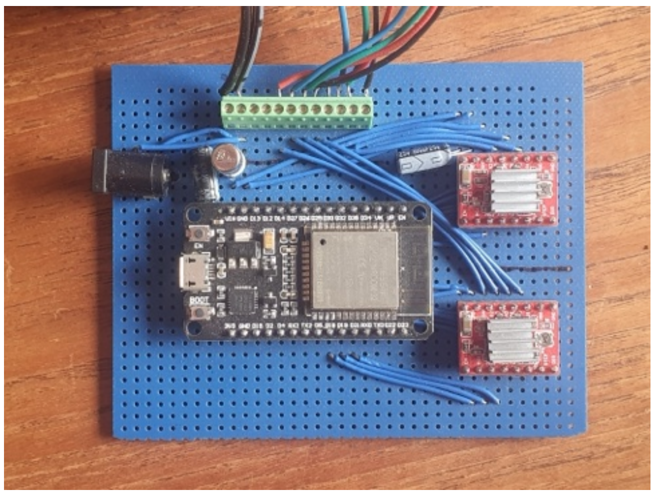


Figure 1: Picture of the final PCB developed during the project.

3.1.1 Layout

For this project, the ESP32[13] microcontroller has been used to establish the connection between the mount and a mobile phone. The microcontroller offers BLE (Bluetooth Low Energy), described in section 3.1.2, which is used for communication between devices. The ESP32 is placed on top of a stripboard, which contains wide parallel strips of metallic copper running in one direction across the board. To ensure electronic flow and avoid short circuits, breaks are made in the tracks to divide the strips into electrical nodes. By doing this, the microcontroller, two stepper motors along with two micro-stepping drivers[17] can be placed on the stripboard, to combine it into a prototype PCB (process circuit board).

As shown in figure 2, the two motor drivers are connected to the ESP through the pins which are recommended for output logic[23]. The motor driver is designed to operate the stepper motor in "*full-, half-, quarter-, eight-, and sixteenth-step* *mode*". This means that the number of steps per full rotation can be regulated to the need of the product. If for example the driver is set to run full steps, then the motor would have 200-steps-per-revolution, meaning that the stepper motor rotates 1.8° for each full step. Equally the motor rotates 0.9° for each half step.

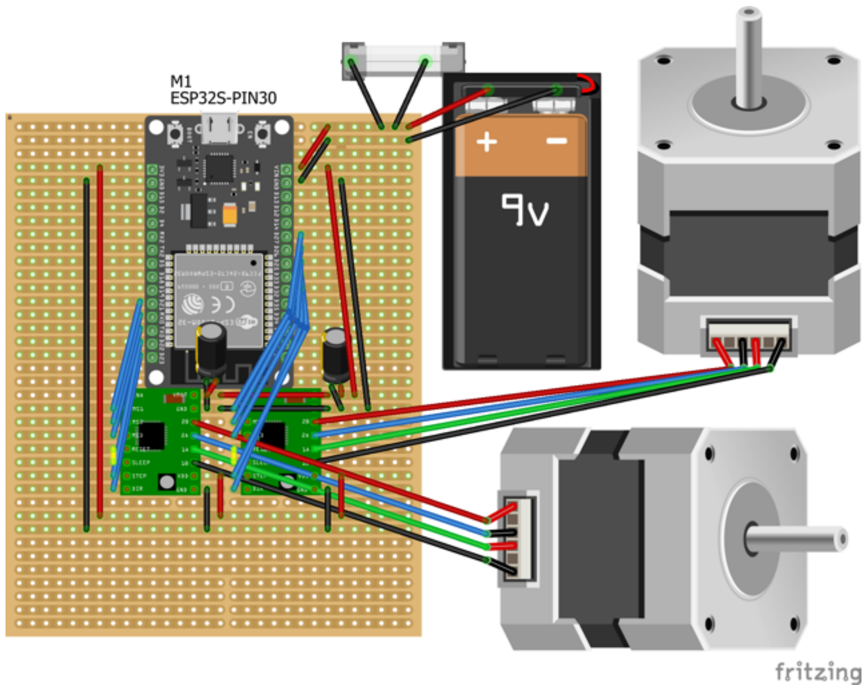


Figure 2: Illustration of the PCB used for operating the mount containing the ESP32 module, two stepper motors(A4988), two motor drivers, power supply, two capacitors and one fuse for the power supply.

Through the development of the prototype, different step modes have been tested. The sixteenth-step mode is the best suited for this product, since the rotation of the motors is significantly smoother than other modes. It is not able to rotate as fast, since the stepper-motors are only reliable up until 1000-steps per second. If the microstep resolution is set below sixteen (eight, quarter, etc.) the mount would pan/tilt too fast in relation to how many frames the object tracking can detect. This would result in the same problem as described in section 3.4.2.

The whole system is powered by a 9 voltage power supply since both motors, and the microcontroller requires a voltage between 6 and 12 to function properly. Furthermore, a fuse is attached between the power supply and the microcontroller to ensure that the circuit is not fried.

3.1.2 BLE (Bluetooth Low Energy)

BLE[22] is a power-saving variation of Bluetooth, which is used for short-distance transmission of low bandwidth (little amount of data). In relation to the original Bluetooth, BLE remains on standby except for when a connection is initiated. This is well suited to the project, since the mobile would never be more than a few inches from the mount when in use and the data packet would never be more than a direction and a speed.

The actual data transfer is established through GATT[21] (generic attributes) which is a hierarchical data structure revealed to interconnected BLE devices. With the GATT profile, connections are exclusive, meaning that the BLE surface is only able to connect to one device at a time. This way, no one can interfere with the connection between the microcontroller and the connected device. It will stop being visible to other devices and unable to connect until the established connection is terminated.

For the BLE devices to be able to communicate, UUID (universally unique identifiers) must be generated for them to know exactly where to send the data. Once these are generated and associated with the characteristics, the mobile and microcontroller are able to use operations such as, read, write and write with no response, to transfer data.

To establish communication between the mount and the mobile, two types of devices have been implemented, a server and a client. The ESP32 can act like both of them. In this case, it acts as the server, to be able to listen for client requests/data packets. When establishing the connection, the client (in this case the mobile) scans to find the server. Once the connection has been established, the server begins to listen for input from the client which is used to operate the PT-mount.

The microcontroller must communicate a digital signal which is either logical high (3.3V) or logical low (0.0V), to the pins which are responsible for the direction and speed (step frequency). Whenever the pin gets a logical high, the step-motor driver will forward the signal to the stepper-motor indicating that it should rotate in the defined direction by the specific pin. Hence, the microcontroller continually reads data inputs from the mobile using a simple loop. Based on the data received, the signals to the pins changes in terms of direction and speed.

3.2 Pan/tilt mount

The purpose of the PT-mount is to carry the smartphone and hardware, combining them into one coherent system. This will accommodate FR-8 defined in section 2.1. A blueprint from Thingiverse[4] was used and modified to accommodate the functional requirements of the bachelor project. The main issue with the original mount from Thingiverse was that it was too compact, limiting the space where the smartphone is placed. One of the requirements of the project was that the mount should be able to work with all kinds of smartphone orientation defined in FR-3 in section 2.1. The group had to increase the diameter between the two supports. The right support carrying one of the stepper motors had to be moved to the edge of the base of the mount and the main bridge had to be extended. A phone holder from GetPivo[18] was incorporated on the bridge to hold the smartphone in place. In order to incorporate the component, a 1/4 inch screw had to be placed in a 3/8 hole, in the bridge. This also gives the user the option to switch out the GetPivo component with their own smartphone holder.

Figure 3: Shows the PT-mount

Even though, the design of the PT-mount is a valid solution, it has a range of limitations due to its design. The stripboard is not able to fit in the base of the PT-mount due to the size of the board and battery. The gear is placed in such a way that it is facing outside with no protection. This may result in the gear getting stuck, if placed on a non flat surface, and getting damaged. Furthermore,

[18] getpivo.com

there is a big difference in the gear ratio (size) between the base and support. The base is responsible for the panning and the support is responsible for tilting. The panning rotation speed is slow due to the gear ratio of the base compared to the tilting rotation speed. Inside the base gear, the rotation mechanism is loose because of a press-fit design. That means the durability of the design is low, since there is a chance for the base gear to fall out when the mount is lifted incorrectly. When the smartphone is placed in portrait mode, it cannot tilt 360 degrees without damaging either the PT-mount or the smartphone. The size and weight of the PT-mount is also problematic in relation to transportation, since it is fragile and unhandy.

3.3 Object Detection

Tensorflow lite (TFLite)[15] is used for object detection. It is the most up-to-date machine learning framework for Flutter, with better community support than other solutions like ML Kit (for advanced machine learning), which have yet to increase its popularity. ML Kit was supposed to handle the machine learning part of the project, but the Flutter package lacked support in terms of object detection and tracking. A package could have been developed to link flutter code to native Android(Java/Kotlin) and iOS(Swift/Objective-C), this would have been a very time-consuming process. On the other hand, TFLite worked right out of the box with a custom pre-trained model like SSD Mobilenet V1[19], which was already converted to a lite format. It balances performance and precision to make it suitable for running on edge devices such as smartphones.

SSD stands for single shot multibox detection. The model accepts 300x300 image input tensors and has an accuracy of 72% on the VOC2007 dataset[14]. Images are provided from the phone camera as a stream of JPEG images with the help of the Flutter camera package. These images are then decoded with the built-in library in flutter called ***Image*** in order to be used with Tensorflow[20]. The images are resized to match the model and are flipped and or rotated according to the current orientation of the device.

A way of improving the current object detection could be to use a different model, for example, YOLO[19]. YOLO accepts 416x416 input images which combined with how the image tensor is passed through the layers of the model, will make the model slower but more precise. The models are similar in their way of processing an image tensor. This is different from models like R-CNN[8]. SSD and YOLO are able to predict in one evaluation, where R-CNN needs at least two evaluations because it finds regions before running the convolutional neural network to find features. This is the difference that makes these single evaluation models appropriate for real-time object detection since they are significantly faster.

[19] SSD V1 - Tensorflow Hub
[20] flutter_tflite binary image code sample

3.3.1 Object Tracking

To get the most out of the detection model, the app implements Multiple Object Tracking(MOT). This enables users to select between multiple detected targets. To track a single target, despite new recognitions interfering, a simple re-ID algorithm is implemented. It uses the boundaries and center offset of the detected objects, compares them, and returns true, if the two boxes match within a certain threshold. To save resources, the re-ID is implemented rather simple, by merely comparing new recognitions to the current tracked bounding box. Usually, re-ID is implemented as a machine learning model, but this would consume too many resources on the phone, as running the current detection model already uses 31 ms to infer on average. This causes the frame rate to drop to around 10 frames per second observed by the Flutter Performance inspector[21]. The app is not able to run on low-end devices due to this requirement of resourceful computation, as even high-end mobiles get performance issues.

The detection model does not always detect the person in the picture on each frame. It results in the mount stopping its rotation, if no other persons were found, or changes detection to another person. A temporary dummy recognition, identical to the missing recognition of the person, is created as a blue box, which will keep the tracking smooth. When the person is detected again in the next frame, the box is no longer needed and, hence it is removed.

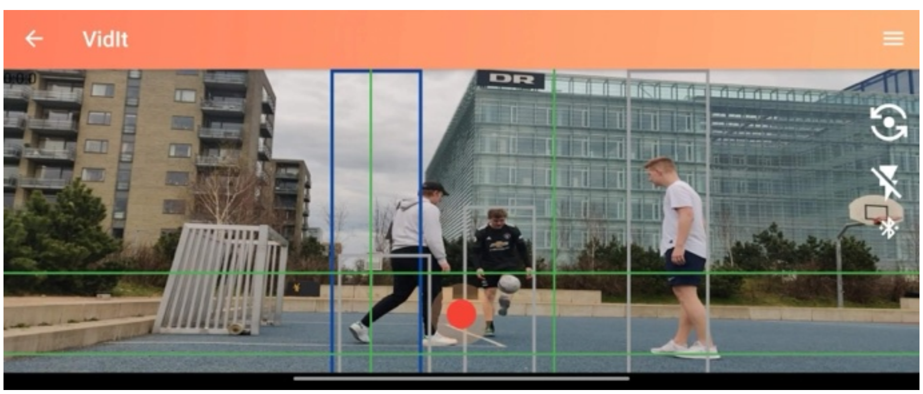


Figure 4: Picture showing horizontal tracking with MOT enabled

On figure 4 a blue smoothing box is seen. On the previous frame of the video the box was orange, which indicated that this was the dimensions and placement of the tracked person. As can be seen on figure 4 the tracking algorithm lost track of the person (the one to the left). Hence the tracking box turned blue in the current frame since the new recognition was too different compared to the previous one. The new recognition is indicated by the left-most gray box. When having the blue box as a replacement, the PT-mount will not begin to tilt

[21]On a OnePlus 7t in debug mode

downwards, and the recognition is restored in the next frame. In the picture, the boxes are drawn too large in terms of height, which is another issue, but it does not affect the tracking until the detected persons get too far away. This is because the box-to-preview calculation is off, making small boxes appear down to the left of the actual person. As seen on figure 5, the preview is cropping the top and bottom of the input picture smaller to make room for an app bar. The bounding boxes need to accommodate this such that the preview displays the boxes correctly on the screen.

Figure 5: The resulting video with preview borders in red

The video preview also appears to be zoomed in due to the cropping. However, it gives a better experience for the user in regards to the display of bounding boxes, but also better tracking as the whole width of the phone is used to make the actual panning more smooth. However, the tilt function suffers from this way of displaying the video preview. The experience might therefore be better when using the app in portrait mode.

3.3.2 Recording video while tracking

The Flutter camera package does not support streaming images to the detection model and recording simultaneously. This causes problems, since these functionalities are an important parts of the main features. A workaround for this problem is to use images from the camera stream, which is already used in the object detection. The camera stream is set to provide JPEG images to easily be able to concatenate frames into a video. If recording is toggled on, the current frame is saved in a temporary directory, before it is concatenated by the use of the FFMPEG package upon stopping the recording. FFMPEG is a comprehensive video and audio editing library. This solution allows the user to record videos while also tracking. However, the post processing of the video takes about half the recorded time, so this solution is not sustainable. A limitation of this is that the user is not able to interact with the phone without turning recording off. If this is done, the video will not be created and all recorded footage is lost. The user should be able to record long videos, where they do not have to worry about the recording or loosing footage. A normal use case is to record a warm up program for about 15 to 30 minutes, which will result in a processing time between 7 and 15 minutes. During these minutes the user would not be able to interact with the phone.

The audio on the video is recorded separately and merged with the image sequence. It might not always be synchronized because the frame rate is non-consistent. Because of this, the frame rate is calculated from images saved and seconds elapsed while recording. The resulting frame rate is provided to FFMPEG as an integer because it is the only option, which will result in a more de-synchronized video the longer the recording is. Most of the time the de-synchronized video is not noticeable, but on longer recordings, the synchronization becomes worse, as the frame rate needs to be more precise the longer the recording is.

3.4 Rotation algorithm

The following section describes the implementation of the rotation algorithm used to operate the PT-mount. Furthermore, this section will describe how the product accommodates the functional requirements FR-2, FR-5, and FR-8 from section 2.1.

3.4.1 Output from Tensorflow model

From the TensorFlow Object Detection API[14], the application can use a pre-trained model to detect the presence and location of different classes of objects. When the model is provided an image, a list of detected objects, the location of a bounding box that contains the detected object, and a score indicating the confidentiality of the detected object are provided.

When given an image the model outputs four arrays[14]:

- Locations: A multidimensional array containing four floating point values, which represents the dimensions of the boundingbox for each object.
- Classes: An array of integers representing the index of a class label from the labels file.
- Scores: An array of floating point numbers between 0 and 1 which represents the confidence in the detected class.
- Number of detections: Integer value representing the number of detections.

3.4.2 Boundaries

To be able to keep the detected object in frame, the PT-mount must be able to rotate and tilt the mobile freely in eight directions. These are defined as: Up, down, left, right, up left, up right, down left and down right.

The idea is, that when the detected object is changing its two-dimensional position the PT-mount, will counter this dimensional change by either rotating, tilting, or both at the same time. An example could be that the detected object is moving left, then the PT-mount will rotate the mobile to the left to keep the object in the defined center of the mobile screen.

To compute in which direction the PT-mount should move, some boundaries have been identified. These boundaries define questions like, what is the middle of the screen? When should the camera pan and tilt? How fast should it be moving? These questions can be quite difficult to answer since a person can vary in size and distance from the camera, making it hard to compute if the camera should move or stay still.

To accommodate these questions a rule has been defined. This rule defines that the center of the bounding box should always be the same no matter the distance between the object and the camera. Following this logic, the center coordinates of the bounding boxes could be calculated and hereby used to determine when the PT-mount should rotate and when it should stay still. When the detected object is outside the center, then the PT-mount should move in the desired direction. This can also be seen on figure 6.

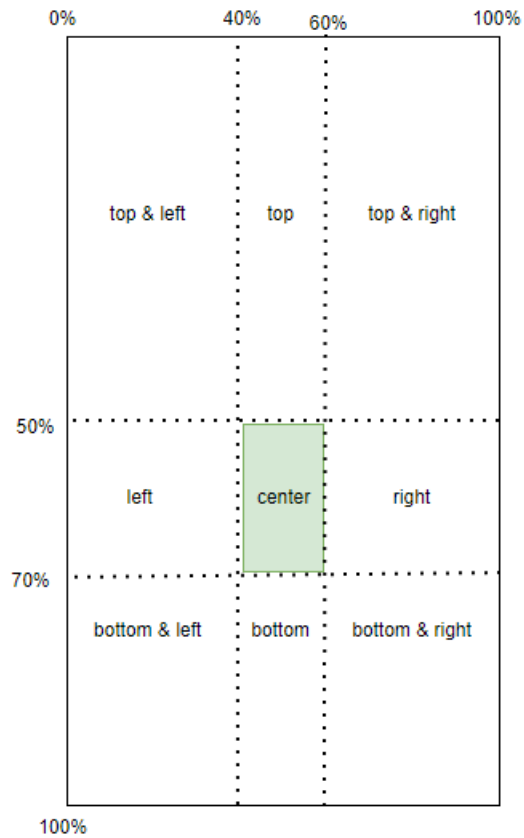


Figure 6: Illustration of the defined screen dimensions which defines the center.

The center of the screen has been defined to be between 40-60% of the screen width and 50-70% of the screen height by default. The reason why it is not completely centered, but slightly lowered, is due to a prioritization of getting the face of the detected person in frame most of the time. Though, the user can change this grid in terms of center placement, if they prefer to have the tracked person elsewhere, e.g. slightly to the left. When the center of the tracking bounding box is crossing these boundaries the rotation is triggered. For example, if the center of the tracked bounding box is between 0-40% of the screen width and between 50-70% of the screen height, it will trigger the mount to rotate left until the bounding box is centered again. By implementing these dimensional boundaries a set of rules have been defined to decide when the PT-mount should perform a pan or tilt action. With this implementation, the product will potentially fulfill the functional requirements FR-2, FR-5, and FR-8 from section 2.1.

Tests with narrower center dimensions have been performed, but as a result of this, it was observed that the tracking was not fast enough to adjust the rotation 9.3. This practically meant that if the detection algorithm tracked an object to be placed left according to the defined center, then the PT-mount would rotate left to counter the position. On the next frame, the tracked object would then be positioned slightly to the right, which the system would counter by rotating right. Hence, the PT-mount would bounce left and right by trying to have the object centered. This is referred to as the bouncing problem.

3.4.3 Speed of rotation

To make the system more intuitive and smooth, multiple rotation speeds have been implemented instead of just having a single constant speed.
Practically, three different speeds have been implemented, which are the following:

- Maximum speed: 750 steps pr. second (marked with red on figure 7)
- Medium speed: 500 steps pr. second (marked with yellow on figure 7)
- Minimum speed: 300 steps pr. second (marked with green on figure 7)

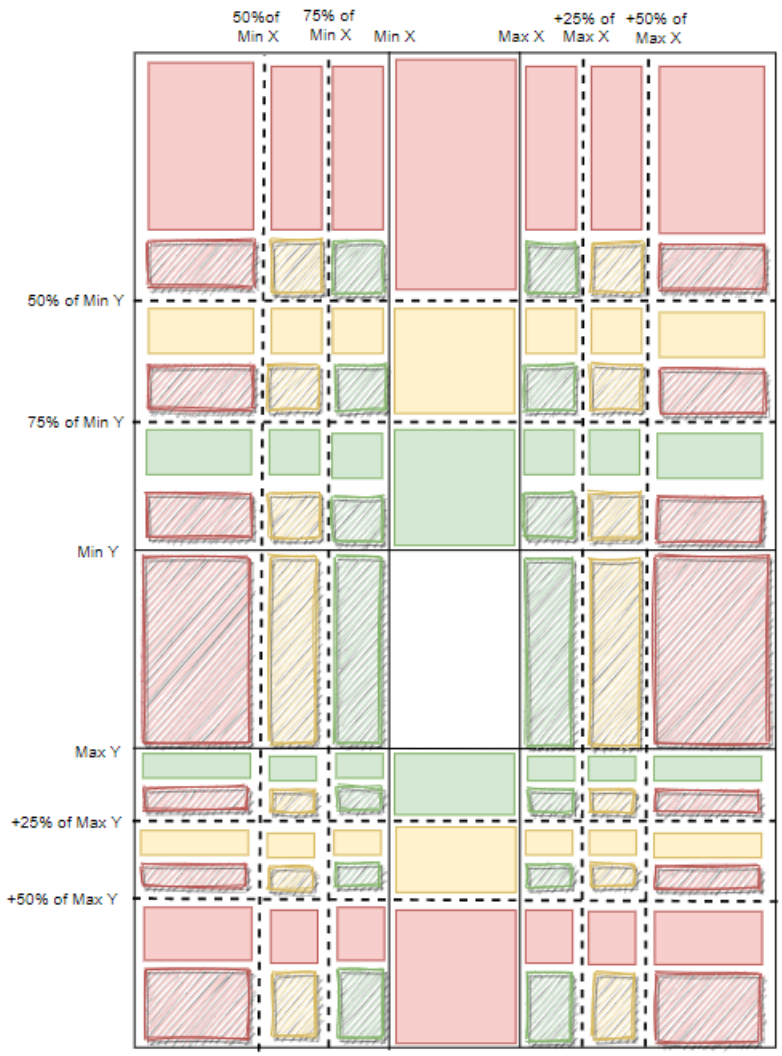


Figure 7: Illustration of rotation speeds and their boundaries on a mobile screen.

An illustration of when these rotation speeds are triggered can be seen on figure 7. The right and left rotations are depicted with scratched boxes. The up and down tilts are depicted with solid boxes. Figure 7 illustrates the same boundaries as figure 6, but here speed boundaries are included. As an example, the PT-mount will rotate left with minimum speed if the center of the tracked object is between "*75% of Min X*" and "*Min X*", and between "*Min Y*" and "*Max Y*". If for example, the center of the tracked object is between "*50% of Min X*" and "*75% of Min X*", and between "*50% of Min Y*" and "*75% of Min Y*", then the mount would rotate left and tilt up with a medium speed.

The Arduino AccelStepper library[3] has been used to be able to rotate the stepper motors. It provides an object-oriented interface with support for multiple simultaneous stepper motors, with each motor having independent concurrent stepping. The library provides two functions which are *setSpeed()* and *runSpeed()*. The Arduino will continue to loop these functions with the information received from the mobile application until interrupted. Hence, the mobile application will continue to send directions in terms of pannings, tilts, and speeds, to the Arduino to keep tracking the desired object. If the object would move out of frame, or the mobile somehow was not able to detect the tracked object, the rotation would stop.

Different step speeds have been tested. It was observed that, if they were set to a slower speed, then the mount was not able to keep track of a moving object. It would simply get out of frame in consequence of the mount not panning or tilting quick enough. If the steps were set to a higher speed, then the PT-mount would pan and tilt too quickly. This results in the same bouncing problem as described in section 3.4.2.

As a different solution to the hardcoded three-step acceleration, the AccelStepper library provides methods used for defining acceleration. The function *setAcceleration()* can be used to set the desired acceleration in steps per second. This function can then be combined with the function *setCurrentPosition()* which resets the current position of the motor. The position that the motor is currently at, is then considered to be the new start position. When these functions are combined the motor would be able to accelerate and decelerate to a defined position, which is considered to be true acceleration compared to the one implemented for this project. However, this method of acceleration was not possible to implement since the tracked object is rarely stationary. The main disadvantage of the *setCurrentPosition()* function is that the stepper motor will accelerate to the defined position, but cannot be given a new position unless starting over from zero acceleration speed. Practically, it is not possible to update the position of the object without interrupting the acceleration.

3.5 Price Estimation

In order to accommodate FR-6 in section 2.1 in relation the the products mentioned in section 1.3, a price estimation was made. The following table shows every part and its estimated price for one prototype:

Part	**Price**
Mount PLA	70 DKK[22]
ESP 32 Arduino	88 DKK[23]
Stepper motor drivers	2 x 12.28 DKK[24]
NEMA 17 Stepper motors	2 x 89 DKK[25]
Velleman Stripboard	43 DKK[26]
608ZZ Miniature Ball Bearings	2 x 15 DKK[27]
Wiring and soldering	10 DKK[28]
Screws and nuts	10 DKK[29]
RC 9.6V Battery	20 DKK[30]
Phone mount	49 DKK[31]
Total	522.56 DKK

The above price estimation is without man-hours used on developing the software, 3d-printer electricity, etc., and is therefore not entirely accurate. The purpose is to get an idea of what the prototype will cost to produce. Most of the materials will get cheaper, if they were to be bought in larger amounts. A part where most money could be saved is on the ESP32 and the stepper motors, which were bought from a danish retailer. The Arduino board can be purchased from China for 46.57 DKK[32], the same applies for most other parts. However, the cheap price may affect the build quality.

Knowing this, a realistic price for the parts of the prototype would be around 400 DKK. This price and the prototype will be discussed in relation to aforementioned related products in section 1.3, and further in section 6.2.

[22] 3DE PLA Estimate based on mount size
[23] ESP32
[24] Geekcreit A4988
[25] NEMA 17
[26] Velleman Stripboard
[27] 608ZZ
[28] Own estimate
[29] Own estimate
[30] Bought used
[31] Universal phone mount - Tripod excluded estimate
[32] ESP32 — Banggood.com

4 Tests and Results

The following section describes the survey, user test, performance test, and present the results.

4.1 Survey

The following will explain and go into depth with a survey the bachelor group has performed. The survey gathered both, knowledge regarding the current state of learning in educations, and opinions about if the product that the bachelor group developed could accommodate some of the issues and increase interactive learning for students. The detailed survey can be found in the appendix. However, the appendix only shows some parts of the survey, due to the disclaimer in the survey.

4.1.1 Survey structure and data

The survey was made to investigate the potential users' opinion and idea to how learning can be improved in an online setting. This survey was divided into two parts. One that was related to the application domain and one that captured the opinion about the prototype developed by the bachelor group. Participants in the survey was anonymous.

The survey's targeted audience were people studying either on universities, high schools or other educations that has been affected by COVID-19 and teaching in an online setting. The survey was shared on social media, a dorm and acquaintances.

A lot of the input fields on the survey was free text, since the bachelor group was very interested in the users' own thoughts. This meant that when the survey was concluded, all the data was processed and put into different categories. This was done to create an overview of the different opinions. The complete survey form can be seen in appendix 9.1. A total of 35 people responded to the survey. 48.6% of the respondents were university students, 42.9% were high school students and 8.6% were from other educations (Appendix: 18).

4.1.2 Survey results

The following will discuss the survey's results divided into application domain and prototype. It is important to mention that the following results only are tendencies and implications. Since there is 35 respondents, there is not statistical evidence to conclude anything with certainty based on the surveys. The results are not representative[9], therefore, the bachelor group cannot conclude that every student have the same opinion about the application domain and prototype.

Application domain

The bachelor group asked the respondents about their opinion to their educations' way of handling teaching in an online setting. The majority of the respondents, on a scale from 1-5, think that the educations are doing a good job by rating them 3 or 4 (Appendix: 19). However, the difficulty of learning new theory or skills during COVID-19 and teaching in an online setting has increased. 80% of the respondents think it is difficult, on the contrary only 20% disagrees (Appendix: 20). The respondents were also asked to clarify and give details on why they think learning is more difficult in an online setting. Two respondents wrote the following:

Respondent 1: *"As my education is largely based on lab work it has been extremely difficult to obtain the needed skills as the professors/uni has not really succeeded in turning a lab course into something that actually brings knowledge and skills online."* (Appendix: 9.1)

Respondent 2: *"It is harder to concentrate for longer periods online than on campus. I find that the length of the sections of a lecture is too long to be able to concentrate enough. I also find it difficult to get help with exercises."*(Appendix: 9.1)

As it can be seen on the respondents answers, they find it particular hard to do activities that is related to their education. For example for someone studying biology or medicine, lab work is a very important part of their understanding and practice of a certain subject. Furthermore, it seems like people who are not used to sit in front of a computer for longer period of times, find it hard to concentrate during online lectures.

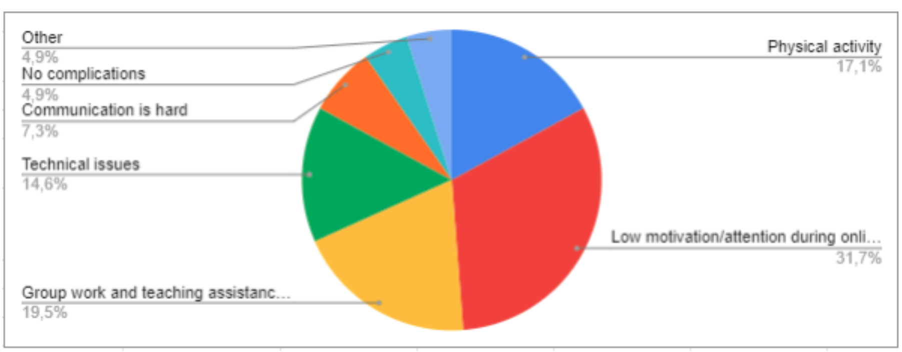


Figure 8: Summary of reasons behind the difficulty of learning in an online setting

On figure 8, it can be seen that there a several reasons to why learning in an online setting is difficult. 31.7% find low motivation as the main reason, whereas 19.5% find group work and teaching assistance hard, and 17.1% finds physical related activities difficult. Furthermore, when the respondents are asked for suggestions to solution or areas to focus on when developing a solution, 30.6% believes that a valid solution focuses on more interactivity, whereas 22.2% believes that teaching form in an online setting is too much the same and more creativity should be used to make teaching more interesting (Appendix: 22).

Based on the data gathered on the survey regarding the application domain, it indicates that there exists an issue with the learning outcome of teaching in an online setting. Students find it hard to learn new theory based on the lack of motivation and the way the teachings are performed. Furthermore, performing and showcasing physical activities as a part of interactive learning is very limited. This could be sports, lab work or normal class teaching. The survey indicates that there is room for improvements to make teaching more creative, but also to introduce more interactivity in the teachings when performed online.

Opinion about product

The bachelor group asked the respondents some questions related to the product. The respondents were provided a picture and a small description of the features of the final product. The respondents were asked *"can you see any advantages in the product"*. 33.3% had the opinion that the product had the most advantage, when used for presentations or teaching on blackboards. 23.5% believes that the advantage in the product is when doing physical activities. Only 9.8% thinks that one of the product's advantages is to contribute to interactivity. 7.7% don't think the product has any advantages. The results can be seen on figure 9.

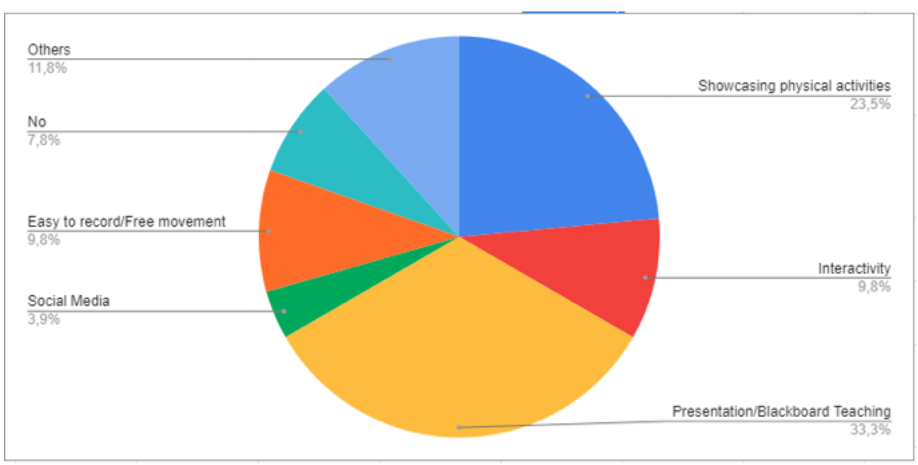


Figure 9: Summary of opinions about the advantages of the product

Two respondents wrote the following to the question regarding advantages:

Respondent 1: *"This product could be very useful for our sport lectures, where we have to film ourselves, this would make some of the exercises easier. One of the problems regarding sport lectures is when we have to record something in the exercises. Here, you use a lot of time as student to move around the camera, because you cannot be in the frame all the time. For example, we should create a swing program, which results in us getting out of frame a lot. This product could be useful for this kind of activity"* (Appendix: 9.1)

Respondent 2: *"This product would be great for teachers, when teaching in front of a blackboard or smartboard, as it would make it easier for both the students and the teacher to follow the writing, as opposed to now, where most teachers write with one hand and film with the other, which makes it very shaky and hard to follow."* (Appendix: 9.1)

The two respondents have different perspectives. Respondent 1 is more focused on lectures where physical activity plays an important role and respondent 2 values teaching on blackboards. It is interesting to see that the respondents have different mindset and think the advantages of the product is in different fields.

When the respondents were asked about the limitation of the product, 42.9% did not see any limitations. 8.6% had concerns regarding that some movements could be too fast to track. 8.6% thought that phone support could be an limitation. 5.7% thought that the product as a whole was too big and unhandy and another 5.7% had concerns regarding the battery lifetime of the phone (Appendix: 24). The respondents were also asked the question *"Where do you see this product being used?"*. 59% find the major usage of the product in lectures, on the contrary, 15.4% believes the major usage will be in regards to physical activities (Appendix: 25).

The final question the respondents were asked was if they thought the product could help improve learning within a course or subject. On figure 10, it can be seen that 60% thinks that the product has an influence on the learning outcome and can improve learning. On the contrary, 11.4% do not think the product has an influence on learning and 28.6% do not know if the product can improve learning.

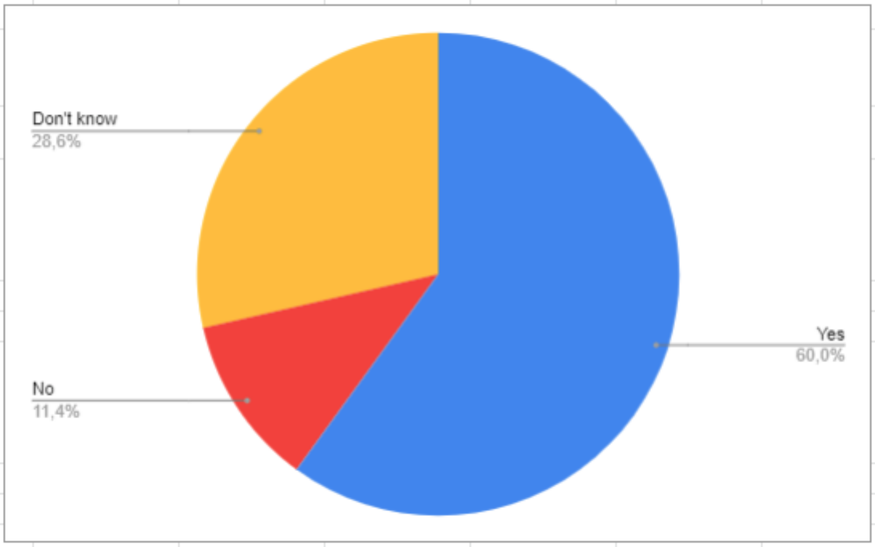


Figure 10: Summary of opinions about if the product can help improve learning

Two of the respondents wrote the following to the question regarding if the product can improve learning

Respondent 1: *"Yes I believe that it could improve learning, and specifically in an online setting, as it would make it easier for the students to follow the teacher and make it less shaky, as it is attached to a mount."* (Appendix: 9.1).

Respondent 2: *"I don't think technology on its own can improve learning significantly. That comes down to the teacher."* (Appendix: 9.1)

It is interesting to read that respondent 1 and respondent 2 disagrees with each other. Respondent 1 think that this product could help with the improvement of learning, since it can be used as an active and important part of the lecture. Respondent 2 disagree because they don't think that technology can improve learning. The question was based on a matter of opinion, so it was expected that people would express their own opinion. However, it is an interesting point to discuss. It will be discussed in a later section whether the developed product can improve learning.

Based on the gathered data regarding opinions about the product, it indicates that the respondents were divided in their opinions. The respondents focused on different aspects of the product, which they found important. This caught the bachelor group's attention, since some of the focus points were something that the group did not focus on. Something that was surprising was that only 9.8% of the respondents, thought that one of the advantages of the product was interactivity. The respondents think that the products main advantages is, when it is used for physical activities such as lab work, yoga, etc. and for presentation and teaching on blackboard. The majority of the respondents find no limitation in the product. A few thinks that the product will be limited on performance wise factors such as battery lifetime, ability to track fast movements, etc. When it comes to usage, the majority sees the product being used within physical activities and lectures. This fits together with the respondents opinion about the advantages of the product. The majority thinks that the product of the bachelor group will help or improve learning. However, when the respondents were asked to elaborate, they focused on different aspects. Some had the opinion that it would be most beneficial for an young audience and some believes that the product would be most beneficial when used in an online setting.

4.2 User Tests

To be able to reflect on whether the product works as expected fulfills FR-7 in section 2.1, a series of user tests were carried out. Usability and user experience is tested, as well as the limitations and shortcomings of the PT-mount. The test is also meant to capture objective opinions about the product, as not all issues can be observed from the perspective of the developer. The tests were carried out on 3 persons, which were acquaintances of the developers. This means that the test will give an idea about how well the app works, but the amount of participation is not enough to conclude anything at this time.

Structure

The test is facilitated by a facilitator which tells the test person what to do, while taking notes about the behavior of the test person on each task that is executed. The detailed template can be seen in appendix 9.2.

Results

The test revealed that the product was usable, with a few flaws, which was the consensus of the participants. The most conspicuous findings in the test was that the app lacked the opportunity for the user to get help, or become properly on-boarded and introduced to the app. When asked to use the grid view to place the test subject on the left side of the screen, the user needed help:

Participant: *"She did not find out. I had to show her."*(Appendix: 9.2)

Another interesting result was that the users did not see many limitations and found many features to be useful and the app to be robust:

Participant: *"It is very nice. Quality is good, and it is easy to use."* (Appendix: 9.2)

When asked to start tracking, the users found the correct button easily. Some of the users had problems figuring out what way to place the phone in the mount. In regards to tracking, the PT-mount tracked the user as they expected, sometimes the tracking would fail if the users made faster movements such as jumping or running around. A person said this about the tracking:

Participant: *"When turned the right way. It tracked very good when I moved at normal speed. When I did some jumping squads it got confused and stopped following me."* (Appendix: 9.2)

The users were asked to use some of the extra features such as MOT, auto zoom and the tracking boxes. The users were satisfied with these features as they found them useful in various situations. They were also easy to find and understand, especially with the explanatory text beneath each option. The downsides with the extra features was that it was never explained how they work (adjusting the grid view with two fingers etc.), and some features needed more fine tuning to work perfectly. An example, is the auto zoom which zoomed a bit too much, such that the tracking would fail. Here is what a participant had to say about it:

Participant: *"Not good. It zooms too close. When I move it could not find me before zooming out."* (Appendix: 9.2)

In general, the system would track well, but in most tests, the mount would loose the tracked person at least one time. Most of the time it was due to other people in the background or sometimes the test user was simply not detected. In order to get tracked again, the user had to enter the camera frame. In most cases this had minor impact on the user experience. A few small tweaks was made to the app based on the test findings, the tests were otherwise satisfying and showed no major downsides with the app.

4.3 Performance Tests

For this bachelor project, a number of performance tests have been carried out to locate limitations of the product. The following tests bring focus to certainty of recognition . Furthermore, the rotation algorithm is tested in different situations. For a more detailed view of the tests, please see appendix 9.3.

4.3.1 Certainty of recognition

A series of tests have been performed to test out the certainty of recognition from different distances between the system and the tracked object. A test was performed where the goal was to see how the application performed indoor. The result of this test can be seen in figure 11.

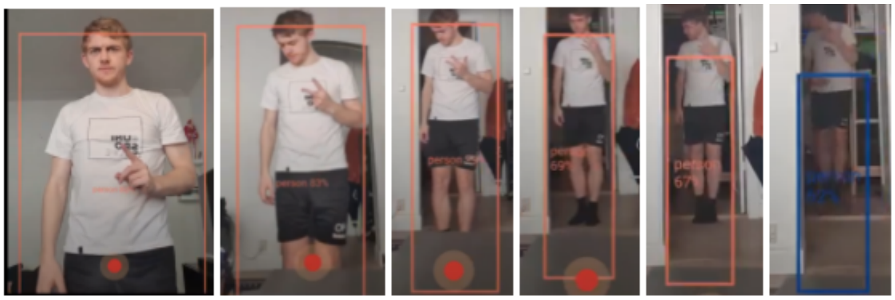


Figure 11: Picture of performed distance test indoor.

Looking at figure 11, it can be seen that the application recognized a person within the frame up to 5 meters. Once the person was more than 5 meters away from the camera, the detection algorithm was not able to detect the person with a certainty of more than 50%. On the following table, it can be seen how certain the algorithm was, that the detected object was a person.

Frame	**Distance**	**Recognition %**
1	1 Meter	80%
2	2 Meters	83%
3	3 Meters	82%
4	4 Meters	69%
5	5 Meters	67%
6	6 Meters	< 50%

The table shows that a drastic change happens between 3 and 4 meters. The detection algorithm is only able to detect that the object is 69% person. This change can be a result of multiple factors. The group assume, that the result could have been different, if the person would have had a bright red shirt on or perhaps if there had been better lighting in the room.

A different approach to recognizing and tracking persons was the use tags. This was tested to check, if a better solution would be for a person to move around with a tag that the system identifies. This can accommodate the issue with multiple persons being within the same camera frame, since the person being tracking is wearing an unique tag. To perform such a test, pieces of paper with a red stop sign have been printed, since the SSD-Mobilenet model can detect stop signs. Here the same test case as described above was performed to test this alternative solution. Figure 12 shows the result of this test.

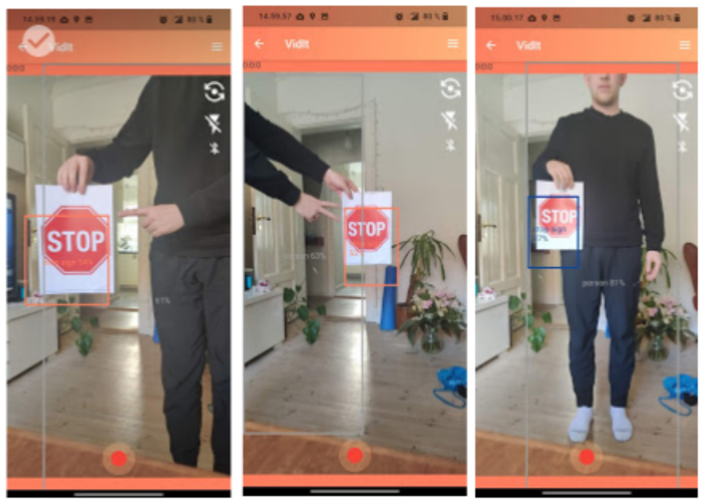


Figure 12: Picture of performed distance test with stop sign indoor.

Figure 12 shows that the detection algorithm was able to detect a stop sign with a certainty of more than 50% within 2 meters from the camera. Once the stop sign was more than 2 meters away from the camera, the algorithm was not able to recognize it with a certainty of more than 50%. If this solution should be implemented, the person wearing the stop sign tag would not be able to move more than 2 meters away from the camera. Which would not be a viable solution for most use cases. Furthermore, the size of the stop sign had an influence on the test results. When tested with a smaller size, the results became less accurate.

4.3.2 Box drawing

As seen on figure 11 and 12, the orange box surrounding the tracked object becomes less accurate in terms of dimensions and placement. As an example, the box is drawn almost perfectly on frame 3 on figure 11. When looking at frame 4 on the same figure, it can be seen that box is drawn slightly lowered compared to the person. This gets worse when looking at frame 5, where the head and shoulders are not within the orange box drawing. This problem is a result of incorrectly sized camera preview. This means that the preview displaying what the camera sees is not showing the full frame, which the model gets as detection input. The tracked person would not be correctly centered, compared to the camera preview, when the detection algorithm decides so.

Figure 13 show two frames of a person in motion. In the first frame, the person is walking casually. In the second frame, the person is doing a fast movement to the right.

Figure 13: Picture of box drawing in motion.

Looking at figure 13, it can be seen that the orange tracking box is drawn almost perfectly around the person on the first frame. It is slightly drawn to the left but has no problems following and detecting the person. On the second frame, it can be seen that the orange tracking box is drawn completely off. This is because the person is performing a quick movement to the side. This shows that the model is not fast enough to update the tracking box for each frame.

4.3.3 Live action

When testing the product in live-action events, different sports have been prioritized, to review how the product performed. Screen recordings of the performed test can be seen in appendix 9.3. Video 1 shows a person playing dart outside. Here the system tracks the person almost perfectly in terms of certainty of recognition and box drawing. The system lost track a few times but caught up with the person quickly which resulted in the person never getting out of frame.

Video 9 shows a person walking around the camera while playing basketball. When comparing this video to video 10 (about 4 min into the video), which shows a person also playing basketball around the camera wearing contrasting colors, a big performance difference in terms of recognition certainty can be seen. The system loses track on the person in video 9, a few times, and the person also gets out of the frame. The person has to step back into the picture for the camera to be able to follow him. Looking at video 10, the camera tracks the person almost perfectly and does not lose track. When comparing the average recognition certainty of the two videos, it shows that the detection algorithm performs better with the person in video 10 than video 9. After reviewing the videos it can be estimated, that video 10 has an average certainty of about 69%, whereas video 9 has an average certainty of about 60%.

4.3.4 YOLO model

To compare the performance of the used detection model (see section 3.3.1) a certainty test of recognition has been performed with the YOLO model.

Figure 14: Picture of certainty tests with the YOLO model.

Figure 14 shows the result of the certainty test performed with the YOLO model. It can be seen, that this model has no problem detecting a person in the frame, as expected, it performed better than the used model. The following table shows the certainty percentage of each frame.

Frame	**Distance**	**Recognition %**
1	1 Meter	68%
2	2 Meters	84%
3	3 Meters	83%
4	4 Meters	91%
5	5 Meters	94%
6	6 Meters	93%

The table shows that this model has a very high certainty that the detected object is a person. It is worth mentioning that this model performs best, if the object is more than 3 meters away. This is completely opposite to the results described in section 4.3.1, where the recognition got significantly worse when the object was more than 3 meters away.

Figure 15 shows the test with a stop sign, that was performed with the YOLO model. It can be seen that the model was able to detect the stop sign up to 4 meters distance from the camera. This is significantly better than what the SSD model was able to perform.

Figure 15: Picture of ”certainty stop sign test” with the YOLO model.

The reason why the YOLO model was not used compared to the SSD-Mobilenet model, is because the detection is a lot slower. When testing motion with the YOLO model, it lost tracking with the object much faster than the used model. If a user was doing motion and were to use a system containing YOLO, he would have to move much slower for the system to be able to track him. This is because the detection algorithm is slower as explained in section 3.3.

5 Limitations and Future Improvements

The following section will focus on limitations discovered during the above described tests and survey. Furthermore, there will be a focus on future improvements, which could be implemented as a result of the discovered limitations.

5.1 Performance

Section 9.3 showed that the final product had some limitations in relation to performance. The system was not able to detect fast enough, which resulted in a low frame-rate. As a result of this, a person was not able to move around too fast, since the system would simply lose tracking. This limits the intensity of physical exercises to for example playing dart, walking, and in general slow paced activities.

A possible improvement to this limitation could be to use the phone's GPU as processing power instead of the CPU if it is supported by the phone. The GPU can be used to run the detection algorithm. Tensorflow states on their website, that the new GPU backend performs two to seven times faster than the current floating-point implementation, which is used for this project[33].

Another limitation in the model could be improved by the support of GPU processing. For this project, the SSD-Mobilenet model has been used. Section 4.3.1 showed that the model performed significantly worse in terms of certainty when the detected person was more than three meters away from the camera. A drastic change occurred between three to four meters. Although, the YOLO model did not seem to have these limitations. If GPU support was implemented, the system could perform object detection faster and the YOLO model could be used with higher frame rates. This was the limitation that made the bachelor group decide not to use it.

5.2 Rotation

The performance test showed, that the PT-mount was not able to rotate fast enough to be able to track a person doing fast movements. This was due to two problems. Firstly, the detection algorithm was not fast enough to detect a person in fast movement. A solution to this was described in section 5.1. Secondly, the gearing of the part, which is responsible for panning the camera, was too big compared to how many steps the stepper motors were able to perform each second as described in section 3.2. This limited the mount to pan at a slow speed compared to how fast it was able to tilt. The reason was that the gearing of the part responsible for tilting were smaller.

A solution would be to print a new PT-mount with a smaller gearing for the

[33]Tensorflow Blog - TFLite GPU

part responsible for panning the phone. To be able to do this, the group have to make a new design which would be very time-consuming. It is worth mentioning, that if the performance is not improved, then an improvement of rotation speed does not make sense. This is because the system would not be able to detect fast enough, resulting in the same problem described in section 3.4.2.

5.3 Movement prediction

To make the mechanical rotation of the PT-mount smooth, a three-step rotation speed was implemented. A movement prediction algorithm could be used to make the mechanical rotation more efficient, being responsible for calculating an exact acceleration speed.

To implement such functionality, the group would have to compare the position of the person on the current frame with the position from the last frame. This way, the group could calculate the expected movement speed of the person, which would benefit the system with an even smoother experience.

5.4 Auto-zoom

Even though a version of an auto-zoom functionality has been implemented, it is still not satisfactory when looking at the results of the user tests. The issue is, that it quickly losses tracking of the person since the system is not able to locate where the person is in the picture.

A possible solution could be to make the system use an extra camera, which is not zoomed in. The idea is that the system uses an extra camera for determining, where the person is in the picture and pan-tilt to keep this person in the middle of the screen. Whereas, the other camera will just be responsible for zooming in and recording the video. This functionality would be like the new function which Xiaomi introduced in their smartphone "*Xiaomi mi 11*"[2]. Figure 16 shows a screenshot from a video displaying the new functionality from Xiaomi. In the upper left corner is a preview that displays what the camera sees without it being zoomed in. The general screen shows the camera preview which is zoomed in.

Figure 16: Picture from a video showing the new functionality in Xiaomi Mi 11.

By implementing this functionality, the group expects that the system would be able to avoid losing track of a person. There is a possibility, that the person gets out of frame, but the system does not have to zoom all the way out to detect the person again.

5.5 User guide

Another result of the user test was that some features were difficult to understand and use. Some of the users wished that the system would have a guiding system or help functionality, which could ease out the process of learning to use the system. One way of implementing such a functionality is to have a startup guide. This startup guide would take the user through the different functionalities of the system, and show them how to use them in different situations.

5.6 Multi Object Tracking

As mentioned in section 3.3.1, the detection algorithm includes a simple re-ID algorithm that uses boundaries and the center offset of the detected object. This functionality could be improved with extra attributes like shirt color and distance from the camera, to be able to detect the desired object with higher certainty. To be able to acquire distance information the system have to use two cameras. The most recent phones on the market comes with at least two back cameras, but some older models might not have multiple cameras. Therefore, the functionality is limited to specific phone models.

Another approach for determining what object to track is to implement a tag solution as mentioned in section 4.3.1. The person whom the object have to track would be equipped with a specific tag. For this solution to work, the model have to be trained with a larger dataset on the specific tag. This is because the detection algorithm is not able to detect the tag when more than three meters away from the camera with the SSD Mobilenet model. Another limitation of this approach is the placement of the tag. An example of this, is a person that has a tag attached to his chest and then laid down or turned around. The system will not be able to detect him.

5.7 Box drawing

Section 4.3.2 mentions a problem where the box is drawn incorrectly due to the dimensions of the preview being incorrectly calculated compared to the frames delivered to the detection algorithm. This issue limits the model in terms of keeping the object in the grid view center. When the orange tracking box is drawn incorrectly (meaning slightly lowered), it will result in the object not being centered correctly. In some cases, this will result in the head of the person not being present in the frame which is an unpleasant user experience. To improve this experience, a way to calculate the preview dimensions corresponding to the frames coming from the camera stream, have to be found.

5.8 Live-streaming

As of today, the final prototype cannot be used by students, teachers and others in real-time. This limits the use of the product to prerecorded videos. Going back to the use case with a student recording herself doing some warm-up program. There is no way for the student to show her teacher the warm-up program in real time, making the product less interactive and interesting.

The group believes that an implementation of a live-stream feature is a functionality that could improve the product the most, when thinking about the research questions. As mentioned in section 1.1 interactive learning can affect students to become emotionally invested in both lectures and presentations. When live-streaming, the students and teachers has to be more aware of their surroundings and pay attention, since they in most cases cannot go back and rewatch the live-stream. Live-streams are distinct from a traditional prerecorded videos because of the currently viewing audience, and the real-time discussion between the audience and the streamer[16].

6 Discussion

The following will discuss and reflect upon the developed product. It will be discussed whether the product can solve the problem statement and issues stated in this bachelor project.

6.1 Increased learning with VidIT

Through the development of this bachelor project, the main goal has been to develop an IT product that can help improve the learning quality in an online setting. Survey data has been gathered, performance and user tests has been utilized, and research has been performed. The remaining question is: "Can the resulting product improve the learning quality in an online setting based on the information the group has obtained?"

Based on the results from the survey, it can be seen that the majority of teachers fulfill the first two theories presented by the University of Vanderbilt, see section 1.1. Teachers use the cognitive load by focusing on the intrinsic load by ensuring a connectivity and relevancy between a subject and the course. They ensure germane load by showing, helping and providing tasks where it is important to master the theory in order to perform a proper analysis. Teachers also reduce the extraneous load by making sure there is no distracting elements in the videos such as unnecessary background music. Focusing on non-cognitive elements, teachers are good at speaking conversational instead of formal, speaking in different speeds and with enthusiasm. Furthermore, visual representations are also used by teachers to emphasize theories or subjects. In general, teachers are good at following the theory of cognitive load and using non-cognitive elements in order to impact the engagement of the students.

Even though teachers are good at fulfilling cognitive load and non-cognitive elements, there is always room for improvements. It can be argued that the product developed by the bachelor group can be used as a non-cognitive element. It would be possible to use the video from the product as a visual representation of some theory or practical tasks, such as theorems shown on a blackboard or experiments in a lab. The survey shows that some students lack motivation, since the lectures are very similar and there is not enough variation, see section 4.1. On one side, the product can provide a video that benefit and provides a more interesting visual representation than a stationary camera. This means that it will be easier to make lectures that varies in visual content, since there can be more dynamic videos where movement is tracked rather than still images. On the other side, it can be argued how big of an impact the product has on learning in an online setting. It can definitely be used to increase student engagement and make a lecture more visually interesting. However, it is questionable whether it has a direct impact on the students understanding and learning of a subject.

The theory presented by the center of teaching at the University of Vanderbilt also substantiates an issue with the interactivity part of educational videos. This is indicated on the results from the surveys, see section 4.1. When teaching in an online setting, it can be difficult to include interactivity in the lectures. Some teachers use guiding questions, or questions the audience, in order to include them in the discussion. However, when used in an online setting, it can be difficult for the students to maintain attention, see section 4.1.

Based on Dongsong Zhang's investigation of interactivity in videos[1] and Eric Mazur's survey regarding interactivity in a normal lecture[5], it can be said that they found evidence to support the fact, that interactivity in lectures and videos has an influence on a students understanding and learning of a subject. Therefore, it is important to include interactivity in lectures whether it is online or not. One of the tools that can be used to increase interactive learning is to give the students control and influence on the actions in the lecture[1]. The product developed in this bachelor project are able to support interactivity in the form of a prerecorded video. Unfortunately, the current state of the product does not allow for real time streaming. This means that the different uses of the product are very limited and this can have an effect on how the product can increase interactivity and learning. If the product allowed live-streaming, teachers would be able to use the product for real time teaching. They could utilize the product in combination with zoom[34], so when doing an experiment in a lab, the teacher could ask the class for input. For example, he could do some experiments with different kinds of liquids, where he walks between different tables in the room. He could then ask the students what liquid he should put in next and why that specific liquid should be used. While the teacher moves around, VidIt tracks him, keeping him in frame at all times. This increases the interactivity, since the students will have a bigger influence on what happens in the lecture and feel more engaged when the teacher is more physical active. Even though, the bachelor product does not provide streaming in real time and have limits to what it can be used for, it can be argued to still have an influence on the learning outcome. One respondent wrote the following in the survey:

"My electronics course that was supposed to be 50% lab work was made 100% theoretical, same for semester project"(Appendix: 9.1).

It can be seen that the respondent's practical part of the electronics course was changed to be only be theoretical. The product developed in this bachelor project would be able to ease the work of teachers and students, and make it easier for them to record themselves. It can increase the learning outcome for the students, since the teachers and students can show practical experiments and work again without any complications.

Another area where the learning outcome can be improved is when the stu-

[34] Zoom.us

dent themselves use the product. An example of this could be when students has to perform some physical activities, where technique has a significant role. They can use VidIT to record themselves easily, without having to adjust the phone camera all the time. They can send their video to their teacher, who can give proper feedback, since the teacher is able to see all the detail and technique, due to VidIT tracking the student.

It requires a lot of extensive tests to confirm with certainty that a tool or IT product can increase learning. Based on the theory presented in the section 1.1 and the bachelor group's survey and tests, it is hard to say with certainty that the developed product can increase learning. The product definitely gives the teachers the option to record more practical parts of the lectures to share with their students, which can increase the learning outcome. However, when talking about interactivity in lectures, it is hard to measure and obtain data due to the fact that the product does not allow for streaming in real time. Ideas to how the learning outcome and interactivity could have been measured more effectively, will be discussed below.

6.1.1 Changes in testing strategy

In retrospective, the bachelor group could have utilized more effective tests to investigate more precisely, if the developed product could increase the learning outcome of students.

The bachelor group could have adapted the same technique used by Eric Mazur[5]. A teacher could have held two online lectures at the same time, where half of the class was using zoom and the other half was using zoom together with VidIT. After the lecture, a survey or some questions have been prepared to check how much the students remembered and understood from the lecture. If the result showed that the students better understood and could use the theory from the lecture where VidIT was involved, it could be concluded that the product had an influence on the learning outcome and the interactivity. However, this form of testing is complicated. In order to get precise results it would require the participants to have the same educational background. People are studying on high schools or university and comes from different elementary schools and education places spread out over the country. This means that the people participating in the test will have different basis knowledge, which could have an influence on the test results. Furthermore, effective learning is different for different students. Some students learn best by doing practical work and others learn best by reading.

Another and more advanced test could be to measure the students' brain waves when participating in a lecture with and without VidIT in use. The results could then be compared to see if there is a difference in brain activity. There is two types of learning. Explicit and implicit learning[10]. The activity, we are interested in are the explicit learning activity in the brain, that is activated when we learn and memorize things such as theory, content in a book etc.[10]. If it can be seen that there is more brain activity from students in a lecture where the product is used, it would be possible to argue that the product can help improve the learning outcome as well as student participation. However, one of the limitations of this approach is the price range of the equipment needed to conduct this test. An electroencephalography (EEG) is needed, which has an average price of 934$ in the USA[35]. According to related work, described in section 1.2, in the subject regarding teaching in an online setting, it was clear that teachers had concerns for the changes to teaching in an online setting. Many could not see any advantages of it. The related work does describe some concrete solutions incorporating IT products. A valid solution to the teachers' concerns could be the implementation and utilization of VidIT.

6.2 Comparison of cost and simplicity

When looking at the related products that is on the marked. The different companies have various tracking methods, most uses machine learning models to track based on image recognition as described in section 1.3. The quality of the detection may vary with the price, but it is hard to determine the better AI implementation even with the product in the hand. DJI focuses a lot on hardware, their build quality is very good and as it is the 4th version, it is a refined product.

The companies mentioned in section 1.3 have put a lot of work into their apps, making it easy to share pro quality videos on social media. The purpose of VidIT was to simply and easily be able to record home videos. To do more than that, the users will have to use more money, as it is safe to say that the VidIT mount was rather cheap to make. However, the development of the AI technology and app is the largest aspect as the most hours were used developing that particular feature. A lot of the prices on the related products are not based on hardware but on the software quality on which VidIT certainly have way to go. Some of the products also support tracking while streaming, which allows the app to run in the background. This would be a very good feature to have, to make VidIT a competitive product. The apps of the products focuses more on editing the captured videos and sharing on social medias like TikTok[36]. Companies like GetPivo briefly shows how their product can be used for teaching in one of their commercials[37]. While the rest promotes tracking capabilities the most[38].

In this field of competition, VidIT fits in well when measuring price and simplicity. VidIT does not have many advanced features, but is much cheaper than the alternatives, and with the mostly positive feedback from the tests done in

[35] How Much Does an EEG Cost? — MDSAVE
[36] TikTok
[37] GetPivo — Facebook
[38] OBSBOT Commercial — Facebook

section 4, the average high school student might want to buy it. The main reason for this is the price, the VidIT system can be produced for a minimum price about 400 DKK, seen in section 3.5, which is roughly 65$. It is therefore safe to say that the VidIT system is very competitive in terms of price, with the closest competitor being GetPivo costing 164$ as mentioned in section 1.3. The VidIT system will still be relevant even when raising the price to accommodate for unexpected expenses or inaccurate calculations of the price estimate.

If comparing VidIT to related work mentioned in section 1.2, the VidIT mount is a rather simple solution. Especially compared to the LMS and AR proposals, these solutions will be expensive to realise, but the learning outcome is better. The LMS provides a more wholesome system for learning remotely, while the VidIT system only solves one delimited issue with remote sport pedagogy. VidIT may be more practical than the ILTC system introduced in section 1.2, as VidIT can also tilt and does not need to run on a stationary computer. This is the biggest difference between VidIT and ILTC. ILTC have a frame rate of 24 while VidIT only achieves around 10 fps as described in section 3.3.1. The difference in fps is of course due to the hardware limitations in smartphones, but the question is whether the tracking would be significantly improved by somehow increasing the frame rate, because ILTC still have problems with rapid movement even with a detection rate on 24 fps. Extra sensors like IR might help VidIT to become more reliable, but the cost compared to the tracking gains might not be enough.

6.3 Future relevance

Since high schools and universities are opting, and sometimes forced, to cancel face-to-face lectures as a way to minimize the spread of the COVID-19 disease, new ways to keep the institutions going have emerged. Most students anno 2021 have experienced teaching in an online setting like Microsoft teams[39] or Zoom[40] as a daily form of lecture. It is difficult to conclude whether educational teaching will continue in this manner, but it is known that online teaching has brought multiple benefits for the students. According to Bernardo Da Silva, this new setting improves class attendance since it is easier for students to access, and because students are more willing to ask questions when they are not in front of their colleagues[6]. However, the preparation for both teachers and students has proven to be more time-consuming in terms of preparing and managing different types of hardware and software as described in section 1.2.

The main goal of the developed prototype was, that it should ease out the process of recording yourself. This process of creating a video single-handedly is very time-consuming, if a lot of movements has to be performed during the video. VidIT can help students and teachers to create such videos in a time-saving manner, since the user does not have to adjust the camera manually

[39] Microsoft teams
[40] Zoom.us

during the recording. Furthermore, the product is cheap compered to other related products as described in section 6.2. This makes the product attractive for both students and teachers.

Based on the result of the survey and the user tests, it can be seen that many of the participants could see a use for VidIT within an online setting. Most users argued that they could use VidIT for presentations, laboratory subjects, and physical activities. Unfortunately, most presentations in high schools and universities happens in real time, meaning that they take place on some live-streaming platform, which VidIT is not able to compete against. If the future improvements and especially the live-stream functionality as described in section 5.8 were implemented, then the product could be a candidate suggestion for a product which potentially could be a time-saving educational technology used by students and teachers. Furthermore, interactive learning could be improved by implementing the live-stream functionality. Giving the user the option to explain theory and subject while showing their actions in real time. An example could be within the subject of physics where the professor can do an experiment and based on the feedback of his students could react or do certain things in a lab, that they request.

Knowing this, online teaching will be of relevance in the future. A system like VidIT will be beneficial to this form.

7 Conclusion

In this thesis, an IT solution called VidIT, with the purpose to improve learning in an online setting, has been developed. The system is a cheap implementation of an automated tracking and recording system, which students and teachers can use for online learning and teaching purposes. A pan-tilt mount keeping a smartphone in place, enables for motorized rotations to record the person and keep the user in frame with no help of others.

The first part of the survey in the report showed a need for change in relation to online teaching. Students found it difficult to acquire knowledge about new theory, due to lack of motivation and badly prepared and performed lectures. The students wished for additional interactivity in order to accommodate these needs. The second part of the survey showed the respondents opinion to VidIT. The respondents found the product most effective in an environment with physical activities, presentations, lectures and blackboard teaching. Furthermore, the respondents had concerns in regards of limitations in battery life time, fast movements etc.

The product was a success when comparing the quality with production price and time used developing. The product invites to record a more visual appealing educational video which both students and teachers can benefit from. After discussing the ability of the product to increase learning in an online setting, it can be concluded that the product needs to support live-streaming. The reason being that VidIT could be most efficient when students are presenting homework or teachers are lecturing. If the online teaching should be enhanced by VidIT, it would be in a scenario where the teacher is physically active. The teacher has to be able to move around, either doing physical exercises or moving between different work stations. With these elements combined, VidIT would be of more relevance to increase learning outcome.

The user tests carried out, showed that the app was simple in relation to usability. However, the users sometimes found it difficult to interact with the PT-mount. The test could conclude that the system worked as intended. However, there were only three participants, which rendered the test unusable to conclude anything with certainty. It could be concluded based on the user tests that the tracking should be faster and more performant. It was explained in the sections about the system architecture, that this was not achieved due to hardware and software framework limitations. Time constraints also played an important role in most cases where the system was lacking. The tracking capabilities of the system was tested in the performance test section. The tests showed beginning limitations when tracking persons who were more than three meters away or was running around. The YOLO object detection model was introduced with tradeoffs compared to the SSD model. It was concluded that the SSD was best for the purpose of VidIT, utilizing faster detection on the cost of accuracy.

Based on the tests conducted in the report, it can be concluded that VidIT will have a tough time competing with existing products on the market, as they have many features and were extensively developed. The tracking of VidIT was the most problematic, while the PT-mount was beneficial, supporting both panning and tilting, which only DJI had implemented.

The functional requirements of VidIT has been fulfilled except for FR-7, since the users found some features complicated to use as explained in section 4.2.

The developed system can be used by both students and teachers for recording videos single-handedly. The current state of VidIT has a small effect on the learning outcome of students, but can be improved significantly by implementing additional as well as improving current features.

8 References

[1] Cynthia J Brame et al. *Effective educational videos*. 2015.

[2] M. Brownlee. *The World's Largest Smartphone Camera!* Youtube. 2021. URL: https://www.youtube.com/watch?v=R5_v0OdFD0A&t=424s&ab_channel=MarquesBrownlee. (Accessed on 05.05.2021).

[3] Neil Cameron. "Servo and Stepper Motors". In: *Arduino Applied*. Springer, 2019, pp. 157–176.

[4] Isaac Chasteau. *DSLR Camera Pan Tilt Mount (Stepper Motor Driven)*. Thingiverse. 2020. URL: https://www.thingiverse.com/thing:4316563. (Accessed on 05.05.2021).

[5] Catherine H Crouch and Eric Mazur. "Peer instruction: Ten years of experience and results". In: *American journal of physics* 69.9 (2001), pp. 970–972.

[6] Bernardo Marques Da Silva. "Will Virtual Teaching Continue After the COVID-19 Pandemic?" In: *Acta medica portuguesa* 33.6 (2020), pp. 446–446.

[7] Hamid Ghasemi and Leonardo Jose Mataruna-Dos-Santos. "An Overview of New Opportunities for Training Sport and Physical Education Courses in the Corona Pandemic". In: (2021).

[8] Ross Girshick et al. "Rich feature hierarchies for accurate object detection and semantic segmentation". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2014, p. 1.

[9] Diego Graglia. *How many survey responses do I need to be statistically valid? Find your sample size*. Survey Monkey. URL: https://www.surveymonkey.com/curiosity/how-many-people-do-i-need-to-take-my-survey/. (Accessed on 23.04.2021).

[10] Becky Ham. *Brain waves reflect different types of learning*. URL: https://news.mit.edu/2017/brain-waves-reflect-different-types-learning-1011. (accessed: 09.05.2021).

[11] *Idræt C*. Børne- og Undervisningsministeriet. 2009. URL: https://www.retsinformation.dk/eli/lta/2009/66. (Accessed on 05.05.2021).

[12] Sule Kirbas. "The Views of Physical Education and Sports Teaching Instructors on Education in the COVID-19 Period." In: *Journal of Education and Learning* 9.6 (2020), pp. 196–205.

[13] Shawn Litingtun. *Introduction to the ESP32 WiFi / Bluetooth Wireless Microcontroller*. Predictable designs. 2019. URL: https://predictabledesigns.com/introduction-to-the-esp32-wifi-bluetooth-wireless-microcontroller/. (Accessed on 05.05.2021).

[14] Wei Liu et al. "Ssd: Single shot multibox detector". In: *European conference on computer vision*. Springer. 2016, pp. 21–37.

[15] Marcia Sahaya Louis et al. "Towards deep learning using tensorflow lite on risc-v". In: *Third Workshop on Computer Architecture Research with RISC-V (CARRV)*. Vol. 1. 2019, p. 1.

[16] Zhicong Lu, Seongkook Heo, and Daniel J Wigdor. "Streamwiki: Enabling viewers of knowledge sharing live streams to collaboratively generate archival documentation for effective in-stream and post hoc learning". In: *Proceedings of the ACM on Human-Computer Interaction* 2.CSCW (2018), pp. 1–26.

[17] Allegro MicroSystems. *DMOS Microstepping Driver with Translator And Overcurrent Protection*. Allegro MicroSystems. 2014. URL: https://www.pololu.com/file/0J450/a4988_DMOS_microstepping_driver_with_translator.pdf. (Accessed on 05.05.2021).

[18] Fred Paas, Alexander Renkl, and John Sweller. "Cognitive load theory and instructional design: Recent developments". In: *Educational psychologist* 38.1 (2003), pp. 1–4.

[19] Joseph Redmon et al. "You only look once: Unified, real-time object detection". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, p. 779.

[20] Tan-Hsu Tan, Tien-Ying Kuo, and Huibin Liu. "Intelligent Lecturer Tracking and Capturing System Based on Face Detection and Wireless Sensing Technology". In: *Sensors* 19.19 (2019), p. 4193.

[21] Kevin Townsend, Carles Cufı, Robert Davidson, et al. *Getting started with Bluetooth low energy: tools and techniques for low-power networking.* " O'Reilly Media, Inc.", 2014.

[22] Random Nerd Tutorials. *ESP32 Bluetooth Low Energy (BLE) on Arduino IDE*. Random Nerd Tutorials. 2018. URL: https://randomnerdtutorials.com/esp32-bluetooth-low-energy-ble-arduino-ide/. (Accessed on 05.05.2021).

[23] Random Nerd Tutorials. *ESP32 Pinout Reference: Which GPIO pins should you use?* Random Nerd Tutorials. 2018. URL: https://randomnerdtutorials.com/esp32-pinout-reference-gpios/. (Accessed on 05.05.2021).

[24] Valeria Varea, Gustavo González-Calvo, and Alfonso Garcıa-Monge. "Exploring the changes of physical education in the age of Covid-19". In: *Physical Education and Sport Pedagogy* (2020), pp. 1–11.

[25] Morgan K Williams. "John Dewey in the 21st century". In: *Journal of Inquiry and Action in Education* 9.1 (2017), p. 7.

[26] Dongsong Zhang et al. "Instructional video in e-learning: Assessing the impact of interactive video on learning effectiveness". In: *Information & management* 43.1 (2006), pp. 15–27.

9 Appendix

9.1 Summary of survey data

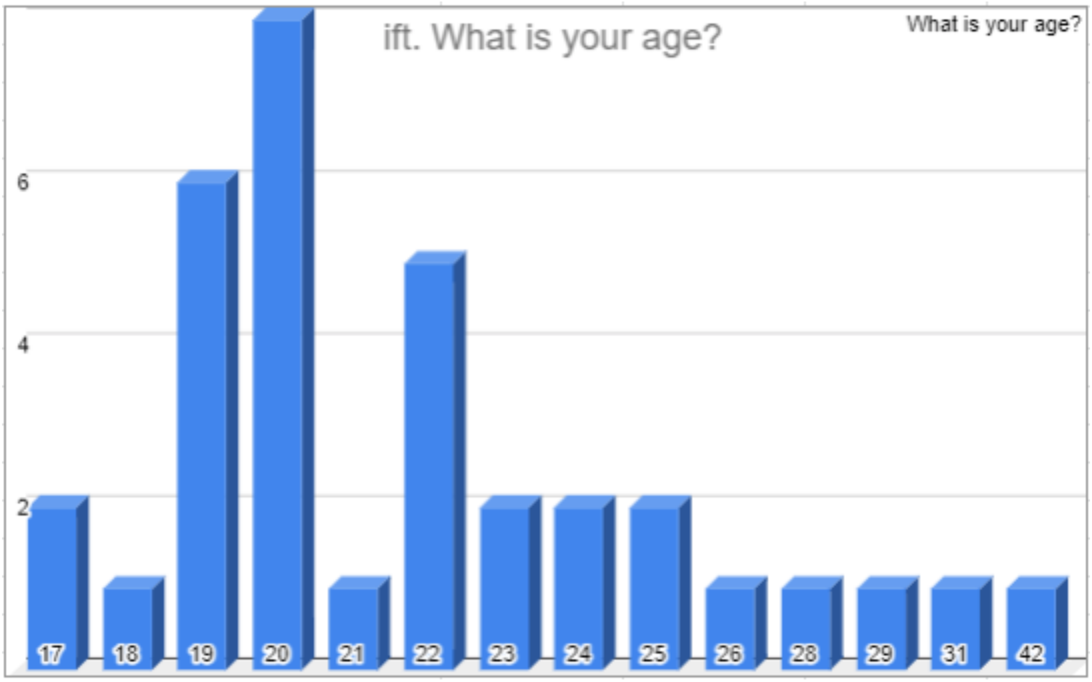


Figure 17: Summary of age of the person's participating in the survey

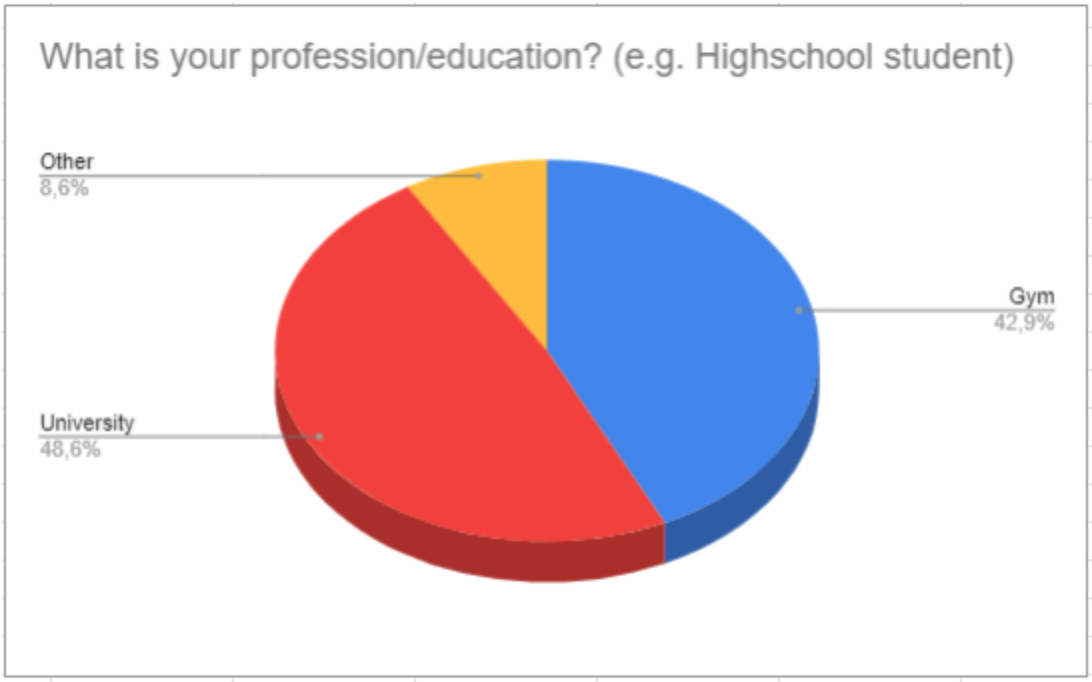


Figure 18: Summary of the profession or education of the person's participating in the survey

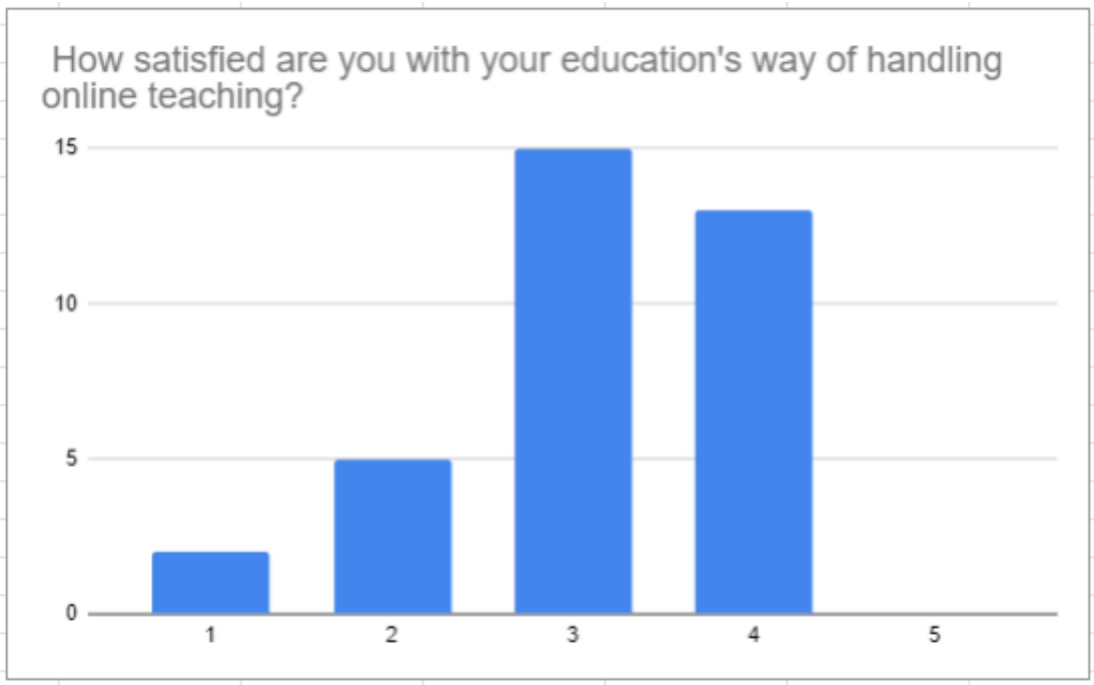


Figure 19: Summary of the satisfaction of educations' way of handling online teaching

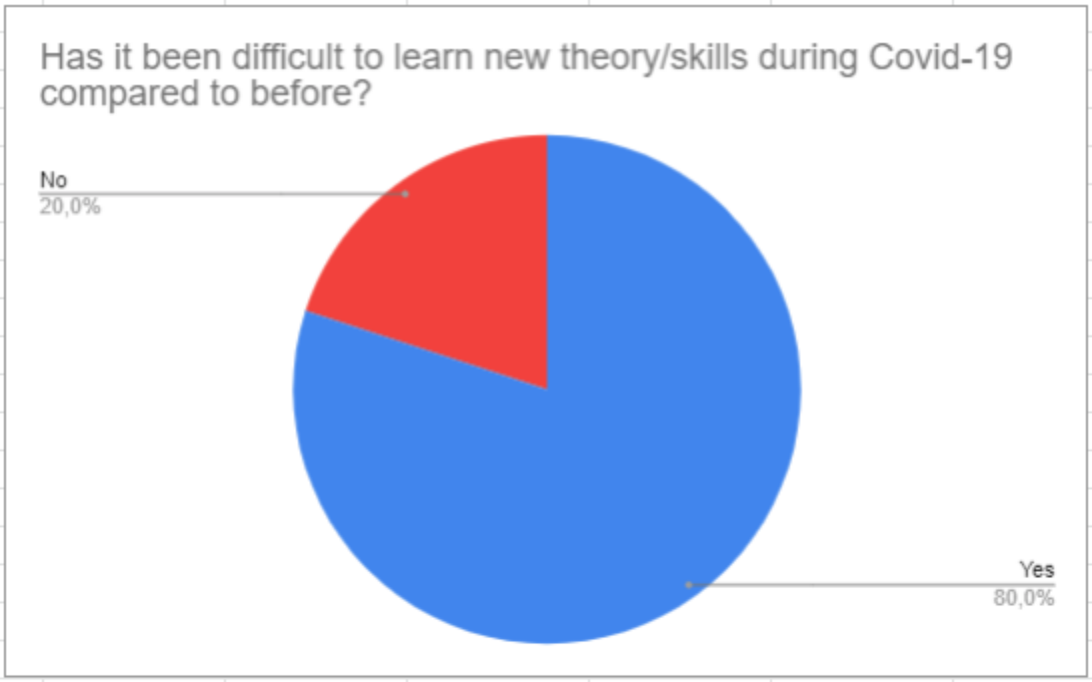


Figure 20: Summary of the difficulty to learn new theory or skills during COVID-19

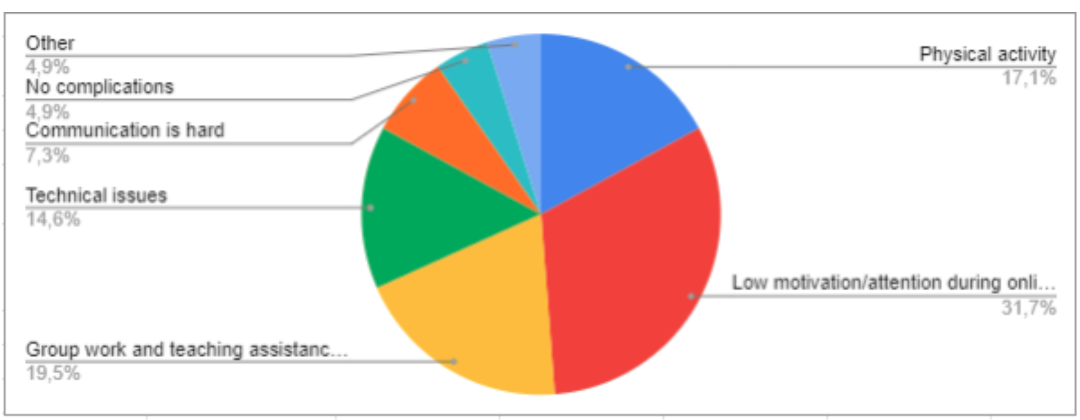


Figure 21: Summary of reasons behind the difficulty of learning in an online setting

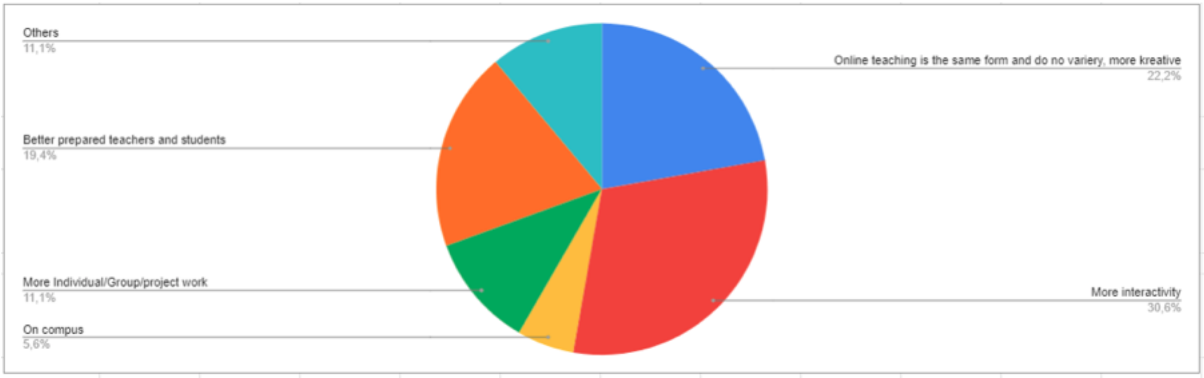


Figure 22: Summary of ideas how to increase learning in an online setting

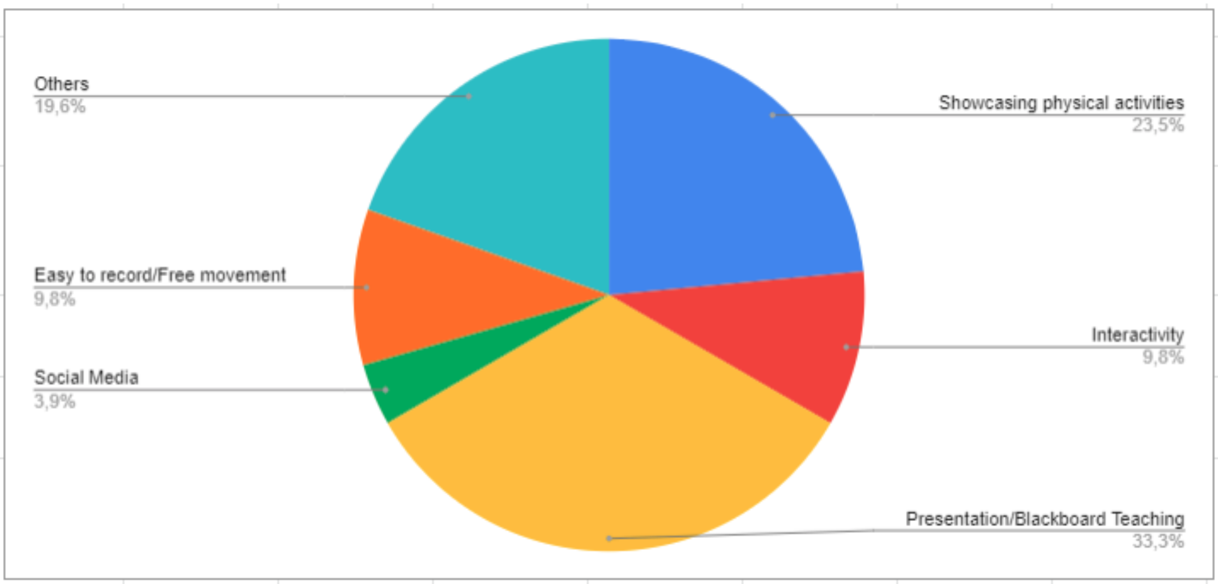


Figure 23: Summary of the opinion about the advantages of our product

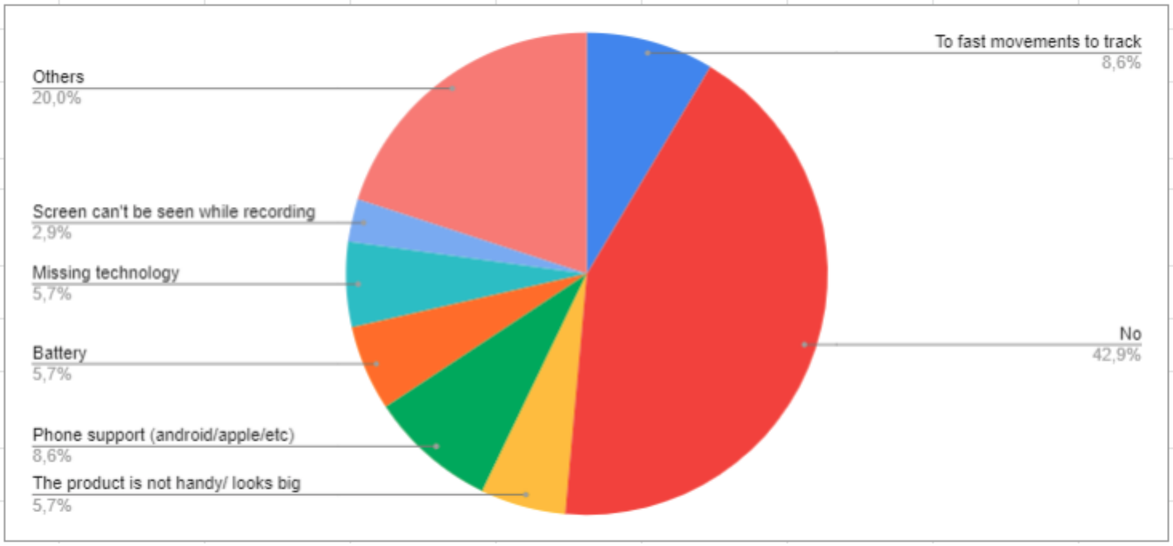


Figure 24: Summary of the opinion about limitation of our product

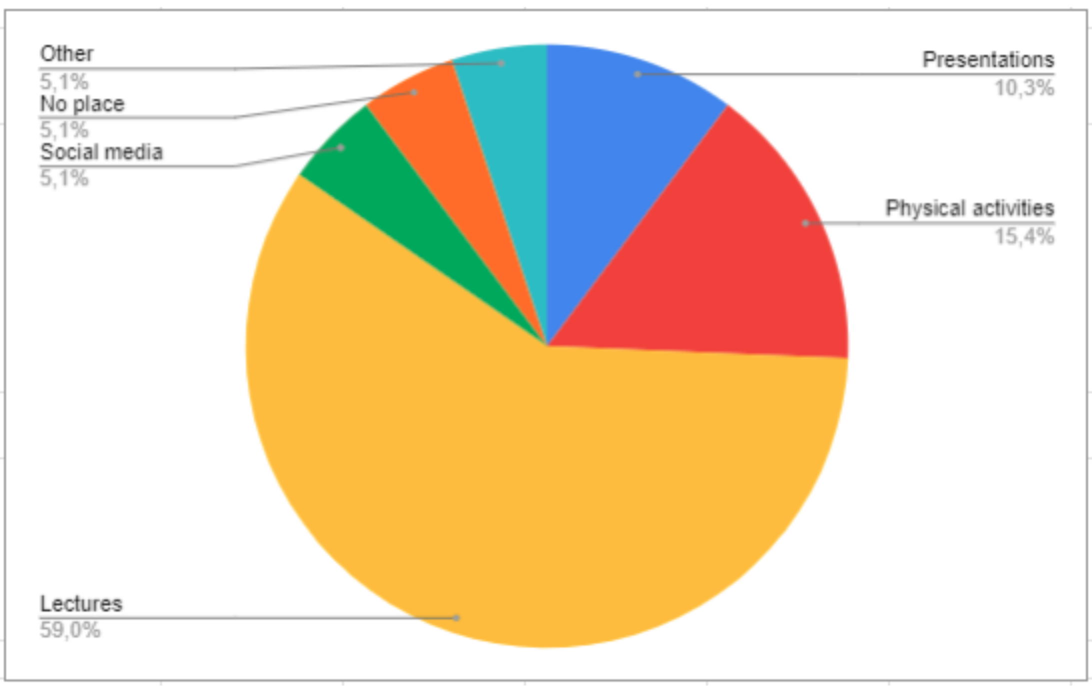


Figure 25: Summary of where they see our product being used

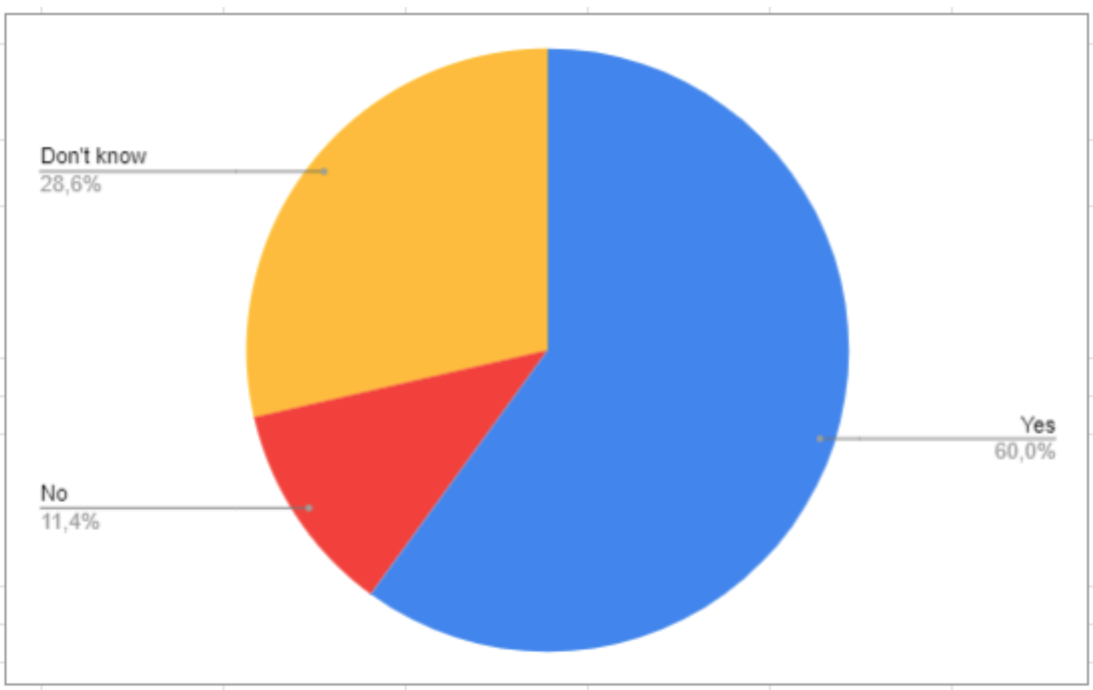


Figure 26: Summary of the opinion about if our product can increase learning

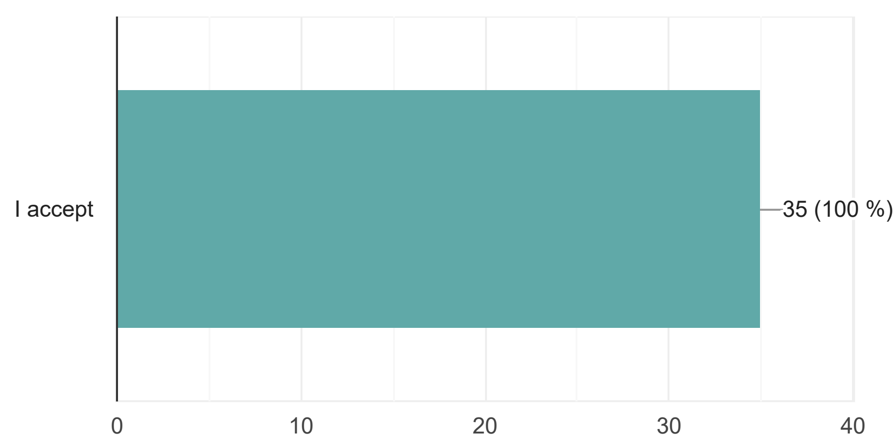
VidIT - Bachelor Project - IT University of Copenhagen

35 svar

Offentliggør analyse

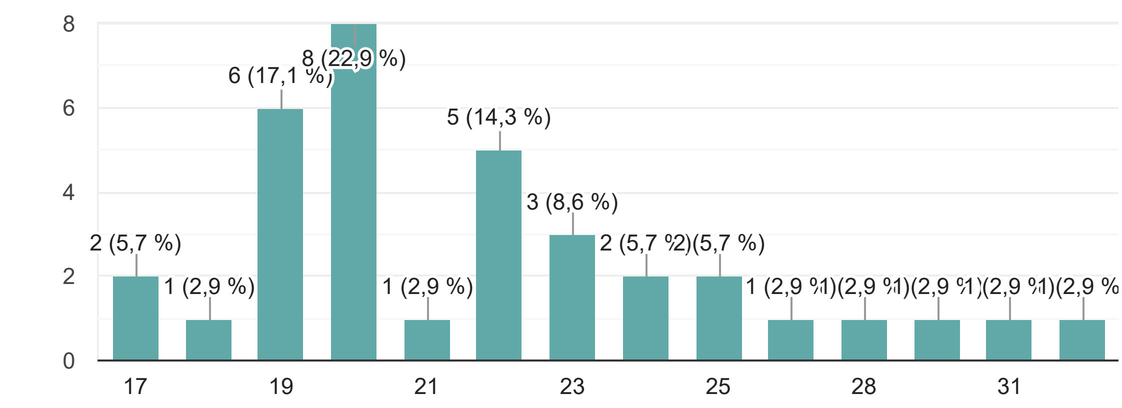
Disclaimer: The questionnaire survey is anonymous and the gathered information will only be used in the development of this Bachelor Project. The only persons who has access to the information is us and only a summary/conclusions will be available to the university. When the Bachelor Project has been completed all information will be deleted. The email provided will only be used to contact you, if you have won the contest. By accepting the disclaimer, you give use permission to use the information that you have submitted in our project.

35 svar



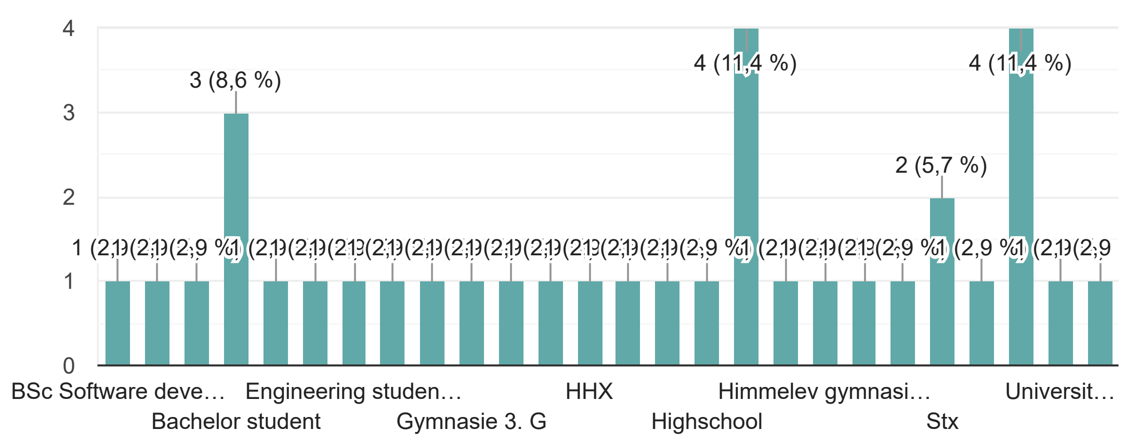
What is your age?

35 svar



What is your profession/education? (e.g. Highschool student)

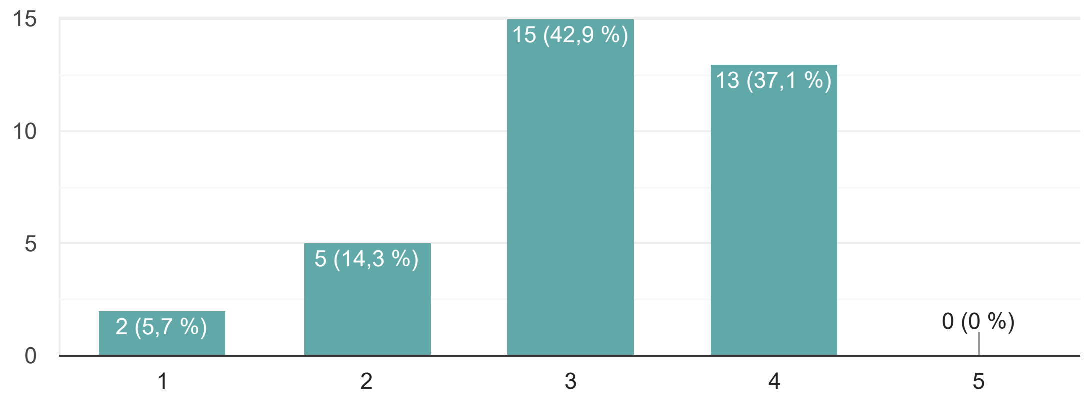
35 svar



Research questions

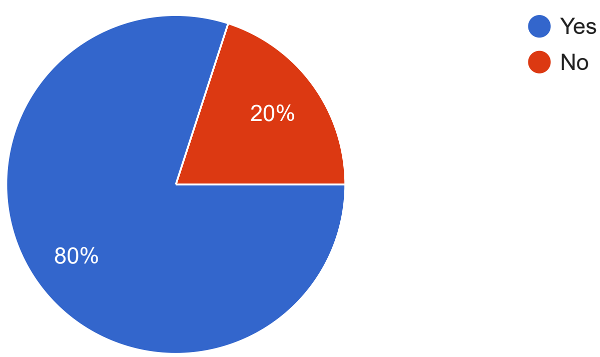
How satisfied are you with your education's way of handling online teaching?

35 svar



Has it been difficult to learn new theory/skills during Covid-19 compared to before?

35 svar



9.2 User test template with answers

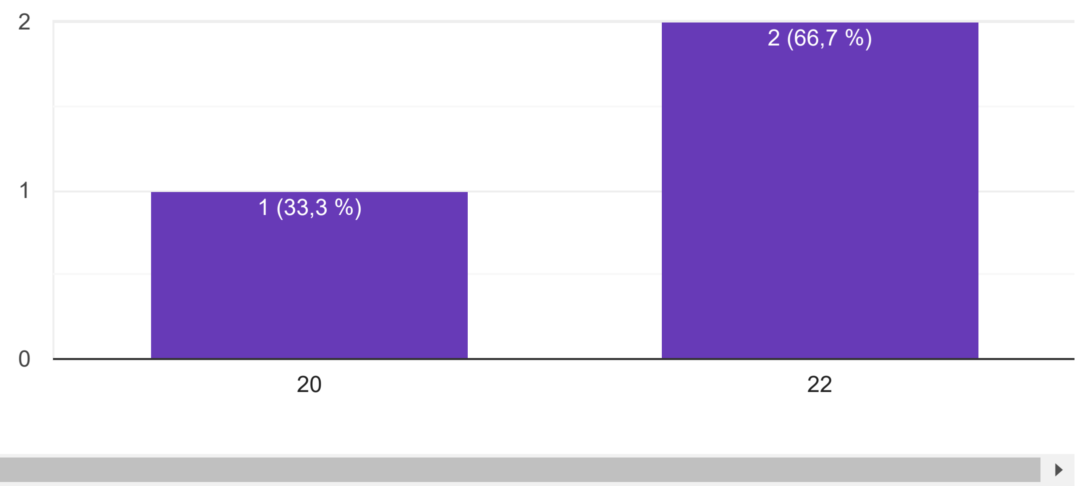
Information

3 svar

Offentliggør analyse

How old are you?

3 svar



What is you profession/education?

3 svar

Student - Biotech

Highschool student (Himmelev)

Student - For PE Teacher and German

Connect the phone to the mount

Is it easy to find and connect to the mount?

3 svar

Easy, but it needs a (better) feature to tell that it is connected.

Yes. The bluetooth logo was nice but may it should change color when connected

No, the bluetooth would not connect, and I had to reset the Arduino servaral times, he could not have found this out himself.

How would you use the mount to track yourself?

3 svar

Just place the phone and click start tracking. When place the phone, she turned it the wrong way.

She would place the phone in landscape mode with the front camera

put the phone in the mount and press start tracking

Start tracking your self

How did it go?

3 svar

The head is not in frame when she is in landscape mode, she switches to portrait.

She pushed the start tracking button. But the tracking was with the back camera. And she could not figure out to turn the cam. I had to do this.

Good, until Bluetooth got disconnected, I reconnected for him.

Did the mount begin to move?

3 svar

Lidt dumt at den tracker med det samme. (Måske knappen kan kaldes noget andet). Yes it did.

Yes. It saw a person which was not Laura. So it turned unintended.

yes

How/When did you expect it to move?

3 svar

She did not expect it to move. Maybe add a pop-up to explain explicitly that the mount will begin to move when "OK" is pressed.

When I move around I guess.

Not right away like this, as he was not ready for it to move, because he was too close to the mount so it would just keep going up, so I had to set it back, so it could track him.

Does the button name make sense?

3 svar

It was easy to find the button, but she did not think the mount would start tracking so NO it did not make sense.

Yes. That was easy to understand.

He thinks so...

Start performing some action (yoga, dance, etc)

How do you think the mount is tracking?

3 svar

She does yoga. The body fills the frame sometimes because she is spreding her legs, and therefore they are not inside the frame entirely. When it comes to speed, the mount tracks really good and follows her just fine.

When turned the right way. It tracked very good when I moved at normal speed. When I did some jumping squads it got confused and stopped following me.

He starts going around in front of the camera, and it tracks as he expects, but a little bit slow

Use the other camera of your phone

Does it make sense to switch camera view?

3 svar

Yes it was easy.

Now she could find it when I showed her earlier.

He made the app crash, because he pressed the switch button too many times, but the issue is only on my (OnePlus) phone.

Was it easy to shift the phone around?

3 svar

Yes, she found easily out how to switch the phone around in the mount.

yes.

he thought so yes.

Remove the orange boxes around the person

Did you find the Menu?

3 svar

Yes. Easy.

Yes. That was easy. It looks like many other apps

no, he tried to swipe/touch directly on the boxes, i told him about the menu in the right corner...

Is it easy to do?

3 svar

yes. Good explanations to the menu settings.

yes

Yes, after he managed to open the menu, he easily found the setting

When would you use the Tracking Boxes option?

3 svar

If she had to use it for physical activity she would use it. If she uses it for something where she has to look at herself (e.g. make up) she would remove it.

hmm always I think. Can not see why I should not use them

When someting (red. tracking) is not working right.

Turn on the grid view

Can you see the purpose of having the grid view on?

3 svar

Yes. But there should be a "first time you open guide" which shows that you can change grid view. It clearly requires some explanation

Yes. It is a nice feature to see when the cam should move

not at first, it is asymetric, which he did not understand, i told him it was to center his head better.

Is it easy to find and enable?

3 svar

Yes.

yes. Just like tracking boxes.

yeah, good explanation on the settings menu

Use two fingers to change the grid view to have the person on the left side of the screen

How did it go?

3 svar

Very easy. Nice feature.

She did not find out. I had to show her.

he thought that the app should say how to do this. or else he would not have found out how to do it. He realized how the grid could help him now.

I'm gonna step into the picture now. - Try out the Multi Tracking option

Was it easy to make the phone track yourself?

3 svar

Yes. The functionality worked fine.

Easy to find the check but it lost me sometimes.

yes, still a bit slow, but it did not track any others (it was just 2 persons) than him.

Turn on auto zoom and test it (by moving away)

Did you find out how to do this?

3 svar

Yes. It took some time to figure out that one had to go further away before it worked.

yes easy to find.

yes

How do you think the auto zoom performs?

3 svar

OK. But there should be a setting so you can decide how much it should zoom.

Not good. It zooms to close. And when I move I could not find me before zooming out.

a bit weird but he sees the purpose of it

Record a video of yourself

How do you think the app performs?

3 svar

Very good. A little annoyed that one has to wait so long. Tracking is smooth and well functioning.

The video looks nice. Good quality.

good, everything works, sometimes he was not detected(I could see) but he did not notice.

Is the recording good?

3 svar

Yes. Very high quality.

yes

yes, good quality

Does the tracking work?

3 svar

Yes.

yes

yes but he went out of the frame a couple of times because it had slow pan-movement

Does it hurt your experience significantly that you have to wait for the recording to be saved?

3 svar

A little. Not much.

no. It did not take so long

not at all, but then i told him it would be longer the longer the recording was, and then he had some concerns, but nothing major he said, he could just do someting else on his phone while saving he said, because he could see by the small circle indicator when it would be done he said.

Go back and try to control the mount manually

Did you succeed?

3 svar

Yes. However, it turns wrong right / left when I film myself in portrait. Annoying that default speed is 0.

yes. But she wondered why it was 0 speed

yeah, it was a piece of cake

When would you use this?

3 svar

You can use it to set your starting position.

hmm she did not know

maybe controlling the mount with another phone on?

Ending points

In what context do you see a use for this product?

3 svar

Yoga and sport. Presentations and formal stuff. Would also be good for a youtuber.

When I do my sport lectures. It followed me good but not when I jumped.

he could well use it when he had to record himself juggling. He could well see it being used in teaching, since the parents are not always there to record the pupils in primary school (where he has been teaching)

Did you find any limitations to the product or something you didn't like?

3 svar

See above. Men der kunne godt være en "Help" feature

Jumping

It is a little slow when panning, the blue boxes could maybe stay a bit longer, as he was lost alot of times when moving too fast.

What is your general opinion about the product?

3 svar

It is very nice. Quality is good and it is easy to use.

It was very nice and easy to use when you know every feature.

It is cool!

9.3 Performance test

The following is the performance test performed both indoor and outside. The videos contains distance tests, motion test and test with an alternative tag solution.

Frederik, Mathias O, Mathias M

Link to videos: https://photos.app.goo.gl/9YV3JofXEGnM1iLB8
Link til billeder med skildt: https://photos.app.goo.gl/B1ezMv7P2EvCWGny8

Video 1

Length: 0.53
Activity: Single person playing dart outside from about 4 meters from the camera
Camera settings: Multitracking, normal grid, no auto-zoom, portrait mode
Person percentage: 69-77%
Box drawing: Dimensions is very good but is drawn slightly lowered
Camera pan/tilt: It follows the object perfectly. The object never gets out of frame.

Video 2

Length: 1.43
Activity: Single person play dart outside from about 3 meters from the camera
Camera settings: Multitracking, normal grid, with auto-zoom, portrait mode
Person percentage: 72-82%
Box drawing: Dimensions are very good and draw close to the object.
Camera pan/tilt: The object moves out of the frame sometimes and the mount zooms out before it can find the object again.

Video 3

Length: 0.53
Activity: Mathias distance test starting at 1 meter going to 8 meters where there is strong back light
Camera settings: Single tracking, normal grid, no auto-zoom, protrait mode, indoor, strong backlight
Person percentage:

- 1 meter: 72%
- 2 meter: 82%
- 3 meter: 79%
- 4 meter: 73%
- 5 meter: 77%
- 6 meter: 63%
- 7 meter: lost tracking of object
- 8 meter: lost tracking of object

Box drawing:

- 1 meter: It is perfect. Person fills out almost the whole screen
- 2 meter: Still perfect. Box top is in the middle of the persons face
- 3 meter: Still good, but it is slightly lowered
- 4 meter: Still good, but it is lowered
- 5 meter: It is ok, but there is to much space under the person
- 6 meter: It is not so good, beginning to draw lowered from person
- 7 meter: lost tracking of object and starts detecting persons at random places
- 8 meter: lost tracking of object and detects persons at random places

Video 4

Length: 0.55
Activity: Mathias distance test starting at 1 meter going to 8 meters where there is no strong back light
Camera settings: Single tracking, normal grid, no auto-zoom, portrait mode, indoor, weak backlight
Person percentage:

- 1 meter: 78-80%
- 2 meter: 82%
- 3 meter: 83%
- 4 meter: 70%
- 5 meter: 63-69%
- 6 meter: 50% and gets blue
- 7 meter: lost tracking of object
- 8 meter: lost tracking of object

Box drawing:

- 1 meter: It is perfect. Person fills out almost the whole screen
- 2 meter: Still perfect. Box top is in the middle of the persons face
- 3 meter: Still good, but it is slightly lowered
- 4 meter: Still good, but it is lowered
- 5 meter: It is ok, but there is to much space under the person
- 6 meter: It is no good. It lost tracking and draws a blue box lowered from the person
- 7 meter: lost tracking of object and starts detecting persons at random places
- 8 meter: lost tracking of object and detects persons at random places

Video 5

Length: 5.10
Activity at timestamp (1.48-2.21): Dribbling with basketball in <u>normal speed</u> outside from about 4 meters distance from the camera
Camera settings: Single tracking, outside, 4 meters from the cam, portrait mode, strong backlight, person is not having contrasty clothes on.
Person percentage: 57-77%
Box drawing: The box is draw in perfect dimensions but is sligthly lowered from the tracked person. It detects pretty good, but sometimes it looses tracking and gets blue for about half a second.
Rotation: The rotation is very good and the object is in the frame all the time. Sometimes it gets blue but catches up on the tracked object.

Activity at timestamp (2.37-3.07): Dribbling with basketball in <u>fast speed</u> outside from about 4 meters distance from the camera
Camera settings: Single tracking, outside, 4 meters from the cam, portrait mode, strong backlight, person is not having contrasty clothes on.
Person percentage: Lost tracking with person because of too fast movements
Box drawing: It is drawing ok, but the person moves out of frame because of too fast movements.
Rotation: The mount is not able to pan/tilt fast enough to keep track of the person.

Video 6

Lenght: 1.57
Activity at timestamp(0.04-0.55): Person dribbling with basketball from side to side (no stop) at normal speed.
Camera settings: Single tracking, outside, 4 meters from the cam, portrait mode, string backlight, the person is not having contrasty clothes on.
Person percentage: 66-77%
Box drawing: It draws perfectly but slightly lowered. One time it lost tracking and gets blue but quickly finds tracking again.
Rotation: The mount is able to keep the tracked object in the frame all the time. Sometimes it is quite close to losing the tracking but is not happening.

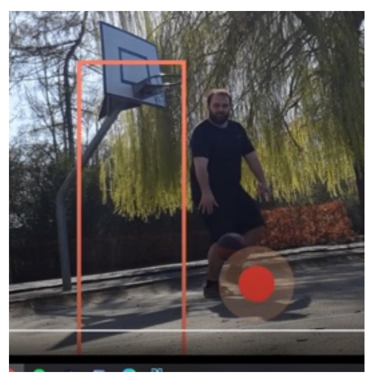
Activity at timestamp(0.55-0.55): Person dribbling with basketball from side to side (no stop) at fast speed.
Camera settings: Single tracking, outside, 4 meters from the cam, portrait mode, string backlight, the person is not having contrasty clothes on.
Person percentage: 62-76% lost tracking 2 times but caught up with the object
Box drawing: Drawing is good but loses tracking with the person a few times but is able to catch the person again and continue the tracking.
Rotation: The mount is able to barely keep the object in the frame. It is very close to losing the tracking but still manages to keep the tracking.

Video 7

Length: 1.25
Activity at timestamp (0.02-0.46): Dribbling with the basketball from side to side in slow tempo
Camera settings: Landscape mode, outside, 5 meters from the cam, strong backlight, the person is not having contrasty clothes on.
Person percentage: 69-74%
Box drawing: Left and right dimensions are perfect, but the box is too high and low (meaning it is drawing too much at the top and too much at the bottom of the screen).
Rotation: The mount keeps up with the person perfectly and there is a lot of room both left and right.

Activity at timestamp (0.46-1.02): Dribbling with the basketball from side to side in normal tempo
Camera settings: Landscape mode, outside, 5 meters from the cam, strong backlight, the person is not having contrasty clothes on.
Person percentage: 69-74%
Box drawing: Left and right dimensions are perfect, but the box is too high and low (meaning it is drawing too much at the top and too much at the bottom of the screen).
Rotation: The mount tracks the person very good and there is still nice room both left and right.

Activity at timestamp (1.02-0.46): Dribbling with the basketball from side to side in fast tempo
Camera settings: Landscape mode, outside, 5 meters from the cam, strong backlight, the person is not having contrasty clothes on.
Person percentage: 53-68%
Box drawing: It draws ok but sometimes struggles a little bit. The movement is very close to being to fast for the mount to keep tracking. It also gets blue very quick sometimes. The drawing also lacks a bit because the person is moving too fast:



Rotation: The mount is barely able to track the person. It gets a little confused sometimes and almost loses tracking of the person but is able to keep the person in the frame.

Video 8

Length: 6.13
Activity at timestamp (0.41-2.13): Shooting with basketball with slow tempo from about 5 meters from the camera
Camera settings: Landscape mode, 5 meters from the cam, strong backlight, person not having contrasty clothes on, single tracking
Person percentage: 57-72%
Box drawing: It is drawing the box ok. Still, the box is too high and low (meaning that it is drawing too much over and under the person). Sometimes it detects a very small person very quickly and then goes back to the actual person.
Sometimes the mount also tracks the other group of people instead of the intended object because multitracking is not enabled.
Rotation: The mount keeps tracking the object very good with lots of space on both sides.

Activity at timestamp (2.13-2.59): Shooting with basketball with normal tempo from about 5 meters from the camera
Camera settings: Landscape mode, 5 meters from the cam, strong backlight, person not having contrasty clothes on, single tracking
Person percentage: 62-74%
Box drawing: It is drawing the box ok. Still, the box is too high and low (meaning that it is drawing too much over and under the person). Sometimes it detects a very small person very quickly and then goes back to the actual person.
Sometimes the mount also tracks the other group of people instead of the intended object because multitracking is not enabled.
Rotation: The mount keeps tracking the object very good with much space on both sides.

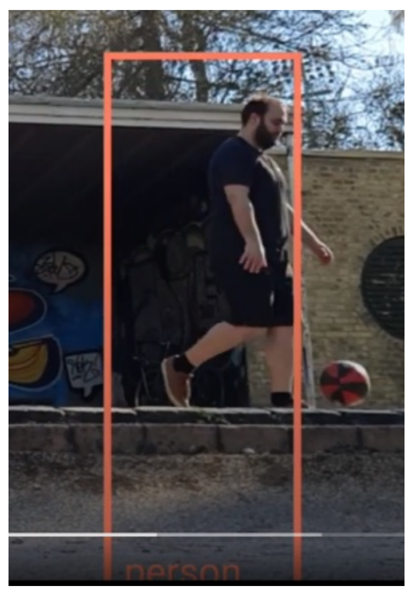
Activity at timestamp (3.18-2.59): Shooting with basketball with fast tempo from about 5 meters from the camera
Camera settings: Landscape mode, 5 meters from the cam, strong backlight, person not having contrasty clothes on, single tracking
Person percentage: 62-74%
Box drawing: The box is draw ok around the object still it is too big. It is also sligthly lacking meaning that it is a little bit behind the person when he is moving to one side.
Sometimes the mount also tracks the other group of people instead of the intended object because multitracking is not enabled.
Rotation: The mount keeps track with the person but is struggling more because the person is moving faster. Sometimes it is close to lose tracking but manages to keep the object in frame.

Video 9

Length: 3.11
Activity at timestamp (0.07-0.31): Dribbling around the camera in slow temp outside with changing light conditions
Camera settings: Landscape, outside, 5 meters from the cam, person not having contrasty clothes on, single tracking
Person percentage: 62-71%
Box drawing: The box drawing is very good still a little to big. Sometimes it loses tracking to some very small object on the ground wich the phone think is a human.
Rotation: The mount keeps track with the object until it loses the tracking to the very small object

Activity at timestamp (0.48-2.06): Dribbling around the camera in slow temp outside with changing light conditions
Camera settings: Landscape, outside, 5 meters from the cam, person not having contrasty clothes on, multi tracking
Person percentage: 62-72%
Box drawing: The box drawing is very good still a little to big. With multi tracking enable the mount does not lose track of the person even with other people in the background.
Rotation: The mount keeps track with the object until it loses the tracking to the very small object

Activity at timestamp (2.06-3.11): Dribbling around the camera in slow temp outside with changing light conditions
Camera settings: Landscape, outside, 7 meters from the cam, person not having contrasty clothes on, multi tracking
Person percentage: 62-72%
Box drawing: The box is not drawing very good when he is this far away, it is very much lowered.
With multi tracking enabled the mount does not lose track of the person even with other people in the background.
Rotation: The mount keeps track with the object with extra space on the left and right. Sometimes the object is moving a little to fast and gets out of frame.



Video 10

Comment: This is a very good showcase video.
Length: 6.52
Activity at timestamp(3.28-4.55): Person with contrasty clothes on dribling around the camera in slow/normal tempo
Camera settings: landscape mode, multitracking, 3-4 meters from the cam
Person percentage: 67-81%
Box drawing: It is drawing the box fine around the object still it is a little to big, but does not lose tracking of the object so often, since the person have contrasty clothes on that makes it easier for the camera to detect the object as a person.
Rotation: The mount tracks the person perfectly. He does not get out of frame even then doing quick shot movements. It gets blue sometimes but is able to catch up with the object.

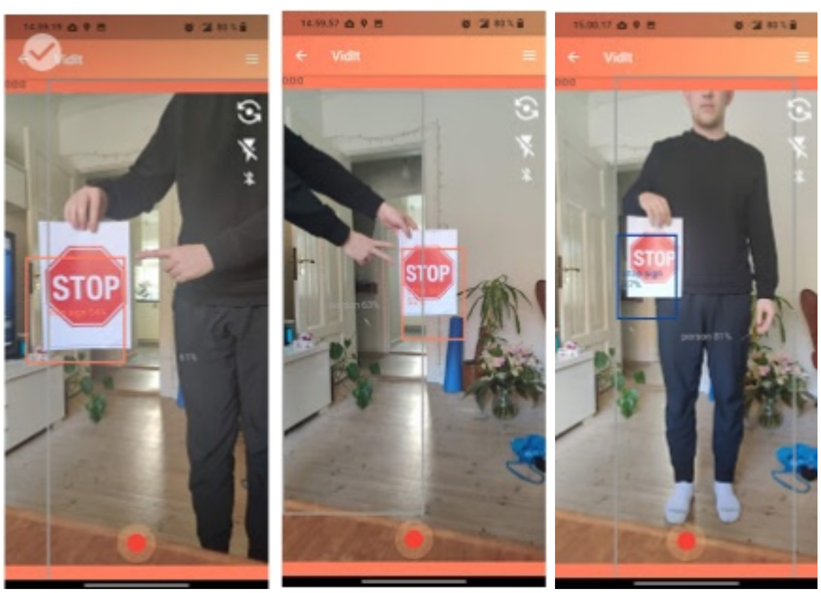
Test with stop sign

Stop sing percentage:

- 1 meter: 54%
- 2 meter: 52%
- 3 meter: >50%
- 4 meter: Not able to detect

Box drawing:

- 1 meter: Fine a little lowered
- 2 meter: Fine a little lowered and to big
- 3 meter: Losing detection
- 4 meter: Not able to detect



YOLO model test

9.4 User Stories

1. As a student who has sport lectures during COVID-19, I would like to record myself with no help of others
2. As a student, who owns a smartphone, I would like to store the video locally on my phone, so i can share it later.
3. As a student with limited money, I would like to use my phone to record myself
4. As a user, who doesn't like wired setups, I would like to connect wireless to the system.
5. As a student of a high school, I would like it to be simple and easy to use.
6. As a student who has sport lectures, I would like the camera to record me even if I move out of the frame.
7. As a student who would like to look at myself while I record, I would like to be able to use the front camera of my phone
8. As a student who records myself for sport lessons, I would like to record in portrait and landscape mode.
9. As a person that moves a lot on my videos, I would like to be able to record 360 degrees.