# P1: Determining the Best Camera Angle of Deafblind Interaction

Project report

Group itc1b117

Aalborg University
Department of Electronic Systems
Fredrik Bajers Vej 7B
DK-9220 Aalborg Ø

**Department of Electronic Systems**
Fredrik Bajers Vej 7B
DK-9220 Aalborg Ø
https://es.aau.dk

**AALBORG UNIVERSITY**
STUDENT REPORT

**Title:**
Determining the Best Camera Angle of Deafblind Interaction

**Theme:**
Networks and programming

**Project Period:**
Fall semester 2018

**Project Group:**
ITC1b117

**Participants:**
Jonathan Velgaard Sørensen
Marius Frilund Hensel
Nicholas Bernth Strømgaard Hansen
Thomas Flø Østergaard Madsen
Victor Büttner

**Supervisors:**
Gilberto Berardinelli

**Copies:** 1

**Number of Pages:** 48

**Date of Completion:**
December 19, 2018

**Abstract:**

This paper investigates and proposes a solution, in a proof of concept state, for a problem encountered by The Center of Deafblindness and Hearing Loss in Aalborg. The problem mainly consists of an analyst wanting to prove that if deafblind students perform adaptive physical activity together, rather than separately, they will thrive. To prove this hypothesis, the analyst needs proof in the form of recordings. The proposed solution is a system consisting of multiple cameras, a server, and an interface to analyze the recordings.
However, only a proof of concept meant to incentivize further development is developed. This proof of concept shows that it is possible to transfer footage to a server using RTSP, transfer recordings to a client using SMB, and provide an interface to analyze recordings using the Qt toolkit. It was not, however, tested whether the quality of a given camera angle using OpenCV and Caffe deep learning framework could be correctly evaluated in a production environment. Since the proof of concept is incomplete, suggestions for further work are also proposed.

# Nomenclature

| *Abbreviation* | *Name* |
|---|---|
| WHO | World Health Organization |
| CDH | Centre for Deafblindness and Hearing Loss |
| GDPR | General Data Protection Regulation |
| EU | European Union |
| SMB | Server Message Block |
| FTP | File Transfer Protocol |
| DVR | Digital Video Recorder |
| NVR | Network Video Recorder |
| PoE | Power over Ethernet |
| FPS | Frames Per Second |
| POC | Proof of Concept |
| RTSP | Real Time Streaming Protocol |
| TCP/IP | Transmission Control Protocol / Internet Protocol |
| NetBEUI | NetBios Extended User Interface |
| IPX/SPX | Internetwork Packet Exchange / Internetwork Sequenced Packet Exhange |
| RFC | Request for Comments |
| UFW | Uncomplicated Firewall |
| APT | Advanced Package Tool |
| CLI | command-line interface |
| DHCP | Dynamic Host Configuration Protocol |
| NTLM | NT LAN Manager |
| SYN | Synchronize |
| ACK | Acknowledge |
| SYN, ACK | Synchronize-acknowledge |
| FIN, ACK | Finish-acknowledge |
| MD4 | Message Digest 4 |
| ARP | Adress Resolution Protocol |
| VPN | Virtual Private Network |
| GUI | Graphical User Interface |
| IDE | Integrated Development Environment |
| UI | User Interface |
| APA | Adaptive Physical Activity |
| UTF | Unicode Transformation Format |
| BGR | Blue Green Red |
| RTP | Real Time Protocol |
| HTTPS | Hypertext Transfer Protocol Secure |
| RGB | Red Green Blue |

# Contents

# Introduction 1

Deafblindness can either be congenital or acquired later in life. Congenital deafblindness can be caused by multiple diagnoses and, in Denmark, Rubella Syndrome is the leading cause. Acquired deafblindness is often due to old age, other times it's caused by being born blind or deaf and acquiring the other disability later in life.

In 1980 the nordic countries collectively made a definition for deafblindness, but was later changed in 2006. It got revised again in 2016 based on FN standards and classifications from World Health Organization (WHO). The new definition made it easier to share data abroad. [1]

**Definition of deafblindness**
"*Deafblindness is a combined vision and hearing impairment of such severity that it is hard for the impaired senses to compensate for each other. Thus, deafblindness is a distinct disability.*" [2]

This project was made in collaboration with the national Centre of Deafblindness and Hearing loss (CDH) of Denmark, which offers education, respite care, accommodation and more.

Therefore, the project will focus on both pre- and adolescent students from the centre, who are described by their supervisor as very different from each other in terms of level of disability and personality.

## 1.1 Task description

An analyst at the CDH, has the thesis that the students with hearing loss actively try to communicate with the deafblind students. Furthermore, he believes these interactions are positive and beneficial for both parties.

In order to prove this thesis, empirical evidence is needed, which is done by recording an Adaptive Physical Activity (APA) lesson. An APA lesson is where the students perform different physical exercises depending on their capabilities. The lesson is done in the gymnasium seen on Figure 1.1 on the following page, where the aforementioned interactions happen.

Post recording the analyst wants to analyze the footage by tagging one or more frames with types of interactions such as eye contact, touch or similar.

**Figure 1.1:** The gymnasium at CDH

According to the analyst, recording with a hand held device either impairs the gymnastics lesson or the footage quality. The reason for this is that the analyst uses a lot of his focus to get a good recording during the gymnastics lesson. In order to solve this, a system with an arbitrary amount of cameras to get decent footage and a corresponding interface to analyze the footage, needs to be developed. For such a system to work, the cameras need to be able to transfer the footage to the client interface. The footage also needs to be processed, determining the best angle at a certain time. This can be done either directly on the cameras, the client, or on one or multiple devices in between. One session of gymnastics is approximately 1 hour and 20 minutes long.

It is preferable that the camera system is mobile, such that it can be shown off at conferences and be used by more institutions than just the CDH in Aalborg.

A possible solution is that cameras could be placed in a room and record from different angles. The video footage will be sent to a server, where the camera with the best angle will be determined as shown on Figure 1.2 on the next page

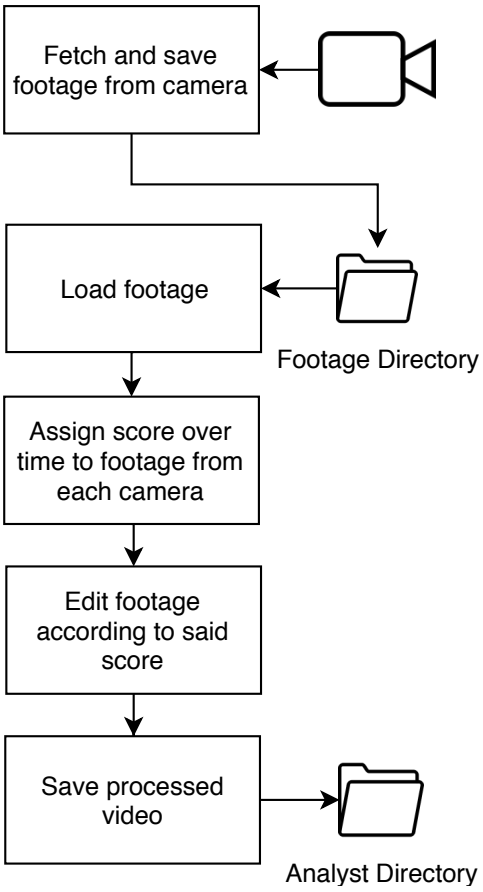



**Figure 1.2:** Example server actions

The score will be assigned by using some type of tracking token (e.g. a hat or a vest) combined with face recognition. An example scenario of the whole solution, can be seen in Figure 1.3. Note that this is merely an example and not a proposed solution.

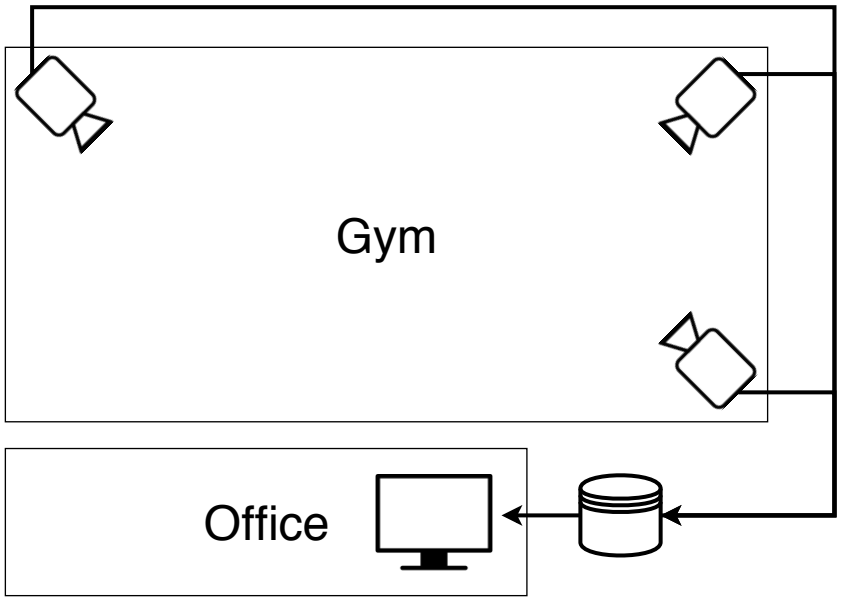



**Figure 1.3:** Example scenario

In this example scenario, the analyst can access the server with his computer and use an application to tag the video with some predefined actions. This application could be reminiscent of Figure 1.4.

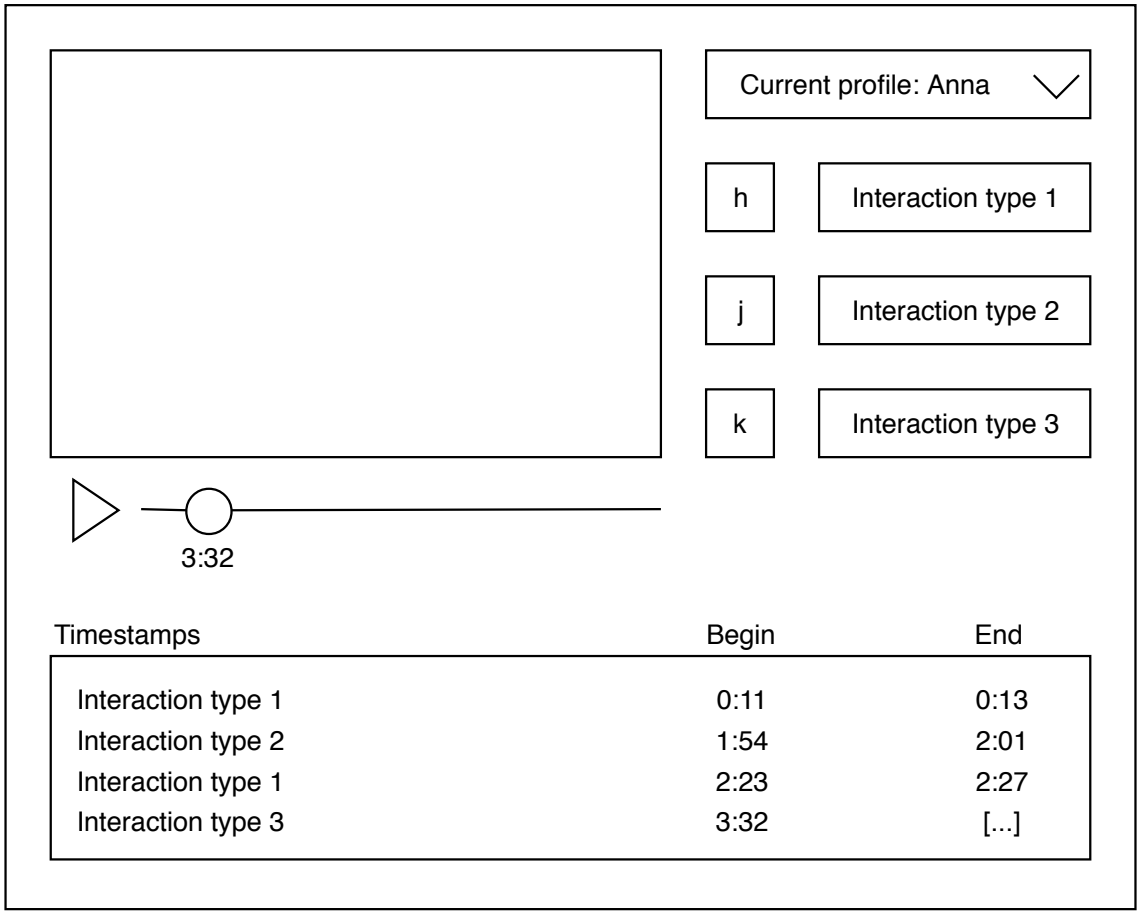



**Figure 1.4:** Example application

# Problem Analysis 2

## 2.1 Background

### 2.1.1 Extent of deafblindness

"*It is estimated that there are 1500 people with deafblindness, including elderly people with age-related vision- and hearing loss problems. [...] Between 75-100 of those individuals are under the age of 18.*
*[...] The difference between congenital- and acquired deafblindness is determined after 2 years from birth. If the individual has made a communicative language, as you do at the age of 1-2, before the loss of both senses, it will be known as acquired deafblindness, if not they suffer from congenital deafblindness.* (Translated from danish)" [3]

#### Congenital deafblindness

According to CDH, as of 2015, there are 173 people with congenital deafblindness in Denmark. 127 of them are adults and 46 are children. Although they do claim that the number of people with congenital deafblindness probably is beyond their knowledge.[4]

**Table 2.1:** Deafblindness in Denmark [5]

| Kind of deafblindness | Individuals | % |
|---|---|---|
| Completely deaf and blind | 33 | 18.30 % |
| Adults | 29 | 87.88 % |
| **Children (under 18)** | **4** | **12.12 %** |
| Blind - functional residual hearing | 37 | 20.60 % |
| Adults | 22 | 59.46 % |
| **Children (under 18)** | **15** | **40.54 %** |
| Deaf - functional residual vision | 50 | 27.80 % |
| Adults | 34 | 68.00 % |
| **Children (under 18)** | **16** | **32.00 %** |
| Functional residual hearing and vision | 60 | 33.30 % |
| Adults | 25 | 41.66 % |
| **Children (under 18)** | **35** | **58.34 %** |

According to Table 2.1, 18.3 % of the deafblind-population are completely deaf and blind. 87.88 % of those are adults, one of the reasons might be that some of them have had either residual hearing, vision or both earlier in their lives, but have lost the ability over the years.

People that are blind, but with functional residual hearing are at 20.6 %, of which 40.54

% of those are children. Functional residual hearing means that they may be able to hear and react to sounds.

People that are deaf, but with functional residual vision are 27.8 %. 32 % of those are children. Functional residual vision means that they are able to see very little or see silhouettes. [5]

A study made in Denmark with 95 adults with congenital deafblindness showed that 34 % of the subjects had a mental retardation. Moreover, only 26 % of the group had no mental or behavioral disorders. [6]

**Acquired deafblindness**

According to The Danish Deafblind Association there are around 1500 people with some sort of deafblindness, hereof 173 of them has congenital deafblindness. This means that there are approximately 1327 people with acquired deafblindness. Most of the people with acquired deafblindness are elderly, because of age-related maculadegeneration and age-related hearing loss.

One of the other main groups are people with Usher syndrome (the 3 types of Usher syndrome combined). Based on a study in Denmark from 1997 the number of individuals with Usher syndrome was estimated to be 5 pr 100.000 inhabitants, with approximately 270 individuals. Another study made by the earlier known "The Knowledge Center for Acquired Deafblindness"[1] a study in Aarhus Municipality in 1995 estimated that around 200 young people in Denmark have acquired deafblindness. [7]

The statistics shows that a large number of people with congenital deafblindness also have either a mental retardation or a behavioral disorder. However, if their teacher can give them better education by analyzing the progress students make, then the students will be more independent. This means that the municipality might not need to hire as much staff and thus, money will be saved.

### 2.1.2 Deafblind support

Deafblind people living in Denmark has the opportunity to get support. People with deafblindness has the opportunity of accommodation at CDH. Children up to the age of 18 gets the option to live at *Bøgehuset* where the staff are making sure that the children feels safe and happy. Young adults aged 18-23 can live in *Ungdomshjemmet* where it is like a dormitory with social activities. Adults over the age of 23 gets the opportunity to either live at *Danalien* or *Kloden* where the focus is to make sure that there is space for everyone, and their needs. [8]

All people in Denmark has the right to education therefore the education for the deafblind people has to be customized to their needs. This means that some of it is one to one even though it also is in groups. The children have the opportunity of special education where the focus is on learning the basic things like math and how to do chores. This means that it's easier than a regular public school because of the communication challenges the deafblind has. After the public school they have the opportunity to go on a 3-year

[1] Videnscentret for Døvblindblevne

education called "Specially Arranged Youth Education"[2] which builds on the previously taught curriculum. Adults are primarily offered fifty hours of education per year but it could be more depending on their needs. Various courses are also available.

### 2.1.3 Life quality of deafblindness

If a person is deafblind then, according to the supervisor, they might have a lower quality of life, because they aren't having many interactions with other people as a result of the difficulty of communication. Therefore The Knowledge Center for Acquired Deafblindness[3] had some people with acquired deafblindness answer a questionnaire about what they think about having a job. The deafblind persons with jobs answered that they primarily were happy to have a job. Most of them also answered that they aren't doing it for the money, but they are doing it because it makes them happy to have something to do. [9]

According to the supervisor, if deaf children and deafblind children are mixed during APA, both groups of children benefits socially because the deaf children sees that someone has it worse than them, leading to the deaf children initiating contact with the deafblind children. This does not only increase the over-all happiness of both parties, but also helps the student's communicative skills. Each year a meeting is held with CDH and different municipalities within Denmark, where the educators at CDH have to show some progress from the students. If there is low or no progress at all the municipalities can stop recommending CDH to the student's parents and thereby the center could lose funding as seen in Article 4, Paragraph 2 of "Publication regarding special needs education in primary education": *"The local government can make decisions regarding referring a student to special needs education in another school in the municipality than the district school. The local government can, in agreement with another local government, refer a student to special class/school enrollment in that municipally, cf. the Danish Folkeskole Act Article 22, Paragraph 1, No 1. The parents' and the student's desires concerning the school location shall, as much as possible, be met."* [10] (Translated from Danish by Jonatan Isachsen, BA in International Business Communication in English and French)

Therefore it is necessary to document the benefits of the weekly APA, which can be done by recording the children with a camera. This can help CDH showcase the progress of the students to both parents and their respective municipality.

## 2.2 Current solution

The solution that are currently used to record a student is where the analyst follows the student with an iPad and records the student doing interactions while simultaneously teaching. When all footage is recorded, the analyst is able to analyze the footage by manually noting every time the student interacts with another student. This is not optimal since the analyst has to spend all his time walking around filming the student and not use his time on educating the students. Furthermore, it might influence the behavior of the

[2] Særligt Tilrettelagt Ungdomsuddannelse
[3] Videnscentret for Døvblindblevne

students with vision, or residual vision, if the camera is too intrusive. Hiring personnel to record the session is also an option, but rather costly.

## 2.3 Camera choice

The first obstacle met is the choice of cameras, since they need to record in a sufficient quality. Type of camera, network protocols, video-encoding, video quality and price have to be considered. Additionally, it has to be considered whether it should be an IP camera or should work over USB. In order to determine further, more specific, requirements, experimentation is needed. Due to the gymnasiums size, however, it is assumed that the higher resolution, the better. The framerate should be 24 per second or higher in order for the footage to appear continuous. [11]
In the following, several options for cameras are presented.

### 2.3.1 Camera with local storage

A camera with local storage will store the recorded footage locally on the camera, e.g. on a SD-card. This adds manual labor to the system, as the SD-cards have to be physically moved from the camera to a computer.

### 2.3.2 Analog Camera

An analog camera must be connected to a DVR (Digital Video Recorder). The analog cameras are connected to the DVR with a coaxial cable. The DVR then encodes and compresses the analog video-signal and stores it digitally on a hard drive. Analog cameras use NTSC, PAL or SECAM analog video standards which serves as a bottleneck with regards to image resolution, that is often 1080p. [12, 13]

### 2.3.3 IP Camera

An IP camera is a digital camera, that works by utilizing the IP-protocol to communicate. Some IP-cameras work by connecting to a NVR (Network Video Recorder) to handle recording and storage of the footage, whilst others can record directly to storage, both local and remote. IP cameras have multiple advantages to analog cameras. IP cameras can be powered by Power over Ethernet (PoE). This means, that cameras which support PoE can be powered over the same cable that transmits the video signal making installation easier. Image resolution of IP cameras are not limited by analog video standards, and thus have higher resolution which means that faces are clearer and they can cover a larger area. [13]

## 2.4 Tracking type

In order to decide which method of tracking to use, three different methods have been researched.

- **Face detection** is used to determine whether a subject is facing the camera or not, however, this presents issues if the subject is facing neither of the cameras or if multiple faces are present. [14]
- **Shape detection** is used to track predefined shapes (e.g. a heart or a star). This could be used to get the best camera angle by placing a token of a certain shape on the subject. This prevents the issue with multiple faces, but the risk that none of the cameras spotting the shape still remains. [15]
- **Color detection** is used to find a predefined color. This could be used in similar fashion to shape detection. However, the subject's clothes and background colors could interfere with the detection. [16]

None of the tracking methods are perfect, though they can seemingly be combined to complement each other, but experimentation is needed to conclude anything.

### 2.4.1 Tracking token

The use of a tracking token is necessary for shape- and/or color detection, however they are not all equally considerate. Another benefit of a token, is that the educators know to give the student with the token(s) more space to interact with other students.

- An **armband** or a **legband** will not create any issues with a deafblind student according to the supervisor, as long as they aren't too tight.
- A **vest** will not be an issue for a deafblind student since it will be like wearing a regular shirt. The vest is also bigger than the other options, which means that it will be easier for the program to track the student with the vest.
- A **hat** has high risk of falling off and will, according to the supervisor, bother some students. This prevents a head mounted camera as a valid solution.

## 2.5 Regulatory considerations

A concern is the General Data Protection Regulation (GDPR) employed as of the 25th of May 2018. It concerns itself with clear language, user consent, increase transparency, stronger rights and stronger enforcement. Since our solution operates within the European Union (EU) it would almost be a necessity to make sure it complies so the organizations that use it won't have to modify it.

### 2.5.1 Right to access and erasure

Each student, or their respective guardian, has the right to access and delete any data where they are involved. This means that if a guardian wants to access the videos, then the guardian can see the videos of the other children also because the children are together for the whole session. This means that if the guardian wants some video footage deleted, then all the video footage has to be deleted. [17]

## 2.6 Product maintenance

In order to make the best product possible, maintenance of the product has to be considered. The term maintenance refers to updates of the system, potential replacement of cameras, and physical service and cleaning of the server.

Updates of the system should be automated to make the maintenance as easy as possible. The cameras should be easy to replace since damages can happen due to it being in a gym. This means that the chosen cameras should follow the standards that is used. Physical service and cleaning should be easy to do on the server because dust will make the chance of overheating a lot bigger.

## 2.7 Requirement specification

In order for the project to be implemented at the center it needs to fulfill certain requirements which are as follows. especially Section 4,

- The system should be able to choose the best angle of a specific person.
- The system should be able to record video footage in a sufficient quality, and transfer it to a server.
- The system should save all recordings from a session, in case of software failure.
- The cameras should be sturdy enough to withstand impacts from general gym activities.
- The system should be mobile.
- The system should include a interface for the analyzer to view the gym session.
  - The interface should have the possibility to specify actions to e.g. keyboard keys.
  - The analyst should be able to tag one or more frames with an action tag, e.g. a hand raise.
  - Statistics and graphs should be made so that it's possible to see progress.
  - It should be possible to make profiles for the students that saves the tag made.
- The necessary actions, if a student or guardian requests deletion of footage, should be automated by the system.
- The necessary actions, if a student or guardian requests access to footage, should be automated by the system.
- It should be possible to activate and deactivate the cameras remotely.

# Development 3

In order to develop a solution that fulfills as many of the requirements stated in the requirement specification as possible, it is split in to three parts.

The first part is the network configuration, which is responsible for getting footage off of the camera and onto the server. It is also responsible for making sure that the analyst can access the processed video files on the server.

The second part is the angle determination and video processing. It is responsible for determining which cameras footage should be used at a given time, and place it wherever the analyst, as dictated by the network configuration, has access to it.

The third part is the Graphical User Interface (GUI), responsible for allowing the analyst to analyze and tag the footage amongst other things.

Each part will be tested separately before being tested together in Chapter 4. Due to limited resources, the solution will not be complete, but rather a proof of concept (POC), which might show that it is indeed possible to solve the problem, incentivizing further development.

## 3.1 Network Configuration

An overview of the network configuration, which lives up to the requirements, can be seen on Figure 3.1.

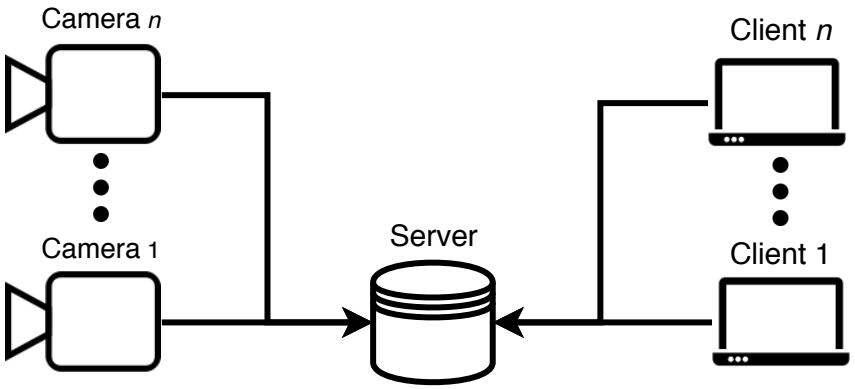



**Figure 3.1:** Network Overview

On Figure 3.1 it can be seen that some, $n$, amount of cameras and clients are connected to, and communicate, with the server. The amount of cameras and clients are not necessarily

the same.

Regarding the cameras, two cameras were available, a HIKVISION DS-2CD2085FWD-I and a TRENDnet TV-IP751WC. The HIKVISION supports uploading footage to a server either through a mounted Network File System (NFS) or Server Message Block (SMB). It also supports uploading through the File Transfer Protocol (FTP) at a very reduced framerate. Furthermore the HIKVISION footage can be fetched through Real Time Streaming Protocol (RTSP), using a RTSP client. The TRENDnet does not support uploading fully fledged footage, but can upload up to three frames per second (FPS) through FTP, but three FPS is not satisfactory.

During nonformal testing it was discovered that the HIKVISION camera could not manage to connect to the SMB server. Due to this, RTSP was used for retrieving footage instead of SMB.

As such, the physical network layout is as shown on Figure 3.2. Each square represents a Network Interface Controller (NIC).

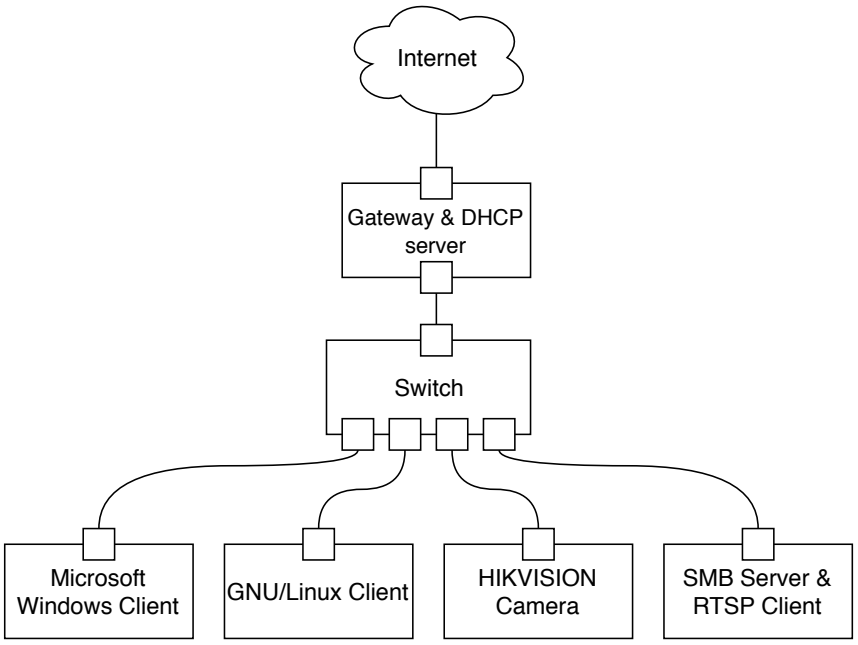



**Figure 3.2:** Physical Network

It is practical to have the SMB server be connected to the internet since that makes it possible to install and update applications without any hassle. The only reason a Dynamic Host Configuration Protocol (DHCP) server is used for assigning IP-adresses is because it is the default configuration for the gateway.

### 3.1.1 SMB

SMB is a client-server, request-response protocol, meaning that the server only responds (i.e. does not initiate communication) to one or multiple clients requests. In most SMB clients, the files appear as if they were in a mounted filesystem like a Universal Serial Bus (USB)-stick or similar.

SMB is ran on top of another, lower level, network protocol such as Transmission Control Protocol / Internet Protocol (TCP/IP), NetBios Extended User Interface (NetBEUI) or Internetwork Packet Exchange / Sequenced Packet Exchange (IPX/SPX). TCP/IP is almost ubiquitous, due to the internet, and since the client should be able to access the samba server across the internet, SMB on TCP/IP is used, which actually runs on Network Basic Input/Output System (NetBIOS) over TCP/IP, which is defined in Request for Comments (RFC) 1001 and RFC 1002. [18]

On Figure 3.1 the layers of SMB on different protocols can be seen in relation to the Open Systems Interconnection (OSI) model and TCP/IP model, with SMB on NetBIOS over TCP/IP colored green.

**Table 3.1:** SMB on different network protocols, inspired by [18]

| OSI Model | TCP/IP Model | SMB on IPX/SPX | SMB on netBEUI | SMB on netBIOS | SMB on TCP/IP |
|---|---|---|---|---|---|
| Application | Application | SMB | SMB | SMB | SMB |
| Presentation | | | | | |
| Session | | NetBIOS | NetBEUI | NetBIOS | NetBIOS |
| Transport | TCP and UDP | IPX/SPX | | DECnet | TCP and UDP |
| Network | Internet Protocol | | | | Internet Protocol |
| Data Link | Network Acces Layer | 802.2, 802.3, 802.5 | 802.2, 802.3, 802.5 | Ethernet V2 | Ethernet V2 |
| Physical | Physical | Out of scope | Out of scope | Out of scope | Out of scope |

Note that the physical layer is not considered, as it is out of scope.

There are multiple implementations, both client- and server side, of the SMB protocol and for this project Samba was used on the server since it is well documented and available in Debian's official repositories [19]. It is not known which implementation of SMB the analyst will use and it may change over time.

To setup Samba it was decided that a script would be beneficial to reduce manual work. Another benefit is that changes to the configuration will be version controlled, since the script was developed under Git version control. Python was chosen due to previous experience and its ease of use [20]. The source code of the script can be found in Appendix A.1. Debian was also chosen due to previous experience and an incidental benefit is its free and stable long term releases [21].

The script is relatively simple and straight forward, as can be seen on Figure 3.3 on the next page.

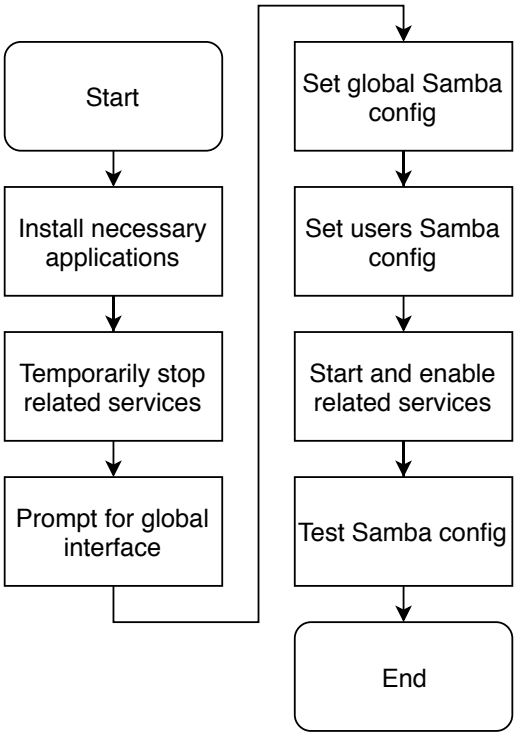



**Figure 3.3:** Debian SMB setup flowchart

As shown on Figure 3.3 the script peforms multiple actions, many of these involve passing commands to the operating system using a library called *os*. It also generates a text configuration file for Samba.

The script is also quite ad hoc since it depends on quite a few command-line interface (CLI)-based frontends, which are not necessarily present on all systems. *Advanced Package Tool* (*APT*) for installing packages, *firewalld* for firewall configuration and *systemd* for managing services. *APT* and *systemd* are default on Debian and were chosen due to that. The default firewall configuration tool, *Uncomplicated Firewall* (*UFW*), was however switched out for *firewalld* due to a lack of experience with *UFW*. The source code can be found in Appendix A.1.

### 3.1.2 Real Time Streaming Protocol

RTSP is a client-server protocol, that simplifies streaming multimedia data over a network using delivery mechanisms based on Real Time Protocol (RTP). RTSP is specified in RFC 2326 and RTP is described in RFC 1889. [22, 23]

For fetching the multimedia stream FFmpeg is used through the following command.

```
ffmpeg -i rtsp://user:pass@address:port/live.sdp -vcodec copy -y file.mp4
```

The different parts of the command works as follows. [24]

- The `-i` option takes an input as a parameter.
- The `-vcodec` option sets the video codec.

- The `-y` option overwrites output files without asking.
- When `copy` is set as a parameter to the `-codec`, `-vcodec` or `-acodec` options, ffmpeg omits decoding and encoding which makes it very fast and lossless.

Using this command the video stream could be saved at a resolution of 1920 by 1080 and at approximately 24 FPS as revealed by the ffmpeg command's output which can be seen in Appendix A.2.

### 3.1.3 Testing

To make sure that the configuration is functional a test was set up. For this, the physical network layout from Figure 3.2 on page 12 was used.

During this test, attempts to access the SMB server from both Microsoft Windows 10's integrated client and *smbclient*, which is a part of Samba, from a GNU/Linux machine, were made. These attempts were successful and a packet capture of *smbclient* retrieving a file was made.

It was also possible to save footage using the method described in subsection 3.1.2 from the HIKVISION camera using FFmpeg and a packet capture of this was also made.

During these packet captures the minimal amount of machines needed to make the capture were connected in order to reduce confusion.

### 3.1.4 Security and Traffic Analysis

As personal data is being processed, it should be secured properly, since a breach or similar would prevent the CDH in staying compliant with the GDPR (see Section 2.5).

In order to assess the security of the system, a traffic analysis of the previously mentioned packet captures was carried out.

For the packet capture of SMB traffic, the roles of the used machines and their respective IP-adresses can be found in Table 3.2.

**Table 3.2:** Roles and IP-addresses for SMB capture

| Role | IP-address |
|---|---|
| SMB Client and Gateway | 10.42.0.1 |
| SMB Server | 10.42.0.61 |

On Figure 3.4 on the next page the first 15 packets of the capture, of which the main part is NT LAN Manager (NTLM) authentication, used by Samba, can be seen. Wireshark was used for both packet capture and analysis.

Note: Keepalive packets will be mostly ignored throughout the analysis and Address Resolution Protocol (ARP) traffic is filtered out, which may cause jumps in packet enumeration.

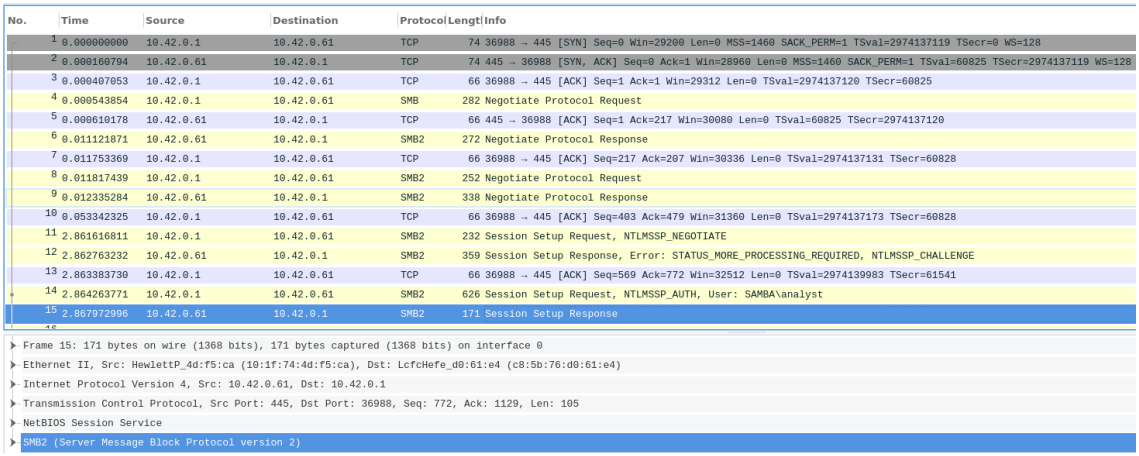



**Figure 3.4:** Packets 1 to 15

Packets one to three is the client establishing a connection through a three way TCP handshake consisting of synchronize (SYN), synchronize-acknowledge (SYN, ACK) and acknowledge (ACK).

In packets four to nine, the client makes a request for protocol negotiation and the server responds, confirming the request-response nature of SMB. Looking at one of the SMB packets, packet 15 in this case, it can be seen that SMB does indeed run on NetBIOS over TCP/IP as mentioned in 3.1.1.

In packets 11 to 15, the NTLM authentication process, which is used by Samba, can be seen. Inspecting the packet reveals, that it is NTLMv2, as shown in Figure 3.5.

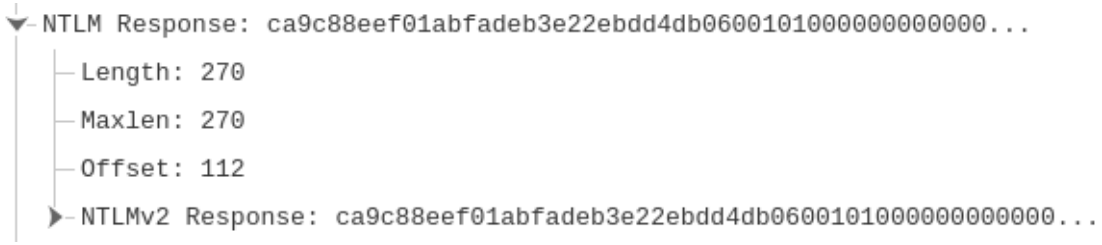



**Figure 3.5:** Cutout of DPI on packet 14

NTLMv2 uses deprecated cryptographic algorithms, such as Message Digest 4 (MD4) [25], and thus its security is questionable [26].

On Figure 3.6 on the next page packet 113 to 128, where a file transfer is initiated, can be seen.

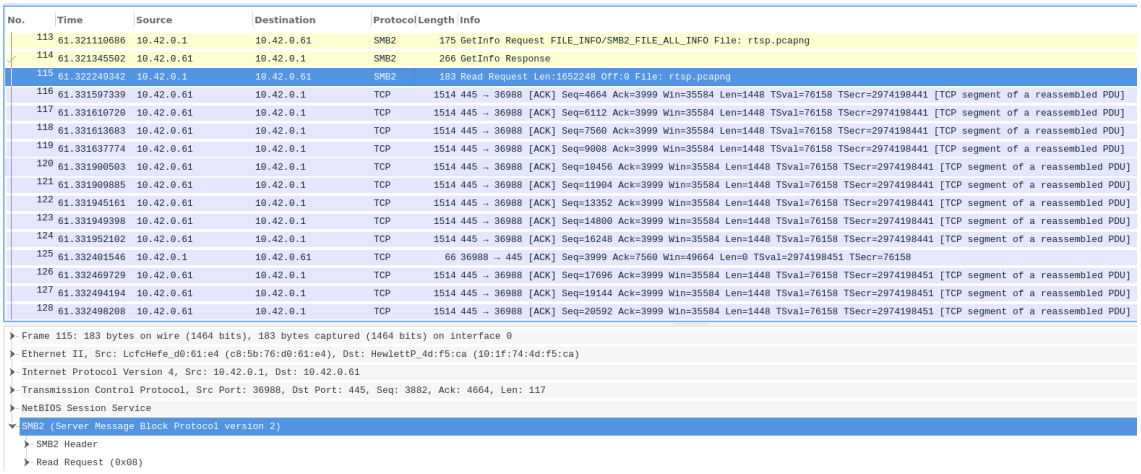



**Figure 3.6:** Packets 113 to 128

In packets 113 to 115 information and the file, "rtsp.pcapng" is requested and a lot of long TCP packets from the SMB server to the client, follow. These TCP packets contain unencrypted pieces of the file and Wireshark can even reassemble it. Since the traffic is not encrypted and the security of the authentication is questionable, traffic should not be transferred over untrusted networks without an extra layer of protection such as that typically offered by a Virtual Private Network (VPN) connection to a trusted network.

On Figure 3.7, the aforementioned file transfer is completed and the connection is terminated.

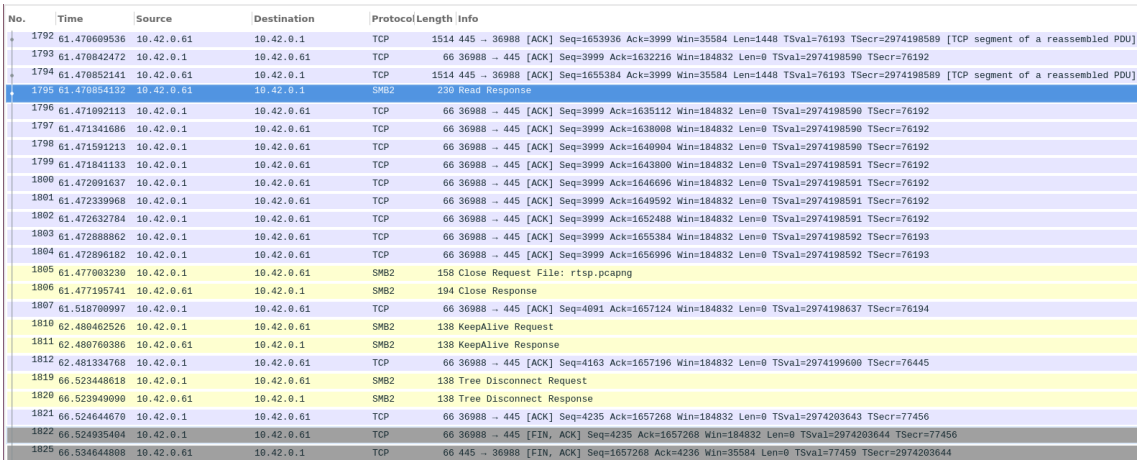



**Figure 3.7:** Packets 1792 to 1825

In packets 1805 and 1806 the SMB file transfer is closed. In packets 1819 and 1820, the SMB connection is terminated and soon after, in packet 1822 the client send a finish-acknowledge (FIN, ACK), indicating that it wants to terminate the connection and the sever complies in packet 1855, with another FIN, ACK.

For the capture of RTSP traffic the roles of each machine and their respective IP-addresses, can be seen in Table 3.3 on the next page.

**Table 3.3:** Roles and IP-addresses for RTSP Capture

| Role | IP-adress |
|---|---|
| IP-Camera | 10.42.0.70 |
| RTSP Client | 10.42.0.61 |

On Figure 3.8 the first 26 packets of fetching the footage stream off of the HIKVISION camera, as described in subsection 3.1.2, is shown.

| No. | Time | Source | Destination | Protocol | Length | Info |
|---|---|---|---|---|---|---|
| 1 | 0.000000000 | 10.42.0.61 | 10.42.0.70 | TCP | 74 | 58256 → 554 [SYN] Seq=0 Win=29200 Len=0 MSS=1460 SACK_PERM=1 TSval=48047762 TSecr=0 WS=128 |
| 2 | 0.000528770 | 10.42.0.70 | 10.42.0.61 | TCP | 74 | 554 → 58256 [SYN, ACK] Seq=0 Ack=1 Win=14480 Len=0 MSS=1460 SACK_PERM=1 TSval=2756685 TSecr=48047762 WS=32 |
| 3 | 0.000622197 | 10.42.0.61 | 10.42.0.70 | TCP | 66 | 58256 → 554 [ACK] Seq=1 Ack=1 Win=29312 Len=0 TSval=48047762 TSecr=2756685 |
| 4 | 0.000637623 | 10.42.0.61 | 10.42.0.70 | RTSP | 153 | OPTIONS rtsp://10.42.0.70:554/live.sdp RTSP/1.0 |
| 5 | 0.000915536 | 10.42.0.70 | 10.42.0.61 | TCP | 66 | 554 → 58256 [ACK] Seq=1 Ack=88 Win=14496 Len=0 TSval=2756685 TSecr=48047762 |
| 6 | 0.001871307 | 10.42.0.70 | 10.42.0.61 | RTSP | 219 | Reply: RTSP/1.0 200 OK |
| 7 | 0.001926753 | 10.42.0.61 | 10.42.0.70 | TCP | 66 | 58256 → 554 [ACK] Seq=88 Ack=154 Win=30336 Len=0 TSval=48047763 TSecr=2756685 |
| 8 | 0.002197825 | 10.42.0.61 | 10.42.0.70 | RTSP | 179 | DESCRIBE rtsp://10.42.0.70:554/live.sdp RTSP/1.0 |
| 9 | 0.003003720 | 10.42.0.70 | 10.42.0.61 | RTSP | 250 | Reply: RTSP/1.0 401 Unauthorized |
| 10 | 0.003435944 | 10.42.0.61 | 10.42.0.70 | RTSP | 370 | DESCRIBE rtsp://10.42.0.70:554/live.sdp RTSP/1.0 |
| 11 | 0.004920193 | 10.42.0.70 | 10.42.0.61 | RTSP/S… | 786 | Reply: RTSP/1.0 200 OK |
| 12 | 0.005411437 | 10.42.0.61 | 10.42.0.70 | RTSP | 418 | SETUP rtsp://10.42.0.70:554/live.sdp/trackID=1 RTSP/1.0 |
| 13 | 0.010067431 | 10.42.0.70 | 10.42.0.61 | RTSP | 274 | Reply: RTSP/1.0 200 OK |
| 14 | 0.010536697 | 10.42.0.61 | 10.42.0.70 | RTP | 54 | PT=ITU-T G.711 PCMU, SSRC=0x0, Seq=0, Time=0 |
| 15 | 0.010553384 | 10.42.0.61 | 10.42.0.70 | RTCP | 50 | Receiver Report |
| 16 | 0.010605503 | 10.42.0.61 | 10.42.0.70 | RTSP | 382 | PLAY rtsp://10.42.0.70:554/live.sdp/ RTSP/1.0 |
| 17 | 0.048391176 | 10.42.0.70 | 10.42.0.61 | TCP | 66 | 554 → 58256 [ACK] Seq=1266 Ack=1173 Win=17696 Len=0 TSval=2756690 TSecr=48047765 |
| 18 | 0.072766741 | 10.42.0.70 | 10.42.0.61 | RTSP | 244 | Reply: RTSP/1.0 200 OK |
| 21 | 0.094115483 | 10.42.0.70 | 10.42.0.61 | RTP | 102 | PT=DynamicRTP-Type-96, SSRC=0x65EC0201, Seq=45606, Time=2507522130 |
| 22 | 0.094187665 | 10.42.0.70 | 10.42.0.61 | RTP | 60 | PT=DynamicRTP-Type-96, SSRC=0x65EC0201, Seq=45607, Time=2507522130 |
| 23 | 0.094322988 | 10.42.0.70 | 10.42.0.61 | RTP | 1482 | PT=DynamicRTP-Type-96, SSRC=0x65EC0201, Seq=45608, Time=2507522130 |
| 24 | 0.094632240 | 10.42.0.70 | 10.42.0.61 | RTP | 1482 | PT=DynamicRTP-Type-96, SSRC=0x65EC0201, Seq=45609, Time=2507522130 |
| 25 | 0.094676415 | 10.42.0.70 | 10.42.0.61 | RTP | 1482 | PT=DynamicRTP-Type-96, SSRC=0x65EC0201, Seq=45610, Time=2507522130 |
| 26 | 0.094693444 | 10.42.0.70 | 10.42.0.61 | RTP | 1482 | PT=DynamicRTP-Type-96, SSRC=0x65EC0201, Seq=45611, Time=2507522130 |

**Figure 3.8:** Packets 3 to 26

In these packets a TCP connection is established through the three way handshake and RTSP is used to start an unencrypted RTP stream. The fact that it is unencrypted does not matter since data transfer should not occur over untrusted networks, because SMB also is unencrypted, anyway.

Everything works in accordance with subsection 3.1.3, though the security remarks should not be ignored.

## 3.2 Angle determination and Video Processing

As mentioned in section 1.1, the analyst requested a system with an arbitrary number of cameras that could capture the best angle of a specific person. The goal of angle determination was to use one single camera and having the software return a score of how good the angle was from the above mentioned camera. By doing this it eases the process of adding multiple cameras to the system. This is done by detecting a colored vest worn by the student in question. By detecting the vest, it is possible to get the coordinates on the video of the vest. Knowing the coordinates of the vest, a nearby face can be found.

### 3.2.1 Object detection

Object detection is done in Python using the OpenCV and NumPy libraries. The object to detect is a colored, in this instance orange, vest. To detect the vest, OpenCV is used

to isolate the color orange and return the coordinates of the vest. NumPy is used to make an array that the opencv method *inRange* takes as an argument. The array in NumPy is used to isolate a color with a set of boundaries in the Blue Green Red (BGR) color space. This can be seen in the two variables *lower* and *upper*, that with the *findNonZero* method will isolate any pixel in a frame that is in-between these two boundaries and detect them as orange.

As can be seen on the implementation in Listing 3.1, the variable *avg* finds the average of all orange points and in *pointInScreen* the middle point is calculated according to a video resolution set to 1920 by 1080 pixels.

```
import numpy as np
import cv2

stream = cv2.VideoCapture("video_path")

resImage = [1920, 1080]
resScreen = [1920, 1080]

while(true):
    (grabbed, frame) = stream.read()

        if not grabbed:
                break
    lower = np.array([20, 10, 150])
        upper = np.array([70, 50, 200])

        orange = cv2.inRange(frame, lower, upper)
        points = cv2.findNonZero(orange)

        if(orange.any()):
                avg = np.mean(points, axis=0)
                #pointInScreen is the middle point of the
                   orange in the frame
                pointInScreen = ((resScreen[0] / resImage[0])
                    * avg[0][0], (resScreen[1] / resImage[1])
                    * avg[0][1] )
```

**Listing 3.1:** Object detection using color

A drawback of using this method of object detection is that before face detection can be done, the color orange needs to be detected first. The reason for this is that orange is used as a reference point for face detection. Another drawback is that there may be other objects of the chosen color (orange in this case), which would create interference. It should also be noted that scoring video footage without orange is avoided, since that could result in the wrong person being filmed.

A quick test to make sure that the middle of an orange mass is indeed found was done by

making the software draw a red dot where it calculates the middle is, as can be seen on Figure 3.9.

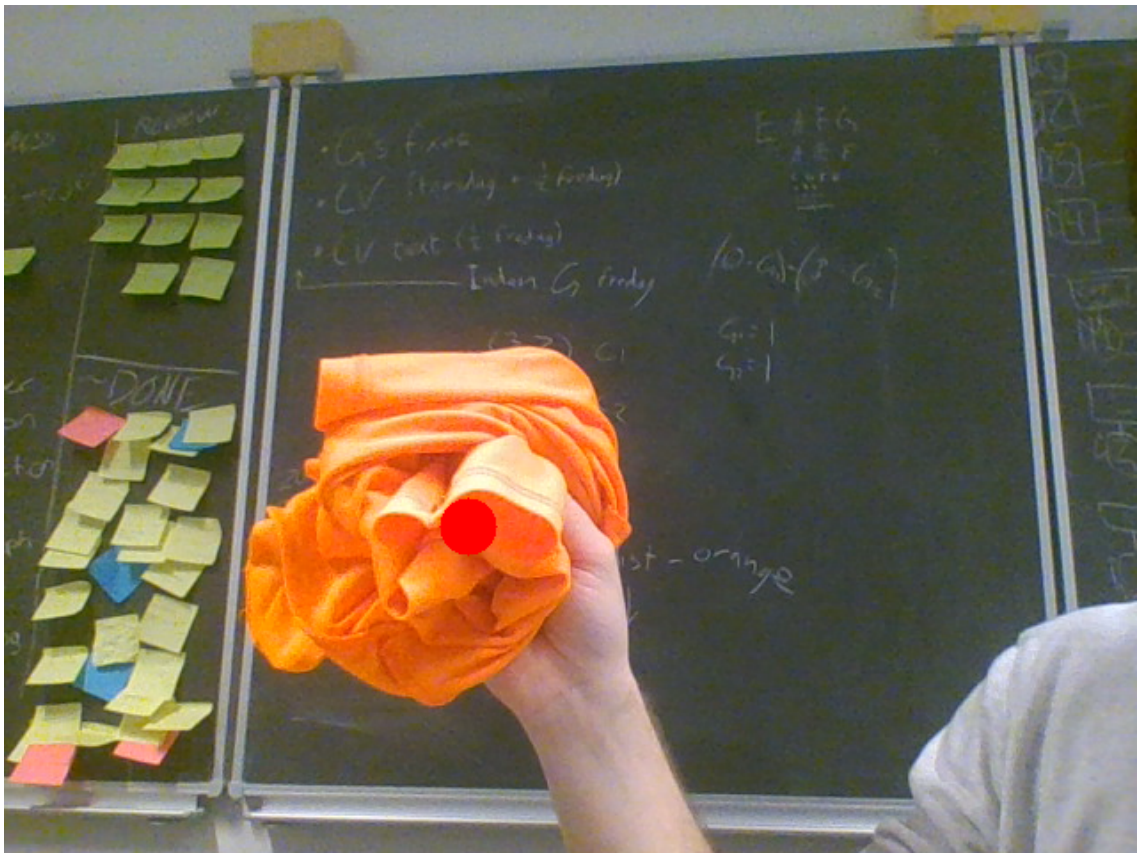


**Figure 3.9:** Finding the middle point of an orange mass.

### 3.2.2 Face detection

In order for the software to determine the best angle it was chosen to use face detection. This was done with the purpose of determining the closest face to the above-mentioned object. By using both face detection and object detection it enhances the possibility of choosing the best possible angle. From face detection a *confidence* percentage is found as seen on Listing 3.2 which is a level of how sure the software is that an actual face is found. It sorts out all faces that is under a *confidence* threshold set to 0.5.

```
net = cv2.dnn.readNetFromCaffe("models/deploy.prototxt.txt",
   "models/res10_300x300_ssd_iter_140000.caffemodel")

stream = cv2.VideoCapture("video_path")
while(true):
    if not grabbed:
                break

    (h, w) = frame.shape[:2]
        blob = cv2.dnn.blobFromImage(cv2.resize(frame, (300,
           300)), 1.0, (300, 300), (104.0, 177.0, 123.0))
```

```
net.setInput(blob)
detections = net.forward()
for i in range(0, detections.shape[2]):

        confidence = detections[0, 0, i, 2]

        if confidence < 0.5:
                continue
```

**Listing 3.2:** Face detection using caffe

The *confidence* level is set to an interval between 0-1, where higher is better, this is used in the scoring system mentioned in section 3.2. A test of the face detection and *confidence* as percentage can be seen on figure 3.10.

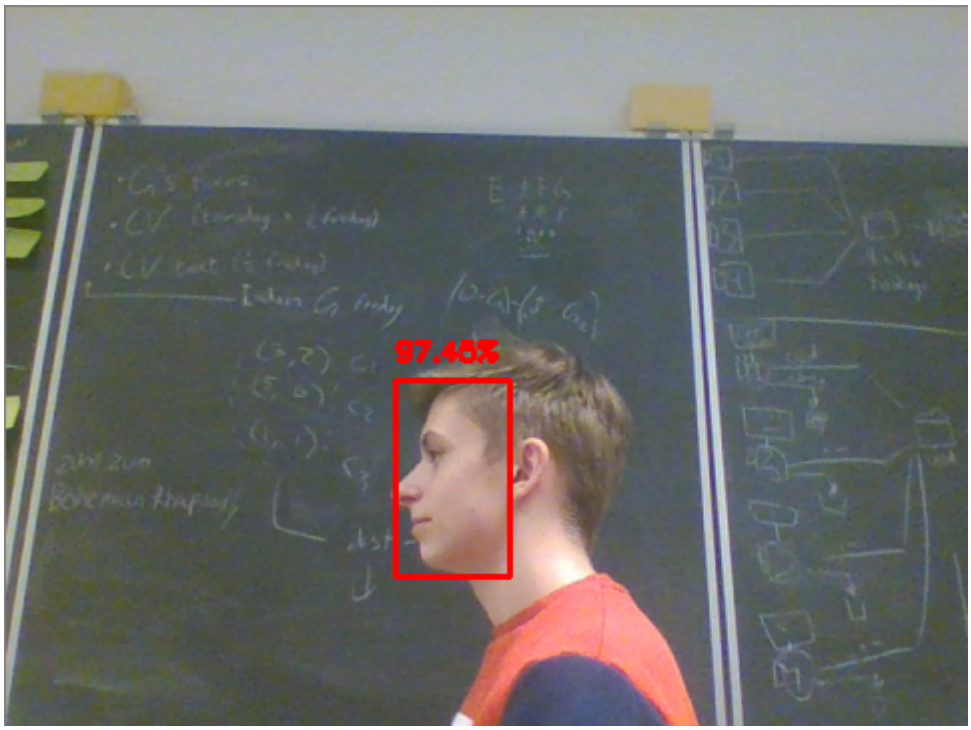



**Figure 3.10:** Side profile of face detection with confidence percentage

### 3.2.3 Final script

In order to determine the best angle a combination of object detection and face detection is used. When both the color orange and a face is present in a frame, the script returns a score. The score is an indication of how good the given angle is.

A flowchart of how the score is returned can be seen in Figure 3.11 on the following page.

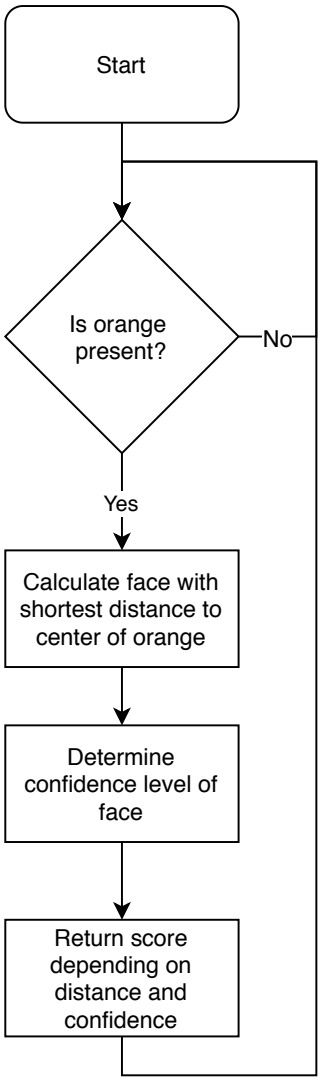



**Figure 3.11:** Flowchart of score determination

As shown in Figure 3.11, the first thing the script does, is look for orange in the frame. If orange is present, the script will look for faces in the frame. If multiple faces are present, it will be calculated which one is closest to the center of the orange pixels in the frame. It then determines the confidence level of the face and uses both the confidence level and the distance to the face to determine the score. During development, it was discovered that assigning scores takes approximately four times as long as real time. Meaning that a 80 minute clip, would take about 320 minutes to process on the test hardware.

In Listing 3.3 an implementation of the distance formula in Python is shown. The implementation is used to calculate the distance from the center of orange, to the center of the nearest face.

```
for i in range(0, len(faceCoordsList)):
        distanceList.append(np.sqrt(np.square(faceCoordsList[
           i][0]-pointInScreen[0])+np.square(faceCoordsList[i
           ][1]-pointInScreen[1])))
```

```
distanceList.sort()
```

**Listing 3.3:** Distance formula implemented in Python

To indicate the center of orange and the face visually, the script was modified to draw a red dot on the center of orange and a red box around all faces. Between the center of orange and the nearest face, a blue line is drawn. An example of the visualization can be seen in Figure 3.12.

**Figure 3.12:** Visualization of face and object detection

## 3.3 Graphical User Interface

It was mentioned by the analyst that an interface for reviewing the footage would be essential to the project. The GUI designed in this project, is based on the requirements listed in Section 2.7. The front end encapsulates all the facets of the software that the user can interact with. The GUI is based on, Qt, a free, open-source, and cross-platform framework, which comes with a tool called Qt Creator. This tool provides an Integrated Development Environment (IDE) with the possibility of designing the User Interface (UI) using drag and drop. The framework contains community supported and maintained libraries. In this project the GUI was written in C++, with the goal of creating an interface that contains the functions requested by the analyst. The Graphical User Interface was created in two phases, first the functional fundamentals of the software was made, after the focus was put onto the maintainability, usability and aesthetics.

### 3.3.1 First phase

The first phase was a basic video player, that could open a file and play/pause the aforementioned file, which can be seen in Figure 3.13 on the following page. The video player used QMediaplayer and QVideoWidget which both are classes in the Qt environment, which needs a codec on the client to play different media files and in this project libravcodec from FFmpeg was used. Therefore, an installer could be beneficial to enhance the usability of this software.

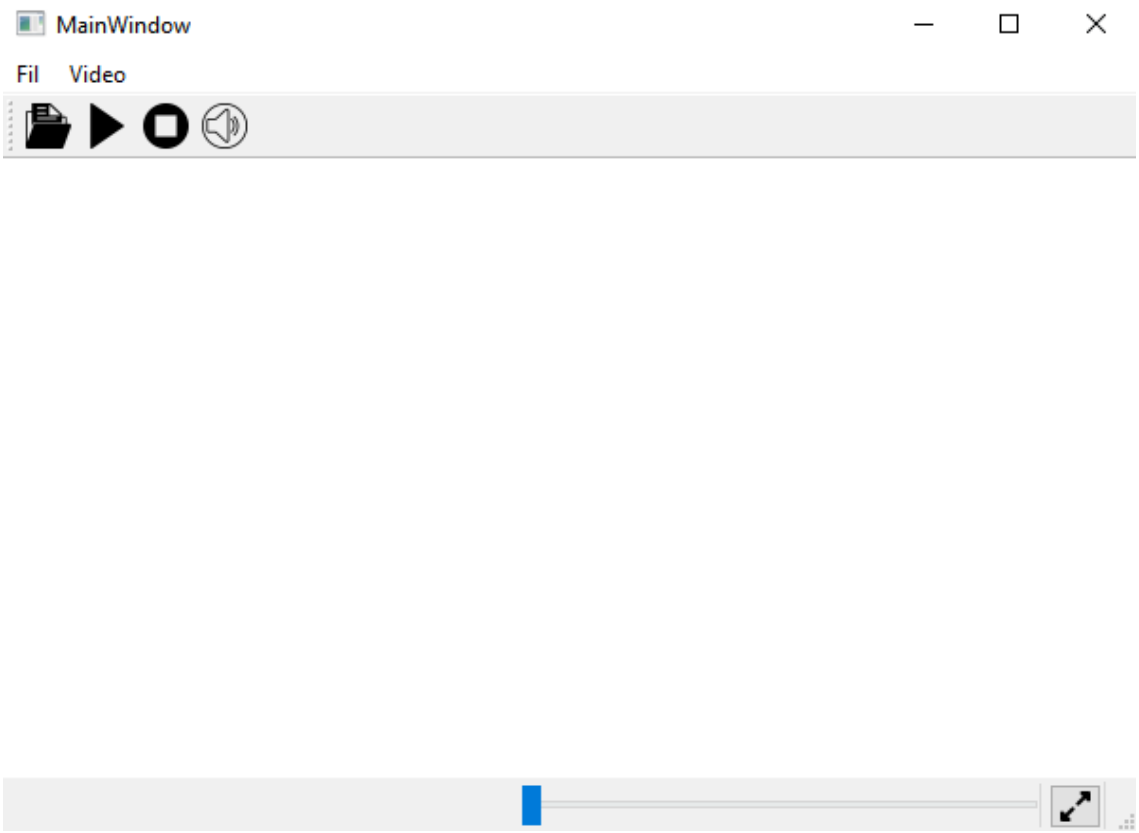
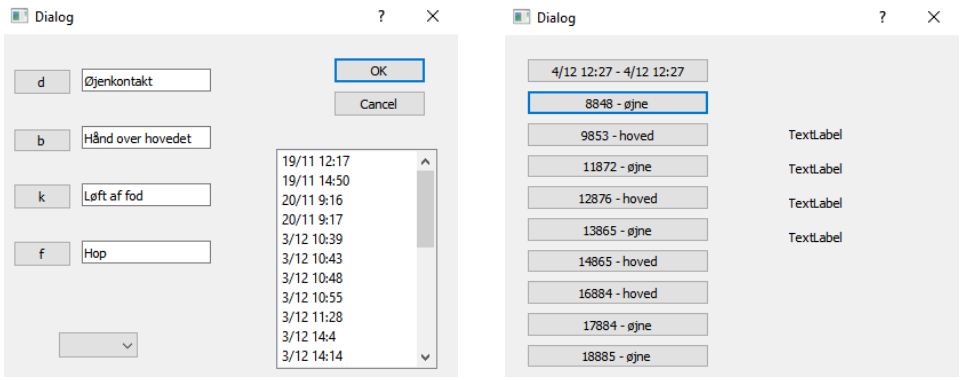



**Figure 3.13:** First phase of the video player

Later the function of being able to bind keyboard keys with tags was implemented, this was done by having buttons that would allow the user to program any keyboard key e.g. *d* together with a "tag/action" e.g. *eye contact*. A list of prior viewing sessions was also added, the list of items were clickable which would open a new window that would show the different tags from that session. These two windows can be seen in Figure 3.14. When one of the four buttons in window (a) were clicked, a window shows up which instructs the user to press a keyboard key, that is used as the new key bind.




**(a)** **(b)**

**Figure 3.14:** (a) Window of keyboard binds with tags, and the list of earlier sessions. (b) When a session in the list is clicked.

When one of the four key binds were used e.g. *d* while watching a video, a new session would be added to a text document with the current date and time. It would also write the tag linked to the key bind e.g. *eye contact* which also included the timestamp in the video. An example of how some of these sessions looked like in a text document can be see in Figure 3.15. In both window (a) and (b) the text document were read in order to fetch the data. In window (b) it was done with the code shown in Listing 3.4.

```
4/12 12:25#21900-hoved#23904-øjne#24908-skulder#25897-knæ#26897-hoved#33904-hoved#35907-øjne#37906
4/12 12:27#8848-øjne#9853-hoved#11872-øjne#12876-hoved#13865-øjne#14865-hoved#16884-hoved#17884-øj
4/12 12:29#4891-hoved#5890-øjne#6891-knæ#7893-hoved#8900-skulder#9890-hoved#11892-knæ#12891-knæ#13
4/12 12:32#5911-hoved#6907-øjne#7906-øjne#8910-hoved#10913-hoved#11915-skulder#12905-knæ#13909-hov
4/12 12:33#3908-hoved#4905-øjne#5896-knæ#6898-hoved#7906-øjne#8896-øjne#9897-øjne#11901-knæ#12904-
4/12 12:37#3931-hoved#4924-skulder#5924-knæ#6924-knæ#7926-skulder#17924-øjne#18918-øjne#19930-øjne
```

**Figure 3.15:** An example text document

```
QStringList ccc = value.split("#");

QPushButton* aBtns[ccc.size()];

for(int x = 0; x<ccc.count(); x++){

    aBtns[x] = new QPushButton(ccc.at(x).split("-").first()+"
        - "+ccc.at(x).split("-").last(), this);
    ui->verticalLayout->addWidget(aBtns[x]);
    aBtns[x]->setObjectName(ccc.at(x).split("-").first());
    connect(aBtns[x], &QPushButton::clicked,this, &itemWindow
       ::pushButtonClicked);
}
```

**Listing 3.4:** How data was fetched from the text document

In the code shown in Listing 3.4 delimiters were used with the *split* function [27], which in this case first took a single line as a *QString* and converted it to substrings each time "#" occurred and put it into a *QStringList* called **ccc**. Then, in a for loop, it goes through each element and creates a new button with the action tag and the time it was pressed.

### 3.3.2 Second phase

In the second phase of the interface it was focused on optimizing the existing functions, such as maintainability, usability and versatility. The GUI was first drawn as shown in Figure 1.4 on page 4. This was used as a reference point when designing the new interface, and the final version of this can be seen in Figure 3.16 on the next page.


**Figure 3.16:** Final version of interface

In the upper-right corner of Figure 3.16 a list is displayed, which finds all .mp4 files in a specific folder. Below the aforementioned list a new function was added, which allowed the user to create new profiles and save the key binds and action tags. Below the slider of the video player a list were added, which gave visual feedback to the user when a key bind were pressed. Moreover, as mentioned earlier maintainability and versatility were a focus for the GUI, therefore instead of writing data to a text document the data was inserted into a local database. This was done with the database engine Sqlite and Qt's SQL module which enables the connection to a database. In Figure 3.17 on the next page the final version of the earlier sessions overview is shown.

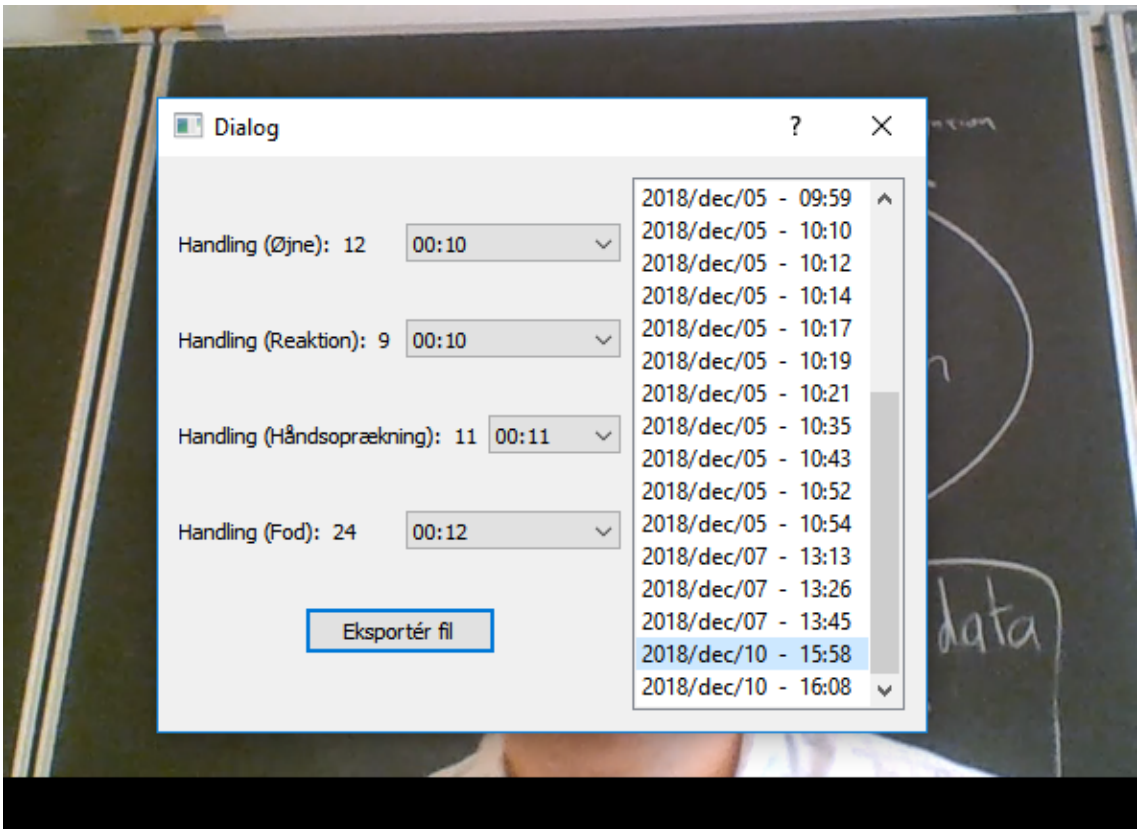



**Figure 3.17:** Revised version of the earlier session window

The separation of different sessions was done by creating a new table for each session, which was done in the video player window as shown in Listing 3.5.

```
void MainWindow::createSessionTable(){
    QSqlDatabase mydb = QSqlDatabase::addDatabase("QSQLITE");
    mydb.setDatabaseName("sessionData.db");

    if(!mydb.open()){
        qDebug() << "Aabning af database fejlede";
        return;
    }
    QTime qt;
     ct = qt.currentTime().toString("hhmm");
    mydb.open();
    QDate qd;
     cd = qd.currentDate().toString("yyyy_MMM_dd");

    QSqlQuery qry;
    qry.prepare("create table session_"+cd+"_"+ct+"(id
       integer primary key,action1_name varchar(20),
       action1_tid varchar(20),action2_name varchar(20),
       action2_tid varchar(20),action3_name varchar(20),
       action3_tid varchar(20),action4_name varchar(20),
```

```
        action4_tid varchar(20))");

    if(qry.exec()){
        qDebug() << "tabel lavet";
        mydb.close();
    }else{
        qDebug() << "tabel fejlede";
        mydb.close();
        mydb.removeDatabase(QSqlDatabase::defaultConnection);
    }
}
```

**Listing 3.5:** Qt creating new table

In most cases the table will have an unique name, but if the user were to create a new session within the same minute, it would not make a new table but instead write to the existing one. In the *qry.prepare* all the column names are declared with a type, which in mostly varchar which is a text string in either Unicode Transformation Format (UTF)-8, UTF-16BE or UTF-16LE) [28]. A table can only hold one primary key which in this case is the column **"id"**, for each new row inserted into the table a new unique auto incremented integer is set into this column. In Listing 3.6 it is shown how the software reads from a table. It is done by using two *for loops* that works by first reading the rows and for each row it reads through all the columns.

```
if(qry.exec(   ("select * from "+tableNameTemp+""))){
        while(qry.next()){

            sqlModel->setQuery(qry);
        }
    }
    int rows = sqlModel->rowCount();
    int columns = sqlModel->columnCount();

    for (int i = 0; i < rows; i++) {
        for (int j = 0; j < columns; j++) {
            if(!sqlModel->data(sqlModel->index(i,j)).isNull()
               ){
                textData += sqlModel->data(sqlModel->index(i,
                   j)).toString();
                textData += ", ";       // for .csv file
                   format
             }
        }
        textData += "\n";
    }
```

**Listing 3.6:** Qt reading from table

# Proof of Concept Test 4

## 4.1 Method

The network configuration from section 3.1 and FFmpeg command from subsection 3.1.2 was reused. Once the recording was finished, the python script was manually executed on the server with video footage from a single camera. When testing Angle Determination it was necessary to set up a static testing environment in order to get consistent results. One of the considerations of a testing environment were the influence of lighting in the footage, therefore it was done in a room without natural light sources. The color calibration of orange were done manually with the Red Green Blue (RGB) color code as a reference point. In the video footage all shades of orange were removed, this also included some shades of the color yellow because it was found troublesome for the chosen color boundaries. The maximum distance tested were approximately 5 metres due to the size of the room. The orientation and angle of the camera were static throughout the process because it was found that a tilted camera could interfere with the face detection.

### 4.1.1 Test setup

The test was conducted with the goal of determining the score of a person (see section 3.2.3) from four different angles: Frontal, dorsal and two side profiles. Each angle was recorded for approximately 5 seconds from a distance of approximately 5 metres and the average score of each of the above-mentioned angles were calculated. First video footage was recorded with the camera, and all data were saved into a folder on the server and a test script was then started manually on the server. The test script then outputs the average score for each angle.

The GUI was tested individually by opening a dummy video with the length of 1 hour and 20 minutes which is the length of APA. Moreover, a presentation of the interface were given to the analyst. In order, to do a full system test it has to be in a production environment.

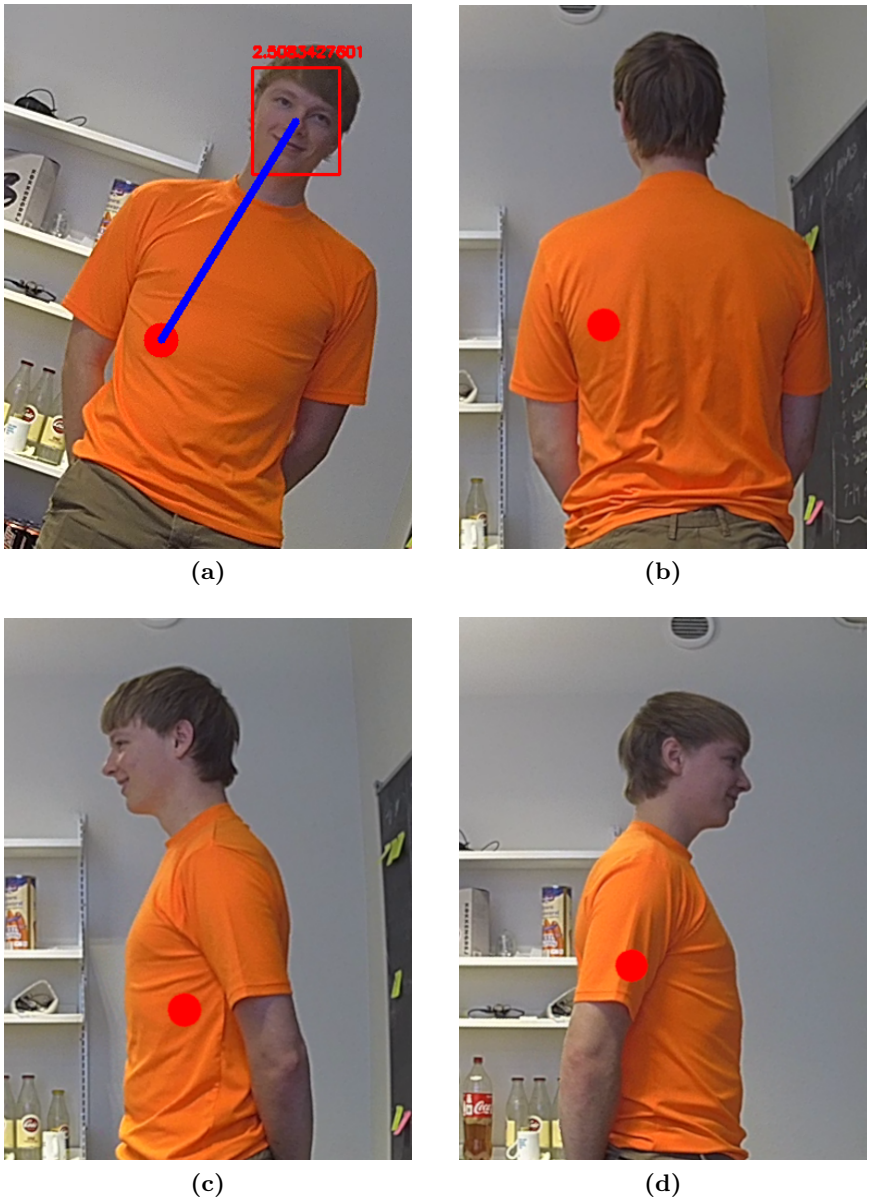
## 4.2 Results

The results of the angle determination showed consistent favor to the frontal side. The second most favored angle were the Left side and both the right and the dorsal side scored lowest. This can be seen on figure 4.1 on the following page.

```
Frontal score:  2.616625798888059
Behind score:  0.5
Right score:  0.5
Left score:  0.8582001464780578
--- 241.57066702842712 seconds ---
```

**Figure 4.1:** The scores from four different angles

A visual depiction of the test and the four angles can be seen on 4.2.



**(a)** **(b)**

**(c)** **(d)**

**Figure 4.2:** Visualization of different angles

The 1 hour and 20 min long recording from the HIKVISION camera resulted in a file size of almost 5GB. It was tested and confirmed that the GUI had no problem opening and playing a file of that size.

To make sure that the analyzing software was usable for the analyst, a review of the interface was set up. The analyst was satisfied with the features of the software so far and requested a possibility to plot the data into a graph of each individual student to get an overview of the progress. This should be taken into consideration if the current system is to be developed further.

# 5 Reflection

The CDH was consulted to determine their problem and their goals for this project. To supplement this, an analysis of the problem was carried out. Based on this analysis, requirements were specified, which, if they were fulfilled by a system, that system could function as a solution to the CDH' problem. With these requirements in mind an attempt at developing a suitable solution was made. The solution was split in to three, somewhat independently, developed parts. During development these three parts were tested independently and since the solution did not reach a state of being automatic, a full system test was not conducted. The testing of the angle determination revealed that the scoring system worked to a certain degree, this means that the e.g. the frontal side gave more points than the dorsal side. Furthermore, it was also found that a side profile sometimes can give points, but with a lower confidence level. However, through informal testing of the angle determination errors were found, this included tracking of the wrong person, incorrect color detection, and at times missing face detection. Examples of these errors can be seen in Appendix B. The GUI testing showed that opening video footage with an equal length to an APA session was possible. Further testing of the GUI is needed, this includes stress tests of both the database engine and the interface as a whole. Testing the system during APA, where one of the students are wearing the tracking token, would be beneficial to evaluate the angle determination and video processing. It would also help to decide how many cameras provide the best cost to benefit ratio and where they should be placed. Putting a larger effort into camera selection through testing of different models would also be beneficial, as it would ensure compatibility and quality. It was not tested if the proposed camera could be used with the angle determination during APA. A comprehensive test of distance with the angle determination algorithm or other could be found necessary to ensure functionality in the gymnasium.

Thorough testing of the GUI should also be performed with footage from APA, which would provide helpful insight regarding potential improvements. Moreover, a usability test with the analyst should also be conducted, which could lead to a reconsideration of the GUI design. A third-party interaction design specialist could be included to enhance the usability and make it more intuitive. Recreating the GUI as a web application could also provide some benefits such as portability across device types and ease of use.

Angle detection should also be revisited. This could involve reevaluating the methods of face detection and potentially having it recognize tilting and upside down faces. Through testing it was found that the execution time of the script were four times higher than the length of the video footage on the testing computer, which could be hardware dependent. Other face detection methods could also be investigated since they might increase the detection rate and/or decrease execution time of the script. An alternative object detection

method should also be investigated due to the shortcomings of the previously mentioned method. Implementing face recognition is also a possibility. Face recognition would make object detection redundant, since the subject would no longer be identified by a token but by their face. A drawback and the reason for omitting face recognition in this project is the need of training data for each individual.

The scoring system was later found to be underdeveloped and inconsistent. The points given for shortest distance from orange to face was unnecessary due to the fact it implicates that another person than the one wearing the orange vest could get points. It was however important to weigh this distance with how the color detection was implemented in the angle detection. Note that if face recognition is implemented, major changes to the point system would have to be made and a similar confidence level, as used in the face detection, could be found beneficial to choose the best angle.

The need for manual intervention when using the system should be reduced and made as automated as possible, which would increase the usability. Moreover, manual interventions are time-consuming and complicated, which would, to a lesser extent, defeat the purpose of this system compared to current solutions. The possibility of turning the cameras on and off with a remote, which was requested by the analyst, should be added. This simplifies an otherwise complicated manual intervention.

Ways to make students and guardians able to delete, and/or see footage relevant to the student without showing the other students should be researched in order to more easily comply with the GDPR. This would require the use of face recognition. Employing some kind of protection (e.g. encryption) of the transferred data would also be favorable, as it would remove the need to trust the local network and thus reduce the complexity of securing the data using a VPN connection or similar. This protection could be added by isolating the IP-Cameras on the local network and using Hypertext Transfer Protocol Secure (HTTPS) if the GUI is recreated as a web application.

# Conclusion 6

It was examined if an angle determination algorithm consisting of face detection and color detection could be used to track a specific person in video footage. Moreover, an investigation of the feasibility of RTSP for retrieving footage and SMB for accessing files on a server was carried out. A GUI was created with the intent to examine the possibilities of creating a cross-platform application that could be used by an analyst to review previously mentioned video footage.

It was not tested whether the angle determination was sufficient for a production environment or not. However, several concerns were raised and while it was deemed better than current solutions, it was not deemed adequate given the requirement specification. Since the angle determination was not sufficient a different method should be used.

The usage of RTSP for retrieving footage and SMB for delivering recordings was tested and deemed acceptable in a testing environment, though further automation should be implemented before being acceptable in production environment.

The GUI was also tested and found to be functionally adequate as a solution in a testing environment. Feedback from the analyst about the GUI suggested improvements, which would make it sufficient in a production environment.

It is speculated that for a fully working system to be developed, a more iterative process which increases the involvement with the CDH should be adopted and testing at the CDH should be adopted.

In conclusion the proof of concept was not compliant with the specified requirements.

# Bibliography


[1] Socialstyrelsen, nov 2018. URL https://socialstyrelsen.dk/handicap/dovblindhed.

[2] Nordens välfärd. Nordic definition of deafblindness, 2016.

[3] Pia Hesse, 2018. Interview with the organization The Danish Deafblind Association (FDDB).

[4] Socialstyrelsen. Medfødt døvblindhed, sep 2018. URL https://socialstyrelsen.dk/handicap/dovblindhed/om-medfodt-dovblindhed/medfodt-dovblindhed-i-tal.

[5] Videnscenter for døvblindfødte. Kortlægning af medfødt døvblindhed - en dansk populationsundersøgelse. PDF, Videnscenter for døvblindfødte, sep 2004. URL https://socialstyrelsen.dk/udgivelser/kortlaegning-af-medfodt-dovblindhed/@@download/publication. Chapter 3.3.2.

[6] Jesper Dammeyer. Mental and behavioral disorders among people with congenital deafblindness. *Research in Developmental Disabilities*, 32(2):571–575, mar 2011. URL https://doi.org/10.1016/j.ridd.2010.12.019.

[7] Socialstyrelsen. Erhvervet døvblindhed, july 2018. URL https://socialstyrelsen.dk/handicap/dovblindhed/om-erhvervet-dovblindhed/erhvervet-dovblindhed-i-tal.

[8] Center for Døvblindhed og Høretab. Afdelinger på døvblindeområdet, aug 2018.

[9] Lis Just, Bettina U. Møller, and Ole E. Mortensen. Døvblinde og beskæftigelse - en artikelsamling. In *Døvblinde og beskæftigelse - En artikelsamling*. Videnscentret for Døvblindblevne, 2008.

[10] Undervisningsministeriet. Bekendtgørelse om folkeskolens specialundervisning og anden specialpædagogisk bistand. Web, jun 2014.

[11] Paul Read and Mark-Paul Meyer. *Restoration of Motion Picture Film*. Butterworth-Heinemann Series in Conservation and Museology. Butterworth-Heinemann, 1. edition, sep 2000. Page 24.

[12] Gordon B. Rose. Nvr vs dvr comparison – difference between them?, 2018. URL https://www.phenomsecurity.com/nvr-vs-dvr/. Phenom Security.

[13] Frontier Technology Team. Top 4 benefits of moving from analog to ip video surveillance, nov 2014. URL http://www.frontier-security.com/blog/top-4-benefits-of-moving-from-analog-to-ip-video-surveillance/. Frontier Security.

[14] Ramiz Raja. Face detection using opencv and python: A beginner's guide, july 2017. URL https://www.superdatascience.com/opencv-face-detection/.

[15] Shermal Fernando. Shape detection & tracking using contours, 2017. URL https://www.opencv-srf.com/2011/09/object-detection-tracking-using-contours.html.

[16] Shermal Fernando. Color detection & object tracking, 2017. URL https://www.opencv-srf.com/2010/09/object-detection-using-color-seperation.html.

[17] European Parliament and of the council of 27. April 2016. General data protection regulation, may 2018. URL https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679. Article 8, 15 and 17.

[18] Richard Sharpe. Samba Documentation, oct 2002. URL https://www.samba.org/cifs/docs/what-is-smb.html.

[19] Debian Wiki, feb 2016. URL https://wiki.debian.org/SambaServerSimple.

[20] Al Lukaszewski. What is python programming language? *ThoughtCo.*, may 2018. URL https://www.thoughtco.com/what-is-python-2813564.

[21] Editor. Redhat vs debian : Administrative point of view. *TecMint*, oct 2013. URL https://www.tecmint.com/redhat-vs-debian-administrative-point-of-view/.

[22] Chunlei Liu. Multimedia Over IP: RSVP, RTP, RTCP, RTSP. Technical report, Washington University - Department of Computer Science & Engineering, jan 1998. URL https://www.cse.wustl.edu/~jain/cis788-97/ftp/ip_multimedia/.

[23] H. Schulzrinne, A. Rao, and R. Lanphie. Real Time Streaming Protocol (RTSP). Technical report, The Internet Society, apr 1998. URL https://www.ietf.org/rfc/rfc2326.txt.

[24] The FFmpeg developers. ffmpeg Documentation. Web, dec 2018. URL https://ffmpeg.org/ffmpeg.html.

[25] Eric Glass. Davenport documentation. Web, 2006. URL http://davenport.sourceforge.net/ntlm.html.

[26] Gaëtan Leurent. MD4 is Not One-Way. Technical report, École Normale Supérieure – Département d'Informatique, 2008. URL https://www.iacr.org/workshops/fse2008/docs/papers/day_3_sess_2/26_md4_oneway.pdf.

[27] Qt. Qstring class. Web, dec 2018. URL https://doc.qt.io/qt-5/qstring.html.

[28] Sqlite. Datatypes in sqlite version 3. Web, dec 2018. URL https://www.sqlite.org/datatype3.html.

# Appendix A

## A.1 Debian Samba Setup

```
#!/usr/bin/env python3

# Based on this guide: https://www.digitalocean.com/community
   /tutorials/how-to-set-up-a-samba-share-for-a-small-
   organization-on-ubuntu-16-04

# Also based on this guide: https://www.linode.com/docs/
   security/firewalls/introduction-to-firewalld-on-centos/

import os
import time

# Install packages
os.system('apt-get install samba')
os.system('apt-get install firewalld')

# nmbd is not required
os.system('systemctl stop nmdb.service')
os.system('systemctl disable nmdb.service')

# Temporily stop samba while unconfigured
os.system('systemctl stop smbd.service')

smb_glo_interface = 'temp'
conf_path = '/etc/samba/smb.conf'
srv_str = 'cdh_cam_server'

def smb_prompt():
    os.system('ip link show')
    smb_glo_interface = input('''
What is the name of your global interface (eg. eth0)?
''')

    smb_conf = '''[global]
        server string = {}
```

```
        server role = standalone server
        interfaces = lo {}
        bind interfaces only = yes
        disable netbios = yes
        smb ports = 445
        log file = /var/log/samba/smb.log
        max log size = 10000
'''.format(srv_str, smb_glo_interface)

    file = open('{}'.format(conf_path), 'w')
    file.write('{}'.format(smb_conf))
    file.close()

def user_setup( client_username ):
    # Configure users, their directories and privileges
    os.system('mkdir /samba/')
    os.system('chown :sambashare /samba/')
    os.system('mkdir /samba/{}'.format(client_username))
    os.system('adduser --home /samba/{} --no-create-home --
       shell /usr/sbin/nologin --ingroup sambashare {}'.
       format(client_username, client_username))
    os.system('chown {}:sambashare /samba/{}'.format(
       client_username, client_username))
    os.system('chmod 2770 /samba/{}'.format(client_username))

    # Add user
    os.system('smbpasswd -a {}'.format(client_username))

    # Enable user
    os.system('smbpasswd -e {}'.format(client_username))

    smb_user_conf = '''
[{}]
        path = /samba/{}
        browseable = no
        read only = no
        force create mode 0660
        force directory mode = 2770
        valid users = {} @admins
'''.format(client_username, client_username, client_username)

    file = open('{}'.format(conf_path),'a')
    file.write('{}'.format(smb_user_conf))
    file.close()
    return
```

```
smb_prompt()
user_setup('analyst')
user_setup('camera_1')
os.system('systemctl start smbd.service')
os.system('systemctl enable smbd.service')

#Configure firewall
os.system('systemctl start firewalld')
os.system('systemctl enable firewalld')
os.system('firewall-cmd --zone=$(firewall-cmd --get-default-
   zone) --add-service=samba')
os.system('firewall-cmd --runtime-to-permanent')
os.system('firewall-cmd --reload')

# Test samba config
os.system('testparm')
```

**Listing A.1:** Debian Samba Setup

## A.2 FFmpeg Output

```
$ ffmpeg -i rtsp://admin:1q2w3e4r@10.42.0.70:554/live.sdp -
   vcodec copy -y frivid.mp4

ffmpeg version 3.3.9 Copyright (c) 2000-2018 the FFmpeg
   developers

  built with gcc 6.3.0 (Debian 6.3.0-18+deb9u1) 20170516

  configuration: --disable-decoder=amrnb --disable-decoder=
     libopenjpeg --disable-mips32r2 --disable-mips32r6 --
     disable-mips64r6 --disable-mipsdsp --disable-mipsdspr2
     --disable-mipsfpu --disable-msa --disable-libopencv --
     disable-podpages --disable-stripping --enable-avfilter
     --enable-avresample --enable-gcrypt --enable-gnutls --
     enable-gpl --enable-libass --enable-libbluray --enable-
     libbs2b --enable-libcaca --enable-libcdio --enable-
     libfdk-aac --enable-libfontconfig --enable-libfreetype
     --enable-libfribidi --enable-libgme --enable-libgsm --
     enable-libilbc --enable-libkvazaar --enable-libmp3lame
     --enable-libopencore-amrnb --enable-libopencore-amrwb --
     enable-libopenh264 --enable-libopenjpeg --enable-
     libopenmpt --enable-libopus --enable-libpulse --enable-
     librubberband --enable-libshine --enable-libsnappy --
     enable-libsoxr --enable-libspeex --enable-libtesseract
     --enable-libtheora --enable-libvidstab --enable-libvo-
```

```
    amrwbenc --enable-libvorbis --enable-libvpx --enable-
    libx265 --enable-libxvid --enable-libzvbi --enable-
    nonfree --enable-opengl --enable-openssl --enable-
    postproc --enable-pthreads --enable-shared --enable-
    version3 --incdir=/usr/include/x86_64-linux-gnu --libdir
    =/usr/lib/x86_64-linux-gnu --prefix=/usr --toolchain=
    hardened --enable-frei0r --enable-chromaprint --enable-
    libx264 --enable-libiec61883 --enable-libdc1394 --enable
    -vaapi --disable-opencl --enable-libmfx --disable-
    altivec --shlibdir=/usr/lib/x86_64-linux-gnu

  libavutil      55. 58.100 / 55. 58.100

  libavcodec     57. 89.100 / 57. 89.100

  libavformat    57. 71.100 / 57. 71.100

  libavdevice    57.  6.100 / 57.  6.100

  libavfilter     6. 82.100 /  6. 82.100

  libavresample   3.  5.  0 /  3.  5.  0

  libswscale      4.  6.100 /  4.  6.100

  libswresample   2.  7.100 /  2.  7.100

  libpostproc    54.  5.100 / 54.  5.100

[rtsp @ 0x5655378a5a80] max delay reached. need to consume
   packet

[rtsp @ 0x5655378a5a80] RTP: missed 3 packets

[h264 @ 0x5655378a9160] error while decoding MB 58 54,
   bytestream -5

[h264 @ 0x5655378a9160] concealing 1671 DC, 1671 AC, 1671 MV
   errors in P frame

Input #0, rtsp, from 'rtsp://admin:1q2w3e4r@10.42.0.70:554/
   live.sdp':

  Metadata:

    title           : Media Presentation
```

```
  Duration: N/A, start: 0.274556, bitrate: N/A

    Stream #0:0: Video: h264 (Main), yuvj420p(pc, bt709,
       progressive), 1920x1080 [SAR 1:1 DAR 16:9], 22 fps, 25
        tbr, 90k tbn, 44 tbc

Output #0, mp4, to 'frivid.mp4':

  Metadata:

    title           : Media Presentation

    encoder         : Lavf57.71.100

    Stream #0:0: Video: h264 (Main) ([33][0][0][0] / 0x0021),
        yuvj420p(pc, bt709, progressive), 1920x1080 [SAR 1:1
       DAR

    16:9], q=2-31, 22 fps, 25 tbr, 90k tbn, 90k tbc

Stream mapping:

  Stream #0:0 -> #0:0 (copy)

Press [q] to stop, [?] for help

[mp4 @ 0x5655378e1c20] Non-monotonous DTS in output stream
   0:0; previous: 0, current: -21110; changing to 1. This may
    result in incorrect timestamps in the output file.

[mp4 @ 0x5655378e1c20] Non-monotonous DTS in output stream
   0:0; previous: 1, current: -17510; changing to 2. This may
    result in incorrect timestamps in the output file.

[mp4 @ 0x5655378e1c20] Non-monotonous DTS in output stream
   0:0; previous: 2, current: -13910; changing to 3. This may
    result in incorrect timestamps in the output file.

[mp4 @ 0x5655378e1c20] Non-monotonous DTS in output stream
   0:0; previous: 3, current: -10310; changing to 4. This may
    result in incorrect timestamps in the output file.

[mp4 @ 0x5655378e1c20] Non-monotonous DTS in output stream
   0:0; previous: 4, current: -3110; changing to 5. This may
   result in incorrect timestamps in the output file.
```

```
frame=   41 fps=0.0 q=-1.0 size=    1858kB time=00:00:01.52
   bitrate=9976.2kbits/s speedframe=   52 fps= 49 q=-1.0 size
   =    2641kB time=00:00:02.04 bitrate=10575.4kbits/s
   speeframe=   64 fps= 40 q=-1.0 size=    3149kB time
   =00:00:02.60

bitrat
e=9902.5kbits/s speedframe=   76 fps= 36 q=-1.0 size=    3546
   kB time=00:00:03.12 bitrate=9294.7kbits/s speedframe=   87
    fps= 33 q=-1.0 size=    3971kB time=00:00:03.64 bitrate
   =8924.1kbits/s speedframe=   99 fps= 31 q=-1.0 size=
   4381kB time=00:00:04.16 bitrate=8615.0kbits/s speedframe=
    110 fps= 30 q=-1.0 size=    4962kB time=00:00:04.68

bitrate=8674.8kbits/s speedframe=  122 fps= 29 q=-1.0 size=
      5369kB time=00:00:05.20 bitrate=8448.7kbits/s
   speedframe=  133 fps= 28 q=-1.0 size=    5731kB time
   =00:00:05.72 bitrate=8200.2kbits/s speedframe=  145 fps=
   27 q=-1.0 size=    6145kB time=00:00:06.28 bitrate=8009.1
   kbits/s speedframe=  157 fps= 27 q=-1.0 size=    6988kB
   time=00:00:06.80

bitrate=8411.8kbits/s speedframe=  168 fps= 27 q=-1.0 size=
      7407kB time=00:00:07.32 bitrate=8283.0kbits/s
   speedframe=  180 fps= 26 q=-1.0 size=    7828kB time
   =00:00:07.84 bitrate=8172.6kbits/s speedframe=  191 fps=
   26 q=-1.0 size=    8222kB time=00:00:08.36 bitrate=8050.3
   kbits/s speedframe=  203 fps= 26 q=-1.0 size=    9100kB
   time=00:00:08.88

bitrate=8388.5kbits/s speedframe=  214 fps= 26 q=-1.0 size=
      9498kB time=00:00:09.40 bitrate=8272.1kbits/s
   speedframe=  226 fps= 25 q=-1.0 size=    9954kB time
   =00:00:09.96 bitrate=8181.9kbits/s speedframe=  238 fps=
   25 q=-1.0 size=   10366kB time=00:00:10.48 bitrate=8097.7
   kbits/s speedframe=  249 fps= 25 q=-1.0 size=   10761kB
   time=00:00:11.00

bitrate=8009.5kbits/s speedframe=  261 fps= 25 q=-1.0 size=
     11681kB time=00:00:11.52 bitrate=8301.6kbits/s
   speedframe=  272 fps= 25 q=-1.0 size=   12159kB time
   =00:00:12.04 bitrate=8268.8kbits/s speedframe=  284 fps=
   24 q=-1.0 size=   12741kB time=00:00:12.60 bitrate=8279.4
   kbits/s speedframe=  296 fps= 24 q=-1.0 size=   13128kB
   time=00:00:13.12
```

```
bitrate=8193.1kbits/s speedframe=  307 fps= 24 q=-1.0 size=
     13713kB time=00:00:13.64 bitrate=8231.8kbits/s
   speedframe=  319 fps= 24 q=-1.0 size=   14230kB time
   =00:00:14.16 bitrate=8228.9kbits/s speedframe=  330 fps=
   24 q=-1.0 size=   14568kB time=00:00:14.68 bitrate=8126.2
   kbits/s speedframe=  331 fps= 24 q=-1.0 Lsize=   14592kB
   time=00:00:14.72

bitrate=8117.5kbits/s speed=1.07x
video:14589kB audio:0kB subtitle:0kB other streams:0kB global
    headers:0kB muxing overhead: 0.021353%
Exiting normally, received signal 2.
```

**Listing A.2:** FFmpeg Output

# Appendix B

## B.1 Error examples

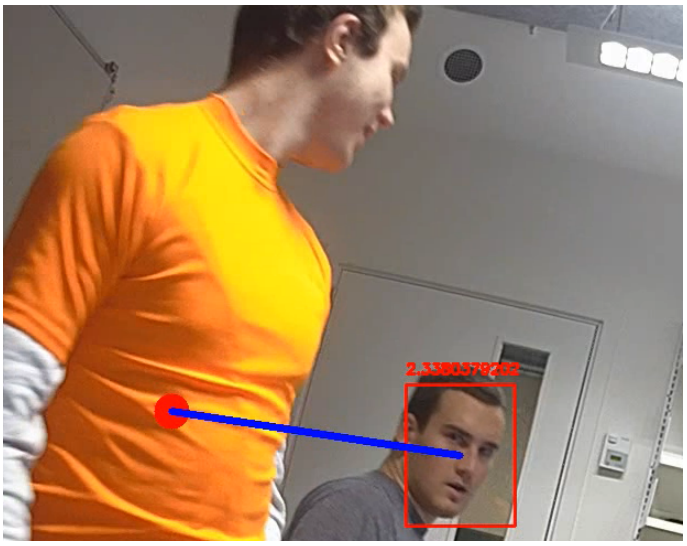


**Figure B.1:** A face is scored but it is not the token wearer

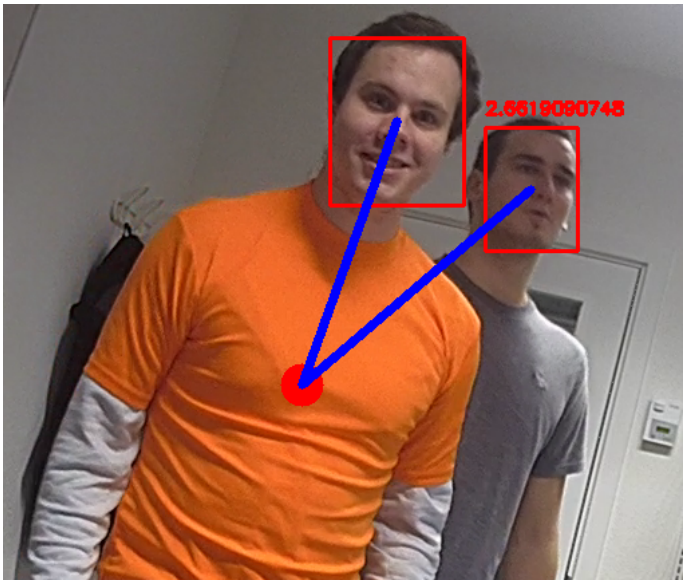



**Figure B.2:** A face is scored but it is not the token wearer

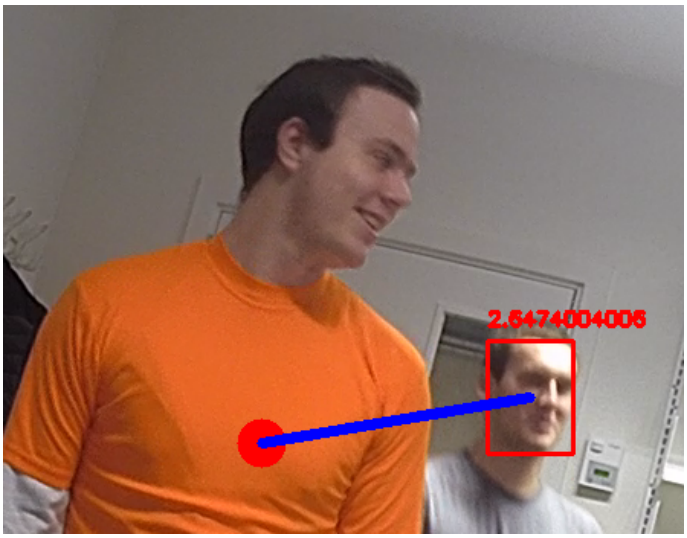



**Figure B.3:** A face is scored but it is not the token wearer